Accepted refereed manuscript of: Neil C, Spyrakos E, Hunter PD & Tyler AN (2019) A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. *Remote Sensing of Environment*, 229, pp. 159-178. DOI: https://doi.org/10.1016/j.rse.2019.04.027

A global approach for chlorophyll-*a* retrieval across optically complex inland waters based on optical water types

Neil, C.[1]*, Spyrakos, E.[1], Hunter, P.D.[1], Tyler, A.N.[1]

[1] Earth and Planetary Observation Centre (EPOC), Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling, United Kingdom

* Corresponding author: claire.neil@stir.ac.uk

Keywords

algorithm validation, chlorophyll-*a*, water quality, inland waters, case 2 waters, remote sensing, optical water type.

Abstract

Numerous algorithms have been developed to retrieve chlorophyll-*a* (Chla) concentrations (mg m^{-3}) from Earth observation (EO) data collected over optically complex waters. Retrieval accuracy is highly variable and often unsatisfactory where Chla co-occurs with other optically active constituents. Furthermore, the applicability and limitations of retrieval algorithms across different optical complex systems in space and time are often not considered. In the first instance, this paper provides an extensive performance assessment for 48 Chla retrieval algorithms of varying architectural design. The algorithms are tested in their original parametrisations and are then retuned using in-situ remote sensing reflectance ($R_{rs}(\lambda)$, sr^{-1}) data (n = 2807) collected from 185 global inland and coastal aquatic systems encompassing 13 different optical water types (OWTs). The paper then demonstrates retrieval performance across the full dataset of observations and within individual OWTs to determine the most effective model(s) of those tested for retrieving Chla in waters with varying optical properties. The results revealed significant variability in retrieval performance when comparing model outputs to in-situ measured Chla for the full in-situ dataset in its entirety and within the 13 distinct OWTs. Importantly, retuning an algorithm to optimise its parameterisation for each individual OWT (i.e. one algorithm, multiple parameterisations) is found to improve the retrieval of Chla overall compared to simply calibrating the same algorithm using the complete in-situ dataset (i.e. one algorithm, one parameterisation). This resulted in a 25% improvement in retrieval accuracy based on relative percentage difference errors for the best performing Chla algorithm. Improved performance is further achieved by allowing model type and specific parameterisation to vary across OWTs (i.e. multiple algorithms, multiple parameterisations). This adaptive framework for the dynamic selection of in-water algorithms is shown to provide overall improvement in Chla retrieval across a continuum of bio-geo-optical conditions. The final dynamic ensemble algorithm produces estimates of (log_{10}-transformed) Chla with a correlation coefficient of 0.89 and a mean absolute error of 0.18 mg m^{-3}. The OWT framework presented in this study demonstrates a unified approach by bringing together an ensemble of algorithms for the monitoring of inland waters at a global scale from space.

1. Introduction

Since the successful launch of the Coastal Zone Color Scanner (CZCS) in 1978, satellite remote sensing (RS) has played an increasingly important role in observing the complex biogeochemical interactions that occur in the global ocean and its response to drivers of environmental change (Gordon et al., 1980; Antoine et al., 1996). Radiometric sensors mounted on satellites have provided the capability to deliver synoptic maps of global chlorophyll-*a* concentration (Chla) (McClain, 2009) which have led to fundamental contributions in oceanographic research, coastal management and climate change studies (Brown & Yoder, 1994; Behrenfeld et al., 2005; Hu et al., 2005; Nair et al., 2008; Yang et al., 2013). Satellite data have also been used in the monitoring of inland waters to provide information on a suite of functionally relevant indicators of water quality and ecosystem condition (Gitelson et al., 1993; Kutser et al., 1998; Lindell et al., 1999; Dekker et al., 2002; Kutser et al., 2005; Simis et al., 2005; Giardino et al., 2010; Tarrant et al., 2010; Hunter et al., 2010; Matthews et al., 2010; Odermatt et al., 2010; Nechad et al., 2010; Dogliotti et al., 2015; Palmer et al., 2015a) however optical complexity in these waters often limits operational use. In this context, Chla is the main bioindicator of water quality retrievable from EO data and its variations over space and time offer unique insight into the changing status of inland waters (Adrian et al., 2009) and the effects of environmental stressors (e.g., nutrient enrichment, hydrological modifications, climate change) at local, regional and global scales.

Various studies have shown promising results for retrieving Chla from inland waters using EO data (Palmer et al., 2015a , Matthews and Odermatt, 2015; Tyler et al., 2016) but the majority of these evaluate performance on individual or small populations of lakes with often limited variability in their optical properties. With a large number of algorithms available to the EO community it can be difficult to ascertain the applicability range and limitations of each method when applied globally (Morel et al., 2007; Matthews, 2011; Odermatt et al., 2012; Blondeau-Patissier et al., 2014; Tilstone et al., 2017). Algorithm performance often varies in response to changes in the optical properties of the water column which in turn are related to the presence of the non-covarying optically active constituents suspended particulate matter (SPM, mgm^{-3}) and coloured dissolved organic material (CDOM, m^{-1}); a simple example is the Case 1 or Case 2 bipartite classification scheme (Morel & Prieur, 1977) which predefines the conditions where standard ocean colour Chla algorithms are expected to break-down (McKee et al., 2007; Moore et al., 2009; Mouw et al., 2015). This paper aims to extend this strategy to assess algorithm performance in relation to a number of distinct Optical Water Types (OWT) with the ambition of not only improving the overall performance of retrieval algorithms across a continuum of optical properties but also improving our ability to select appropriate algorithms and parameterisations for a given scenario. The accuracy of a number of Chla algorithms will be assessed over a diverse range of OWTs derived from inland (and some transitional) waters in support of the UK's Natural Environment Research Council funded GloboLakes' project, which is developing a global observatory for inland waters using archived and near real-time processing of ocean colour imagery (Envisat MERIS and Sentinel-3 OLCI). In the context of this research, the performance methodology presented here will inform the robust selection of an ensemble of candidate algorithms capable of accurately retrieving concentrations of Chla in approximately 1000 lakes globally (and > 50% of the Earth's surface water by area) (Politi et al., 2016; Tyler et al., 2016). The overarching idea is not to advocate a single algorithm for global application, but to combine several retrieval models in an ensemble and use the OWT framework to dynamically select optimal models in space and time and thereby improve the overall accuracy of Chla retrieval across a wider range of water bodies. To this end, the study was partitioned into the following subtasks: (1) existing (hereafter denoted original) algorithms and their parameterisations were tested against an extensive database of in-situ reflectance and Chla measurements; (2) algorithms were calibrated by empirically adjusting model coefficients where applicable using in-situ measurements as a training dataset; (3) calibration was applied using in-situ data grouped by OWT cluster; (4) the performance of original (ORG), calibrated (CAL) and cluster (CLUS) retuned algorithms was compared and ranked to benchmark suitable Chla retrieval algorithms for each defined OWT.

2. Methods

2.1 Data

The validation and training dataset used to investigate Chla retrieval algorithms consists of 17 individual datasets collected from lakes and other inland water bodies worldwide (https://www.limnades.org/home.psp). The number of lakes and samples per dataset is shown in Table 1. A full description of the individual datasets with corresponding measurement and processing protocols are provided in Spyrakos et al. (2018a). The database comprises in-situ measurements of inherent and apparent optical properties (IOPs and AOPs respectively) and biogeochemical constituents collected from 185 aquatic systems representing a variety of bio-geo-optical conditions. The primary input to the Chla algorithms considered in this study is the remote sensing reflectance, $R_{rs}(\lambda)$ (sr^{-1}) which can be defined as the wavelength dependent ratio of water-leaving radiance and downwelling irradiance just above the water surface. $R_{rs}(\lambda)$ collected in-situ above the water surface is essentially the spectral distribution of reflected radiation a satellite sensor would detect with no atmospheric contribution and is considered reference data for RS algorithm development and radiometric validation. The validation dataset comprised 2807 hyperspectral $R_{rs}(\lambda)$ (sr^{-1}) measurements (interpolated to a common 1 nm spectral resolution) with corresponding concentrations of Chla. Measurements were obtained following generally accepted methods originating from more than 40 published studies

(Spyrakos et al., 2018a). Approximately 73% of Chla estimates used in this study were obtained spectrophotometrically. Of the remaining estimates, 13% were determined from HPLC-based methods, 7% fluorometrically and 7% were calculated from absorption coefficients using the equation of Ritchie (2008). It is known that variability in Chla quantification methods and interlaboratory protocols may contribute to uncertainty in the final the Chla estimate (Claustre et al., 2004; Hooker et al., 2005; Sørensen et al., 2007; McKee et al., 2014). Often refinement and optimization of measurement procedures are required in inland waters to tackle extreme optical complexities, thus prohibiting the standardisation of protocols. Nonetheless, Sørensen et al. (2007) suggests that discrepancy due to measurement variability is particularly consequential when monitoring case 1 waters. Furthermore, spectrophotometric methods, which account for a majority of Chla samples analysed in this study, have been shown to produce more consistent results between laboratories when compared to HPLC estimates (Sørensen et al., 2007). All of the datasets used in this study were validated by the individual data providers and then quality checked before inclusion in the LIMNADES database. Hyperspectral $R_{rs}(\lambda)$ measurements were spectrally resampled to the wavebands of MERIS (412, 442, 490, 510, 560, 620, 665, 681, 708, 753 nm) using the sensor spectral response function (https://earth.esa.int). No radiometric resampling was performed. The measurement range of corresponding biogeochemical constituent concentrations is shown in Figure 1. A mean Chla concentration of approximately 33.9 mg m^{-3} (median=16 mg m^{-3}) indicates a slight over representation in the dataset of high-biomass eutrophic systems relative to current global estimates (e.g., Sayers et al., 2015). However, the dataset also included measurements from a number of oligotrophic or hypereutrophic (Chla up to 1000 mg m^{-3}) systems as well as humic-rich and mineral-laden systems.

Table 1. Summary of the datasets used for algorithm development and validation

Dataset	Number of Lakes	Number of Samples
1	1	71
2	3	251
3	63	131
4	44	181
5	5	218
6	5	301
7	1	29
8	1	38
9	3	190
10	6	144
11	3	48
12	41	543
13	2	192
14	3	10
15	1	14
16	1	243
17	2	203
Total	185	2807

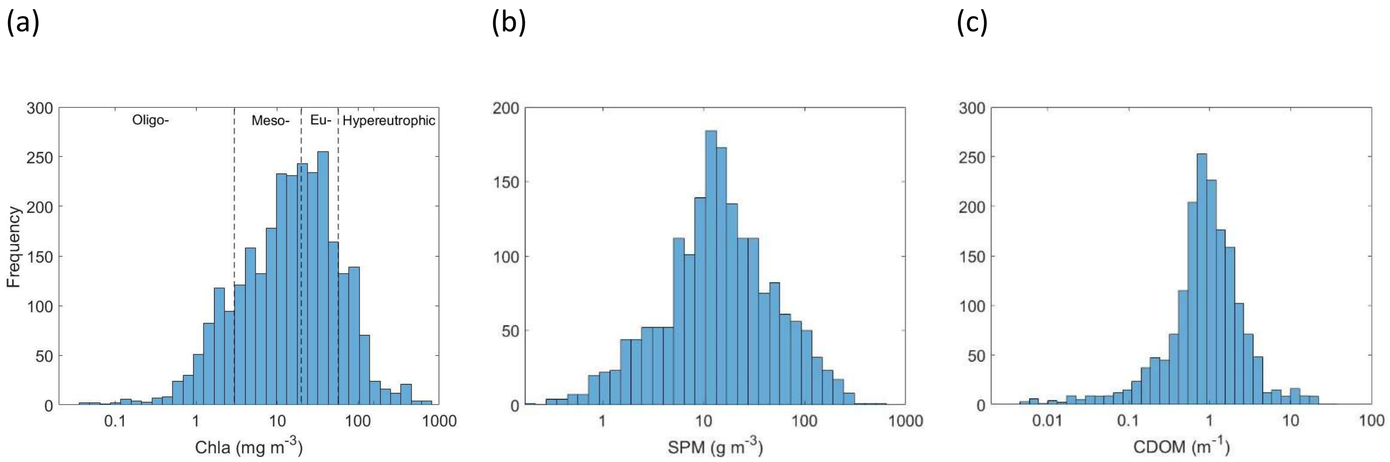
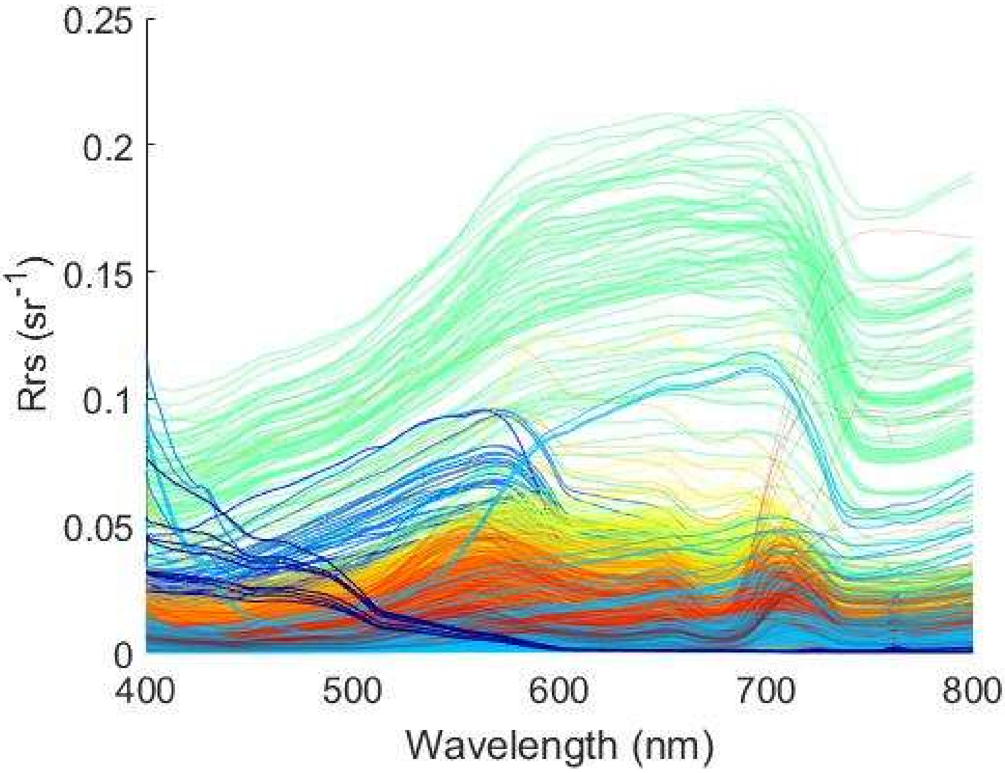


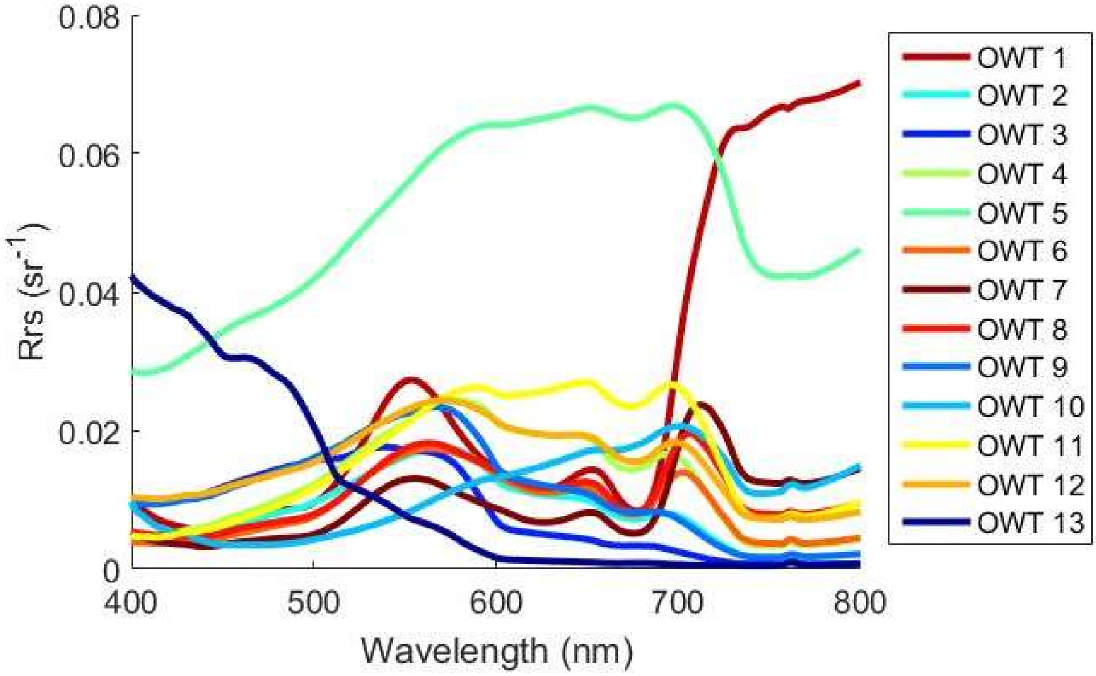
Figure 1. Biogeochemical constituent range of water bodies included in the validation dataset. Trophic class divisions (based on Carlson et al., 1996) are indicated with dashed lines on the Chla constituent histogram (a).

2.2 Optical water type framework

Previous work has been done to formally classify the R_{rs} spectra contained within the validation dataset into optical water typologies. In Spyrakos et al. (2018b), a k-means classification was adopted to identify and categorise OWTs based upon the differences in magnitude and shape of the hyperspectral R_{rs} spectra. This procedure identified 13 distinct OWTs, each corresponding to a specific combination of bio-geo-optical characteristics. R_{rs} spectra coloured according to OWT are shown in Figure 2a (and the mean R_{rs} spectra for each OWT is shown in Figure 2b). There are obvious differences in spectral shape *and* magnitude of R_{rs} for each defined OWT suggesting the applied classification scheme broadly captures the unique characteristics of the in-situ reflectance measurements. Median values of the optical constituent components Chla, SPM and CDOM are shown for each OWT group in Figure 3. The highest median Chla concentrations are observed in OWT 7, whilst SPM and CDOM occur in the highest concentrations in OWTs 5 and 1 respectively. Differences in the descending order of OWT group median values for each constituent confirm that OWTs have not been derived from a simple Chla concentration threshold and instead rely on a mixed combination of each optical constituent.


0.25
0.2
0.15
0.1
0.05
0
Rrs (sr-1)
400
500
600
700
800
Wavelength (nm)

(a)


0.08
0.06
0.04
0.02
0
Rrs (sr-1)
400
500
600
700
800
Wavelength (nm)
OWT 1
OWT 2
OWT 3
OWT 4
OWT 5
OWT 6
OWT 7
OWT 8
OWT 9
OWT 10
OWT 11
OWT 12
OWT 13

(b)

Figure 2. (a) R_{rs} spectra used in the validation dataset coloured by classified optical water type. (b) Average R_{rs} spectra per optical water type.

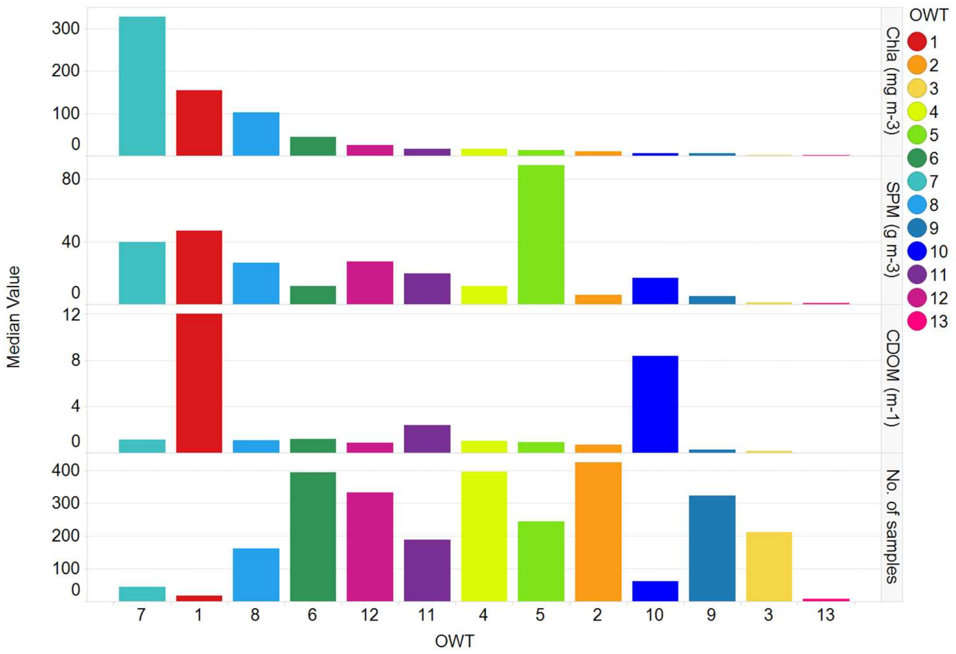


Figure 3. In-situ biogeochemical constituent median values for OWT groups ordered by median Chla concentration.

2.3 Chlorophyll algorithms

Based on bio-optical theory, $R_{rs}(\lambda)$ is related to water IOPs such as absorption, *a,* and backscattering, b_b (Gordon et al., 1988; Kirk, 1994; Mobley, 1999; Maritorena et al., 2002). Total IOPs, which are determined by the additive contribution of individual optically active constituents found in a water body, can be calculated by multiplying the concentration of each constituent by the appropriate specific inherent optical property (SIOP). As such the spectral signature of R_{rs} varies according to changes in constituent composition and concentration. Algorithms developed for the quantitative assessment of in-water constituents exploit the bio-optical model in different ways. Empirical methods establish a relationship between optical measurements and concentrations of constituents based on experimental data. They are simple to develop and implement, yet their intrinsic design make them particularly sensitive to changes in the composition of water constituents. An alternative analytical approach is to first infer IOPs from the reflectance signal and solve the radiative transfer equation to produce simultaneous estimates of optically active water constituents (Gordon et al., 1988; Mobley, 1994). The relationships between IOPs and the constituent concentrations are empirically derived, and thus these algorithms are said to be semi-analytical. There are a number of different approaches to semi-analytical modelling which include spectral matching or look-up-table methods (Kutser et al., 2001; Louchard et al., 2003; Mobley et al., 2005; Brando et al., 2009), non-linear optimization (Kuchinke et al., 2009), matrix inversion (Hoogenboom et al., 1998; Brando & Dekker, 2003), and direct inversion methods such as the multiband quasi-analytical model (QAA) (Lee et al., 1999) and the GSM semi-analytical model (GSM) (Maritonera et al., 2002). Semi-analytical methods have shown varying performance in case 2 waters (Shanmugam et al., 2010; Dekker et al., 2011; Odermatt et al., 2012). While based on solid physical principles, the general assumptions and simplifications of the semi-analytical methods, along with empiricism in the relations between IOPs and AOPs, often lead to ambiguities in water constituent retrieval (Bricaud et al., 1995; Defoin-Platel & Chami, 2007; McKee et al., 2014). Advanced analytical methods such as neural networks (Doerffer, 2007) also retrieve simultaneous combinations of biogeochemical constituents but these rely heavily, from a coverage and performance standpoint, on the quality of the spectral libraries employed in the training data sets. In this study, we assess the efficacy of empirical 1, 2 and 3 band algorithms, semi-analytical bio-optical models and a neural network which focus on the retrieval of Chla concentration. All algorithms included in the validation exercise, as summarised in Table 2 and described in the following section, were openly published (proprietary models were excluded from the exercise), well documented and developed for a range of optically variable environments. The tested algorithms can be generally categorised by their architectural designs as: (i) empirical methods which exploit ratios between R_{rs} collected remotely at blue and green wavelengths typically used for open ocean waters (O'Reilly et al., 1998); (ii) empirical NIR-red band ratio methods which are typically employed in turbid or eutrophic coastal and inland waters where Chla concentrations exceed 3 mg m^{-3} (Gitelson, 1992) and red reflectance may be relatively high; (iii) peak height methods which quantify the reflectance peak in relation to a standard baseline (Letelier and Abbott, 1996; Huot et al., 2005) and use the resulting relationship to empirically evaluate Chla; (iv) neural networks; and (v) semi-analytical methods.

Derived model coefficients have been denoted *a*, *b*, *c…* in each method where applicable. For models estimating the coefficient of absorption by phytoplankton (a_{ph}) as an output parameter (Model R and Model S), Chla was calculated as a function of a_{ph} using the expression (Bricaud et al., 1998);

$$Chla = \left(\frac{a_{ph}(443)}{a}\right)^{\frac{1}{b}} \qquad (1)$$

where *a* and *b* are derived empirically from the calibration dataset.

Model A

Model A refers to the two-band ratio algorithm of Dall' Olmo et al. (2003), Moses et al. (2009) and Gitelson et al. (2011), originally proposed by Gitelson and Kondratyev (1991) and later adapted to MERIS bands. This is an empirical formula based on a linear relationship between in-situ Chla and the ratio of MERIS satellite remote sensing reflectance, measured at NIR, $R_{rs}(708)$, and red, $R_{rs}(665)$, wavelengths;

$$Chla\_A = a \times \left(\frac{R_{rs}(708)}{R_{rs}(665)}\right) + b \quad (2)$$

where a = 61.324 and b = -37.94 are determined empirically.

Model B

Model B refers to the three-band algorithm developed by Moses et al. (2009) and adapted by Gitelson et al. (2011) to include R_{rs} measured at 753 nm, $R_{rs}(753)$;

$$Chla\_B = a \times \left(\frac{R_{rs}(753)}{(R_{rs}(665) - R_{rs}(708))}\right) + b \quad (3)$$

where a = 232.329 and b = 23.174 are determined empirically. In theory, the combination of three bands alters the model sensitivity to the presence of optically active constituents by removing the effects of SPM and CDOM ($R_{rs}(665)$ and $R_{rs}(708)$ are comparably influenced by SPM and CDOM and $R_{rs}(753)$ is mainly driven by backscattering).

Model C

Model C refers to the two-band empirically derived ratio algorithm of Gurlin et al. (2011);

$$Chla\_C = a \times \left(\frac{R_{rs}(708)}{R_{rs}(665)}\right)^2 + b \times \left(\frac{R_{rs}(708)}{R_{rs}(665)}\right) + c \qquad (4)$$

where a = 25.28, b = 14.85 and c = -15.18.

Model D

Model D refers to the three-band ratio algorithm of Gurlin et al. (2011) which was calibrated using field measurements of R_{rs} and Chla taken from Fremont lakes Nebraska;

$$Chla\_D = a \times \left(\frac{R_{rs}(753)}{(R_{rs}(665) - R_{rs}(708))}\right)^2 + b \times \left(\frac{R_{rs}(753)}{(R_{rs}(665) - R_{rs}(708))}\right) + c \qquad (5)$$

where a = 315.50 , b = 215.95 and c = 25.66.

Model E

Model E refers to the advanced two-band semi-analytical algorithm proposed by Gilerson et al. (2010). While this is governed by the ratio of NIR to red reflectance, model coefficients are determined analytically from individual absorption components contributing to the total IOPs of the water body. It is assumed that the water term dominates (at red – NIR wavelengths) where Chla concentration is greater than 5 mg m^{-3}, and that the contribution to absorption by CDOM and backscattering terms are significantly smaller to give the following expression;

$$Chla\_E = \left[a_{w708} \times \left(\frac{R_{rs}(708)}{R_r(665)}\right) - a_{w665}\right] / a^*_{ph665} \quad (6)$$

where a_{w708} = 0.7864 m^{-1} and a_{w665} = 0.4245 m^{-1} are absorption by water at the specified wavelengths (Pope & Fry, 1997) and phytoplankton specific absorption (a^*_{ph}) at 665 nm, a^*_{ph665}=0.022 x $Chla^{-0.1675}$. Substituting a^*_{ph665} into eq. 6 gives;

$$Chla\_E = \left[35.75 \times \left(\frac{R_{rs}(708)}{R_{rs}(665)}\right) - 19.30\right]^{1.124} \quad (7)$$

which can also be modified to allow for regional calibration of the a^*_{ph665} variable;

$$Chla\_E = \left[\frac{0.7864}{a} \times \left(\frac{R_{rs}(708)}{R_{rs}(665)}\right) - \frac{0.4245}{a}\right]^{1/b} \quad (8)$$

Here *a* may be determined empirically and *b* is parameterised to fit the data. The water term becomes less dominant when Chla < 5 mg m^{-3}, and therefore the assumed negligibility of the influence of CDOM and SPM is no longer valid under these conditions.

Model F

Model F refers to a simplified version of Gilerson et al. (2010) which relates Chla to the NIR-red reflectance band ratio through a simple power function;

$$Chla\_F = a \times \left(\frac{R_{rs}(708)}{R_{rs}(665)}\right)^b \quad (9)$$

where *a* and *b* are defined empirically as opposed to analytically as per Model E.

Model G

Model G refers to the advanced three-band semi-analytical algorithm proposed by Gilerson et al. (2010). As per Model E, the three-band model is based on a semi-analytical expression for the red-NIR ratio of reflectance in combination with water absorption and a^*_{ph665} (=0.022 x $Chla^{-0.1675}$);

$$Chla\_G = \left[a_{w753} \times \left(\frac{R_{rs}(753)}{(R_{rs}(665)-R_{rs}(708))}\right) - a_{w708} + a_{w665}\right] / a^*_{ph66} \quad (10)$$

where a_{w753} = 2.494 m^{-1} (Pope & Fry, 1997). Substituting the expression for a^*_{ph665} gives;

$$Chla\_G = \left[113.36 \times \left(\frac{R_{rs}(753)}{(R_{rs}(665)-R_{rs}(708))}\right) - 16.45\right]^{1.124} \quad (11)$$

Or for regional calibration of a^*_{ph665};

$$Chla\_G = \left[\frac{2.494}{a} \times \left(\frac{R_{rs}(753)}{(R_{rs}(665)-R_{rs}(708))}\right) - \frac{0.7864}{a} + \frac{0.4245}{a}\right]^{1/b} \quad (12)$$

where *a* and *b* are determined empirically. This expression is valid under the same conditions as defined by Model E.

Model H

Model H refers to the semi-analytical algorithm presented by Gons et al. (2002, 2005 and 2008) which incorporates information on water absorption and backscattering in relation to the MERIS red-NIR reflectance ratio;

$$Chla\_H = \left[\left(\frac{R_{rs}(708)}{R_{rs}(665)}\right) \times (a_{w708} + b_b) - a_{w665} - b_b^p\right]/a^* \qquad (13)$$

where water absorption coefficients are approximated as a_{w708} = 0.7 m^{-1}, a_{w665} = 0.4 m^{-1} (Pope & Fry, 1997), Chla specific absorption coefficient a^* = 0.016 m^2 g^{-1} , empirical constant p = 1.063 and b_b is related to R_{rs} at 778 nm by conversion factor;

$$b_b = 1.61 \times \pi R_{rs}(778)/(0.082 - 0.6\pi R_{rs}(778)) \qquad (14)$$

The algorithm may be recalibrated by adjusting a^* and p, denoted a and b respectively in eq. 15 for model parameterisation brevity to give;

$$Chla\_H = \left[\left(\frac{R_{rs}(708)}{R_{rs}(665)}\right) \times (0.7 + b_b) - 0.4 - b_b^a\right]/b \qquad (15)$$

Model I

Model I refers to the band index algorithm presented by Yang et al. (2010), which is based on a conceptual model (Gitelson et al., 2008) that adopts relevant wavebands according to their sensitivity to water absorption properties;

$$Index = \frac{\left(R_{rs}^{-1}(665) - R_{rs}^{-1}(708)\right)}{\left(R_{rs}^{-1}(753) - R_{rs}^{-1}(708)\right)} \qquad (16)$$

where it is assumed $R_{rs}(665)$ has maximum sensitivity to phytoplankton absorption, $R_{rs}(708)$ is insensitive to phytoplankton absorption but comparably sensitive to CDOM and $R_{rs}(753)$ is insensitive to phytoplankton and CDOM absorption and is mainly influenced by backscattering. Chla is estimated from a three-band index using a simple empirical formula;

$$Chla\_I = a \times Index + b \tag{17}$$

where coefficients a = 161.24 and b = 28.04 have been calibrated for lakes in Japan and China.

Model J

Model J refers to the normalized difference chlorophyll index (NDCI) proposed by Mishra et al. (2012). This uses a two-band difference ratio to predict Chla concentration in estuarine and coastal turbid waters;

$$Chla\_J = a + b \times \left(\frac{R_{rs}(708) - R_{rs}(665)}{R_{rs}(708) + R_{rs}(665)}\right) + c \times \left(\frac{R_{rs}(708) - R_{rs}(665)}{R_{rs}(708) + R_{rs}(665)}\right)^2 \tag{18}$$

where a = 42.197 , b = 236.5, c = 314.97. This version of the model has been calibrated using modelled R_{rs} spectra.

Model K

Model K refers to the normalized difference chlorophyll index (NDCI) proposed by Mishra et al.(2012) calibrated using field data collected from Chesapeake and Delaware Bay;

$$Chla\_K = a + b \times \left(\frac{R_{rs}(708) - R_{rs}(665)}{R_{rs}(708) + R_{rs}(665)}\right) + c \times \left(\frac{R_{rs}(708) - R_{rs}(665)}{R_{rs}(708) + R_{rs}(665)}\right)^2 \quad (19)$$

where *a* = 14.039, *b* = 86.115, *c* = 194.325.

Model L

Model L refers to the NASA OC 4-band ratio algorithm set at MERIS wavebands (OC4E) (O'Reilly et al., 2000) which relates log transformed Chla concentration to the maximum ratio of blue (443, 490, 510) to green (560) R_{rs};

$$Chla\_L = 10^{(a+bX+cX^2+dX^3+eX^4)} \quad (20)$$

where

$$X = log10(R_{rs}(443) > R_{rs}(490) > R_{rs}(510)/R_{rs}(560)) \quad (21)$$

and coefficients *a* = 0.3255, *b* = -2.7677, *c* = 2.4409, *d* = -1.1288, *e* = -0.4990 have been derived empirically from the NASA bio-Optical Marine Algorithm Data set (NOMAD) (Werdell et al., 2005).

Model M

Model M refers to a previous version of the NASA OC 3-band ratio algorithm set at MERIS wavebands (OC3E) which employs the maximum R_{rs} ratio of two blue wavebands (443, 490) and green (560) to determine Chla concentration;

$$Chla\_M = 10^{(a+bX+cX^2+dX^3+eX^4)} \tag{22}$$

where

$$X = log10(R_{rs}(443) > R_{rs}(490)/R_{rs}(560)) \tag{23}$$

and coefficients a = 0.2424, b = -2.2146, c = 1.5193, d = -0.7702, e = -0.4291 have been derived from the NOMAD dataset.

Model N

Model N refers to the earliest version of the NASA OC 2-band ratio algorithm set at MERIS wavebands (OC2E) where the ratio of blue (490) to green (560) R_{rs} is used to determine Chla concentration;

$$Chla\_N = 10^{(a+bX+cX^2+dX^3+eX^4)} \tag{24}$$

where

$$X = log10(R_{rs}(490)/R_{rs}(560)) \tag{25}$$

and coefficients *a* = 0.2389, *b* = -1.9369, *c* = 1.7627, *d* = -3.0777, *e* = -0.1054 have been derived from NOMAD.

Model O

Model O refers to the NASA fluorescence line height (FLH) algorithm presented by Gower et al. (1999). It produces an estimate of the magnitude of sun induced chlorophyll fluorescence (SICF) at 681 nm above a baseline interpolated between 665 and 708 nm;

$$FLH = R_{rs681} - \left[R_{rs}(708) + (R_{rs}(665) - R_{rs}(708)) \times \left(\frac{\lambda_{708} - \lambda_{681}}{\lambda_{708} - \lambda_{665}}\right)\right] \quad (26)$$

As the output of FLH is a difference in R_{rs}, the algorithm requires empirical calibration to convert to Chla concentration;

$$Chla\_O = a + b \times FLH \quad (27)$$

The operational range of FLH depends, among other factors, on the concentrations of optically active constituents present in the water column.

Model P

Model P refers to the maximum peak height (MPH) algorithm presented by Matthews et al., 2012. This is designed with a conditional peak position selector, which searches for the maximum radiance over three bands, as opposed to one fixed peak as seen in model N. The baseline is calculated over a larger spectral range, 664 to 885 nm, and the maximum peak intensity and position is determined from the maximum radiance measured at wavelengths 681, 709 or 753 nm. SICF is then estimated as follows;

$$MPH = brr_{max} - brr_{664} - \left[(brr_{885} - brr_{664}) \times \left(\frac{\lambda_{max}-\lambda_{664}}{\lambda_{885}-\lambda_{664}}\right)\right] \quad (28)$$

where brr_{max} and λ_{max} are magnitude and position of the greatest in magnitude Bottom of Rayleigh reflectance from bands 681, 709 or 753 nm. In this context, *brr* is assumed to be generally consistent with in-situ measured R_{rs}. Concentration of Chla was then determined in waters identified as non-cyanobacteria dominant;

$$Chla\_P = 5.24 \times 10^9 MPH^4 - 1.95 \times 10^8 MPH^3 + 2.46 \times 10^6 MPH^2 + 4.02 \times 10^3 MPH + 1.97 \quad (29)$$

Model P was not recalibrated in this study as it is based on *brr* and not R_{rs}.

Model Q

Model Q refers to the Garver-Siegel-Maritorena (GSM) semi-analytical inversion model that was developed by Garver and Siegel in 1997 and updated by Maritorena et al. (2002). It is based on an underlying quadratic relationship relating R_{rs} to the IOPs of the water body at a given wavelength (λ);

$$R_{rs}(\lambda) = \frac{t^2}{n_w^2} \sum_{i=1}^{2} g_i \left(\frac{b_b(\lambda)}{b_b(\lambda)+a(\lambda)}\right)^i \quad (30)$$

IOPs are partitioned into their contributing components where $b_b(\lambda) = b_{bw}(\lambda) + b_{bp}(\lambda)$ for water and SPM and $a(\lambda) = a_w(\lambda) + a_{ph}(\lambda) + a_{cdom}(\lambda)$ for water, phytoplankton and CDOM. The IOP spectra are then parameterized as a known shape with an unknown magnitude using the following expressions;

$$a_{ph}(\lambda) = Chla \times a^{*}_{ph}(\lambda), \quad (31)$$

$$a_{cdom}(\lambda) = a_{cdom}(\lambda) \times \exp(-S(\lambda - \lambda_0)), \quad (32)$$

$$b_{bp}(\lambda) = b_{bp}(\lambda_0) \times \left(\frac{\lambda_0}{\lambda}\right)^{Y} \quad (33)$$

Originally designed for SeaWiFS, the GSM model uses wavebands that overlap with available MERIS wavelengths. Inversion of the model produces simultaneous estimates of the unknown quantities of Chla, CDOM and b_{bp} from R_{rs} by application of a nonlinear least square optimisation routine. Global parameters, $a_w(\lambda)$, $b_{bw}(\lambda)$, n_w, t, and g_i were taken from the literature (Pope & Fry, 1997; Smith & Baker, 1981; Gordon et al., 1988), while a^*ph, S and Y were derived empirically from the SeaWiFs Bio-Optical Algorithm Mini-Workshop (SeaBAM) in-situ dataset.

Model R

Model R refers to the QAA method devised by Mishra et al. (2013 & 2014). This was developed primarily for the retrieval of cyanobacteria in turbid waters, however produces estimates of Chla as a routine by-product. As a first step, total absorption and particulate backscattering are estimated from subsurface R_{rs} (r_{rs}, sr^{-1}) at a given wavelength;

$$a(\lambda) = \frac{(1-u(\lambda))\left(b_{bw}(\lambda)+b_{bp}(\lambda)\right)}{u(\lambda)} \quad (34)$$

where

$$u(\lambda) = \frac{-g_0+\sqrt{(g_0)^2+4g_1\times rrs(\lambda)}}{2\times g_1} \quad (35)$$

and g_o = 0.089, g_1 = 0.125. The absorption signal is then decomposed into CDOM and phytoplankton components using known relations and empirical estimations;

$$a_{cdom}(\lambda) = a_{cdom}(443) \times \exp(-S(\lambda - 443)), \quad (36)$$

$$a_{ph}(\lambda) = a(\lambda) - a_w(\lambda) - a_{cdom}(\lambda) \quad (37)$$

The slope of CDOM, *S*, was derived empirically from samples collected from aquaculture ponds in Mississippi.

Model S

Model S refers to the artificial neural network (NN) model presented by Ioannou et al. (2013) which was developed to retrieve IOPs from R_{rs} at available MODIS (or similar satellite) wavelengths. This is based on a synthetic dataset of R_{rs}, where IOPS a_{ph}, a_{cdom}, b_{bp}

(and subsequently Chla) are computed directly from the R_{rs} signal. The model was trained for Chla concentrations ranging from 0.02 – 70 mg m^{-3}, and as such is only expected to perform within these conditions. Model S produces Chla as a standalone product (Model S) and Chla derived from IOPs (Model S2).

Table 2. Summary of validated models including their original Chla training range.

Model	Architectural approach	Chla training range (mgm^{-3})	Reference
Model A	NIR-red band ratio	0 - 70	Moses et al., 2009
Model B	NIR-red band ratio	0 - 70	Moses et al., 2009
Model C	NIR-red band ratio	2.3 - 200.8	Gurlin et al., 2011
Model D	NIR-red band ratio	2.3 - 200.8	Gurlin et al., 2011
Model E	Semi-analytical	0 - 80	Gilerson et al., 2010
Model F	NIR-red band ratio	0 - 1000	Gilerson et al., 2010
Model G	Semi-analytical	0 - 80	Gilerson et al., 2010
Model H	Semi-analytical	0 - 100	Gons et al., 2002
Model I	NIR-red band ratio	0 - 100	Yang et al., 2010
Model J	NIR-red band ratio	0 - 30	Mishra et al., 2012
Model K	NIR-red band ratio	0 - 30	Mishra et al., 2012
Model L	Blue-green band ratio	0.012 - 77	O'Reilly et al., 2000
Model M	Blue-green band ratio	0.012 – 77	O'Reilly et al., 2000

Model N	Blue-green band ratio	0.012 - 77	O’Reilly et al., 2000
Model O	Peak height	1 - 10	Gower et al., 1999
Model P	Peak height	0 - 350	Matthews et al., 2012
Model Q	Semi-analytical	0 - 100	Maritorena et al., 2002
Model R	Semi-analytical	59 - 1376	Mishra et al., 2013
Model S	Neural network	0.02 - 70	Ioannou et al., 2013

2.4 Model version denotations for algorithm calibration and validation

Models were denoted as ORG, CAL or CLUS according to the parameterisation of the model coefficients. In the first case, the ORG algorithm form represents the original published parameterisation of the algorithm. Here, model coefficients have been taken directly from the literature. In CAL form, model coefficients were reparametrized using the best-fit model for entire in-situ training data set. In CLUS form, model coefficients were determined for each OWT by sub-setting the training dataset into OWT groups before refitting the models using the subset data. Coefficients a, b, c, d and/or e correspond to those presented in models A to R (equations 1-33).

2.5 Analysis of performance

Standard statistical metrics were used to formally evaluate and describe the performance of selected Chla algorithms. These were combined as error metrics in a quantitative scoring system designed to objectively rank each algorithm according to the collective average performance (based on a modified version of the methodology proposed by Brewin et al., 2015). Points were assigned based on the median value calculated for each error metric whereby one point was awarded where an algorithm's error statistics were shown to be similar to the median error statistic for all models, and two and zero points were awarded where the calculated metrics were statistically better or worse respectively. The total number of metric points were then summed for each algorithm and performance rank was allocated based on the total point score. Consequently, a high score corresponds to the best performing models whilst comparatively low scores indicate poor model performance. To encourage a fairer representation of the validation data and limit bias towards the larger individual datasets contained within the combined validation dataset, a jack-knife routine was used to randomly subset 50 percent of the validation dataset 1000 times before calculating error metrics. In the case of OWT groups, a leave-*p*-out cross validation method was used to randomly subset data, where *p* was defined as 10 percent of the OWT grouped data. This produced a probability distribution of error statistics for each algorithm, from which the mean value was used to determine the final algorithm score. Metrics used as objective performance indicators are described in the following section along with the corresponding scoring criteria. All error metrics were applied to $\log_{10}$-transformed values of Chla concentration, which follows an approximate lognormal distribution (Campbell, 1995). Transformation to log-log space was aimed primarily to improve symmetry and heteroscedasticity of skewed regression residuals for statistically compliant metric calculations (in terms of residual distributions) and to reduce the influence of high concentration independent variable extremities (within an OWT) on metric results.

Root Mean Square Error

The absolute root mean square error (RMSE) was used to provide a general description of the difference between measured ($Chla_{meas}$) and predicted ($Chla_{mod}$) Chla concentration (units in mg m^{-3}) (Antoine et al., 2008). It is defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(log10Chla_{mod} - log10Chla_{meas})^2} \qquad (38)$$

where *N* is the number of model retrievals. The 95% confidence intervals for RMSE were also calculated to determine similarity between models. These were defined as statistically different where the confidence intervals did not overlap for two or more models. As such, the scoring system was defined as:

- 0 points awarded where RMSE is higher than median RMSE and 95% confidence levels do not overlap.
- 1 point awarded where RMSE 95% confidence levels overlap with median RMSE 95% confidence levels.
- 2 points awarded where RMSE is lower than median RMSE and median 95% confidence levels do not overlap.

Mean absolute error

The mean absolute error (MAE) was used in this study to quantify the difference between the modelled and measured Chla variables (Willmot et al., 2005 and Seegers et al., 2018). MAE was calculated using the following expression;

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|log10Chla_{mod} - log10Chla_{meas}| \quad (39)$$

where *N* is the number of model retrievals. The 95% confidence intervals for MAE were used to determine similarity between models which were defined as statistically different where confidence intervals did not overlap for two or more models. As such, the scoring system was defined as:

- 0 points awarded where MAE is higher than median MAE and 95% confidence levels do not overlap.
- 1 point awarded where MAE 95% confidence levels overlap with median MAE 95% confidence levels.
- 2 points awarded where MAE is lower than median MAE and median 95% confidence levels do not overlap.

Slope and intercept of type-II linear regression

Least squares linear regression was used to calculate the slope (*m*) and intercept (*c*) of a best fit line plotted between $Chla_{mod}$ and $Chla_{meas}$ (units in mg m^{-3}). Type II regression was used to account for uncertainty in the in-situ data by calculating the perpendicular offsets between $Chla_{meas}$ and the linear fit:

$$log10Chla_{mod} = m \times log10Chla_{meas} + c \quad (40)$$

and assumes that residuals are normally distributed. The scoring system for *m* and *c* was based on the median and standard deviation calculated for each parameter individually such that:

- 0 points awarded where the standard deviation of *m* is greater than median standard deviation of m for all models and *m* ± its standard deviation does not overlap with 1 ± two times the median standard deviation of m for all models.
- 1 point awarded where the standard deviation of *m* is less than median standard deviation of m for all models or *m* ± its standard deviation overlaps with 1 ± two times the median standard deviation of m for all models.
- 2 points awarded where the standard deviation of *m* is less than median standard deviation of m for all models and *m* ± its standard deviation overlaps with 1 ± two times the median standard deviation of m for all models.

- 0 points awarded for a particular model where the standard deviation for *c* is greater than median standard deviation of *c* for all models and c ± its standard deviation does not overlap with zero ± two times the median standard deviation of *c*.
- 1 point awarded where the standard deviation of c for a particular model is less than the standard deviation of c for all model or c ± its standard deviation overlaps with zero ± two times the median standard deviation of c for all models.
- 2 points awarded where the standard deviation of c for a particular model is less than the standard deviation of c for all model and c ± its standard deviation overlaps with zero ± two times the median standard deviation of c for all models.

Pearson's correlation coefficient

The Pearson's correlation coefficient *r*, is a useful statistic for determining the strength of a linear relationship between measured and predicted variables (Doney et al., 2009) In this study, *r* was used in combination with the z_{score} to determine if a model value of *r* was statistically higher or lower than the mean *r*-value for all models. z_{score} was calculated using the following expression;

$$z_{score} = \frac{z_{mod} - z_{mean}}{\{[1/(N_{mod}-3)] + [1/(N_{mean}-3)]\}^{1/2}} \quad (41)$$

where

$$z_{mod} = 0.5\log\left(\frac{1+r_{mod}}{1-r_{mod}}\right) \quad (42)$$

$$z_{mean} = 0.5\log\left(\frac{1+r_{mean}}{1-r_{mean}}\right) \quad (43)$$

and r_{mod} is the model *r*-value, r_{mean} is the mean *r*-value for all models, N_{mod} and N_{mean} are the number of model retrievals and the mean number of retrievals for all models respectively. z_{score} was converted to a *p*-value assuming a normal probability distribution and statistical difference was defined where p-value < 0.05. The scoring system for *r* was then based on the determined *p*-value and the location of the model *r*-value in relation to mean *r* such that;

- 0 points were awarded where *r* is lower than mean *r* and is statistically different.
- 1 point awarded where model *r* and mean *r* were statistically similar.

- 2 points awarded where model *r* was statistically higher than the mean *r*-value for all models.

Average absolute percent difference

Uncertainty between modelled and measured variables (Antoine et al., 2008) was determined using the average absolute (unsigned) relative percent difference (RPD) defined as;

$$RPD = 100 \times \frac{1}{N}\sum_{i=1}^{i=N}\left(\frac{|Chla_{mod} - Chla_{meas}|}{Chla_{meas}}\right) \quad (44)$$

The scoring system for RPD was again based on a mean value of RPD calculated across all algorithms with the inclusion of the 95% confidence interval. This accounts for lower confidence in retrieved estimates where a low value of RPD is observed. As such, the RPD scoring classification was defined as;

- 0 points awarded where RPD for a particular model is greater than mean RPD and RPD ± its 95% confidence interval does not overlap with mean 95% confidence interval for all models.
- 1 point awarded where RPD ± its 95% confidence interval overlaps with the mean 95% confidence interval for all models.
- 2 points awarded where RPD for a particular model is less than mean RPD and RPD ± its 95% confidence interval does not overlap with mean 95% confidence interval for all models.

Bias

Calculation of the bias was used to assess the likelihood of systematic errors in algorithm outputs (units in mg m^{-3}) (Seegers et al., 2018);

$$bias = 100 \times \frac{1}{N}\sum_{i=1}^{i=N}(log10Chla_{mod} - log10Chla_{meas}) \quad (45)$$

A value close to zero indicates the algorithm corresponds well with in-situ measurements. As such, the bias scoring system was defined as follows;

- 0 points awarded where the bias confidence interval for a particular model is greater than median bias ± its 95% confidence interval for all models plus the model bias confidence interval does overlap with zero ± median confidence interval.
- 1 point awarded where the model bias confidence interval overlaps with median bias ± its 95% confidence interval or the model bias overlaps with zero ± median confidence interval for all models.
- 2 points awarded where the model bias confidence interval overlaps with median bias ± its 95% confidence interval and the model bias overlaps with zero ± median confidence interval for all models.

Percentage of retrievals

The percentage of possible retrievals (*%n*) was included as a statistical indicator to assess an algorithm's capability of producing global estimates of Chla and not, therefore, contributing to data gaps. This was calculated as follows;

$$\%n = \frac{N_{mod}}{N_{meas}} \quad (46)$$

where N_{mod} is the number of algorithm retrievals and N_{meas} is the number of in-situ measurements. The scoring system for %*n* was based on the average number of retrievals for all algorithms such that;

- 0 points awarded where %*n* is less than mean %*n* for all models.
- 1 point awarded where %*n* is greater than mean %*n* for all models but less than 99%.
- 2 points awarded where %*n* is greater than 99%

3. Results

Error metrics were determined for two arrangements of the validation data. In the first case, objective performance scores were calculated per model for the entire R_{rs} dataset converted to Chla in ORG, CAL and CLUS algorithm forms (2807 sample points). In the case of the CLUS form, coefficients derived for an OWT group subset were used to estimate Chla from corresponding OWT group spectra. All subsets were then recombined (number of rows equivalent to ORG and CAL outputs) to calculate error metrics on the entire validation dataset. In the second validation arrangement, Chla concentrations derived from ORG, CAL and CLUS algorithm forms were subset into groups defined by their assigned OWT and performance scores were calculated for each model within the OWT subset group.

3.1 Full dataset comparison

Figure 4 shows a quantitative comparison of Chla generated from each of the examined models against the in-situ measurements. Corresponding error metrics are presented in Figure 5. Scatterplots in Figure 4 demonstrate the high variability of algorithm performance generated across the range of tested models. Several algorithms are shown to perform poorly, some are simply unable to retrieve Chla at the concentrations observed. Apparent failures occur with three-band Models B, D and G which may be attributable to the elevated values of $R_{rs}(708)$ leading to negative estimates of the independent ratio variable. Nonetheless, several models perform reasonably well in terms of the accuracy of the Chla retrieval when considering the significant range of constituent concentrations included in the validation dataset (Figure 1). Most notably, empirical Models A, C and J produce *r*-values in excess of 0.85 and regression slopes close to 1 when compared to in-situ measurements. For all models, error residuals are heteroscedastic and vary as a function of Chla concentration, with the most obvious spread of data observed at low concentrations of Chla. This suggests a targeted water type specific algorithm could improve performance across the Chla concentration continuum and is further implied by the notable differences in performance produced by ORG, CAL and CLUS algorithm forms. In almost every case, the CLUS model form produced more accurate estimates of Chla, as demonstrated in Models E, F, I and R. For some models, reparametrizing model coefficients with the entire training dataset (CAL version) causes algorithm performance to degrade, as is the case with Models B, D and G. No obvious differences in overall algorithm performance were observed between empirical and semi-analytical model architectural approaches (Table 2 for architectural summary). Ignoring OWT classification, i.e. ignoring CLUS model forms, the most accurate retrievals of Chla were obtained using Models A_ORG, C_ORG, H_ORG and J_CAL, which each estimate log transformed Chla with a MAE of less than 0.27 mg m^{-3}.

735

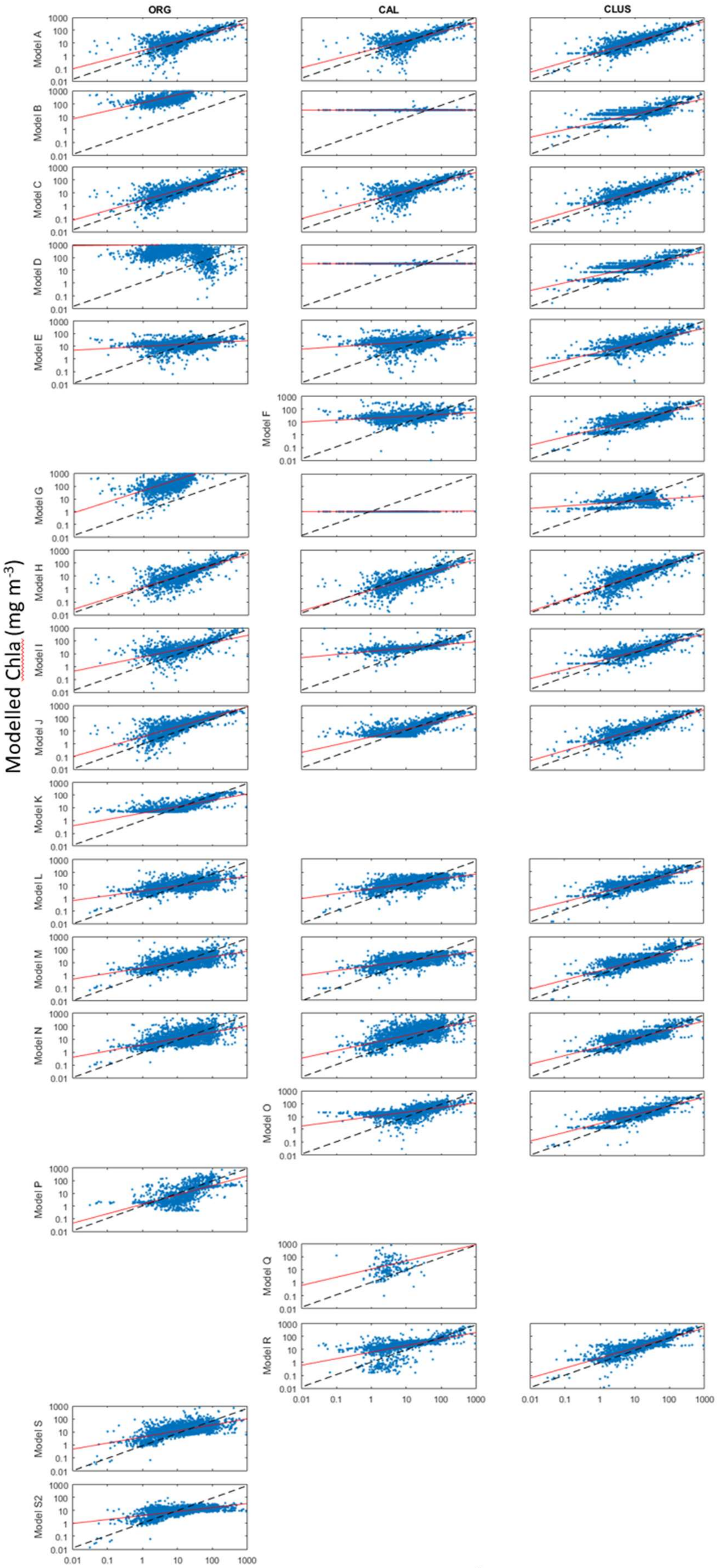

ORG
CAL
CLUS
Model A
Model B
Model C
Model D
Model E
Model F
Model G
Model H
Model I
Model J
Model K
Model L
Model M
Model N
Model O
Model P
Model Q
Model R
Model S
Model S2
0.01
0.1
1
10
100
1000
Modelled Chla (mg m-3)
Measured Chla (mg m-3)

Figure 4. Comparison of model derived and in-situ measured Chla concentrations. The 1:1 relationship between measured and modelled Chla is represented by a dashed line.

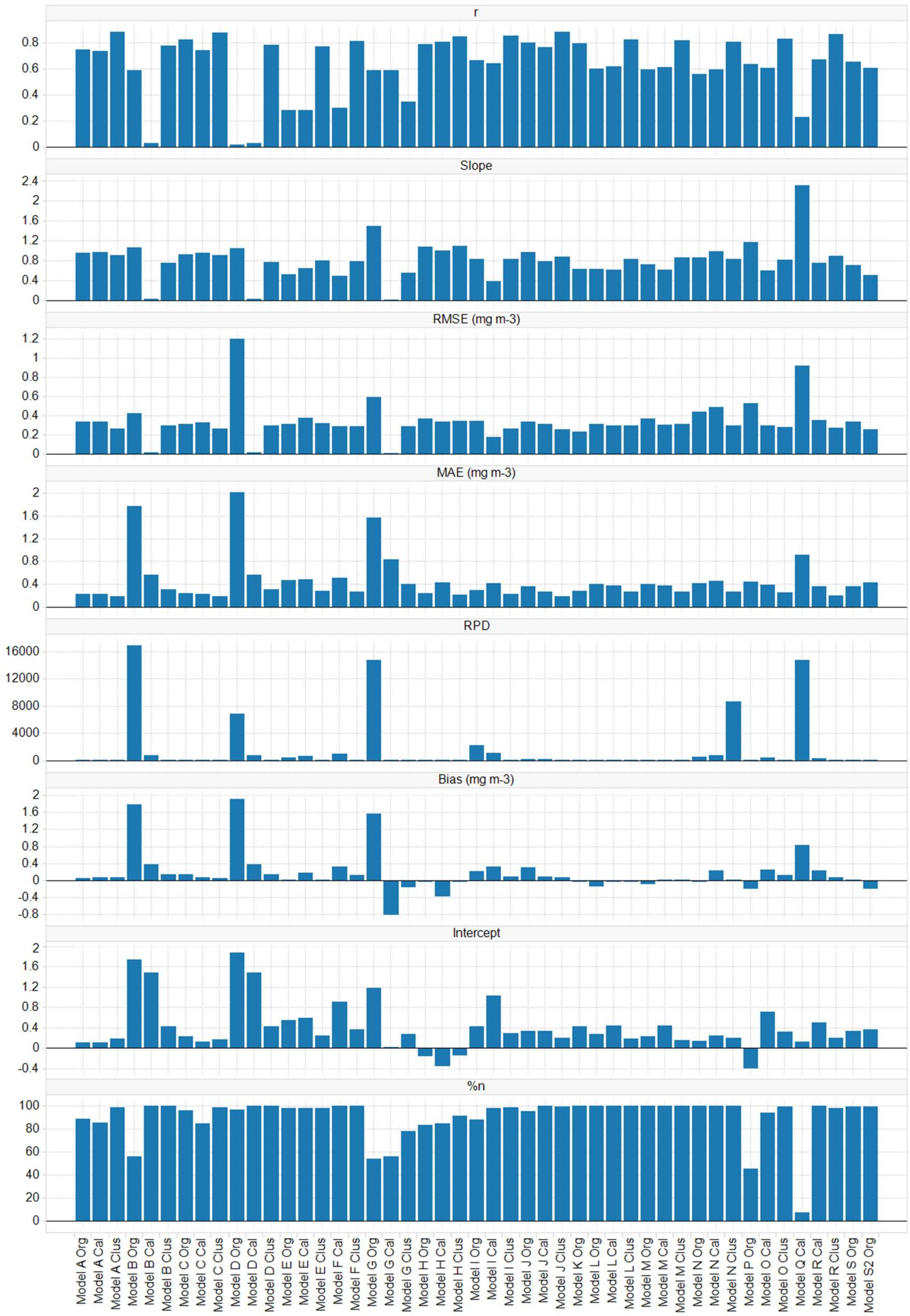


Figure 5. Error metrics calculated when comparing model outputs of Chla with in-situ Chla concentrations.

The corresponding performance scores as determined by the objective scoring system are shown in Figure 6. These are ranked according to total error score, with a maximum score denoting the best performing model of those tested for producing accurate estimates of Chla concentration. Results are consistent with conclusions inferred from the scatter plots presented in Figure 4, indicating the objective scoring system is capable of accurately classifying algorithm performance. The highest scoring algorithms are Models A, C, J, L,M and R which translate to 3 red-NIR band ratio based algorithms, 2 blue-green ratio based algorithms and a QAA model. Error statistics for the top-ranked models are shown in table 3. In almost every case, the CLUS version of the algorithm, coloured by light blue on Figure 5, produced a greater score when compared to ORG and CAL counterparts (dark and mid blue respectively). With a total error score of 14, these are the best performing algorithms when comparing modelled and measured Chla for the full validation dataset. Those models exhibiting high discrepancies between modelled and measured Chla are represented with low scores, such as ORG and CAL versions of Models B, D, E and F. The objective scoring systems also identifies the apparent poorly performing algorithms of Models G and Q which produce a zero score in one or more algorithm form. Again, it is shown that recalibrating a model with the entire calibration dataset (CAL form) using a best-fit approach does not always improve model performance. The Chla constituent range of the training dataset may be too large to effectively calibrate the evaluated models (as shown in Figure 1) and as such produce a detrimental effect on error metrics. This is particularly obvious in the MAE calculated for Models A, H and N, where an error increase of approximately 2%, 83% and

11% respectively is observed when comparing ORG and CAL outputs. Nonetheless, significant improvement in terms of performance score is achieved when converting to the CLUS algorithm form for low scoring models.

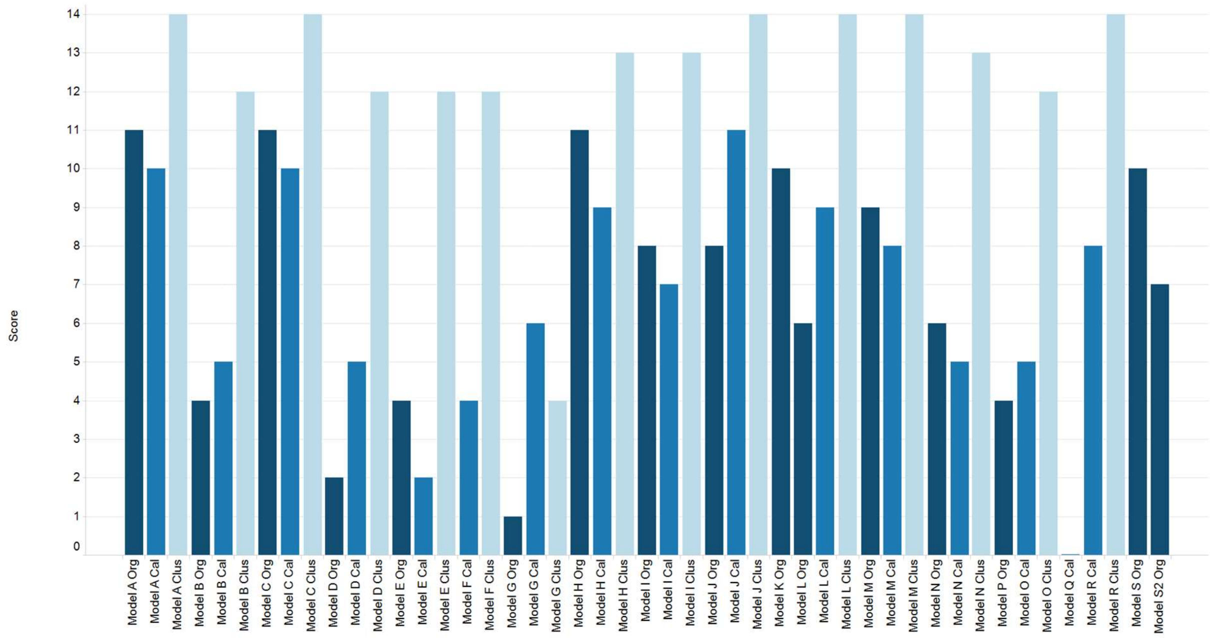


Figure 6. Performance scores generated from Chla single model comparisons using the complete validation dataset. Colours represent form of model, i.e. dark-blue is original form (ORG), mid-blue is full dataset calibrated form (CAL) and light-blue is OWT calibrated form (CLUS).

Table 3. Error statistics generated when comparing modelled log_{10} Chla with in-situ measurements for each of the first ranked models, ordered by mean absolute error.

Model	*r*	Slope	RMSE (mgm^{-3})	MAE (mgm^{-3})	RPD	Bias ($mg\ m^{-3}$)	Intercept	%n
C_CLUS	0.885	0.914	0.256	0.188	79.49	0.057	0.156	98.82
J_CLUS	0.888	0.885	0.248	0.189	81.92	0.066	0.200	99.46
A_CLUS	0.883	0.909	0.260	0.191	85.20	0.068	0.173	98.64
R_CLUS	0.872	0.889	0.267	0.205	91.48	0.079	0.206	98.21
L_CLUS	0.825	0.839	0.291	0.261	91.04	0.001	0.184	100

M_CLUS	0.823	0.875	0.306	0.266	96.16	0.015	0.157	100

3.2 Performance per OWT

The second stage of the algorithm validation focussed on performance within a specific OWT group. Chla concentrations generated using the OWT training subsets (CLUS parameterisations) were compared to outputs from CAL and ORG models for only the corresponding OWT assigned spectra. For example, 425 of 2807 spectra were assigned as OWT 2 and error metrics were calculated for these 425 points in ORG, CAL and CLUS versions to determine algorithm performance within OWT 2. Results from objective scoring are shown in Figure 7. Each model/OWT combination was assigned a performance score based on the median value calculated for a metric within an OWT and as such, scores are independent of OWT group. Algorithm performance is highly variable across the tested models, with scores ranging from zero to 13 or 14 in several of the OWTs. Several models are shown to perform reasonably well across several OWTs, for example, Model J displays a relatively high score (jointly ranked first) in OWT 2, 4, 5, 6, 11 and 12. Conversely, Models D, G and O perform poorly in every OWT. One noticeable difference when generating error statistics based on OWT subsets as opposed to the entire validation dataset is the performance per algorithm version. We now have several cases where the ORG or CAL version of a model produces more accurate estimates of Chla when compared to those derived from the refined OWT CLUS reparametrisation. For example, the CAL version of Model J was found to be a leading candidate model in three OWT groups. This result may be a consequence of unsuitable model parametrisation in under-sampled OWTs with comparatively small training datasets, i.e. OWTs 1, 7, 10 and 13.

Corresponding error and regression statistics for the maximum OWT model scores are shown in Table 4. It is clear that significant variability in performance is observed across each water type, even for maximum scoring algorithms. In five of the 13 OWTs, one or more models produce a correlation coefficient between measured and modelled Chla which is greater than 0.7. This indicates only a proportion of the validation dataset is sufficiently characterised by the algorithms tested; these are OWTs 2, 4, 8, 9 and 12, and collectively they comprise 58.4% of the total validation dataset (1639 spectra from 2807). These OWTs mainly lie within the mid-range of the Chla concentration distribution, with median values of Chla per OWT ranging from 4.2 mg m^{-3} to 102 mg m^{-3}. The higher section of the Chla concentration range (OWTs 7, 1, 8 and 6) is retrieved reasonably well with model outputs producing *r*-values in excess of 0.5 for each OWT. The lower section of the Chla range is shown to be the most challenging, with maximum *r*-values for OWTs 3 and 13 calculated as 0.372 and 0.595 respectively. In these waters where median concentrations of Chla are less than 1.5 mg m^{-3}, a number of validation sample points lie outside the original training range for the tested models which may influence the derived error metrics.

Optical Water Type	1	2	3	4	5	6	7	8	9	10	11	12	13
Model A Org	9	10	4	8	8	9	9	13	4	8	9	10	4
Model A Cal	9	8	4	8	8	8	9	13	4	8	9	10	5
Model A Clus	9	14	9	9	8	10	9	11	8	6	12	11	10
Model B Org	1	5	6	3	2	1	1	1	5	1	4	3	5
Model B Cal	7	4	6	4	8	7	6	5	5	8	6	5	7
Model B Clus	8	6	9	6	7	7	9	9	6	8	7	5	11
Model C Org	8	11	6	10	8	8	9	10	9	8	9	10	6
Model C Cal	9	8	3	9	8	8	9	13	4	8	9	10	1
Model C Clus	11	13	9	10	10	10	9	11	8	8	11	10	10
Model D Org	7	4	5	2	3	2	3	2	5	4	4	3	6
Model D Cal	7	4	6	4	8	7	6	5	4	8	7	5	7
Model D Clus	9	6	9	6	8	6	8	9	7	8	8	5	9
Model E Org	9	8	6	7	8	6	7	5	6	4	7	6	6
Model E Cal	8	7	6	6	6	7	7	5	6	3	7	5	6
Model E Clus	9	8	9	8	8	7	10	7	8	7	8	6	9
Model F Cal	10	6	6	8	7	7	8	6	6	5	8	4	6
Model F Clus	9	8	9	8	7	7	9	8	8	8	8	5	10
Model G Org	1	4	4	2	2	1	1	1	4	1	4	2	6
Model G Cal	1	6	8	5	5	1	1	1	7	1	7	5	9
Model G Clus	1	6	9	5	7	1	1	1	8	8	8	5	9
Model H Org	4	9	6	10	4	9	9	12	6	9	12	9	7
Model H Cal	8	9	6	7	4	7	9	8	6	6	9	7	7
Model H Clus	4	13	5	8	6	10	10	14	8	7	11	8	6
Model I Org	1	6	2	7	5	8	4	10	6	5	9	10	1
Model I Cal	1	6	4	7	7	8	5	10	4	7	8	9	3
Model I Clus	8	7	9	8	8	10	8	10	6	6	11	9	9
Model J Org	8	10	5	9	6	6	9	7	7	7	8	8	6
Model J Cal	9	9	5	11	10	8	9	11	7	8	11	11	7
Model J Clus	9	14	9	10	10	10	9	11	9	8	14	10	10
Model K Org	9	9	5	10	10	8	8	11	8	8	12	10	7
Model L Org	8	7	8	7	6	5	5	5	9	6	9	7	10
Model L Cal	8	6	7	7	7	5	7	6	8	8	8	8	12
Model L Clus	11	8	12	9	9	9	9	9	11	10	10	9	11
Model M Org	8	6	8	6	7	5	5	4	8	8	7	6	10
Model M Cal	8	6	7	7	8	5	7	6	8	8	7	6	10
Model M Clus	11	8	12	9	9	9	9	7	11	10	10	8	12
Model N Org	7	5	8	6	8	4	4	3	8	4	7	7	10
Model N Cal	8	4	8	5	7	2	5	4	7	4	6	6	9
Model N Clus	11	8	12	9	9	9	9	8	11	10	10	9	1
Model O Cal	9	4	4	7	7	7	7	8	2	7	8	6	7
Model O Clus	8	6	9	7	7	7	8	8	6	7	8	6	10
Model P Org	5	3	5	4	5	3	5	3	4	8	6	4	7
Model Q Cal	1	1	7	1	2	4	1	1	4	1	2	2	1
Model R Cal	9	8	5	8	7	10	11	9	6	6	9	9	5
Model R Clus	11	13	9	10	7	8	9	10	8	7	9	10	10
Model S Org	8	7	10	7	6	7	5	5	8	6	6	5	12
Model S2 Org	6	6	9	7	6	4	8	5	8	6	7	5	12

Figure 7. Performance scores for models tested within each OWT determined from objective scoring. A high score indicates better performance relative to all tested models within an OWT. Highest ranking scores (i.e. joint ranked first or second) have been coloured blue.

Table 4. Error statistics generated when comparing modelled Chla with in-situ measurements for each of the first and second jointly ranked models in an OWT.

Cluster	Model	*r*	Slope	RMSE (mgm^{-3})	MAE (mgm^{-3})	RPD	Bias (mg m^{-3})	Intercept	%n
1	C_CLUS	0.629	0.757	0.205	0.167	67.65	-0.054	0.593	100
1	F_CAL	0.443	0.495	0.153	0.268	52.49	0.191	0.931	100
1	L_CLUS	0.596	0.575	0.159	0.213	53.49	0.009	0.936	100
1	M_CLUS	0.621	0.738	0.200	0.246	75.19	-0.115	0.698	100
1	N_CLUS	0.632	0.738	0.199	0.245	75.03	-0.116	0.699	100
1	R_CLUS	0.625	0.612	0.165	0.175	66.42	-0.080	0.942	100
2	A_CLUS	0.816	0.958	0.211	0.158	50.36	-0.029	0.072	96.71
2	C_CLUS	0.789	0.991	0.239	0.165	52.27	-0.028	0.036	98.58
2	H_CLUS	0.813	0.978	0.218	0.160	49.13	-0.014	0.036	96.70
2	J_CLUS	0.782	0.948	0.235	0.160	52.08	-0.030	0.082	100
2	R_CLUS	0.774	1.035	0.258	0.171	53.39	-0.023	-0.012	98.35
3	L_CLUS	0.268	0.270	0.084	0.230	80.96	-0.001	0.079	100
3	M_CLUS	0.352	0.450	0.136	0.219	80.13	-0.029	0.088	100
3	N_CLUS	0.372	0.375	0.112	0.217	75.15	0.000	0.067	100
4	C_ORG	0.777	0.682	0.175	0.207	109.9	-0.157	0.510	100
4	C_CLUS	0.705	0.891	0.254	0.194	82.75	-0.047	0.170	98.23
4	H_ORG	0.742	0.976	0.256	0.198	70.05	0.034	-0.007	95.45
4	J_CAL	0.790	0.852	0.213	0.187	62.13	0.018	0.146	100
4	J_CLUS	0.671	0.806	0.243	0.207	83.75	-0.047	0.263	100
4	K_ORG	0.782	0.617	0.157	0.221	67.40	0.074	0.351	100
4	R_CLUS	0.772	0.845	0.211	0.171	66.94	-0.045	0.222	94.70
5	C_CLUS	0.442	0.574	0.196	0.239	134.2	-0.092	0.582	99.58
5	J_CAL	0.474	0.422	0.141	0.246	137.4	-0.118	0.783	100
5	J_CLUS	0.450	0.544	0.184	0.239	131.2	-0.093	0.618	100
5	K_ORG	0.477	0.335	0.111	0.244	103.4	-0.002	0.767	100
5	L_CLUS	0.286	0.257	0.094	0.268	116.8	-0.030	0.885	100
5	M_CLUS	0.336	0.315	0.112	0.262	113.6	-0.030	0.819	100
5	N_CLUS	0.352	0.331	0.118	0.261	111.7	-0.030	0.799	100
6	A_ORG	0.462	0.564	0.122	0.127	57.71	-0.058	0.770	100
6	A_CLUS	0.461	0.564	0.123	0.128	55.41	-0.035	0.749	100
6	C_CLUS	0.466	0.520	0.113	0.126	54.41	-0.037	0.823	100
6	H_ORG	0.476	0.633	0.136	0.136	51.94	-0.017	0.617	100
6	H_CLUS	0.474	0.589	0.127	0.134	52.70	-0.030	0.702	100
6	I_CLUS	0.535	0.598	0.123	0.127	49.89	-0.033	0.691	100
6	J_CLUS	0.463	0.526	0.114	0.126	54.67	-0.037	0.812	100
6	L_CLUS	0.205	0.202	0.048	0.171	53.75	-0.003	1.308	100
6	M_CLUS	0.160	0.160	0.039	0.175	54.25	0.002	1.372	100
6	N_CLUS	0.162	0.161	0.039	0.175	54.58	0.000	1.373	100
6	R_CAL	0.521	0.360	0.075	0.134	47.53	-0.014	1.061	100
7	E_CLUS	0.256	0.631	0.192	0.252	79.22	0.020	0.862	100
7	H_CLUS	0.563	0.642	0.141	0.152	43.89	4.455	0.866	97.73
7	R_CAL	0.608	0.759	0.191	0.178	42.79	0.040	0.537	100
8	A_ORG	0.663	0.841	0.129	0.119	26.86	0.038	0.278	100
8	A_CAL	0.667	0.856	0.131	0.118	29.89	-0.011	0.300	100

8	C_CAL	0.668	0.850	0.130	0.118	30.32	-0.017	0.316	100
8	H_CLUS	0.712	0.891	0.129	0.108	26.62	0.006	0.210	100
9	L_CLUS	0.646	0.644	0.192	0.221	71.56	0.001	0.245	100
9	M_CLUS	0.708	0.704	0.194	0.212	61.34	0.002	0.203	100
9	N_CLUS	0.713	0.714	0.200	0.211	61.61	-0.001	0.199	100
10	H_ORG	0.539	0.316	0.166	0.720	989.1	-0.716	1.309	90.16
10	L_CLUS	0.495	0.489	0.256	0.433	152.7	0.001	0.441	100
10	M_CLUS	0.382	0.380	0.210	0.458	169.1	0.001	0.534	100
10	N_CLUS	0.509	0.504	0.259	0.417	161.2	-0.004	0.431	100
11	J_CLUS	0.697	0.910	0.301	0.233	75.12	-0.043	0.147	100
12	A_ORG	0.794	0.686	0.180	0.177	60.74	-0.005	0.429	99.70
12	A_CAL	0.796	0.730	0.188	0.169	62.00	-0.026	0.391	99.10
12	A_CLUS	0.778	0.824	0.221	0.175	68.34	-0.059	0.300	99.10
12	C_ORG	0.801	0.635	0.165	0.183	72.11	-0.065	0.557	100
12	C_CAL	0.792	0.764	0.199	0.168	61.39	-0.022	0.342	99.10
12	C_CLUS	0.763	0.911	0.249	0.182	68.42	-0.046	0.166	98.49
12	I_ORG	0.700	0.969	0.270	0.191	66.61	-0.048	0.092	92.17
12	J_CAL	0.796	0.856	0.225	0.190	54.87	0.077	0.118	100
12	J_CLUS	0.768	0.777	0.213	0.184	71.74	-0.074	0.373	100
12	K_ORG	0.788	0.648	0.174	0.264	58.24	0.175	0.297	100
12	R_CLUS	0.792	0.602	0.160	0.204	83.87	-0.107	0.642	100
13	B_CLUS	0.262	0.404	0.233	0.620	328.8	-0.431	-0.190	100
13	L_CAL	0.528	0.381	0.193	0.320	71.72	-0.020	-0.622	100
13	L_CLUS	0.595	1.238	0.594	0.581	72.72	0.436	-0.190	100
13	M_CLUS	0.595	1.238	0.594	0.581	72.72	0.436	-0.190	100
13	S_ORG	0.272	0.381	0.219	0.343	53.30	0.220	-0.862	100
13	S2_ORG	0.500	0.432	0.223	0.398	48.34	0.362	-0.951	100

3.3 Recommendation for a dynamic OWT switching algorithm

Based on objective scoring and individual error statistics, the recommended algorithm selection for inland waters exhibiting water-leaving reflectance characteristics similar to those described by Spyrakos et al., (2018b) is shown in table 5. The confirmed choice of models over the OWT range is varied and complex. Of the 13 groups, eight models identified as high performers in terms of Chla retrieval appear in their CLUS form, where the re-calibration of the model was based on the OWT group subset data. Two OWTs are more accurately characterised in terms of Chla retrieval by their original published algorithms (ORG) and three OWTs by parameterising the proposed models using the complete training dataset (CAL). This observation is contrary to results presented in the full dataset comparison, where the performance of almost every model was improved by switching to CLUS algorithm form, and may be a consequence of the variation in the number of observations per OWT which in turn affects the retuning of the algorithm. Furthermore, the process of defining OWTs will inevitably generate extremes where no algorithm will perform satisfactorily. A general pattern relating to the architecture of the best performing models per OWT is also evident, whereby OWTs consisting of predominantly low concentrations of Chla are better characterised by the blue-green ratio Models M and N, which were developed for low chlorophyll, open ocean conditions. Conversely, Model R, which was originally developed for regions of mid to high Chla concentration, has been identified as a leading candidate for eutrophic OWTs 6 and 7. Model H was the leading performer for OWT 10 where concentrations of Chla are often high but the optical signal is primarily dominated by the presence of other optically active constituents. The remaining mid-range Chla concentrations have been captured by several versions of the red-NIR band ratio algorithm.

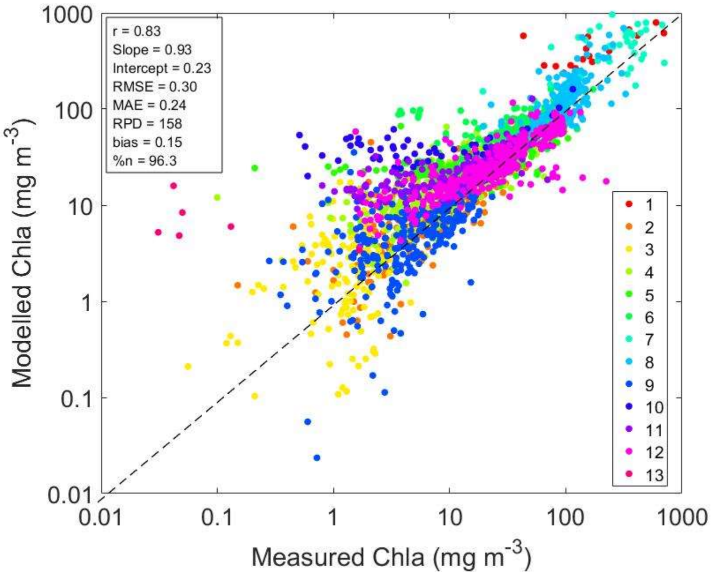
Table 5. Recommended model for each defined OWT (Spyrakos et al., 2018b) ordered by OWT group median Chla concentration (Figure 3). Calibration coefficients for each model have been highlighted in bold.

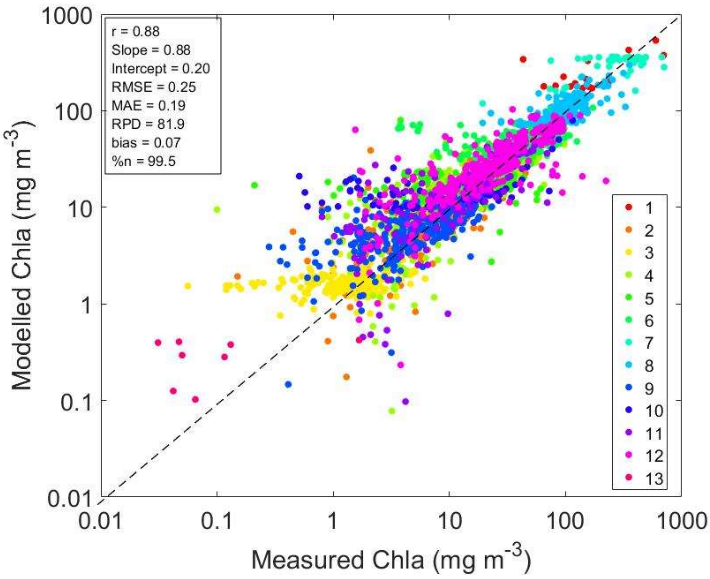
OWT	Model	Architectural approach	Equation	*a*	*b*	*c*	*d*	*e*
7	R_CAL	Semi-analytical	$a_{ph}(\lambda) = \boldsymbol{a}(\lambda) - a_w(\lambda) - a_{cdom}(\lambda)$	0.0135				

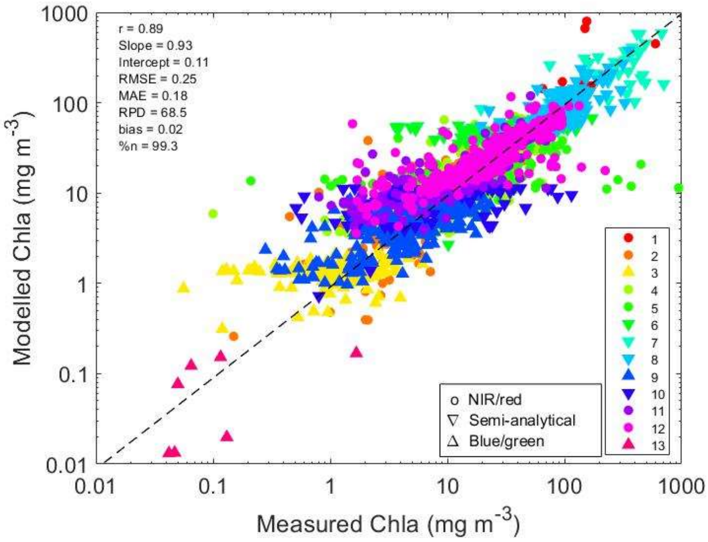
			$a_{cdom}(\lambda) = a_{cdom}(443)\exp(-\boldsymbol{a}(\lambda - 443))$					
1	C_CLUS	NIR-red band ratio	$Chla\_C = \boldsymbol{a} \times \left(\frac{R_{rs708}}{R_{rs665}}\right)^2 + \boldsymbol{b} \times \left(\frac{R_{rs708}}{R_{rs665}}\right) + \boldsymbol{c}$	86.09	-517.5	886.7		
8	H_CLUS	Semi-analytical	$Chl\_H = \left[\left(\frac{R_{rs708}}{R_{rs665}}\right) \times (0.7 + b_b) - 0.4 - b_b^{\boldsymbol{a}}\right] / \boldsymbol{b}$	1.25	0.0174			
6	R_CAL	Semi-analytical	$a_{ph}(\lambda) = \boldsymbol{a}(\lambda) - a_w(\lambda) - a_{cdom}(\lambda)$ $a_{cdom}(\lambda) = a_{cdom}(443)\exp(-\boldsymbol{a}(\lambda - 443))$	0.0135				
12	A_CLUS	NIR-red band ratio	$Chla\_A = \boldsymbol{a} \times \left(\frac{R_{rs708}}{R_{rs665}}\right) + \boldsymbol{b}$	80.7	53.18			
11	J_CLUS	NIR-red band ratio	$Chla\_J = \boldsymbol{a} + \boldsymbol{b} \times \left(\frac{R_{rs708} - R_{rs665}}{R_{rs708} + R_{rs665}}\right) + \boldsymbol{c} \times \left(\frac{R_{rs708} - R_{rs665}}{R_{rs708} + R_{rs665}}\right)^2$	19.31	153.5	105.4		

4	J_CAL	NIR-red band ratio	$Chla\_J = \boldsymbol{a} + \boldsymbol{b} \times \left(\frac{R_{rs708} - R_{rs665}}{R_{rs708} + R_{rs665}}\right) + \boldsymbol{c} \times \left(\frac{R_{rs708} - R_{rs665}}{R_{rs708} + R_{rs665}}\right)^2$	18.44	149.2	374.9		
5	K_ORG	NIR-red band ratio	$Chla\_K = \boldsymbol{a} + \boldsymbol{b} \times \left(\frac{R_{rs708} - R_{rs665}}{R_{rs708} + R_{rs665}}\right) + \boldsymbol{c} \times \left(\frac{R_{rs708} - R_{rs665}}{R_{rs708} + R_{rs665}}\right)^2$	14.039	86.115	194.33		
2	A_CLUS	NIR-red band ratio	$Chla\_A = \boldsymbol{a} \times \left(\frac{R_{rs708}}{R_{rs665}}\right) + \boldsymbol{b}$	53.29	-30.08			
10	H_ORG	Semi-analytical	$Chla\_H = \left[\left(\frac{R_{rs708}}{R_{rs665}}\right) \times (0.7 + b_b) - 0.4 - b_b^{\boldsymbol{a}}\right] / \boldsymbol{b}$	1.063	0.016			
9	N_CLUS	Blue-green band ratio	$Chla\_N = 10^{(\boldsymbol{a}+\boldsymbol{b}X+\boldsymbol{c}X^2+\boldsymbol{d}X^3+\boldsymbol{e}X^4)}$ $X = log10(R_{rs490}/R_{rs560})$	0.0536	7.308	116.2	412.4	463.5
3	M_CLUS	Blue-green band ratio	$Chla\_M = 10^{(\boldsymbol{a}+\boldsymbol{b}X+\boldsymbol{c}X^2+\boldsymbol{d}X^3+\boldsymbol{e}X^4)}$ $X = log10(R_{rs443} > R_{rs490}/R_{rs560})$	0.1098	-0.755	-14.12	-117	-17.76
13	M_CLUS	Blue-green band ratio	$Chla\_M = 10^{(\boldsymbol{a}+\boldsymbol{b}X+\boldsymbol{c}X^2+\boldsymbol{d}X^3+\boldsymbol{e}X^4)}$ $X = log10(R_{rs490}/R_{rs560})$	-5020	2.9e+04	-6.1e+04	5.749e+04	-2.026e+04

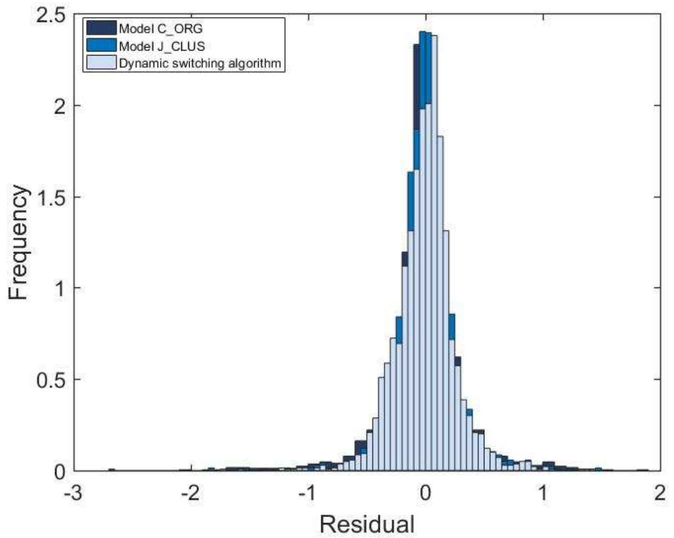
OWT recommended algorithms (as shown in table 5) were combined to form a dynamic switching algorithm, which selects the optimum Chla model for a given OWT. Estimates generated by the dynamic switching algorithm are compared to in-situ measurements of Chla concentration in Figure 8 (c). Points are coloured according to OWT group and shaped according to the chosen algorithm architectural approach. In order to qualitatively compare overall performance, scatterplots of the output from the best performing original form algorithm Model C_ORG (8 a) and the top performing single (non-dynamic) algorithm Model J_CLUS (8 b) are also shown in this figure. The corresponding histogram of residuals for Model C_ORG, Model J_CLUS and the dynamic switching algorithm are shown in Figure 8 d. It is clear that overall improvement in Chla retrieval accuracy is achieved by focussing on an OWT framework. Firstly, retuning model coefficients within an OWT group (Figure 8 b) improved the overall RPD calculated between measured and modelled Chla from 158% for Model C_ORG to 81.9% for the optimised Model J_CLUS. Next, dynamically altering the chosen algorithm per OWT (Figure 8 c) further reduced the RPD to 68.5%. The final version of the dynamic switching algorithm estimates log-transformed Chla from R_{rs} with a MAE of 0.18 mg m^{-3} (Figure 8 c). In terms of objective scoring, improvements in the final Chla outputs generated by the dynamic switching algorithm produced a total score of 15, which was the highest recorded score from all 48 algorithms tested.


r = 0.83
Slope = 0.93
Intercept = 0.23
RMSE = 0.30
MAE = 0.24
RPD = 158
bias = 0.15
%n = 96.3
Modelled Chla (mg m-3)
Measured Chla (mg m-3)
1000
100
10
1
0.1
0.01
0.01
0.1
1
10
100
1000
1
2
3
4
5
6
7
8
9
10
11
12
13
(a)


r = 0.88
Slope = 0.88
Intercept = 0.20
RMSE = 0.25
MAE = 0.19
RPD = 81.9
bias = 0.07
%n = 99.5
Modelled Chla (mg m-3)
Measured Chla (mg m-3)
1000
100
10
1
0.1
0.01
0.01
0.1
1
10
100
1000
1
2
3
4
5
6
7
8
9
10
11
12
13
(b)



(c)



(d)

Figure 8. Validation of Chla estimated from Model C_ORG (a), Model J_CLUS (b) and a combination of the best performing models per OWT group denoted the dynamic switching algorithm (c).Points have been coloured according to their OWT group classification and shaped according to the chosen model architectural approach (8 c only). A dashed 1:1 line representing a perfect modelled relationship has been annotated for reference. Regression statistics of correlation coefficient, linear slope and intercept, root mean square error

(RMSE), mean absolute error (MAE), relative percentage difference (RPD) and bias are shown for each model. A histogram of residuals for model outputs are shown in (d).

4. Discussion

4.1 Implications for remote sensing

In this validation exercise we have shown how accuracy in the retrieval of Chla from R_{rs} can be improved by targeting specific OWTs in algorithm development. In a comparison of single model retrievals, in other words, a single architecture across the entire dataset, it was shown that accuracy of the models was highly variable (Figure 4 and 5). Models A, C, J and R were highlighted as leading performers based on their objective score and error statistics. The empirical three-band ratio algorithms were shown to perform poorly when compared their two-band counterparts. For almost every model validated, the statistically tuned per OWT CLUS version produced the most accurate results. When comparing results on an OWT basis, the validation metrics were also highly variable (Figure 7). Model J appeared as a high scorer from ranked objective scoring the greatest number of times, and 15 of the 48 tested models were identified as leading performers in at least one OWT. Improvement in the final objective score and the corresponding error statistics was made by unifying an ensemble of the top performing algorithms for each OWT, as presented in Table 5. The resulting dynamic switching algorithm produced a relative percentage improvement in log-transformed MAE of 25% when compared to the top performing algorithm in its original form (Model C_ORG) (Figure 8 a and c). This result demonstrates that overall improvement in retrieval performance can be achieved by focusing on distinct OWTs during algorithm development. There remains uncertainty in the accuracy of retrievals obtained from several of the associated OWTs, particularly where low concentrations of Chla are observed in the presence of highly variable CDOM. However, this adaptive method allows for a directed effort in improving algorithms over specific OWTS and could be used to prioritise water bodies for future validation and algorithm development exercises.

This study did not attempt to validate every algorithm developed for retrieving concentrations of Chla from water colour observations. Instead, the intention was to test a range of empirical, semi-analytical and neural network model types to determine those best suited in terms of performance statistics for RS in complex and optically deep inland waters and to present those models in an adaptive framework that delivered accurate retrievals across a global continuum of environmental conditions. It was therefore interesting to observe apparent clustering in the high-scoring-model architectural approaches, whereby many OWTs were represented by similar models, as shown in Figure 8 c. These clusters appear at defined positions on the Chla concentration continuum (independent of other optically active constituents which also affect the R_{rs} signal), with typical blue-green ratio ocean colour algorithms representing clearer oligotrophic conditions, red-NIR band ratios capturing the mid-range meso- to eutrophic concentrations and more complex semi analytical models covering the hypereutrophic events. This result has some physical meaning as the validated algorithms have known capabilities and limitations in optically complex waters (see table 2 for algorithm training ranges). For example, the blue-green ratio algorithms are more sensitive to changes in Chla concentration at low reflectance levels due to the dominance of blue wavelength absorption for chlorophyll pigments, whilst semi-analytical models such as QAA are better equipped at dealing with additional optical complexity instigated by the presence of independently varying concentrations of other optically active constituents. Apparent architectural clustering suggests the defined OWTs may fall into higher-level groupings that could be used to further simplify algorithm selection. This result was also demonstrated by Spyrakos (2018b) using phylogenetic trees to identify sub groups within an OWT cluster and it reaffirms our understanding of the limitations of the tested RS algorithms in optically challenging aquatic systems. To this end, the number of required algorithms and/or parameterisations may be collapsed without significantly compromising overall performance and could lead to a decision tree for algorithm selection based around dominant and commonly occurring optical features, as shown in Figure 9. Where formal OWT classification is unachievable, the recommendation for a switching algorithm based on biological conditions would be; blue-green band ratio methods such as Model M in oligotrophic environments where Chla concentrations normally fall below 3 mgm^{-3}, NIR-red band ratio methods such as Model C where Chla is frequently in excess of 3 mgm^{-3} but less than 155 mg m^{-3} and the semi-analytical method of Model R in hypereutrophic conditions where Chla concentrations commonly exceed 160 mgm^{-3}. If no prior knowledge of water colour or Chla variability is known, Model C_ORG would be the recommended method.

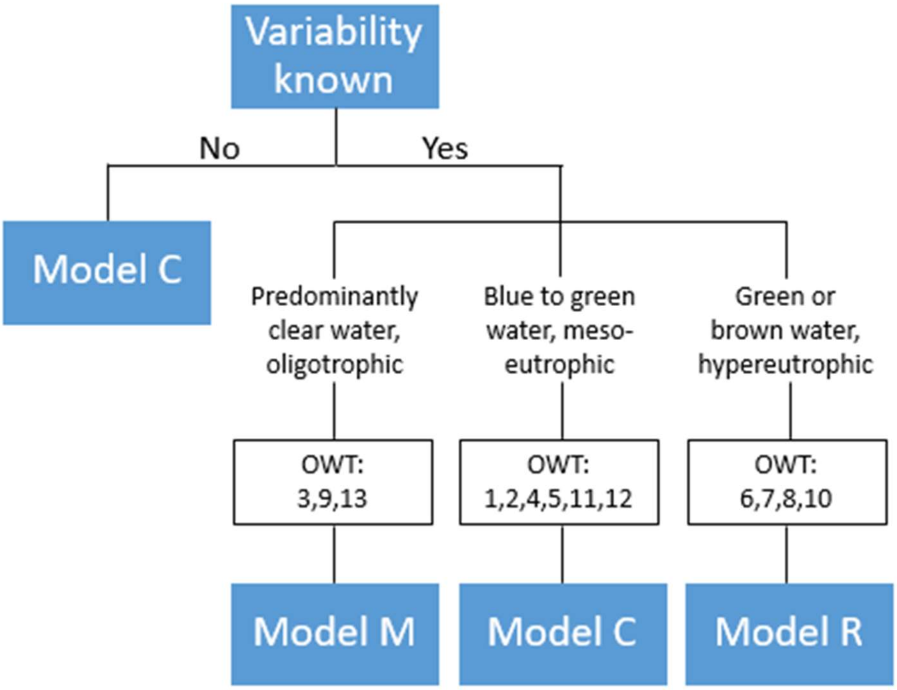


Figure 9. Decision tree depicting recommended algorithms for lakes where formal OWT is unknown.

Improvement in the accuracy of the Chla retrieval was demonstrated by changing not only the architectural type of model used in the retrieval but also by calibrating the chosen model with OWT specific coefficients. In the case of single model top performer Model J, the RPD decreased from 278 to 81 percent (correlation coefficient increased from 0.80 to 0.88 and MAE decreased from 0.36 to 0.26) simply by fitting the model to data collected for a specific OWT. This method allows for improved characterisation of regions with significant optical variability, both temporally and seasonally, which in turn improves the accuracy and effectiveness of the method overall. As a consequence, the recommended dynamic switching algorithm is more accurate than a general algorithm and more effective than a regionally developed algorithm. Several of the OWTs were under-represented by (a) our in-situ data (b) the models tested in this study, resulting in poor error statistics for OWTs 3, 5, 6, 10 and 13. While this appears a large number of misrepresented spectra, it equates to 32.6% percent of the entire in-situ validation dataset and embodies the very low or very high extreme Chla concentrations. A useful by-product of the dynamic switching framework is therefore its effective exposure of areas that require further attention. Our results can be used to identify OWT-specific modelling requirements for RS applications and highlight gaps in knowledge and data needs. Additionally, with widely variable validation results found across OWTs, the dynamic switching framework can act as a flagging system to express confidence in the Chla retrieval. In this context, Chla concentration determined in OWTs identified as poorly characterised could be flagged as ambiguous in a manner similar to atmospheric correction failures, hence providing a better insight into realistic uncertainty budgets. Furthermore, this framework also allows for better error characterisation by providing estimates of OWT-specific error which are potentially more useful to end-users with interests in specific water types.

4.2 Methodologies

This paper investigated the accuracy of several algorithms designed to retrieve concentrations of Chla from measurements of water colour in optically complex aquatic environments. Chla was calculated from an extensive database of in-situ R_{rs} measurements resampled at MERIS wavebands (Spyrakos et al., 2018b). To our knowledge, this is the most comprehensive dataset of inland water reflectance spectra that covers a continuum of optical water types both spatially and temporally. Many of the observations contained within this dataset have been used in previous studies to parameterise algorithms tested in this paper (e.g., Matthews & Odermatt, 2015). Ideally, a validation exercise of this magnitude would be conducted with an independent dataset of in-situ observations to avoid the influence of data dependency on performance results. This is currently not feasible due to a lack of systematic validation work covering the range of OWTs considered in this study and most existing data have been acquired specifically for algorithm development. Moreover, removing data that have been used to derive the original algorithms would make it difficult to undertake a fair comparison. Potential bias was dealt with to some extent by testing a variety of model types, as well as varying model coefficients. Moreover, a jack-knife method was applied to the validation dataset to subsample data and determine error statistics as a distribution. The diversity of performance results suggest model data dependency may not be strongly influencing the outcome. For example, the blue-green ratio ocean colour models (L, M and N) were identified as top performers in several OWT clusters however, no oceanic observations have been included in the R_{rs} validation dataset. This is further demonstrated by comparing the number of observations with the RMSE calculated for the top performing models in each OWT (table 4), as shown in Figure 10. Excluding OWT 13 (which is the most difficult type to model in terms of Chla due to extreme low concentrations) a regression slope of $-6.9e^{-5}$ indicates no trend exists between the number of observations within an OWT group and RMSE calculated for the OWT candidate model and as such the final result is not influenced by data bias.

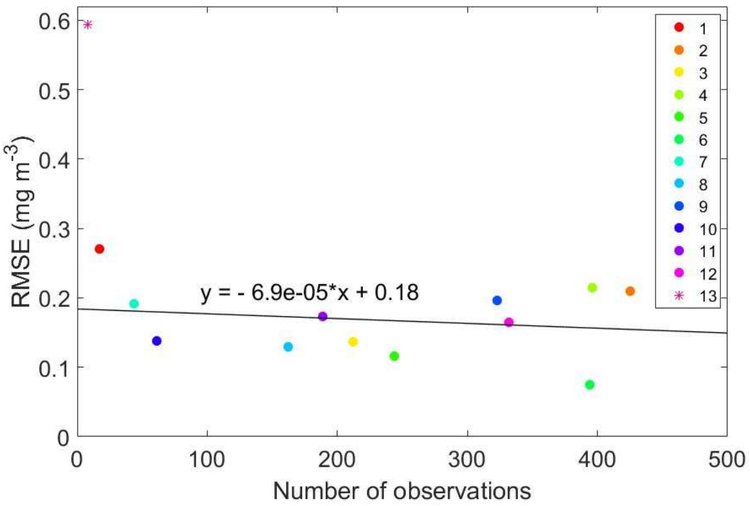


Figure 10. Scatterplot showing number of observations within an OWT against root-mean-square-error (RMSE) calculated for OWT top performing models. Points have been coloured according to their OWT group classification. A regression slope of -6.9^{-5} *(black line, excluding OWT 13) suggests no significant trend exists between these parameters.*

While this study focussed on the OWTs defined by Spyrakos et al. (2018b), it is recognised that these are unlikely to represent all water types occurring in natural waters and that OWTs may be defined by alternative methods (McKee et al., 2007; Moore et al., 2009) or underrepresented by the availability of in-situ data e.g. OWTs 1 and 13. Furthermore, uncertainty in OWT classification is expected, particularly at class member boundaries where component optical properties overlap (Spyrakos et al., 2018b). However, the OWT framework presented in this paper is the most comprehensive to date in terms of data size and range. The proposed method of subdividing data into optical water typologies before applying algorithms or assessing algorithm performance is a key message of this study and is entirely transferrable to other water environments or classification schemes.

Additional data may improve OWT classification accuracy and overall OWT coverage. It may also improve individual algorithm performance by further refining model coefficients (Salama et al., 2012). However it would not necessarily change the final recommendations for the dynamic switching algorithm. Most importantly, it would not alter the general efficacy of an ensemble method built on an OWT framework. We encourage the algorithms and parameterisations published here to be further refined and validated against new bio-optical datasets as they are collected.

The objective scoring system developed by Brewin et al. (2015) and modified for this study proved an effective tool for automatically generating an overview of algorithm performance. It provided a means of objectively ranking models based on their performance relative to average error statistics as demonstrated in Figures 6 and 7. However, it is important to acknowledge that the resulting score does not represent absolute performance and as such, the objective scoring system should be used in conjunction with standard error statistics to determine the most effective algorithms where equal scores are generated. The objective scoring system is also extremely effective for highlighting the underperforming algorithms and the trophic conditions under which the tested algorithms break down. This is particularly relevant when validating a large number of models over a wide range of optical and/or biological conditions.

4.3 Future work

In this paper, we have validated a host of algorithms to determine those capable of accurately retrieving Chla concentration remotely in inland waters. The ultimate ambition in this context is the production of accurate global products that can be exploited for status assessment and climate studies. Advances in computer processing power have allowed the development of machine learning and artificial intelligence procedures in satellite applications (Kim et al., 2014; Bilx et al., 2018; Ceccaroni et al., 2018). Whilst these are relatively under-validated on a global scale, their approach to algorithm selection with limited *a priori* knowledge of environmental condition is an exciting prospect, and has similarities to the framework presented in this paper. Our tractable goal requires a unified approach to data processing routines that are able to adapt to the optical complexities of inland waters and the framework presented in this study is a step forwards in achieving this (Palmer et al., 2015b & Mouw et al., 2015). The analysis has focussed on an extensive database of in-situ R_{rs} with the assumption that these correspond to the true water-leaving reflectance at the bottom of the atmosphere. The next stage of this research is to transfer results to reflectance obtained from satellites including archived Envisat MERIS and Sentinel-3 OLCI data. The presented methodology adopted for image processing will guide a focussed effort in developing an operational RS application suitable for optically complex inland waters.

5. Conclusions

With an ever-increasing number of published algorithms designed for retrieving Chla concentration from space, it has never been more crucial to report realistic limitations and uncertainties of well documented methods and to ensure that all algorithms are comprehensively validated and benchmarked against each other using datasets that incorporate the complexity of lake OWTs found globally. In this study, a series of Chla retrieval models have been validated using a comprehensive dataset of in-situ measurements collected from over 185 inland waters to determine those capable of recovering accurate concentrations of Chla in optically complex environments. A total of 48 algorithms were explored and an objective scoring system was developed to automatically rank models based on their relative statistical performance. From this study, several key conclusions can be made:

- The most suitable and accurate models of those assessed for estimating Chla within the biogeochemical range presented where OWT is uncategorised were Model A_ORG, Model C_ORG, Model H_ORG and Model J_CAL (Moses et al., 2009, Gurlin et al., 2011, Gons et al., 2005 & Mishra et al., 2012 respectively) which produced a MAE for $\log_{10}$ Chla of 0.23, 0.24, 0.23 and 0.27 mg m^{-3} respectively.
- The variable performance of the algorithms tested emphasises the importance of model selection and validation and caution should always be exercised when implementing models across a wide range of water bodies. The presented dynamic switching algorithm attempts to resolve performance variability by altering the selected model for a given OWT. This produces estimates of Chla concentration from reflectance measurements with a final correlation coefficient of 0.89 and a MAE of 0.18 mg m^{-3}.
- An objective scoring system is an extremely useful method for automatically determining performance for a wide range of models. It promotes confidence in the result and insurance for reporting purposes. However, it is not sufficient for making informed decisions regarding algorithm choice and in these cases, results should be considered in conjunction with error statistics.
- Overall performance was improved by focussing algorithm development within distinct OWT clusters. This was demonstrated in two ways; by calibrating models for

specific OWTs and by adjusting the model architecture to better represent an OWT. This framework should be exploited in the design of future operational models.

- This research is helping us progress towards a unified approach for global monitoring of chlorophyll in inland waters from space.

Acknowledgments

Authors gratefully acknowledge funding from the UK NERC-funded GloboLakes project (REF NE/J024279/1) and EC FP7 INFORM (Grant number 606865). We also thank LIMNADES in-situ data providers;

- Stefan G. H. Simis, Plymouth Marine Laboratory, Plymouth, United Kingdom
- Claudio C. F. Barbosa, Image Processing Division, Nacional Institute for Space Research-INPE, Sao Jose dos Campos, Sao Paulo, Brazil
- Caren E. Binding, Water Science and Technology Directorate, Environment and Climate Change Canada, Burlington, Ontario, Canada
- Shane Bradt, Department of Biological Sciences, University of New Hampshire, Durham, New Hampshire
- Mariano Bresciani, Institute for Electromagnetic Sensing of the Environment, CNR-IREA, Milano, Italy
- Giorgio Dall'Olmo, Plymouth Marine Laboratory, Plymouth, United Kingdom
- Claudia Giardino, Institute for Electromagnetic Sensing of the Environment, CNR-IREA, Milano, Italy
- Anatoly A. Gitelson, Department of Civil and Environmental Engineering, Israel Institute of Technology, Technion City, Haifa, Israel
- Tiit Kutser, Estonian Marine Institute, University of Tartu, Tallinn, Estonia

- Lin Li, Planetary and Environmental Remote Sensing Lab, Indiana University-Purdue University at Indianapolis, Indiana
- Bunkei Matsushita, Faculty of Life & Environmental Sciences, University of Tsukuba, Tsukuba, Ibaraki, Japan
- Victor Martinez-Vicente, Plymouth Marine Laboratory, Plymouth, United Kingdom
- Mark W. Matthews, CyanoLakes (Pty) Ltd, Cape Town, South Africa
- Igor Ogashawara, Planetary and Environmental Remote Sensing Lab, Indiana University-Purdue University at Indianapolis, Indiana
- Antonio Ruiz-Verdu, Image Processing Laboratory (IPL), Universitat de Valencia Catedratico Jose Beltran, Paterna, Valencia, Spain
- John F. Schalles, Department of Biology, Creighton University, Omaha, Nebraska
- Emma Tebbs, Department of Geography, King’s College London, London, United Kingdom
- Yunlin Zhang, Taihu Lake Laboratory Ecosystem Research Station, State Key Laboratory of Lake Science and Environment, Nanjing Institute of Geography and Limnology, Chinese Academy of Sciences, Nanjing, PR China.

References

Adrian R, O'Reilly CM, Zagarese H, Baines SB, Hessen DO, Keller W, et al., 2009. Lakes as sentinels of current climate change. Limnology & Oceanography; 54:2283–97.

Antoine, D., André, J-M., Morel, A., 1996. Oceanic primary production 2. Estimation at global scale from satellite (coastal zone color scanner) chlorophyll. Global Biogeochem. Cycles, 10: 57-69.

Antoine, D., F. D'ortenzio, F., Hooker, S.B., Bécu, G., Gentili, B., Tailliez, D., and Scott, A.J., 2008. Assessment of Uncertainty in the Ocean Reflectance Determined by Three Satellite Ocean Color Sensors (MERIS, Seawifs and MODIS-A) at an Offshore Site in the Mediterranean Sea (BOUSSOLE Project). *Journal of Geophysical Research* 113: C07013.

Behrenfeld, M. J., Boss, E., Siegel, D. A., Shea, D. M., 2005. Carbon-based ocean productivity and phytoplankton physiology from space. Global Biogeochem.

Blix, K., Pálffy, K., R. Tóth, V., Eltoft, T., 2018. Remote Sensing of Water Quality Parameters over Lake Balaton by Using Sentinel-3 OLCI. Water, 10, 1428.

Blondeau-Patissier, D., Gower, J.F.R., Dekker, A.G., Phinn, S.R., Brando, V.E., 2014. A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. Prog. Oceanogr. 123, 123–144.

Brando, V., Anstee, J., Wettle, M., Dekker, A., Phinn, S., Roelfsema, C., 2009. A physics based retrieval and quality assessment of bathymetry from suboptimal hyperspectral data. Remote Sensing of Environment, 113, pp. 755-70.

Brando, V. E., & Dekker, A. G., 2003. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. IEEE Transactions on Geoscience and Remote Sensing, 41(6), 1378–1387.

Brewin, R.J., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P.-Y., Devred, E., Doerffer, R., Fomferra, N., Franz, B., Grant, M., 2015. The ocean colour climate change initiative: Iii. A round-robin comparison on in-water bio-optical algorithms. Remote Sens. Environ. 162, 271–294

Bricaud, A., Babin, M., Morel, A., Claustre, H., 1995. Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: analysis and parameterization. J. Geophys. Res. Oceans 100, 13321–13332.

Bricaud, A., Morel, A., Babin, M., Allali, K., & Claustre, H., 1998. Variations of light absorption by suspended particles with chlorophyll a concentration in oceanic (Case 1) waters: Analysis and implications for bio-optical models. Journal of Geophysical Research, 103, 31033−31044.

Brown, C.W., Yoder, J.A., 1994. Coccolithophorid blooms in the global ocean. J. Geophys. Res. 99 (C4), 7467–7482.

Campbell, J. W., 1995. The lognormal distribution as a model for biooptical variability in the sea, J. Geophys. Res., 100, 13,237 – 13,254.

Carlson, R.E., and Simpson, J., 1996. A Coordinator's Guide to Volunteer Lake Monitoring Methods. North American Lake Management Society. 96 pp.

Ceccaroni L., Velickovski F., Blaas M., Wernand M.R., Blauw A., Subirats L., 2018. Artificial Intelligence and Earth Observation to Explore Water Quality in the Wadden Sea. In: Mathieu PP., Aubrecht C. (eds) Earth Observation Open Science and Innovation. ISSI Scientific Report Series, vol 15. Springer, Cham

Claustre, H., et al., 2004. An intercomparison of HPLC phytoplankton methods using in situ samples: Application to remote sensing and data- base activities, Mar .Chem., 85(1–2), 41–61.

Dall'Olmo G., Gitelson A. A., Rundquist D. C., 2003. Towards a unified approach for remote estimation of chlorophyll-*a* in both terrestrial vegetation and turbid productive waters. Geophys. Res. Lett. 30 1938.

Defoin-Platel, M., Chami, M., 2007. How ambiguous is the inverse problem of ocean color in coastal waters?. Journal of Geophysical Research, 112.

Dogliotti A.I., Ruddick K.G., Nechad B., Doxaran D., Knaeps E., 2015. A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters. Remote Sens Environ. 156:157-68.

Doerffer, R., Schiller, H., 2007. The MERIS case 2 water algorithm. International Journal of Remote Sensing, 28, 517–535.

Doney, S. C., Lima, I., Moore, J. K., Lindsay, K., Behrenfeld, M. J., Westberry, T. K., Mahowald, N., Glover, D. M., and Takahashi, T., 2009. Skill metrics for confronting global upper ocean ecosystembiogeochemistry models against field and remote sensing data, J. Mar. Systems, 76, 95–112.

Dekker, A. G., Vos, R. J., Peters, S. W. M., 2002. Analytical algorithms for lake water TSM estimation for retrospective analyses of TM and SPOT sensor data. International Journal of Remote Sensing, 23: 15–35.

Dekker, A.G., Phinn, S.R., Anstee, J.M., Bissett, W.P., Brando, V.E., Casey, B., Fearns, P.R., Hedley, J.D., Klonowski, W., Lee, Z., 2011. Intercomparison of shallow water bathymetry, hydro-optics, and benthos mapping techniques in Australian and Caribbean coastal environments. Limnol. Oceanogr. Methods 9, 396–425.

Giardino C., Bresciani M., Villa P., Martinelli A., 2010. Application of Remote Sensing in Water Resource Management: The Case Study of Lake Trasimeno, Italy. Water Resources Management. 24: 3885-3899.

Gilerson, A., Zhou, J., Hlaing, S., Ioannou, I., Gross, B., Moshary, F., Ahmed, S., 2008. Fluorescence component in the reflectance spectra from coastal waters. II. Performance of retrieval algorithms. Opt. Express 16(4), 2446–2460.

Gilerson, A.A., Gitelson A.A., Zhou, J., Gurlin, D., Moses, W., Loannou, L., Ahmed, S.A., 2010. Algorithms for remote estimation of chlorophyll-*a* in coastal and inland waters using red and near infrared bands. Opt. Express, 18, 24109–24125.

Gitelson, A. A., Kondratyev, K.Y., 1991. Optical models of mesotrophic and eutrophic water bodies, International J. Remote Sensing, 12, 373 – 385.

Gitelson, A., 1992. The peak near 700 nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll concentration. International Journal of Remote Sensing 13 (17), 3367–3373.

Gitelson, A., 1993. Algorithms for remote sensing of phytoplankton pigments in inland waters. Advanced Space Research 13 (5), 197–201.

Gitelson, A., Gurlin, D., Moses, W., Yacobi, Y., 2011. Remote estimation of chlorophylla concentration in inland, estuarine, and coastal waters, Chapter 18. In Q. Weng (Ed.), Advances in environmental remote sensing: sensors, algorithms, and applications (pp. 449–478). : CRC Press, Taylor and Francis Group, 610 p. ISBN:9781420091755.

Gons, H. J., Rijkeboer, M., Ruddick, K. G., 2002. A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. Journal of Plankton Research, 24, 947−951.

Gons, H. J., Rijkeboer, M., Ruddick, K. G., 2005. Effect of a waveband shift on chlorophyll retrieval from MERIS imagery of inland and coastal waters. Journal of Plankton Research, 27, 125−127.

Gons, H.J., Auer, M.T., Effler, S.W., 2008. MERIS satellite chlorophyll mapping of oligotrophic and eutrophic waters in the Laurentian Great Lakes. Remote Sens. Environ. 112, 4098–4106.

Gordon, H.R., Clark, D.K., Mueller, J.L., Hovis, W.A., 1980. Phytoplankton pigments from the Nimbus-7 Coastal Zone Color Scanner: comparisons with surface measurements. Science, 210: 63-66.

Gordon, H. R., Brown, O. B., Evans, R. H., Brown, J. W., Smith, R. C., Baker, K. S., et al., 1988. A semianalytic radiance model of ocean color. Journal of Geophysical Research, 93(D9), 10909−10924.

Gower, J.F.R., Doerffer, R., Borstad, G.A., 1999. Interpretation of the 685 nm peak in water-leaving radiance spectra in terms of fluorescence, absorption and scattering, and its observation by MERIS, Int. J. Remote Sensing, 9, 1771-1786.

Gurlin, D., Gitelson, A. A., Moses, W. J., 2011. Remote estimation of chl-a concentration in turbid productive waters — Return to a simple two-band NIR-red model? Remote Sensing of Environment, 115(12), 3479–3490.

Hoogenboom, H.J., Dekker, A.G., De Haan, J.F., 1998. Retrieval of chlorophyll and suspended matter in inland waters from CASI data by matrix inversion. Can. J. Remote Sens., vol. 24, pp. 144–152.

Hooker, S.B., Van Heukelem, L., Thomas, C.S., Claustre, H., Ras, J., Barlow, R., Sessions, H., Schluter, L., Perl, J., Trees, C., Sutuart, V., Head, E., et al., 2005. The Second SeaWiFS HPLC Analysis Round-robin Experiment (SeaHARRE-2). TM-2005-212785, NASA Goddard Space Flight Center, Greenbelt, MD.

Hu, C., Muller-Karger, F.-E., Taylor, C., Carder, K.L., Kelble, C., Johns, E., Heil, C.A., 2005. Red tide detection and tracing using MODIS fluorescence data: a regional example in SW Florida coastal waters. Remote Sens. Environ. 97, 311–321.

Hunter P.D., Tyler A.N., Carvalho L., Codd G.A., Maberly S.C., 2010. Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. Remote Sens Environ. 114:2705-18.

Huot, Y., Brown, C.A., Cullen, J.J., 2005. New algorithms for MODIS Sun-induced chlorophyll fluorescence and a comparison with present data products, Limnol. Oceanogr. Methods 3, 108 –130.

Ioannou, I., Gilerson, A., Gross, B., Moshary, F., Ahmed, S., 2013. Deriving ocean color products using neural networks. Remote Sens. Environ. 134, 78e91.

Kim, Y., Im, J., Ha, H., Choi, J., Ha, S., 2014. Machine Learning Approaches to Coastal Water Quality Monitoring Using GOCI Satellite Data. GIScience and Remote Sensing 51: 158–174.

Kirk, J. T. O., 1994. Light and photosynthesis in aquatic ecosystems (2nd edition). : Cambridge University Press.

Kuchinke, C. P., Gordon, H. R., Harding, L. W., Voss, K. J., 2009. Spectral optimization for constituent retrieval in Case 2 waters II: Validation study in the Chesapeake Bay. Remote Sensing of Environment, 113, 610–621.

Kutser T., Pierson D., Kallio K., Reinart A., Sobek S., 2005. Mapping lake CDOM by satellite remote sensing. Remote Sensing of Environment. 94: 535-540.

Kutser, T., Herlevi, A., Kallio, K., Arst, H., 2001. A hyperspectral model for interpretation of passive optical remote sensing data from turbid lakes, The Science of the Total Environment 268, pp. 47-58.

Kutser, T., Arst, H., Maekivi, S., Kallaste, K., 1998. Estimation of the water quality of the Baltic Sea and some lakes in Estonia and Finland by passive optical remote sensing measurements on board a vessel. Lakes and Reservoirs; Research and Management, 3, pp. 53–66.

Lee, Z., Carder, K.L., Arnone, R., 2002. Deriving inherent optical properties from water color: a multiband quasi analytical algorithm for optically deep waters. Appl. Opt. 41 (27), 5755–5772.

Lee, Z. P., Carder, K.L., Mobley, C.D., Steward, R.G., Patch, J.S., 1999. Hyperspectral remote sensing for shallow waters: 2. Deriving bottom depths and water properties by optimization, Appl. Opt., 38, 3831–3843.

Letelier, R.M., Abbott, M.R., 1996. An analysis of chlorophyll fluorescence algorithms for the Moderate Resolution Imaging Spectrometer (MODIS), Rem. Sensing Environ., 58, 215-223.

Lindell, T., Pierson, D., Premazzi, G., Zilioli, E., 1999. Manual for monitoring European lakes using remote sensing techniques. Luxembourg, Office for Official Publications of the European Communities, EUR Report n. 18665 EN.

Louchard, E. M., Reid, R.P., Stephens, F.C., Davis, C.O., Leathers, R.A., Downes, T.V., 2003. Optical remote sensing of benthic habitats and bathymetry in coastal environments at Lee Stocking Island, Bahamas: A comparative spectral classification approach. Limnol. Oceanogr. 48:511-521.

Maritorena, S., Siegel, D. A., Peterson, A. R., 2002. Optimization of a semianalytical ocean color model for global-scale applications. Applied Optics, 41, 2705–2714.

Matthews M.W., Bernard S., Winter K., 2010. Remote sensing of cyanobacteria-dominant algal blooms and water quality parameters in Zeekoevlei, a small hypertrophic lake, using MERIS. Remote Sens Environ. 114:2070-87.

Matthews, M.W., 2011. A current review of empirical procedures of remote sensing in inland and near-coastal transitional waters. Int. J. Remote Sens. 32 (21):6855–6899.

Matthews, M.W., Bernard, S., Robertson, L., 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.

Matthews, M.W., Odermatt, D., 2015. Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters. Remote Sens. Environ. Vol. 156, pp. 374-382.

McClain, C.R., 2009. A decade of satellite ocean color observations. Annu. Rev. Mar. Sci. 1, 19–42

McKee, D., Cunningham, A., Dudek, A., 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, 827–834.

McKee, D., Röttgers, R., Neukermans, G., Calzado, V.S., Trees, C., Ampolo-Rella, M., et al., 2014. Impact of measurement uncertainties on determination of chlorophyll-specific absorption coefficient for marine phytoplankton. J. Geophys. Res. Oceans 119 (12), 9013–9025.

Mishra, S., Mishra, D. R., 2012. Normalized difference chlorophyll index: a novel model for remote estimation of chlorophyll-*a* concentration in turbid productive waters. Remote Sensing of Environment, 117, 394–406.

Mishra, S., Mishra, D.R., Lee, Z., 2013. Quantifying cyanobacterial phycocyanin concentration in turbid productive waters: a quasi-analytical approach. Remote Sens. Environ. 133, 141–151.

Mishra, S., Mishra, D.R., Lee, Z., 2014. Bio-optical inversion in highly turbid and cyanobacteria dominated waters. IEEE Trans. Geosci. Remote Sens. 52, 375–388.

Mobley, C.D., 1994. Light and Water: Radiative Transfer in Natural Waters. Academic, San Diego.

Mobley, C. D., 1999. Estimation of the remote-sensing reflectance from above-surface measurements. Applied Optics, 38, 7442–7455.

Mobley, C.D., Sundman, L.K., Davis, C.O., Bowles, J.H., Downes, T.V., Leathers, R.A., Montes, M.J., Bisset, W.P., Kohler, D.D.R., Reid, R.P., Louchard, A.M., Gleason, A., 2005. Interpretation of hyperspectral remote-sensing imagery by spectrum matching and look-up tables. Appl. Opt, 44, 3576–3592.

Moore, T. S., Campbell, J.W., Dowell, M.D., 2009. A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. Remote Sens. Environ. 113: 2424-2430, doi:10.1016/j.rse.2009.07.016.

Morel, A., Prieur, L., 1977: Analysis of variations in ocean color. Limnology & Oceanography, 22(4), 709-722.

Morel, A., Huot, Y., Gentili, B., Werdell, P.J., Hooker, S.B., Franz, B.A., 2007. Examining the consistency of products derived from various ocean color sensors in open ocean (case 1) waters in the perspective of a multi-sensor approach. Remote Sens. Environ. 111 (1), 69–88

Moses, W., Gitelson, A., Berdnikov, S., Povazhnyy, V., 2009. Satellite estimation of chlorophyll-*a* concentration using the red and NIR bands of MERIS — The Azov Sea case study. IEEE Geoscience and Remote Sensing Letters, 4(6), 845–849.

Mouw, C. B., Greb, S., Aurin, D., DiGiacomo, P. M., Lee, Z., Twardowski, M., et al., 2015. Aquatic color radiometry remote sensing of coastal and inland waters: challenges and recommendations for future satellite missions. Remote Sens. Environ. 160, 15–30. doi: 10.1016/j.rse.2015.02.001.

Nair, A., Sathyendranath, S., Platt, T., Morales, J., Stuart, V., Forget, M.-H., Devred, E., Bouman, H., 2008. Remote sensing of phytoplankton functional types, Remote Sens. Environ., 112, 3366– 3375.

National Research Council. 2011. Assessing the Requirements for Sustained Ocean Color Research and Operations. Washington, DC: The National Academies Press. https://doi.org/10.17226/13127.

Nechad B., Ruddick K.G., Park Y., 2010. Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. Remote Sens Environ. 114:854-66.

Odermatt D., Giardino C., Heege T., 2010. Chlorophyll retrieval with MERIS Case-2-Regional in perialpine lakes. Remote Sens Environ. 114:607-17.

Odermatt, D., Gitelson, A., Brando, V.E., Schaepman, M., 2012. Review of constituent retrieval in optically deep and complex waters from satellite imagery. Remote Sens. Environ. 118, 116–126.

O’Reilly, J.E., Maritorena, S., Mitchell, B.G., Siegel, D.A., Carder, K.L., Garver, S.A., Kahru, M., McClain, C., 1998. Ocean color chlorophyll algorithms for seawifs. J. Geophys. Res.: Oceans (1978-2012) 103, 24937–24953.

O'Reilly, J.E., and 24 Co-authors, 2000. SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3. NASA Tech. Memo. 2000-206892, Vol. 11, S.B. Hooker and E.R. Firestone, Eds., NASA Goddard Space Flight Center, 49 pp.

Palmer, S.C.J., Hunter, P.D., Lankester, T., Hubbard, S., Spyrakos, E., Tyler, A.N., 2015a. Validation of Envisat MERIS algorithms for chlorophyll retrieval in a large, turbid and optically-complex shallow lake. Remote Sens. Environ. 157, 158–169.

Palmer, S.C.J., Kutser T., & Hunter P.D., 2015b. Remote sensing of inland waters: Challenges, 746 progress and future directions, Remote Sensing of Environment, 157, 1-8, doi: 747 10.1016/j.rse.2014.09.021.

Politi, E., MacCallum, S.,Cutler, M.E.J., Merchant, C.J., Rowan, J.S., Dawson, T.P ., 2016. Selection of a network of large lakes and reservoirs suitable for global environmental change analysis using Earth Observation, International Journal of Remote Sensing, 37:13, 3042-3060.

Pope, R. M., & Fry, E. S., 1997. Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements. Applied Optics, 36, 8710–8723.

Ritchie, R., 2008. Universal chlorophyll equations for estimating chlorophylls a, b, c, and d and total chlorophylls in natural assemblages of photosynthetic organisms using acetone, methanol, or ethanol solvents, Photosynthetica, 46(1), 115–126.

Salama, M.S., van der Velde, R., van der Woerd, H.J., Kromkamp, J.C., Philippart, C.J.M., Joseph, A.T., O'Neill, P.E., Lang, R.H., Gish, T., Werdell, P.J., Su, Z., 2012. Technical notes: Calibration and validation of geophysical observation models. Biogeosciences, 9, 6, 2195-2201.

Sathyendranath, S., 2000. Remote sensing of ocean colour in coastal, and other optically complex, waters. Reports of the Ocean Colour Coordinating Group, 3, Dartmouth, Canada: IOCCG.

Sayers, M.J., Grimm, A.G., Shuchman, R.A., Deines, A.M., Bunnell, D.B., Raymer, Z.B., Rogers, M.W., Woelmer, W., Bennion, D.H., Brooks, C.N., 2015. A new method to generate a high-resolution global distribution map of lake chlorophyll. Int. J. Remote Sens. 36, 1942–1964.

Schalles, J. F., 2006. Optical remote sensing techniques to estimate phytoplankton chlorophyll a concentrations in coastal waters with varying suspended matter and CDOM concentrations, p. 27-79. In L. Richardson and E. Ledrew [ed.], Remote Sensing of Aquatic Coastal Ecosystem Processes: Science and Management Applications. Springer.

Seegers, B., Stumpf, R., Schaeffer, B., Loftin, K., and Werdell, P., 2018. Performance metrics for the assessment of satellite data products: an ocean color case study. Opt. Express 26, 7404-7422.

Shanmugam, P., Ahn, Y.H., Ryu, J.H., Sundarabalan, B., 2010. An evaluation of inversion models for retrieval of inherent optical properties from ocean color in coastal and open sea waters around Korea. Journal of Oceanography, 66(6), 815–830.

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

Smith, R. C., & Baker, K. S., 1981. Optical properties of the clearest natural waters (200–800 nm). Applied Optics, 20, 177–184.

Sørensen, K., Grung, M., Röttgers, R., 2007. An intercomparison of in vitro chlorophyll a determinations for MERIS level 2 data validation, Int. J. Remote Sens., 28, 537–554.

Spyrakos, E., Hunter, P., Simis, S., Neil, C., Martinez, V., Barbosa, C., Binding, C., Bradt, S., Bresciani, M., Dall'Olmo, G., Giardino, C., Gitelson, A., Kutser, T., Li, L., Matsushita, B., Matthews, M., Ruiz-Verdu, A., Schalles, J., Tebbs, E., Zhang, Y., Tyler, A.N., 2018a. Relationships between biogeochemical and inherent optical properties in inland waters. Submitted to Journal of Geophysical Research.

Spyrakos, E., O'Donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis. S.G.H., Neil, C., Barbosa, C.C.F., Binding, C.E., Bradt, S., Bresciani, M., Dall'Olmo, G., Giardino. C., Gitelson, A.A., Kutser, T., Li, L., Matsushita, B., Martinez Vicente, V., Matthews, M.W., Ogashawara, I., Ruiz-Verdu, A., Schalles, J.F., Tebbs, E., Zhang, Y., Tyler, A.N., 2018b. Optical types of inland and coastal waters. Limnology & Oceanography. 10.1002.

Tarrant P.E., Amacher J.A., Neuer S., 2010. Assessing the potential of Medium-Resolution Imaging Spectrometer (MERIS) and Moderate-Resolution Imaging Spectroradiometer (MODIS) data for monitoring total suspended matter in small and intermediate sized lakes and reservoirs. Water Resources Research. 46.

Tilstone, G., Mallor-Hoya, S., Gohin, F., Belo Couto, A., Sa, C., Goela, P., Cristina, S., Airs, R., Icely, J., Zühlke, M., Groom, S., 2017. Which Ocean colour algorithm for MERIS in North West European waters? Remote Sens. Environ. 189, 132—151, http://dx.doi.org/ 10.1016/j.rse.2016.11.012.

Tyler, A.N., Hunter, P.D., Spyrakos, E., Groom, S., Constantinescu, A.M., Kitchen, J., 2016. Developments in Earth observation for the assessment and monitoring of inland, transitional, coastal and shelf-sea waters. Sci. Total Environ. 572:1307–1321. http:// dx.doi.org/10.1016/j.scitotenv.2016.01.020.

Werdell, P.J. and Bailey, S.W., 2005. An improved bio-optical data set for ocean color algorithm development and satellite data product validation. Remote Sensing of Environment, 98(1), 122-140.

Willmott, C. & Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Clim. Res., 30, 79–10 82,

Yang, W., Matsushita, B., Chen, J., Fukushima, B., Ma, R., 2010. An enhanced three-band index for estimating chlorophyll-*a* in turbid case-II waters: Case studies of Lake Kasumigaura, Japan, and Lake Dianchi, China. IEEE Geosci. Remote Sens. Lett. 7, 655–659.

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., Dickinson, R., 2013. The role of satellite remote sensing in climate change studies. Nature Clim. Change. 3: 875-883.