ISPRS Journal of Photogrammetry and Remote Sensing 190 (2022) 279–300

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Characterising retrieval uncertainty of chlorophyll-*a* algorithms in oligotrophic and mesotrophic lakes and reservoirs

Mortimer Werther [a,*], Daniel Odermatt [b], Stefan G.H. Simis [c], Daniela Gurlin [d], Daniel S.F. Jorge [e], Hubert Loisel [e], Peter D. Hunter [a], Andrew N. Tyler [a], Evangelos Spyrakos [a]

[a] Earth and Planetary Observation Sciences (EPOS), Biological and Environmental Sciences, Faculty of Natural Sciences, University of Stirling, Stirling, United Kingdom
[b] Swiss Federal Institute of Aquatic Science and Technology, Department of Surface Waters – Research and Management, Dübendorf, Switzerland
[c] Plymouth Marine Laboratory, Plymouth, United Kingdom
[d] Wisconsin Department of Natural Resources, Madison, WI, United States
[e] Université du Littoral Côte d'Opale, CNRS, Univ. Lille, UMR 8187 – LOG – Laboratoire d'Océanologie et de Géosciences, F-62930 Wimereux, France

ARTICLE INFO

Keywords:
Chlorophyll-a
Lakes
Uncertainties
Shapley additive explanations
Machine learning

ABSTRACT

Remote sensing product uncertainties for phytoplankton chlorophyll-*a* (chla) concentration in oligotrophic and mesotrophic lakes and reservoirs were characterised across 13 existing algorithms using an *in situ* dataset of water constituent concentrations, inherent optical properties (IOPs) and remote-sensing reflectance spectra ($R_{rs}(\lambda)$) collected from 53 lakes and reservoirs (346 observations; chla concentration < 10 mg m^{-3}, dataset median 2.5 mg m^{-3}). Substantial shortcomings in retrieval accuracy were evident with median absolute percentage differences (MAPD) $> 37\%$ and mean absolute differences (MAD) > 1.82 mg m^{-3}. Using the Hyperspectral Imager for the Coastal Ocean (HICO) band configuration improved the accuracies by 10–20% compared to the Ocean and Land Colour Instrument (OLCI) configuration. Retrieval uncertainties were attributed to optical and biogeochemical properties using machine learning models through SHapley Additive exPlanations (SHAP). The chla retrieval uncertainty of most semi-analytical algorithms was primarily determined by phytoplankton absorption and composition. Machine learning chla algorithms showed relatively high sensitivity to light absorption by coloured dissolved organic matter (CDOM) and non-algal pigment particulates (NAP). In contrast, the uncertainties of red/near-infrared algorithms, which aim for lower uncertainty in the presence of CDOM and NAP, were primarily explained through the total absorption by phytoplankton at 673 nm ($a_\phi(673)$) and variables related to backscatter. Based on these uncertainty characterisations we discuss the suitability of the evaluated algorithm formulations, and we make recommendations for chla estimation improvements in oligo- and mesotrophic lakes and reservoirs.

1. Introduction

Optical remote sensors are used to observe the major optically active water constituent (OAC) concentrations in lakes and reservoirs. The phytoplankton pigment chlorophyll-*a* (chla) is used as a proxy of phytoplankton biomass (Gitelson et al., 1993; Mittenzwey et al., 1992) and net primary production (Carlson, 1977; Huot et al., 2007; Poikane et al., 2010) and thereby to assess the ecological integrity of aquatic ecosystems. In order to map chla consistently in time and space, algorithms are needed to relate the observed signal at the sensor to the concentration of this pigment. Several studies have shown that multiple algorithms in switching and blending (Liu et al., 2021; Neil et al., 2019; Schaeffer et al., 2022; Spyrakos et al., 2018) or ensemble schemes (Werther et al., 2021) outperform individual algorithms across the optical-biogeochemical diversity of natural waters. There is, therefore, a common need to express algorithmic uncertainty associated with each technique, to inform appropriate algorithm selection and to further their development. This requires the assessment of a range of approaches against a representative data set.

Remote-sensing reflectance, $R_{rs}(\lambda)$, is determined by inherent optical properties (IOPs) of water and OACs (Gordon et al., 1988; Maritorena et al., 2002; Mobley, 1999). The estimation of chla from $R_{rs}(\lambda)$ can be analytically approached by exploiting this relationship (through forward and inverse modelling), or chla can be empirically associated with

* Corresponding author.
E-mail address: mortimer.werther@stir.ac.uk (M. Werther).

https://doi.org/10.1016/j.isprsjprs.2022.06.015
Received 22 February 2022; Received in revised form 21 June 2022; Accepted 23 June 2022
0924-2716/© 2022 The Authors. Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

the variability of the $R_{rs}(\lambda)$ signal in one or more wavebands (Schalles, 2006). With increasing phytoplankton abundance in ocean waters, the main observed effect is a shift of the reflectance maximum from blue to green wavebands caused by a combination of increased absorption and particulate backscattering processes (Morel and Prieur, 1977). These empirical observations were translated into the widely used ocean colour (OC) algorithms (O'Reilly et al., 1998; O'Reilly and Werdell, 2019), of which several variations exist for specific sensor configurations and regional effects. Another empirical approach relates chla concentration to sun-induced fluorescence of chla associated with photosystem II in the region around 685 nm (Gower, 1980; Gupana et al., 2021; Neville and Gower, 1977). The height of the fluorescence peak above a baseline between 650 and 730 nm is then assumed to correlate linearly with chla concentration (Gower et al., 2004). Several other empirical and semi-empirical algorithms relate chla concentration to variability in the red/near-infrared (NIR) area of the reflectance spectrum, which is caused by expression of chla absorption in a narrow region around 675 nm. This expression can be offset against NIR reflectance near 700 nm, which has weak combined absorption by phytoplankton and water (Dall'Olmo et al., 2003; Gilerson et al., 2007, 2010; Gitelson, 1992; Moses et al., 2009). A further group of empirical algorithms were specifically designed for chla estimation in waters with high algal biomass and the identification of algae blooms, based on peak height methods like the fluorescence line height, exploiting variability in the spectral shape of the reflectance peak variation near 700 nm (Binding et al., 2013; Matthews et al., 2012). Besides band arithmetic approaches, machine learning (ML) algorithms are increasingly trained for use in inland waters. ML algorithms are based on non-linear regression models and fit within the empirical algorithm category, as they are developed with training datasets generated from field and/or simulated observations (Bricaud et al., 2007; Hieronymi et al., 2017; Pahlevan et al., 2020).

Contrary to purely empirical algorithm formulations, analytical and semi-analytical chla algorithms start from physics-based inversion of IOPs from $R_{rs}(\lambda)$, such as the absorption by phytoplankton $(a_\phi(\lambda))$, which can be subsequently scaled to chla concentration (Lee et al., 2002). Most contemporary semi-analytical algorithms (SAA) invert subsurface remote-sensing reflectance $\left(r_{rs}(\lambda, \mathrm{sr}^{-1})\right)$ obtained through conversion of above-water $R_{rs}(\lambda)$. SAA show many variations in the inversion for $a_\phi(\lambda)$. They vary in their (empirical) definition of the $a_\phi(\lambda)$ spectral shape, the method to calculate the magnitude of $a_\phi(\lambda)$ and the defined relationship between $r_{rs}(\lambda)$ and $a_\phi(\lambda)$ (Werdell et al., 2018).

If retrieval errors were constant across a wide chla concentration range, larger relative errors would occur in the low concentration part of the considered range. Several studies have shown that the issue with low chla concentration estimates is also more pressing in absolute terms, with widely applied empirical algorithms and SAA revealing large uncertainties in oligo- and mesotrophic systems. For example, Mouw et al. (2013) evaluated seven SAAs in Lake Superior (North America), an oligotrophic system with OAC absorption dominated by CDOM. The evaluated SAA algorithms were unable to effectively retrieve chla concentration due to inaccurate $a_\phi(\lambda)$ estimation ($r^2 < 0.2$). On a large *in situ* dataset (>2900 observations), Neil et al. (2019) evaluated 47 chla algorithms, which revealed that the largest retrieval uncertainties for the dataset were associated with oligo- and mesotrophic conditions. More accurate chla estimation was achieved for Swiss perialpine lakes. Odermatt et al. (2010, 2008) demonstrated that for low chla concentrations (chla < 5 mg m^{-3}), lake-parameterised empirical and SAA algorithms can provide precise estimates (correlation coefficients between 0.58 and 0.94). Comparable outcomes for Italian perialpine lakes were achieved by (Bresciani et al., 2018; Giardino et al., 2001; Pepe et al., 2001).

The results of previous studies show that chla retrieval accuracy in oligo- and mesotrophic waters is inconsistent across regions. To improve the chla retrieval across different regions, it is thus necessary to characterise the fundamental causes of chla retrieval uncertainty. Uncertainty of chla estimates (hereafter retrieval uncertainty) can originate from three sources (Merchant et al., 2017). First, observation uncertainty caused by random and systematic errors associated with the physical process of making measurements using optical sensors (IOCCG, 2019). Second, uncertainty introduced in the calculation of $R_{rs}(\lambda)$ from the top-of-atmosphere signal measured by a satellite sensor through an atmospheric correction (AC) algorithm (Moses et al., 2017). And third, algorithmic uncertainty stemming from the empirical and/or physical assumptions made and translated into a procedure to retrieve chla concentration (Brewin et al., 2015; Salama and Stein, 2009).

To aid the study of algorithmic uncertainties over oligo- and mesotrophic systems, we focus on the retrieval of chla concentration from *in situ* measured $R_{rs}(\lambda)$. It is known that significant observation uncertainties are intrinsic to *in situ* $R_{rs}(\lambda)$ (Mélin et al., 2016; Mélin and Franz, 2014; Zibordi and Voss, 2014) and chla quantification methods (Claustre et al., 2004; Hooker et al., 2005; McKee et al., 2014; Sørensen et al., 2007). It is, however, common practice to validate chla retrieval algorithms against *in situ* reference measurements whose observation uncertainty is considered low in comparison to satellite-derived estimates (Sørensen et al., 2007). Yet, these uncertainties are rarely specified in large inland and coastal water *in situ* datasets, and clarification of their origin is not an established practice for algorithm development and validation (McKinna et al., 2021).

Observation and algorithmic uncertainty sources are commonly unknown, thus additional information that can be attributed to the retrieval uncertainty is necessary. The accuracy of chla algorithms can be assessed through calculation of the residual value, which is the difference between an algorithm estimate and a measured *in situ* reference value. The frequency of the obtained residual values forms a residual value distribution (RVD). To improve our understanding of the underlying sources of algorithmic uncertainty, an algorithm's RVD must first be attributed to measurement conditions or the presence of specific substances which may hinder the intended function of the algorithm. The measurement conditions and substances can be represented by a set of explanatory variables that were co-measured with $R_{rs}(\lambda)$ and chla concentration. Example explanatory variables are IOPs, OACs other than chla or spectral band ratios.

To enable characterisations of retrieval uncertainty, we build ML regression models (herein uncertainty models) through a set of explanatory variables. The constructed uncertainty models quantify the retrieval uncertainty from a chla algorithm expressed through the RVD. Subsequently, we investigate the impact of the explanatory variables on an uncertainty model to characterise algorithmic uncertainty through specific observation conditions. For this purpose, we use the SHapley Additive exPlanations (SHAP) procedure (Lundberg and Lee, 2017; Lundberg et al., 2020, 2018). Through SHAP we aim to identify uncertainty drivers of 13 chla algorithms tested against a dataset of 53 waterbodies.

In summary, the research objectives of this study are to: 1) determine sensitivities between empirical and semi-analytical algorithms when applied to the varying spectral resolutions of the Hyperspectral Imager for the Coastal Ocean (HICO) and Ocean and Land Colour Instrument (OLCI), 2) characterise the retrieval uncertainties of the tested algorithms and ultimately, 3) reveal if any of the evaluated algorithms are sufficiently suitable for application to specific observation conditions under the presence of CDOM and NAP absorption across oligo- and mesotrophic lakes and reservoirs.

2. Data

A dataset comprised of oligo- and mesotrophic lakes and reservoirs was compiled from three data sources: (i) U.S. Wisconsin Department of Natural Resources (DNR), (ii) the global LIMNADES database (Lake Bio-optical Measurements and Matchup Data for Remote Sensing (https://limnades.stir.ac.uk/)) and (iii) and an internal dataset from the

Table 1
Datasets and their waterbodies included in this study.

Dataset name	Inland water system name (n = 53)	Number of observations (n = 346)	Prior data use
U.S. Wisconsin DNR	Big Saint Germain Lake, Big Sand Lake, Butternut Lake, Fence Lake, Geneva Lake, Green Lake, Lac Courte Oreilles, Lac Vieux Desert, Lake Chippewa, Lake Mendota, Lake Wissota, Metonga Lake, Pelican Lake, Rainbow Flowage, Rock Lake, Round Lake, Shawano Lake, Trout Lake	46	(Pahlevan et al., 2021b, 2021a, 2020) and see Appendix A Description of Wisconsin DNR in situ data
LIMNADES	Branched Oak Lake, Crescent Lake, Cuerda del Pozo, Diamond Pond, East Twin Lake, Lake Erie, Ewill Lake, Fivemile Pond, Forest Lake, Fresmond, Lake Garda, Ginger Cove, Goose Pond, Granite Lake, Great Salt Lake, Groton Pond, Iznajar Reservoir, Loskop Reservoir, Lake Maggiore, Lake Mantova, Lake Okoboji, Lake Ontario, Lake Paeijaenne, Lake Peipsi, Lake Pyhäjärvi, Lake Trasimeno, Lake Vesijärvi, Lake Winnipeg	276	(Binding et al., 2013, 2011, 2010, 2008; Bresciani et al., 2011; Giardino et al., 2015, 2014a, 2014b, 2005; Gitelson et al., 2009; Guanter et al., 2010; Gurlin et al., 2011; Kallio et al., 2015; Kutser et al., 2013; Manzo et al., 2015; Matthews, 2014; Matthews and Bernard, 2013; Ruiz-Verdú et al., 2008, 2005; Schalles, 2006)
University of Stirling	Lake Biel, Coniston Water, Derwent Water, Lake Geneva, Loch Lomond, Loch Ness, Lake Vänern	24	(Aulló-Maestro, 2019)

University of Stirling (UoStirling) (Table 1).

Only samples including information on $R_{rs}(\lambda)$, chla concentration, total suspended matter (TSM) dry weight, absorption by CDOM at 443 nm ($a_{\mathrm{CDOM}}(443)$) and water transparency (Z_{SD}) were selected, with chla concentration ≤ 10 mg m^{-3} and TSM dry weight ≤ 4 g m^{-3}. We limit this study to oligo- and mesotrophic lakes and reservoirs whose chla concentration does not exceed 10 mg m^{-3} because up to this threshold chla estimation uncertainties were found to vary and increase most substantially (Gilerson et al., 2010; Mouw et al., 2013; Neil et al., 2019; Shi et al., 2013; Smith et al., 2018). Across the U.S. state of Wisconsin *in situ* measurements were made at 18 lakes between spring 2014 and autumn 2016 (n = 46). The dataset was used in recently conducted algorithm development and inter-comparison studies (Pahlevan et al., 2021b, 2021a, 2020). Collection methods are detailed in Appendix A. The datasets and methods used to measure the water constituent parameters and IOPs in LIMNADES are described in Spyrakos et al. (2018) and publications associated with individually contributed datasets (see references in Table 1). IOP measurements were not available for all observations in this study and often only at 443 and 673 nm. $R_{rs}(\lambda)$ was measured just above the water surface at varying spectral resolutions, and provided at a common 1 nm resolution, interpolated from the contributed $R_{rs}(\lambda)$ rather than individual radiometric components. Further *in situ* samples were collected in seven lakes in the United Kingdom, Sweden and Switzerland between 2013 and 2016 (Table 1). A detailed description of the measurements is provided in Appendix B. $R_{rs}(\lambda)$ from the Wisconsin DNR and UoStirling datasets were interpolated to 1 nm spectral resolution to match the LIMNADES measurements. All three datasets were spectrally convolved to the band configurations of the hyperspectral HICO and multispectral OLCI sensors. Each chla algorithm tested in this study was used with the OLCI band configuration. HICO was included to test algorithms designed for hyperspectral configurations for the influence of the additionally available bands on both retrieval performance and associated uncertainty. The resulting dataset consisted of 346 observations from lakes and reservoirs located in the boreal and tundra zones of North America, Canada, and Europe. In addition, moderate-altitude and perialpine lakes (Italy and Switzerland) and clear water reservoirs of Spain and South Africa were included (Fig. 1).

3. Methods

The analysis of algorithm retrieval uncertainty consists of three parts (Fig. 2). First, explanatory variables are selected to construct an uncertainty model. Second, chla concentration is estimated for the collated dataset and the RVD is derived per algorithm. The uncertainty model is then fit to the RVD of the algorithms to quantify the retrieval uncertainty. Finally, the uncertainty model outcome is forwarded to SHAP to explain the variable impact on the model and to identify drivers of the quantified retrieval uncertainty.

3.1. Explanatory variables and uncertainty model definition

The initial set of explanatory variables consisted of the following measurements and indices: $a_{\mathrm{CDOM}}(443)$, $a_{\mathrm{NAP}}(443)$, $a_{\phi}(443)$ and $a_{\phi}(673)$, $\frac{R_{rs}(400)}{R_{rs}(673)}$, TSM and ISM (the inorganic fraction of suspended matter), $\frac{\mathrm{chla}}{\mathrm{TSM}}$ and Z_{SD}. This set of explanatory variables was based on 89 out of 346 dataset records for which all the variables were available. These components of the dataset characterise the relative magnitude of IOPs in a sample, and partially its biogeochemical composition, as well as how biological components could be acclimated to ambient light conditions. Furthermore, they relate to water column clarity and phytoplankton variability.

ML in aquatic remote sensing is conventionally used to build a predictive model to estimate some quantity from $R_{rs}(\lambda)$ such as phytoplankton biomass (Pahlevan et al., 2020; Werther et al., 2021). Here we used a non-linear RandomForest (RF) ML approach to model expected complex RVDs. Given the explanatory variables as inputs we fit a unique RF to the RVD of each chla algorithm. We then investigated the RVD quantification by the RF and how the explanatory variables, used as the input, impacted the RF model. Therefore, RFs in this study were used as regression models that relate the explanatory variables to the retrieval uncertainty of each chla algorithm.

A RF model is based on decision trees (DT), which split a variable space consisting of c variables $c_1, c_2, \cdots, c_n$ into j distinct, unique regions $R_1, R_2, \cdots, R_j$ (Breiman, 2001). DTs are constructed from a sample drawn with replacement (also known as bagging) (Breiman, 1996). A RF exposes several hyper-parameters, most notably the number and maximum depth of the constructed trees, which were tuned as they guide the model construction procedure (Bergstra et al., 2012). The aim was to fit a chla algorithm RVD as accurately as possible through a RF. Since the RFs are not used for prediction, the fitting was optimised through v-fold cross validation (where v was optimised for each RF model) and a search over a large pre-defined grid of the hyper-parameters. Appendix C features a comparison of ML approaches for the uncertainty modelling task.

To reduce multicollinearity between explanatory variables (e.g., $a_{\phi}(443)$ and $a_{\phi}(673)$), the initial set of explanatory variables was filtered using the variance inflation factor (VIF). This procedure regressed each

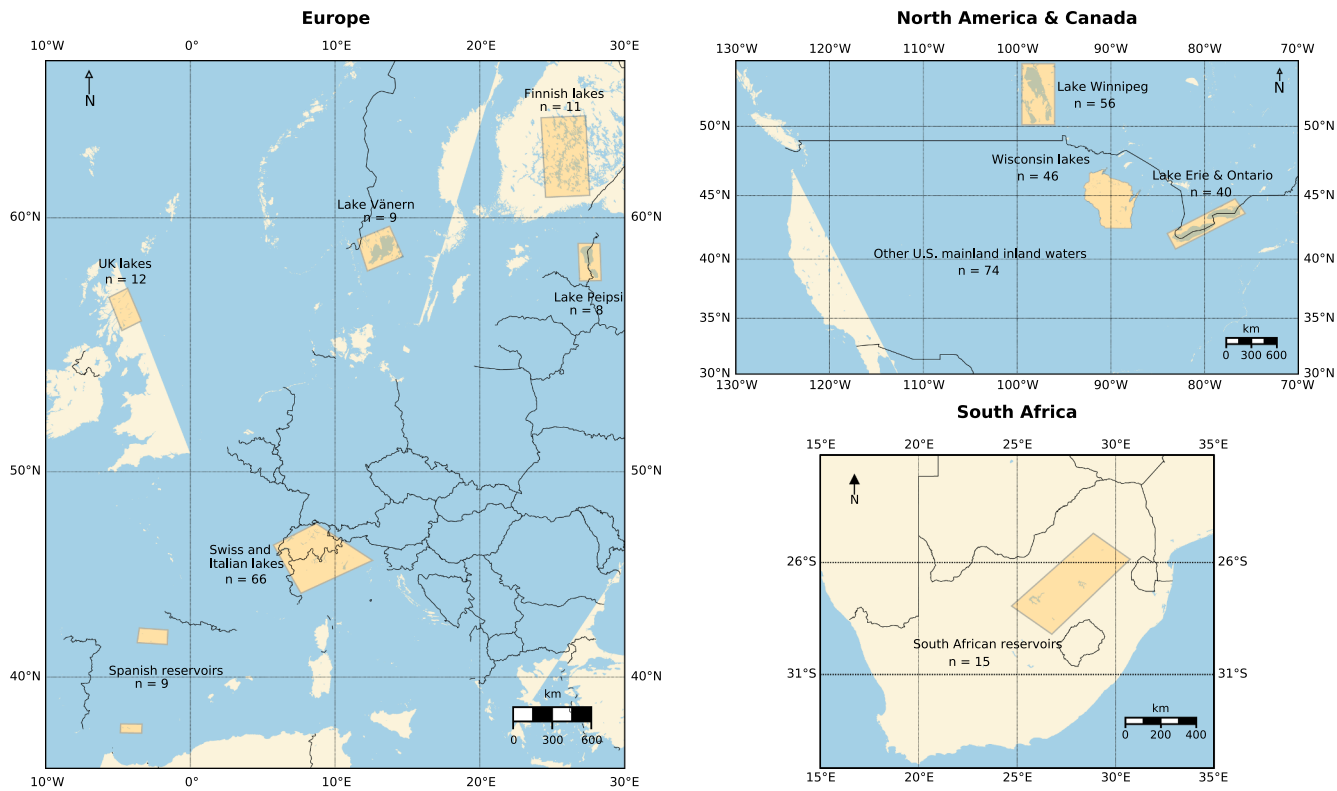

Fig. 1. Locations of the measurement sites of this study. *n* is the number of observations taken in the indicated area. See Table 1 for details about the dataset.

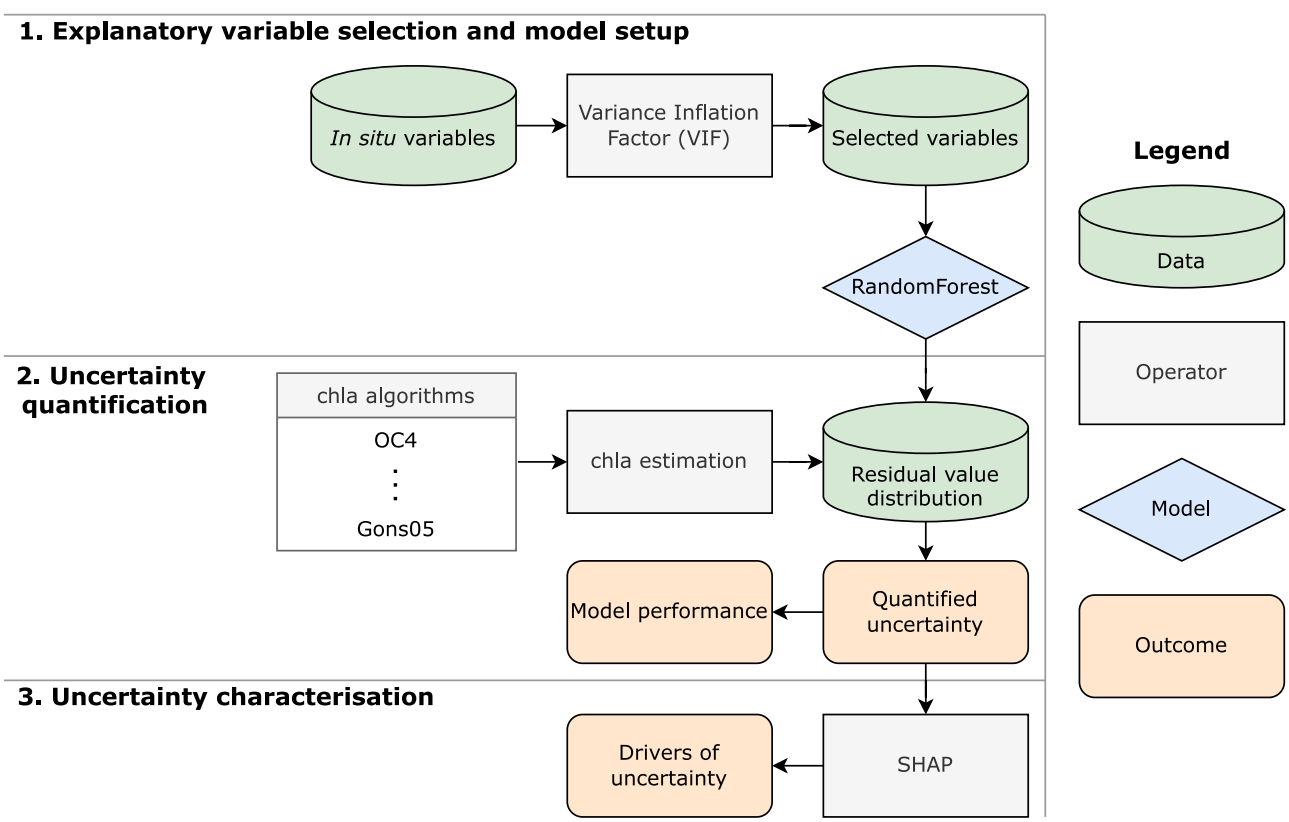

Fig. 2. Uncertainty analysis scheme to characterise the retrieval uncertainties of the tested chla algorithms.

Table 2
Chlorophyll-*a* algorithms used in this study.

Algorithm number	Algorithm abbreviation	Equation	k	l	m	n	o	Reference(s)
I	OC3 -opt	$\text{Chla_OC3} = 10^{(k+lX+mX^2+nX^3+oX^4)}$ $X = log_{10}\left(\frac{max[R_{rs}(443), R_{rs}(490)]}{R_{rs}(560)}\right)$	0.088171	−0.65440	8.56389	17.80220	5.89746	(O'Reilly et al., 1998; O'Reilly and Werdell, 2019)
II	OC4 -opt	$\text{Chla_OC4} = 10^{(k+lX+mX^2+nX^3+oX^4)}$ $X = log_{10}\left(\frac{max[R_{rs}(443), R_{rs}(490), R_{rs}(510)]}{R_{rs}(560)}\right)$	0.136257	−1.084227	9.14503	4.79545	−46.36124	
III	OC6 -opt	$\text{Chla_OC6} = 10^{(k+lX+mX^2+nX^3+oX^4)}$ $X = log_{10}\left(\frac{max[R_{rs}(412), R_{rs}(443), R_{rs}(490), R_{rs}(510)]}{mean[R_{rs}(560)/R_{rs}(665)]}\right)$	0.48767	−1.84059	−0.27403	7.55868	−3.59798	
IV	G11 -opt	$\text{Chla_G11} = kX^2 + lX + m$ $X = \frac{R_{rs}(708)}{R_{rs}(665)}$	12.0259	−4.9116	1.2115			(Gurlin et al., 2011)
V	FLH	$\text{FLH} = R_{rs}(681) - \left(R_{rs}(708) + (R_{rs}(665) - R_{rs}(708)) \times \left(\frac{(708-681)}{(708-665)}\right)\right)$						(Gower et al., 1999)
VI	MPH	$\text{Chla_MPH_0} = 5.24 \times 10^9 MPH_0^4 - 1.95 \times 10^8 MPH_0^3 + 2.46 \times 10^6 MPH_0^2 + 4.02 \times 10^3 MPH_0 + 1.97$ $\text{Chla_MPH_1} = 22.44 \times exp(35.79\ MPH_1)$						(Matthews et al., 2012; Matthews and Odermatt, 2015)
VII VIII	MDN HICO MDN OLCI	$\text{Chla}^e\text{_MDN} = \mu_i(R_{rs}) : i = \text{argmax}\ \sigma(R_{rs})$						(Pahlevan et al., 2021b, 2020)
IX	QAAv6 (443) QAAv6 (673)	$a_\phi(\lambda) = a(\lambda) - a_{dg}(\lambda) - a_w(\lambda)$						(Lee et al., 2002)
X XI	GSM HICO GSM OLCI	$a_\phi(\lambda) = chla \times k$ $a_{CDOM}(\lambda) = a_{CDOM}(\lambda_0) exp(-l(\lambda - \lambda_0))$ $b_{bp}(\lambda) = b_{bp}(\lambda_0)\left(\frac{\lambda}{\lambda_0}\right)^m$	0.029 (443 nm) and 0.018 (673 nm)	0.017	1.0337			(Garver and Siegel, 1997; Maritorena et al., 2002)
XII	3SAA	$\text{Chla_3SAA} = q_i/(q_2 + q_3 + q_4)$, where $i = 2, 3$ and 4 represent pico-, nano- and microplankton absorption, respectively						(Jamet et al., 2012; Jorge et al., 2021; Loisel et al., 2018; Loisel and Stramski, 2000)
XIII	Gons05	$\text{Chla_Gons05} = \left[\left(\frac{R_{rs}(708)}{R_{rs}(665)}\right) \times (0.7 + b_b) - 0.4 - b_b^k\right] \Big/ l$ $b_b = \frac{0.6 \times a_w(778) \times R_{rs}(778)}{0.082 - 0.6 \times R_{rs}(778)}$	0.016	1.063				(Gons et al., 2008, 2005, 2002)

Table 3
$a_\phi(\lambda)$ to chla concentration scaling coefficients used in Eq. (2) for the QAAv6, GSM and 3SAA algorithms derived from the dataset of this study.

Algorithm number	Algorithm configuration	v	w
IX	QAAv6 (443)	0.0150	2.1117
IX	QAAv6 (673)	0.0033	2.8459
X	GSM HICO (443)	0.0015	3.4482
XII	3SAA (443)	0.0027	3.0552

variable against all others, with the VIF calculated as:

$$VIF = \frac{1}{1 - r^2}, \tag{1}$$

where r^2 is the coefficient of determination. A VIF value > 10 for an explanatory variable indicates high multicollinearity, in which case the variable that resulted in the strongest decrease of VIF was excluded (Miles, 2014). The number of variables excluded/kept depends on the considered dataset and the explanatory variables which ultimately determine the degree of multicollinearity. In practice other variables such as the spectral slopes of IOPs or other spectral ratios than $\frac{R_{rs}(400)}{R_{rs}(673)}$ can be used as explanatory variables. Variables should provide new knowledge about the observation conditions not already covered and must be sufficiently de-correlated to pass the VIF selection.

3.2. Chlorophyll-a retrieval

A complete list of the chla algorithms evaluated in this study is provided in Table 2. We optimised the coefficients of the algorithms where possible (denoted as -opt) with the present dataset (denoted as k, l, m, n, o in Table 2). Several red/NIR approaches exist that use the 665 and 708 nm bands (e.g., Gilerson et al., 2010; Mishra and Mishra, 2012; Moses et al., 2009). For this study we selected the Gons05, G11 and MPH algorithms as they are commonly used over inland waters. Where previous studies already suggest merging of approaches (Schaeffer et al., 2022; Smith et al., 2018), we have only included their component algorithms to enable attribution of uncertainty to algorithm-specific causes.

The $a_\phi(443)$ and $a_\phi(673)$ estimates of the Quasi-Analytical Algorithm Version 6 (QAAv6), Garver-Siegel-Maritorena (GSM) and the 3-Step Semi-Analytical Algorithm (3SAA) algorithms were scaled to chla concentration using the relationship by Bricaud et al. (1998):

$$chla = \left(\frac{a_\phi(\lambda)}{v}\right)^{\frac{1}{w}}, \tag{2}$$

where v and w were estimated from the dataset of this study through non-linear least squares fitting. SAA performance highly depends on known parameter values. The slope of CDOM absorption S (0.017 m^{-1}) was estimated from our dataset using the method described in Stedmon et al. (2000). The mass-specific chla absorption coefficients $a^*_\phi(443)$ and $a^*_\phi(673)$ were estimated from observations where $a_\phi(\lambda)$ was available (n

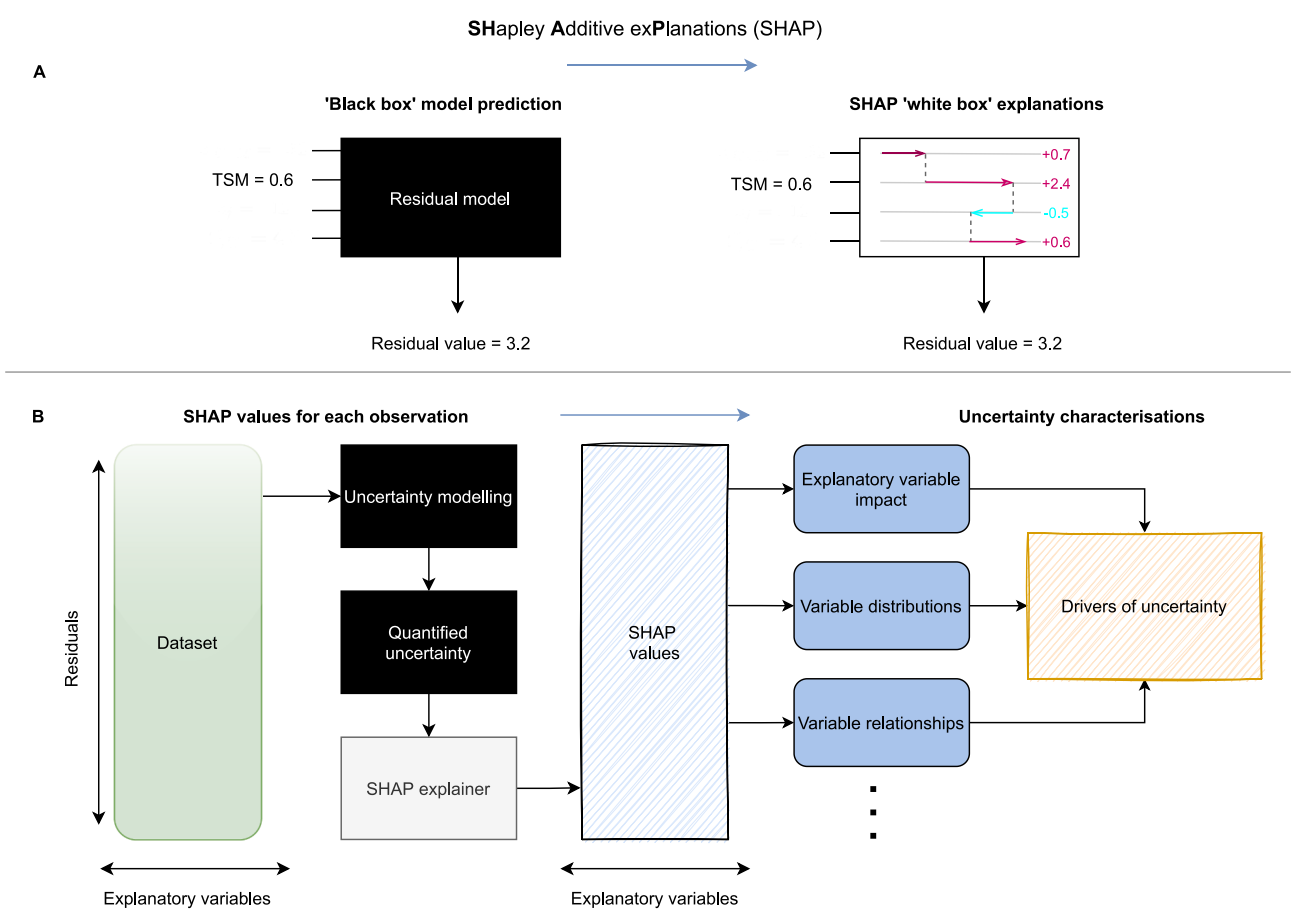

Fig. 3. SHapley Additive exPlanations (SHAP) procedure. (A) Conceptual difference between standard black box ML model estimates and SHAP white box explanations. (B) Retrieval uncertainty is quantified through uncertainty models. The model is then provided to the SHAP explainer to calculate SHAP values. SHAP values provide the impact an explanatory variable had on an uncertainty model, enable to draw variable distributions and to investigate explanatory variable relationships. These drivers of uncertainty are used to characterise retrieval uncertainties.
Adapted from Lundberg et al. (2020)

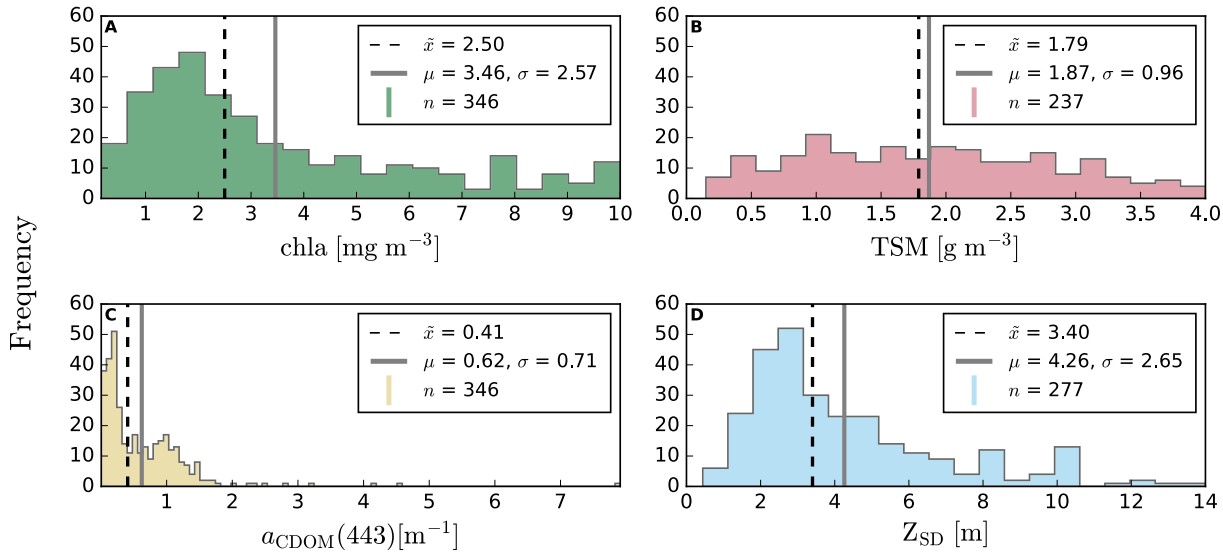

Fig. 4. Parameter distributions of the dataset. (A) chla, (B) TSM, (C) $a_{CDOM}(443)$ and (D) Z_{SD}. Denoted are the median ($\tilde{x}$), mean (μ) and standard deviation (σ) of the respective parameter.

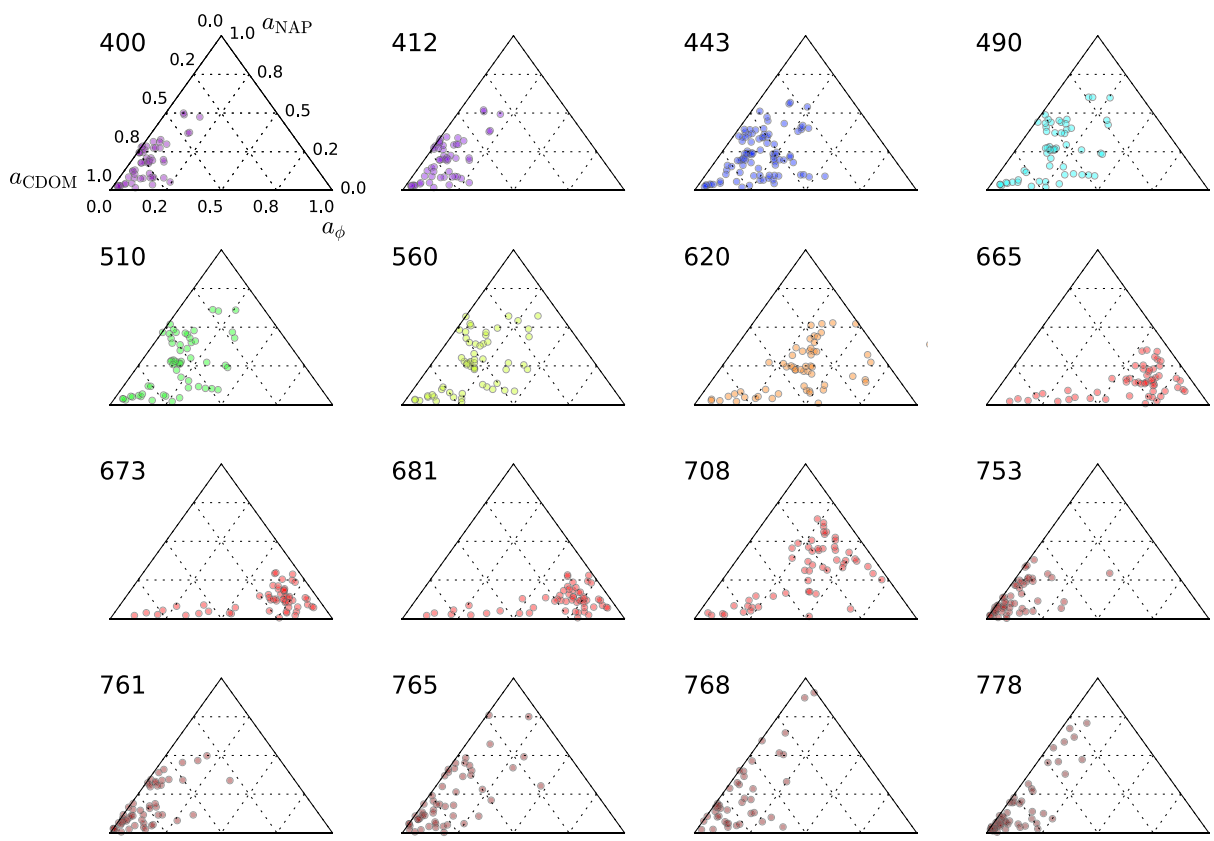

Fig. 5. Ternary diagrams of absorption $\left(a_{\phi}(\lambda),\ a_{\mathrm{NAP}}(\lambda)\ \text{and}\ a_{\mathrm{CDOM}}(\lambda)\right)$ for OLCI bands. Measurements where $a_p(\lambda) < 0.01$ were excluded.

= 89), as 0.029 m^{2}g^{-1} and 0.018 m^{2}g^{-1}, respectively. Table 3 lists the scaling coefficients used in Eq. (2) for the SAA algorithms.

In contrast to the empirical algorithms, as well as to the Gons05 algorithm, the QAAv6, GSM and 3SAA algorithms were not primarily developed to assess chla concentration but IOPs such as $a_{\phi}(\lambda)$. These three algorithms have been thoroughly assessed over a large range of predominantly ocean and coastal water datasets. The latest IOP (not chla) retrieval exercise of the QAAv6, GSM and 3SAA algorithms was performed over 1020 observations and showed MAPD values of 28%, 40% and 31%, respectively (Jorge et al., 2021). Since chla concentration retrieval via SAA depends on $a_{\phi}(\lambda)$, for these three algorithms we also display the SAA $a_{\phi}(\lambda)$ estimates versus the respective *in situ* reference $a_{\phi}(\lambda)$ values. For the uncertainty analysis we use the GSM and 3SAA standard chla outputs as described in the original publications.

The Mixture Density Networks (MDNs) and Maximum Peak Height (MPH) algorithms were only applied to a subset of the dataset. Part of

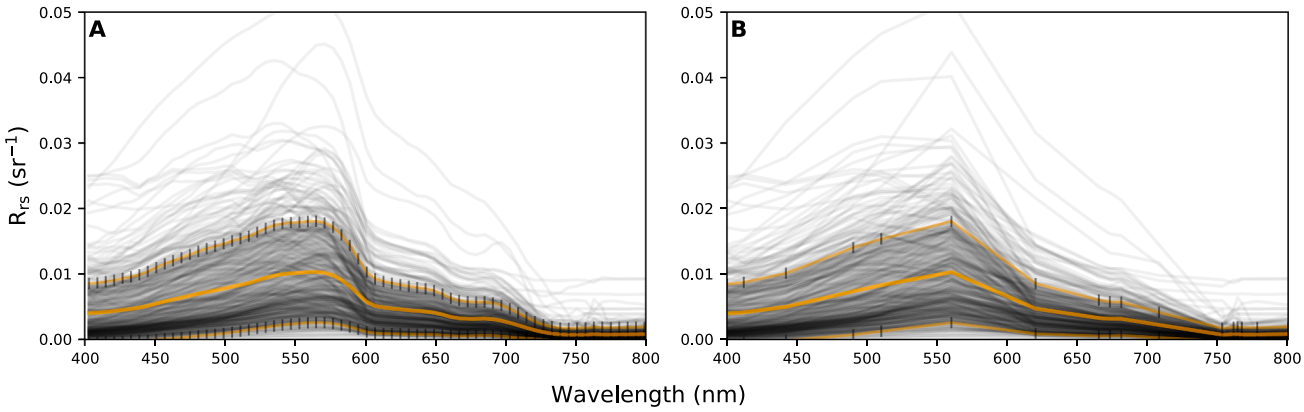

Fig. 6. $R_{rs}(\lambda)$ of the dataset. (A) Hyperspectral HICO resolution. (B) Multispectral OLCI resolution. Solid orange lines depict mean +- standard deviation and black vertical lines the band positions of the sensors.

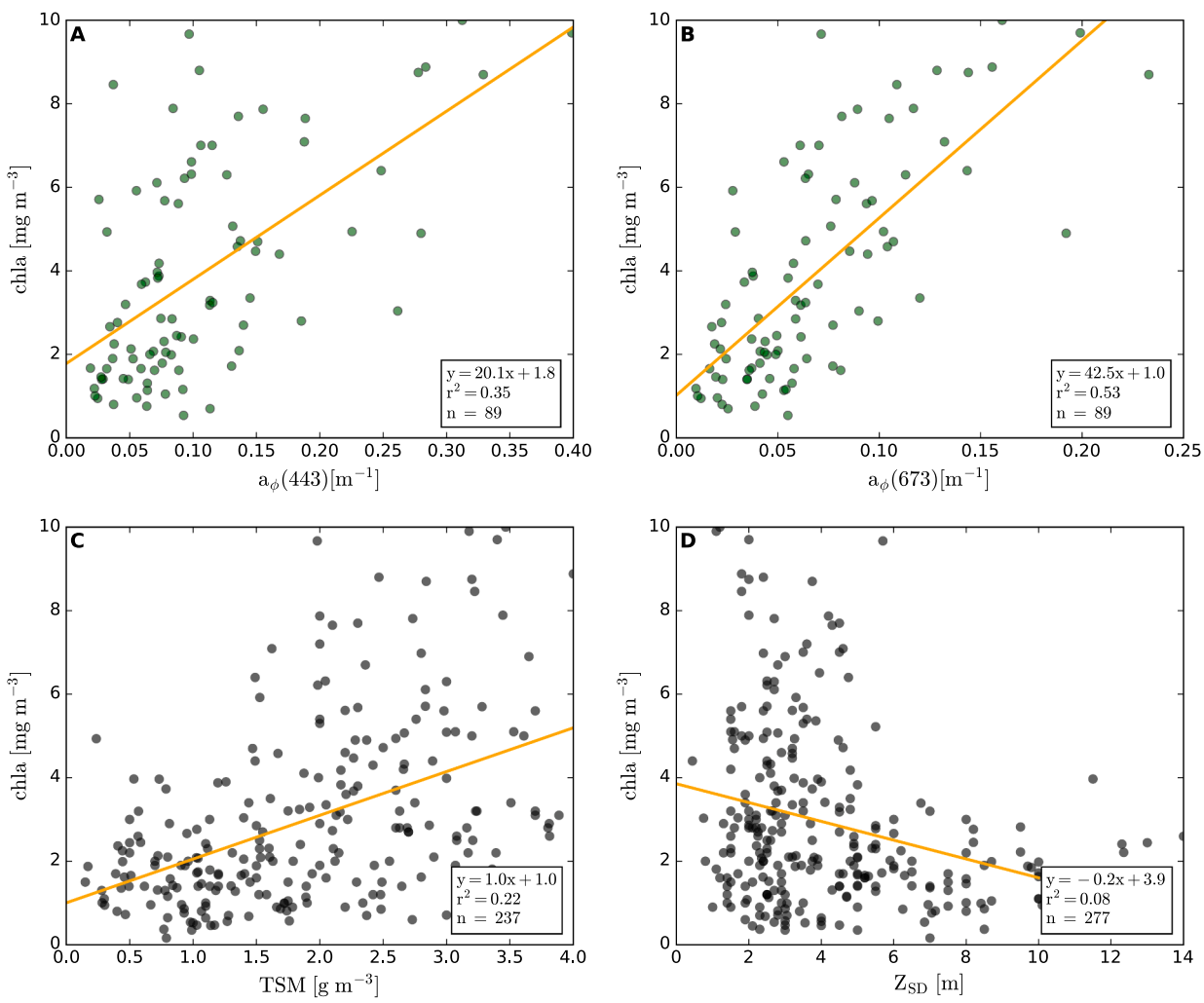

Fig. 7. Relationships of chla concentration to other dataset parameters. Chla versus (A) $a_\phi(443)$, (B) $a_\phi(673)$, (C) TSM, (D) Z_{SD}.

the dataset used in the training of the MDNs overlaps with the Wisconsin DNR and LIMNADES datasets of this study. We thus removed overlapping observations from this study as the evaluation would otherwise not be independent. After removal, 246/346 (71%) observations remained for application of the MDNs. MPH requires $R_{rs}(885)$, however for most observations in the dataset the wavelength range is restricted to 800 nm, hence limiting the evaluation of the MPH to 91 out of 346 observations.

3.3. Uncertainty characterisation

To overcome the drawbacks of standard variable importance measures (Altmann et al., 2010; Grömping, 2009; Strobl et al., 2008) we use the SHAP procedure (Lundberg and Lee, 2016; Lundberg et al., 2020,

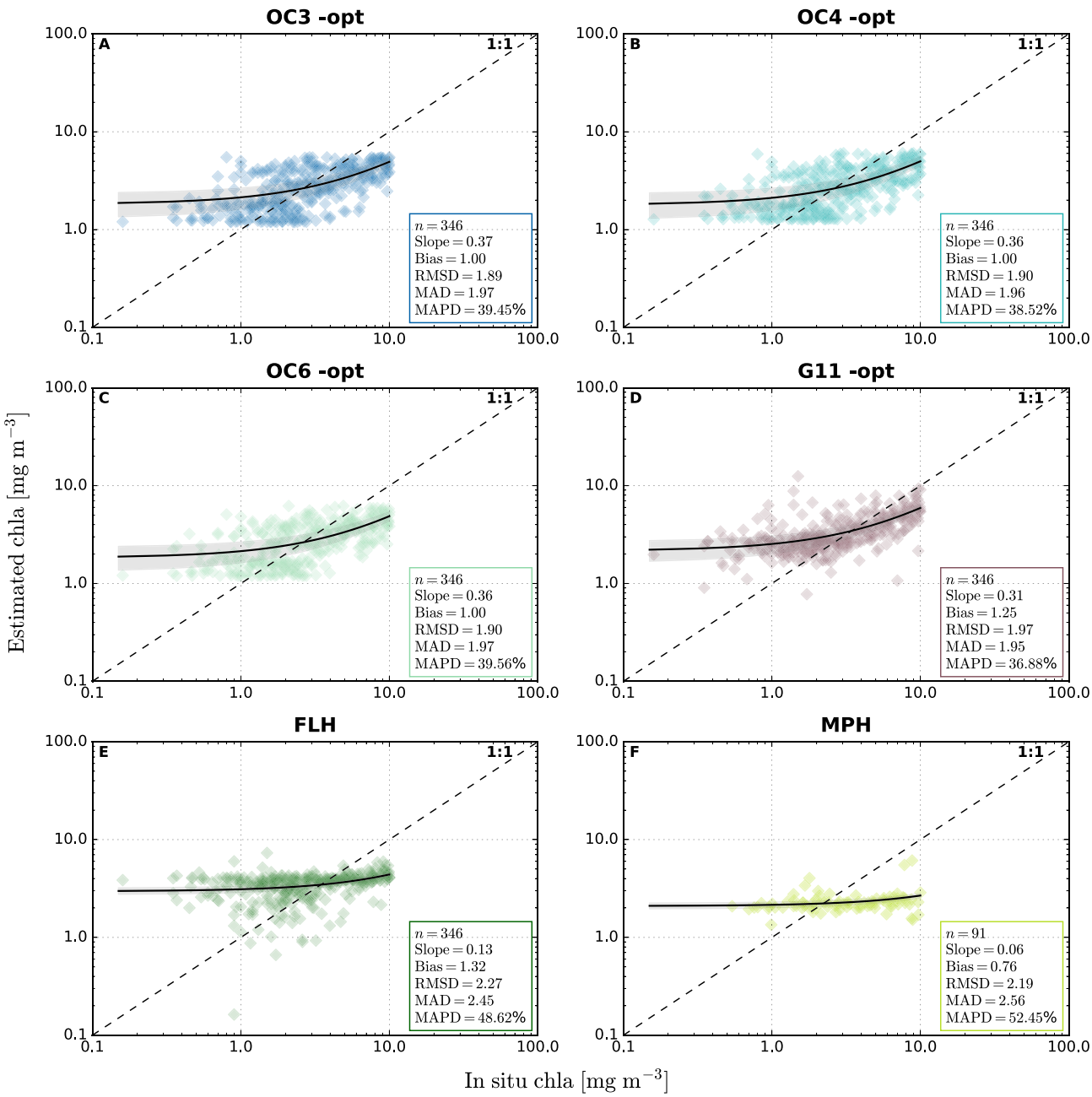

Fig. 8. Chla concentration estimates of the OC -opt (I – III), G11 -opt (IV), FLH (V) and MPH (VI) chla algorithms. Black solid lines represent regression lines with 95% confidence intervals.

2017). Technical details about SHAP and standard variable importance measures are provided in Appendix D.

We quantify the RVD of a chla algorithm with an uncertainty model $f(c^*)$, whose input is an explanatory variable vector $c = c^*$ (Fig. 3). In Shapley game theory, $S \subseteq n = \{1, \ldots, n\}$ is an ordered subset that consists of $|S|$ variables (the coalition). Herein, a contribution function $v(S)$ maps subsets of variables to their real number, which is defined as the marginal contribution of an ordered subset of variables to the model's estimate. A Shapley value $\phi_c(v)$ is the contribution that an explanatory variable c provides to the model:

$$\phi_c(v) = \phi_c = \sum_{S \subseteq n \setminus \{c\}} \frac{|S|!(n - |S| - 1)!}{n!} (v(S \cup c) - v(S)), c = 1, \cdots, n. \quad (3)$$

In Eq. (3), the ordered subsets (coalitions) of variables are formed one at a time. The marginal contribution of each variable to the uncertainty model towards the estimation of the output value is calculated through $(v(S \cup c) - v(S))$. Then, for each variable, the average of this contribution over all orderings of all possible ordered subsets is computed. This yields the weighted mean over the uniquely ordered subsets. The sign of ϕ_c provides information about an explanatory variable's effect on the output. A positive value indicates that a variable c increases the model's estimated output value for an input observation. Conversely, a negative value indicates a decrease of the output value (see white box in Fig. 3A). The values of a variable may thus be related to both low and high residual values. The magnitude of ϕ_c indicates the impact a variable has on a model. Variable candidates can be identified by ranking the variables in order of impact on the model.

3.4. Performance metrics

Chla algorithm performance was expressed in terms of standard

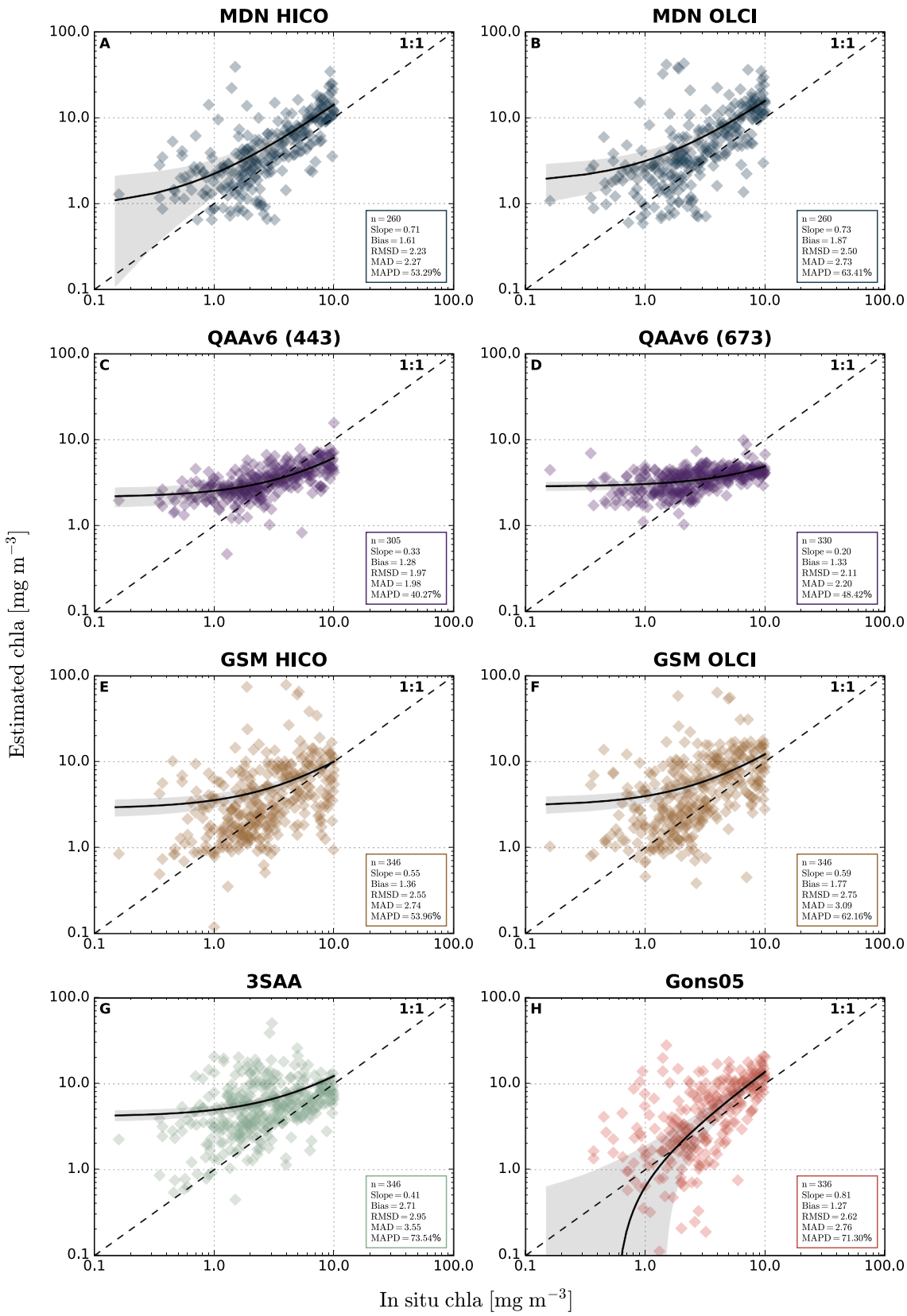

Fig. 9. Chla concentration estimates of the MDN (VII - VIII), QAAv6 (IX), GSM (X - XI), 3SAA (XII) and Gons05 (XIII) algorithms. (A, B) MDN HICO and OLCI. (C) QAAv6 (443). Scaling of $a_\phi(443)$ to chla concentration via Eq. (2) with dataset coefficients (see Table 3 for the coefficients). (D) QAAv6 (673). Scaling of $a_\phi(673)$ to chla concentration with dataset coefficients. (E) GSM HICO, (F) GSM OLCI, (G) 3SAA and (H) Gons05 standard chla concentration estimates.

Table 4
Performance metrics of the chla algorithms applied to the full dataset (n = 346). A retrieval failure means that no chla concentration was returned for an observation.

Algorithm number	Algorithm configuration	Retrieval failures	Slope	Bias	RMSD	MAD	MAPD
I	OC3 -opt	0	0.37	1	1.89	1.97	39.45%
II	OC4 -opt	0	0.36	1	1.90	1.96	38.52%
III	OC6 -opt	0	0.36	1	1.90	1.97	39.56%
IV	G11 -opt	0	0.31	1.25	1.97	1.95	36.88%
V	FLH	0	0.13	1.32	2.27	2.45	48.62%
VI	MPH	0	0.06	0.76	2.19	2.56	52.45%
VII	MDN HICO	0	0.71	1.61	2.23	2.27	53.29%
VIII	MDN OLCI	0	0.73	1.87	2.50	2.73	63.41%
IX	QAAv6 ($a_\phi(443)$ to chla)	41	0.33	1.28	1.97	1.98	40.28%
	QAAv6 ($a_\phi(673)$ to chla)	16	0.20	1.33	2.11	2.20	48.42%
X	GSM HICO	0	0.55	1.36	2.55	2.74	53.96%
XI	GSM OLCI	0	0.59	1.77	2.75	3.09	62.16%
XII	3SAA	0	0.41	2.71	2.95	3.55	73.54%
XIII	Gons05	10	0.81	1.27	2.62	2.76	71.30%

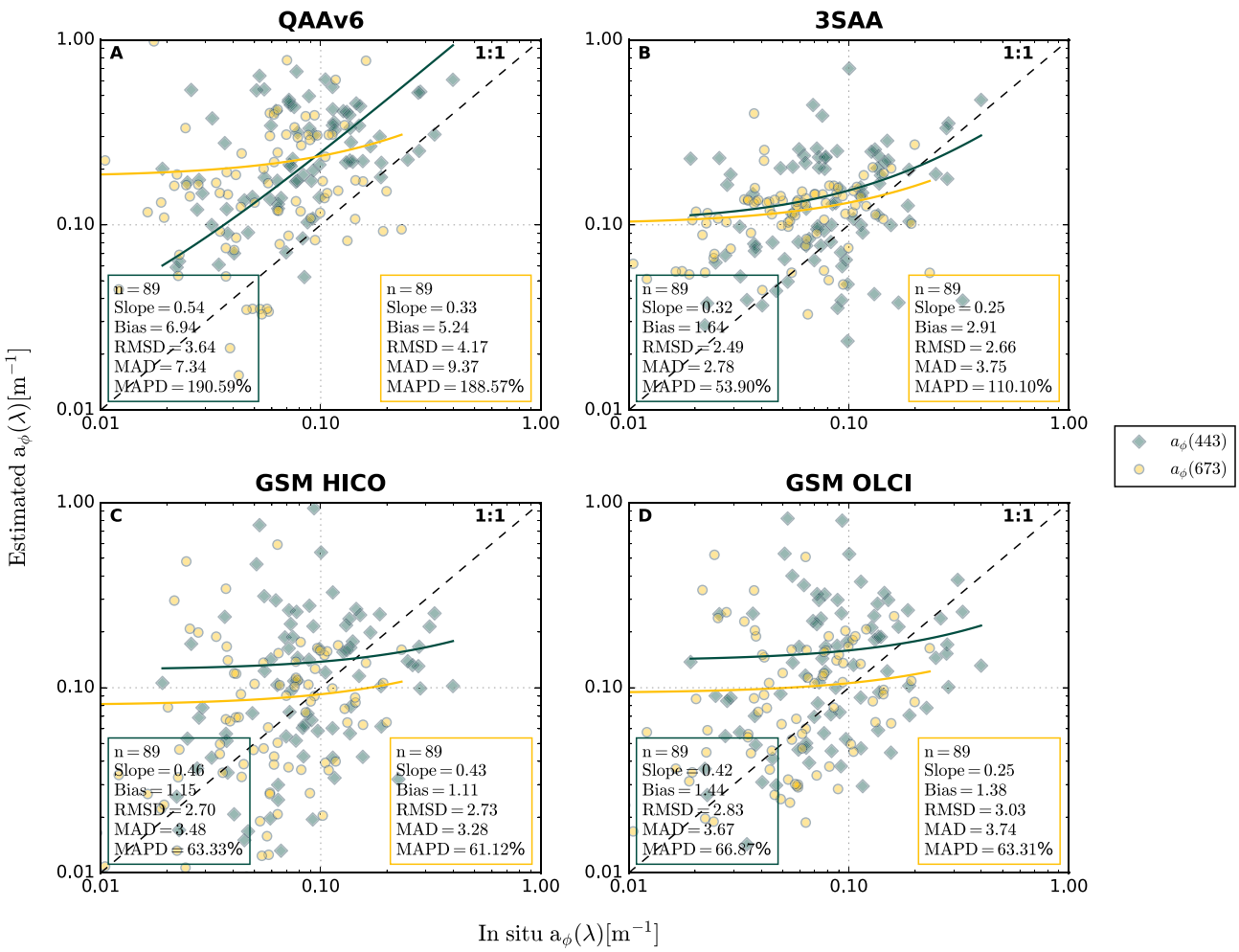

Fig. 10. $a_\phi(443)$ and $a_\phi(673)$ estimates (green and yellow, respectively) of (A) QAAv6 (IX), (B) 3SAA (XII), (C, D) GSM HICO and OLCI algorithms (X - XI).

metrics such as bias, mean absolute difference (MAD), median absolute percentage difference (MAPD) and root mean square difference (RMSD) between the algorithm estimates (E) and the *in situ* chla observations (O) (Morley et al., 2018; Pahlevan et al., 2020; Seegers et al., 2018). We also report the slope of the regression to infer whether an algorithm systematically over- or underestimated chla or $a_\phi(\lambda)$ over the entire range of values. The MAD, MdAD (median absolute difference, no percentage) and RMSD metrics were used to measure the RVD fit accuracy of the RFs. All metrics were first calculated in logarithmic space and then transformed back to linear scale for display following recommended practice by Seegers et al. (2018). Linear scale enables to directly compare the chla performance metrics to the derived RVDs of the algorithms and fit accuracy of a RF uncertainty model.

4. Results

4.1. Optical and biogeochemical properties of the dataset

In situ data distributions show tailed distributions of chla (limited to 10 mg m^{-3}) with a median of 2.5 mg m^{-3} and standard deviation (std.

Table 5
RF uncertainty quantification accuracies of the RVD from each chla retrieval algorithm included in the uncertainty analysis.

Chla algorithms	MdAD [mg m^{-3}]	MAD [mg m^{-3}]	RMSE [mg m^{-3}]
OC4 -opt	0.50	0.79	1.08
OC6 -opt	0.58	0.84	1.13
G11 -opt	0.68	1.05	1.51
MDN HICO	0.64	0.96	1.26
QAAv6 (443)	0.62	0.96	1.39
GSM HICO	1.56	3.08	6.02
3SAA	0.92	1.56	2.42
Gons05	2.17	3.47	4.84

dev) of 2.57 mg m^{-3} (Fig. 4A). TSM concentrations (Fig. 4B) covered a broad range (limited to 4 g m^{-3}) with a median of 1.87 g m^{-3} (0.96 g m^{-3} std. dev). $a_{\mathrm{CDOM}}(443)$ was distributed around a median of 0.41 m^{-1} (0.71 m^{-1} std. dev) (Fig. 4C), while Z_{SD} had a median of 3.9 m (2.65 m std. dev) (Fig. 4D). Absorption contributions of $a_{\phi}(\lambda)$, $a_{\mathrm{NAP}}(\lambda)$ and $a_{\mathrm{CDOM}}(\lambda)$ attribute the highest fraction of absorption to $a_{\mathrm{CDOM}}(\lambda)$ and $a_{\mathrm{NAP}}(\lambda)$ from 400 through 560 nm (Fig. 5). $a_{\phi}(\lambda)$ only contributed about 20% to the total non-water absorption budget but was still the major absorption fraction in red OLCI bands. $R_{rs}(\lambda)$ of the band configurations of HICO and OLCI are displayed in Fig. 6, with a mean reflectance magnitude <0.01 sr^{-1} (Fig. 6A, B). Reflectance peaks in the red area of the spectrum were minor at 685 nm (fluorescence domain) and at longer wavebands (absorption and scattering domain). The relationships of $a_{\phi}(443)$ and $a_{\phi}(673)$ to chla concentration for this dataset were weak - moderate (Fig. 7A, B). Low TSM concentrations were weakly associated with low chla concentrations (Fig. 7C), whereas higher TSM concentrations did not exhibit a linear relationship to increasing chla concentration. Similarly, chla concentration was not correlated to Z_{SD} below <4–5 m. A clearer trend emerged between records with low chla concentration and increased transparency at $Z_{SD} > 6$ m (Fig. 7D).

4.2. Chlorophyll-a algorithm performance

Chla concentration was estimated from all records in the dataset using the algorithms listed in Table 2 (Figs. 8, 9). Performance metrics for each chla algorithm are detailed in Table 4. In addition, $a_{\phi}(443)$ and $a_{\phi}(673)$ values produced by the QAAv6, GSM and 3SAA algorithms are compared against $a_{\phi}(\lambda)$ *in situ* reference observations (Fig. 10).

The GSM HICO and 3SAA algorithms produced a chla standard estimate via their non-linear optimisation routine. For these algorithms we also provide the results of scaling estimated $a_{\phi}(443)$ to chla concentration. This enables a direct comparison to the $a_{\phi}(443)$ estimates of the QAAv6 algorithm.

Chla retrieval through the OC3 -opt, OC4 -opt and OC6 -opt band ratios was only partially accurate across the considered range (MAD 1.97 mg m^{-3}, MAPD > 38%). The OC algorithms overestimated chla up until 5 mg m^{-3} and consistently underestimated it thereafter. G11 -opt slightly outperformed the OC algorithms (MAD 1.95 mg m^{-3} and MAPD of 36.88%) through more accurate retrieval > 5 mg m^{-3}, but like the OC algorithms also overestimated chla in the range below 5 mg m^{-3}. Similar to G11 -opt, Gons05 chla estimates started to become more accurate towards higher chla concentrations (>5 mg m^{-3}). The FLH algorithm was unable to detect a fluorescence signal for most observations in the dataset which resulted in a poor relationship to chla. Like the FLH, the MPH algorithm lacked sensitivity over the chla range considered here and was unable to reproduce chla concentrations > 3 mg m^{-3}. The MDN HICO configuration outperformed its MDN OLCI counterpart by 10–20% (depending on the considered metric). Until 5 mg m^{-3} of chla, both MDN HICO and OLCI over- or underestimated chla concentration and then consistently overestimated it for chla higher than 5 mg m^{-3} by 1 – 3 mg m^{-3}.

The QAAv6 algorithm yielded $a_{\phi}(443)$ and $a_{\phi}(673)$ (hereafter QAAv6 (443) and QAAv6 (673)) which included 41 and 16 negative $a_{\phi}(443)$ and $a_{\phi}(673)$ values, respectively. The negative values could not be used in the power law function of Eq. (2) and were therefore excluded from the algorithm performance metrics but kept in the $a_{\phi}(\lambda)$ comparison (Fig. 10). For the 305 remaining observations, QAAv6 (443) estimated chla similarly to the OC -opt and G11 -opt algorithms. Like the MDNs, the GSM HICO configuration outperformed the OLCI configuration by approximately 10%, but the algorithm was biased towards overestimation of chla concentration. The retrieval differences of the multispectral 3SAA algorithm were higher than the ones of the hyperspectral GSM HICO configuration (MAD 3.55 mg m^{-3} and 2.74 mg m^{-3}, respectively). For the QAAv6, GSM and 3SAA algorithms large retrieval differences (MAPD > 53%) were found in the comparison of algorithm estimated $a_{\phi}(\lambda)$ versus *in situ* $a_{\phi}(\lambda)$ (Fig. 10).

4.3. Uncertainty quantification

Uncertainty budgets were derived for all algorithms with some exceptions. OC3 -opt was excluded because of nearly identical performance and a similar RVD as OC4 -opt (not shown). FLH and MPH algorithm estimates had very low retrieval sensitivity over the tested chla concentration range and were therefore omitted. For the remaining algorithm set, the configuration with the lowest retrieval difference in terms of MAD and MAPD was selected. The MDN HICO compared favourably to its OLCI version, and QAAv6 (443) was more precise than QAAv6 (673). The Gons05 algorithm was not optimised and was thus included in its original form. For the GSM algorithms, we used the default GSM HICO chla estimates as they were more accurate than the OLCI configuration. For 3SAA, we used the chla concentrations resulting from addition of the three phytoplankton size classes.

The VIFs calculated for ISM and TSM (11.75 and 7.64), $a_{\phi}(443)$ (6.74) and $a_{\phi}(673)$(5.43) were higher than any of the other explanatory variables (not shown). The exclusion of ISM and $a_{\phi}(443)$ reduced the overall VIFs of the variables the most. Thus, seven explanatory variables remained that were used as the input for the uncertainty models. Results of the RF retrieval uncertainty quantification are depicted in Table 5. The RVDs of GSM HICO, 3SAA and Gons05 contain high residual values (see Fig. 11F, G, H). Retrieval quantification accuracies of the uncertainty models are thus to be viewed with respect to the RVD of each algorithm. For example, the MdAD of 1.56 mg m^{-3} for the GSM HICO uncertainty model can be considered accurate given the large residual value range (maximum 55.61 mg m^{-3} overestimate). Further, the Gons05 algorithm overestimated 11 observations by > 9 mg m^{-3}, resulting in the least accurate uncertainty model with a MdAD of 2.17 mg m^{-3} (see Fig. 11H). The RVDs of the OC4 -opt, OC6 -opt, G11 -opt, MDN HICO and OLCI and QAAv6 (443) algorithms ranged from −8 mg m^{-3} to 6 mg m^{-3}. Since the range of their RVDs is smaller than from the GSM HICO, 3SAA and Gons05 algorithms, the uncertainty models were able to quantify them with higher precision (e.g., OC4 MdAD of 0.50 mg m^{-3}, MAD of 0.79 mg m^{-3} and RMSE of 1.08 mg m^{-3}).

4.4. Uncertainty characterisation

SHAP values were calculated to assess the explanatory power of each explanatory variable within each RF uncertainty model (Fig. 12). The SHAP value magnitudes vary between chla algorithms because the associated RVDs of the chla algorithms are unique. An explanatory variable had impact on an uncertainty model if it was used in the estimation of an algorithm's RVD. Consequently, the variable impact is expressed by the corresponding SHAP values as either smaller or greater than 0. We examined each explanatory variable to identify drivers of uncertainty of the tested algorithms. Except for MDN HICO and Gons05, the chla/TSM ratio was an impactful driver in all algorithms (Fig. 13). Likewise, $a_{\phi}(673)$ was a highly impactful model variable for most models except for OC4 -opt, MDN HICO and Gons05. TSM was relevant for the OC4 -opt, OC6 -opt, G11 -opt, QAAv6 (443), GSM HICO, 3SAA

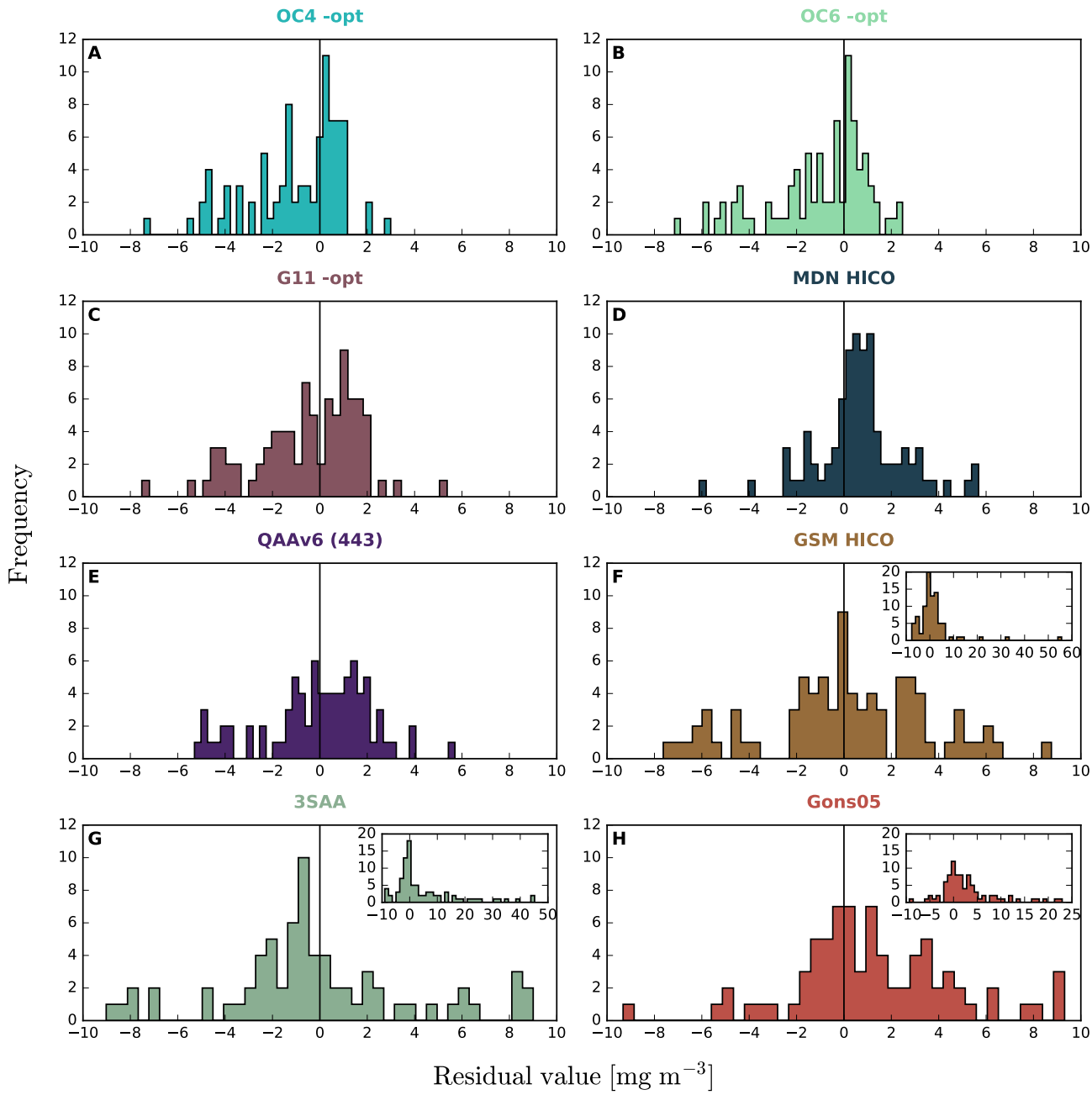

Fig. 11. Residual value distributions (RVDs) of the chla algorithms for the observations included in the uncertainty analysis (n = 89). For GSM HICO, 3SAA and Gons05 the RVDs were limited to 10 mg m^{-3} to facilitate a visual comparison of the distributions. The complete RVD of these three algorithms is displayed as an inset (F, G, H) and used in the uncertainty models.

and Gons05 uncertainty models. Especially for the OC4 -opt, G11 -opt and QAAv6 (443) algorithms, the only variables with a marked influence were the chla/TSM ratio, TSM and $a_{\phi}(673)$. In contrast, in the MDN uncertainty model $a_{\mathrm{NAP}}(443)$ had significant impact. $a_{\mathrm{NAP}}(443)$ had a marginal effect on the OC6 -opt uncertainty model and was not effective in the OC4 -opt and G11 -opt models. In contrast to the other variables of the OC6 -opt uncertainty model, high values of $a_{\mathrm{NAP}}(443)$ $a_{\mathrm{CDOM}}(443)$ decreased the model output variables. The other explanatory variables did not have a demonstrable effect on the OC6 -opt. $a_{\mathrm{CDOM}}(443)$ values also contributed to explain retrieval uncertainty of the MDN, with low absorption values having the largest impact on the model. Otherwise, $a_{\mathrm{CDOM}}(443)$ was only an impactful predictor for the QAAv6 (443), 3SAA, and Gons05 uncertainty models. Z_{SD} values were the least impactful considering the variable impact order across all models. Yet, for the uncertainty models of GSM HICO and Gons05, Z_{SD} was a relevant variable.

For GSM HICO low Z_{SD} values increased the model output, whereas medium to high values decreased it. The $\frac{R_{rs}(400)}{R_{rs}(673)}$ ratio was unimpactful for most models, except for GSM HICO, 3SAA and Gons05. For GSM HICO, low values of the $\frac{R_{rs}(400)}{R_{rs}(673)}$ ratio, i.e., where $R_{rs}(673) > R_{rs}(400)$, reduced the magnitude output, whereas high values, i.e., where $R_{rs}(400) > R_{rs}(673)$ increased the model output. For the 3SAA model, the ratio $\frac{R_{rs}(400)}{R_{rs}(673)}$ was highly impactful together with other variables related to phytoplankton absorption variability and composition, such as $a_{\phi}(673)$ and the chla/TSM ratio. The impact of $a_{\mathrm{CDOM}}(443)$, $a_{\mathrm{NAP}}(443)$ and Z_{SD} on the respective 3SAA uncertainty model was low, which relates to the 3SAA algorithm formulation. The combined

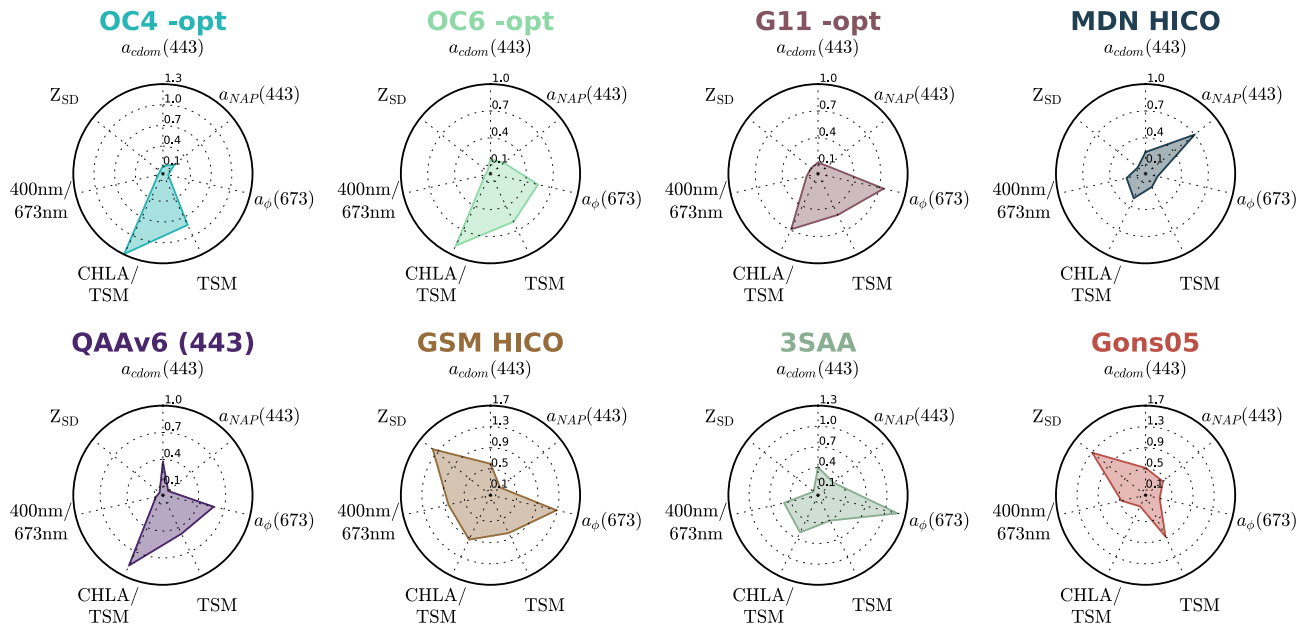

Fig. 12. Average SHAP values of the explanatory variables for each chla algorithm. The average SHAP value was calculated as the average of all the absolute SHAP values for each observation associated with an explanatory variable.

absorption of CDOM and NAP $(a_{dg}(\lambda))$ was estimated in the first step of the algorithm and after removal of non-water absorption $(a_{nw}(\lambda))$ the algorithm retrieved $a_{\phi}(\lambda)$. Hence, most of the influence of $a_{\mathrm{CDOM}}(443)$ and $a_{\mathrm{NAP}}(443)$ was incorporated by the algorithm formulation and the retrieval uncertainty of the algorithm was expectedly not related to these IOPs. $a_{\mathrm{CDOM}}(443)$ and $a_{\mathrm{NAP}}(443)$ were also irrelevant for the GSM HICO uncertainty model. The two SAA algorithms 3SAA and GSM HICO accounted for $a_{\mathrm{CDOM}}(443)$ and $a_{\mathrm{NAP}}(443)$ in their inversion procedure, but their retrieval uncertainty was driven by other factors such as phytoplankton absorption in red and variability (see Fig. 13). For Gons05, Z_{SD} and TSM had the strongest model impact. The values of Z_{SD} and TSM impacted the model in contrary directions: whereas high Z_{SD} values (i.e., higher water transparency) increased the model's output values, high TSM values (i.e., lower water transparency) decreased them. Contrary to the other models, especially G11 -opt, $a_{\phi}(673)$ and chla/TSM were the least impactful variables.

5. Discussion

While it is widely known that atmospheric disturbance on the water-leaving radiance and its correction contribute a large part to product uncertainty (Pahlevan et al., 2021a; Warren et al., 2019), it is equally important to understand algorithmic uncertainty arising from the translation of above-surface reflectance to the concentration of target substances. The approach presented in this study achieves this analysis using a highly heterogeneous collection of *in situ* observations and could be extended to inspect end-to-end product uncertainty (including AC effects) by starting from satellite observations.

The RF uncertainty models quantified the total uncertainty budget of a chla algorithm and did not distinguish between observational and algorithmic sources of uncertainty. Although most data used in this study have been previously published, an unknown degree of observational uncertainty is likely part of the *in situ* reference measurements. This may also include biases between datasets due to use of similar but not identical methodology. The observed systematic and large chla retrieval uncertainties by the algorithms imply that the primary cause of retrieval uncertainty were the algorithmic formulations. The information represented by the explanatory variables was accurately used by the RF models to quantify chla algorithm uncertainties (see Table 5), thereby demonstrating that retrieval uncertainties can be modelled through other *in situ* observations of the same dataset. We use the results of our uncertainty characterisation to discuss the algorithmic formulations tested in this study and highlight new findings about their retrieval uncertainty.

5.1. Blue/green algorithms

The reflectance shape and magnitude in the blue-green part of the spectrum in the evaluated dataset was largely influenced by CDOM and NAP. This contrasts the core assumption of the OC algorithms, namely that $R_{rs}(\lambda)$ co-varies with phytoplankton as the dominant optical component. Under- and overestimation of chla was large (residual values ranging from −5 mg m^{-3} to +4 mg m^{-3}). $a_{\mathrm{CDOM}}(443)$ and $a_{\mathrm{NAP}}(443)$ values were not used by the RF uncertainty models to estimate the RVDs of the OC4 and OC6 algorithms. In fact, the chla/TSM ratio and TSM were the most impactful variables. Thus, under dominating CDOM and NAP absorption, phytoplankton sources that influence the reflectance magnitude and shape variability, including varying phytoplankton cell size or the concentration of accessory photosynthetic and photoprotective pigments, could not be sensed by the OC algorithms.

5.2. Red/NIR algorithms

We observe that the G11 algorithm produced the most accurate chla values out of all the tested algorithms. G11 underestimated chla for concentrations higher than 4–5 mg m^{-3} but the estimates were relatively accurate when compared to the other algorithms within this concentration range. Conversely, for concentrations lower than 5 mg m^{-3} the G11 algorithm systematically overestimated chla values. This lack of sensitivity for concentrations lower than 5 mg m^{-3} has also been observed in other studies (Pahlevan et al., 2021b, 2020). The red/NIR area of the spectrum is the least affected by CDOM and NAP absorption. The RVD of the G11 algorithm was primarily modelled through the $a_{\phi}(673)$, TSM and chla/TSM ratio variables. The SHAP summary of the individual variables emphasised that high values of $a_{\phi}(673)$ were used to model the residual values of the G11 algorithm. We find that the red/NIR band-ratio of G11 does not suffice when the 665 and 708 nm bands hold low information about chla concentration. The Gons05 algorithm

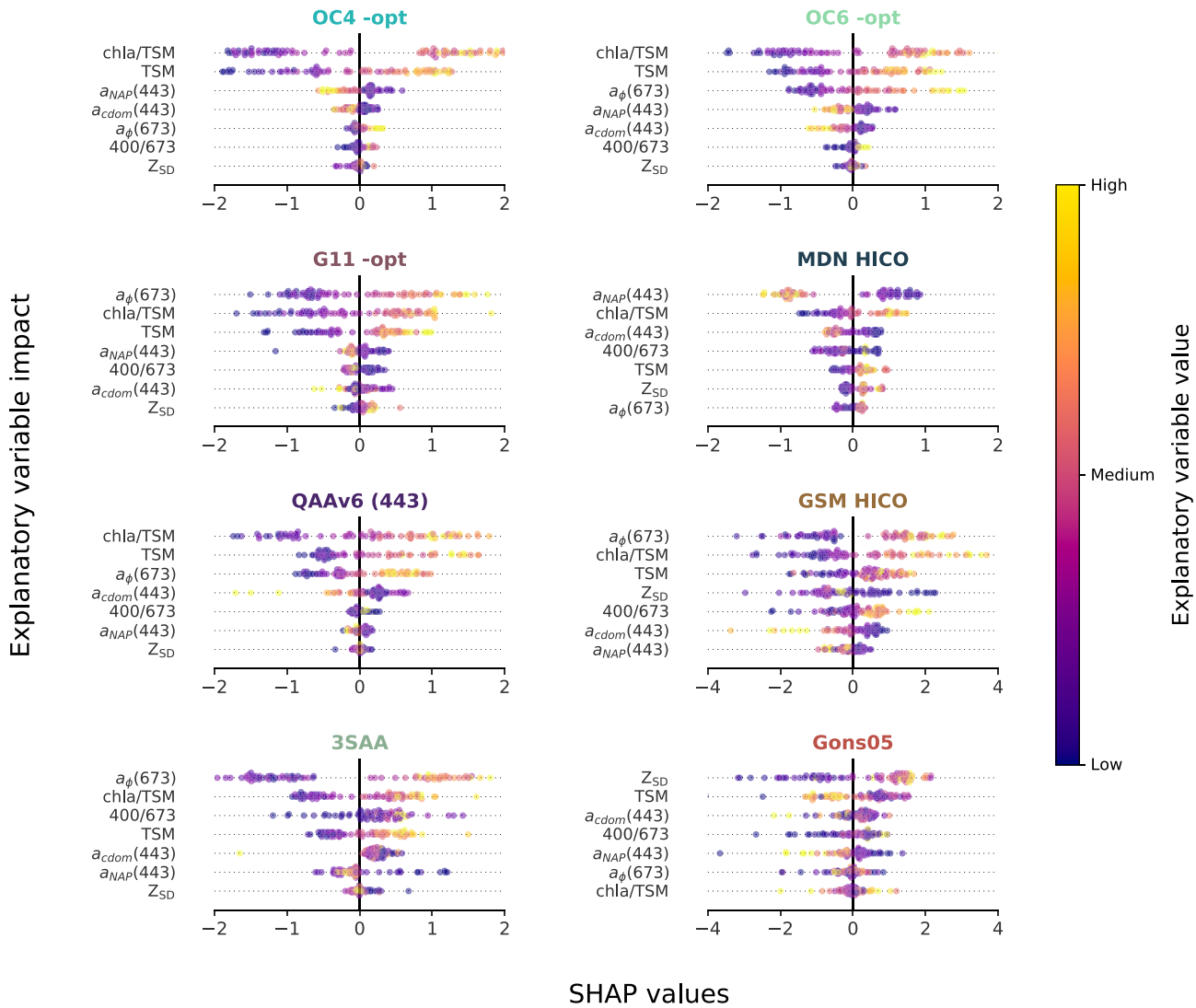

Fig. 13. Individual SHAP value impact on the RF uncertainty models. The individual SHAP values associated with each observation of each explanatory variable and algorithm are shown. The impact of an explanatory variable on the RF uncertainty model is plotted in descending order of impact. The colour represents the explanatory variable value (yellow high, blue low). The model impact of a variable can be both positive and negative (see Fig. 3A for more conceptual details). For display purposes, the variable $\frac{R_{rs}(400)}{R_{rs}(673)}$ is denoted as 400/673. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

was overall less accurate than G11, which was also found by Neil et al. (2019). The performance of the algorithm in this low – moderate chla concentration range confirms findings described in an earlier application of the algorithm to some of the Laurentian Great Lakes (Gons et al., 2008), where the retrieval differences were large for mesotrophic waters. The Gons05 retrieval uncertainty was most effectively modelled through Z_{SD} and TSM. $a_{\phi}(673)$ and chla/TSM only impacted the Gons05 uncertainty model marginally, however these two variables were substantial for the G11 uncertainty model. Both G11 and Gons05 are based on the 665 and 708 nm band ratio, thus reversed variable importance may seem surprising. For Gons05, Z_{SD} values were effective to estimate low residual values. In contrast, low-medium TSM values increased the uncertainty model output magnitude to model higher residual values (see Fig. 13). Z_{SD} and TSM are proxies of backscattering, which in turn is correlated to the calculation of the b_b term using 778 nm in the Gons05 algorithm. 778 nm is however not part of the G11 algorithm, thus explaining the difference in uncertainty model variable impact. Our analysis of variable contributions revealed that inaccurate backscatter approximation is a cause for the encountered Gons05 retrieval uncertainty. Optimisation of the backscatter term in the Gons05 algorithm might lead to performance improvements over the considered concentration range. An alternative pathway is to investigate whether the backscatter coefficient in the red/NIR range is too sensitive to low signal versus noise at 778 nm in oligo- and mesotrophic water bodies.

5.3. Machine learning algorithms

The MDN HICO and OLCI algorithms were previously trained with a semi-global dataset of inland and coastal systems. The dataset consisted of primarily eutrophic waters (chla dataset mean 21.7 mg m^{-3}, median 8.9 mg m^{-3} and a std. dev of 47.5 mg m^{-3}). Seemingly, oligo- and mesotrophic waters with low chla concentrations and variable levels of CDOM and NAP absorption were under-represented during training. Although the algorithm had a positive bias (>1.61), the slope of the regression from the MDN HICO model was moderate (0.71), which implies that chla estimation was fairly consistent across the considered concentration range. This finding is further supported by the algorithm's RVD (Fig. 11), in which most chla values were over- or underestimated by 1 – 3 mg m^{-3}. Through the uncertainty quantification (MdAD 0.64 mg m^{-3}, MAD 0.96 mg m^{-3}) and modelling of the MDN HICO RVD, we

discovered that $a_{NAP}(443)$ and $a_{CDOM}(443)$ characterise the retrieval uncertainties. An addition of observations characterised by the conditions of this study to the training measurement pool of the MDN could lead to direct retrieval improvements. Changes of the MDN training dataset would however not demonstrate whether the mixture density network is superior to other ML approaches for these waters, as we did not test other ML algorithms. New ML algorithms should put a larger emphasis on oligo- and mesotrophic datasets to facilitate a broader evaluation of diverse ML approaches to retrieve chla concentration in these conditions.

5.4. Semi-analytical algorithms

The retrieval uncertainty for the QAAv6, GSM and 3SAA algorithms was characterised by the variables $a_\phi(673)$ and the two ratios chla/TSM and $\frac{R_{rs}(400)}{R_{rs}(673)}$. Our analysis emphasises that phytoplankton absorption and composition variability could not be sufficiently accounted for in the retrieval procedures of the three SAA algorithms tested in the lakes. $a_\phi(\lambda)$ or combined phytoplankton size class absorption must be precisely estimated in the first place to accurately retrieve chla concentration through SAA. $a_\phi(\lambda)$ at 443 and 673 nm only has a weak - moderate relationship to chla concentration in this dataset (Fig. 7). For algorithms that rely on $a_\phi(\lambda)$ to estimate chla, such as SAA (but also empirical fluorescence or IOP neural network algorithms), these weak relationships directly affect their accuracy. Consequently, empirical algorithms are more suited than SAA in scenarios where the uncertainty introduced by $a_{CDOM}(\lambda)$ and $a_{NAP}(\lambda)$ is smaller than by the $a_\phi(\lambda)$/chla ratio. For SAA prior information about $a_\phi(\lambda)$ variability is required. Without calibration, these models default to mass-specific phytoplankton absorption ($a^*_\phi(\lambda)$) values which are likely not appropriate for oligo- and mesotrophic lake conditions. In open ocean and coastal waters, $a^*_\phi(\lambda)$ exhibits a natural variability of a factor of 4 for any given chla value in addition to pigment packaging effects (Bricaud et al., 2004). This pattern stresses that the retrieval of chla concentration through $a_\phi(\lambda)$ may be significantly altered in the studied inland water conditions where $a^*_\phi(\lambda)$ variability is likely even higher. Thus, if uncalibrated, inappropriate $a^*_\phi(\lambda)$ values are used by the inverse SAA algorithms.

6. Conclusions

For oligo- and mesotrophic lakes and reservoirs, chla uncertainties were repeatedly shown to be highest when compared to higher biomass waters (Liu et al., 2021; Neil et al., 2019; Werther et al., 2021). The uncertainties persisted for the dataset of this study, as none of the tested algorithms accurately retrieved chla concentration over the considered range (0 – 10 mg m^{-3}). The reasons for the retrieval uncertainty are algorithm-specific and concluded on in the following.

The OC algorithms retrieved chla imprecisely, because the magnitude and shape of the used blue and green bands was mostly driven by CDOM and NAP absorption rather than by phytoplankton pigments. The results of the uncertainty analysis suggest that the OC algorithms are insensitive to phytoplankton absorption and composition variability in oligo- and mesotrophic lakes and reservoirs under the occurrence of CDOM and NAP absorption. The OC algorithms can therefore only be reliably used in scenarios where it is known that CDOM and NAP are not dominating the IOP absorption budget. In contrast, red-band algorithms such as G11 and Gons05 were partially able to retrieve chla accurately because the signal in their employed bands was not significantly influenced by CDOM and NAP absorption. The uncertainty modelling however revealed other causes for their retrieval inaccuracy, most of them related to low signal to noise ratios in these waters. As the uncertainty analysis demonstrated, for G11 the signal of $a_\phi(\lambda)$ in the red bands 665 and 708 nm was too low to be accurately relatable to chla concentration. For the Gons05 algorithm the explanatory variables were highly related to backscatter which could not be precisely retrieved from the red/NIR bands. Moreover, chla retrieval by red-band peak height algorithms such as FLH and MPH was the least successful of all algorithms because the required sun-induced fluorescence signal was mostly not available or detectable.

The results of the uncertainty modelling for the tested MDN OLCI and HICO ML algorithms revealed that their consistent overestimation of chla concentration was related to CDOM and NAP absorption. The causes however were not comparable to the OC algorithms. Instead, measurements of chla and $R_{rs}(\lambda)$ from oligo- and mesotrophic lakes under high CDOM and NAP absorption were underrepresented during MDN model training, as the MDNs were built with a global dataset in which inland waters with low chla concentrations under varying CDOM and NAP absorption were under sampled (Filazzola et al., 2020; Pahlevan et al., 2020).

The SAA algorithms QAAv6, GSM and 3SAA consistently showed higher retrieval uncertainties than empirical algorithms. Our analysis revealed that SAA algorithms are at a disadvantage when the relationship of $a_\phi(\lambda)$ to chla concentration is only weak or moderate, as it was the case for this dataset at 443 and 673 nm. Moreover, SAA make explicit assumptions about $a_\phi(\lambda)$ intensity and variability, which are likely too static when considering the phytoplankton variability and dynamic across multiple oligo- and mesotrophic lakes and reservoirs. It is thus likely that only with known or switching $a^*_\phi(\lambda)$ parametrisations SAA will produce accurate chla concentrations in the studied conditions.

Possible causes of retrieval inaccuracy over oligo- and mesotrophic waters are still poorly handled by existing algorithms. Significant improvements for the retrieval of chla concentration in oligo- and mesotrophic lakes and reservoirs can therefore only be expected through a reduction of the impact of the identified and algorithm-specific uncertainty drivers. In addition, upscaled hyperspectral resolution could lead to direct benefits for chla estimation (Dierssen et al., 2021). For MDN and GSM, the HICO configuration outperformed OLCI by 10–20% suggesting that higher information content improves the retrieval accuracy. ML algorithms specifically can harness the increased information content of hyperspectral data, but as this study shows, require a representative dataset. *In situ* datasets are now starting to be of sufficient size to inform new algorithm development for optically complex inland water bodies with low chla concentrations. As most current chla retrieval algorithms lack confidence intervals, new developments should indicate if their use over the target conditions is appropriate.

CRediT authorship contribution statement

Mortimer Werther: Writing – original draft, Conceptualization, Data curation, Investigation, Methodology, Visualization. **Daniel Odermatt:** Data curation, Investigation, Funding acquisition, Writing – review & editing. **Stefan G.H. Simis:** Investigation, Funding acquisition, Writing – review & editing. **Daniela Gurlin:** Data curation, Writing – review & editing. **Daniel S.F. Jorge:** Methodology, Writing – review & editing. **Hubert Loisel:** Investigation, Writing – review & editing. **Peter D. Hunter:** Data curation, Writing – review & editing. **Andrew N. Tyler:** Data curation, Funding acquisition, Writing – review & editing. **Evangelos Spyrakos:** Conceptualization, Data curation, Investigation, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study has received funding from the European Union's Horizon

2020 research and innovation programme under grant agreement No. 776480 (MONOCLE). We thank two anonymous reviewers for their constructive and thorough comments to improve the manuscript. We owe our gratitude to the following data providers of LIMNADES for their measurements used in this study: Stewart Bernard, Caren E. Binding, Mariano Bresciani, Claudia Giardino, Anatoly A. Gitelson, Luis Guanter, Kari Y.O. Kallio, Tiit Kutser, Ciro Manzo, Mark W. Matthews, John F. Schalles and Antonio Ruiz-Verdú. We additionally thank Steven R. Greb for his efforts in collecting the Wisconsin DNR data. The authors gratefully acknowledge support from the UK NERC Field Spectroscopy Facility for instrument loans and calibrations supporting the collection of field data.

Appendix A. Description of Wisconsin DNR *in situ* data

Radiometric measurements

Radiometric measurements were made with a pair of inter-calibrated Ocean Optics USB series UV-NIR spectrometers (Ocean Insight, Orlando, FL, USA), which measured sub-surface upwelling radiance, $L_u(\lambda)$ and total downwelling irradiance above the surface, $E_d(\lambda)$ at the same time (Gitelson et al., 2008). $R_{rs}(\lambda)$ was derived through:

$$R_{\mathrm{rs}}(\lambda) = \frac{L_{\mathrm{u}}(\lambda)}{E_d(\lambda)} \times \frac{E_{\mathrm{ref}}(\lambda)}{L_{\mathrm{ref}}(\lambda)} \times R_{\mathrm{ref}}(\lambda) \times \frac{1}{100} \times \frac{t}{n_w^2} \times F_0(\lambda) \times \frac{1}{\pi}, \tag{A.1}$$

where $E_{ref}(\lambda)$ and $L_{ref}(\lambda)$ are the total downwelling irradiance and upwelling radiance from a reflectance target, $R_{ref}(\lambda)$ is the respective spectral reflectance provided by the manufacturer, t is the water-to-air transmittance ($t = 0.98$), n_w is the refractive index of water ($n = 1.33$ at 20 °C), $F_0(\lambda)$ is the spectral immersion factor (Ohde and Siegel, 2003), and π is introduced to transform the intermediately calculated water-leaving radiance reflectance $\rho_w(\lambda)$ into $R_{rs}(\lambda)$.

The spectrometers used for the radiometric measurements were customised with a 25-μm slit, a 600-line grating, which was blazed at 500 nm to optimise the instrument response for the bandwidth from 350 nm to 850 nm, and a standard ILX-511B detector, which resulted in a 1.39 nm spectral resolution. One of the spectrometers was connected to a 25° field-of-view optical fibre which was taped to a 4-m long, handheld Unger Opti-Loc extension pole. The pole was pointed away from the boat and the tip of the fibre was kept just beneath the water surface to measure $L_u(\lambda)$ at nadir on the sun-lit side of the boat. The second spectrometer was connected to an optical fibre fitted with a cosine collector to create a 180° field-of-view, taped to a 2-m pole. This pole was attached to a location on the boat free from adjacent influences to measure $E_d(\lambda)$ at zenith concurrently with the upwelling radiance without interference from surface structures. The spectrometers were inter-calibrated at the start of each set of measurements through six calibration scans of a white Spectralon™ reflectance target (Labsphere, Inc., North Sutton, NH, USA), which was calibrated annually at the manufacturer. The calibration scans were followed by six to twelve scans of optically deep water which were processed to %-reflectance with the CDAP-2 software package (CALMIT, University of Nebraska-Lincoln, USA) and the median spectrum was considered representative for the station. System noise was removed with a smoothing spline function.

Field measurements and laboratory analysis

Water transparency measurements (Z_{SD}) were collected using a standard 20-cm diameter Secchi disk. Surface water samples were collected with a Kemmerer Bottle at a depth of 0.5 m below the surface and stored cold in the dark. Upon completion of the filtering of the water samples shortly after their collection, the membranes for the laboratory analysis of chla were frozen and transferred to the Wisconsin State Laboratory of Hygiene (WSLH) where they were analysed, while the CDOM absorption, particulate absorption ($a_p(\lambda)$), and TSM samples were analysed in-house. The absorption measurements were taken as laboratory triplicates and the final parameter values correspond to the median of the measurements. $a_{\mathrm{CDOM}}(\lambda)$ samples were filtered through acid washed MF-Millipore GSWP mixed cellulose ester membranes while chla samples were filtered through DAWP membranes with a 0.65 μm pore size. For TSM and particulate absorption samples Whatman™ GF/F binder-free glass microfiber filters were used. Chla was extracted from the concentrated algal samples (U.S. EPA 445.0) in a solution of aqueous 90% acetone aided by bath type sonication. Therefore, the sample racks with tubes were suspended in an ultrasonic bath, covered to exclude light, sonicated for 25 min, and stored at a temperature of<4 °C to complete the extraction overnight. The samples were centrifuged for 30 min at 500 XG to clarify the extracts on the next day and the chla concentrations were measured fluorometrically (Welschmeyer, 1994) with a PerkinElmer LS-55 Fluorescence Spectrometer (PerkinElmer, Inc., Waltham, MA). The fluorescence spectrophotometer was calibrated with pure chla standards of a known concentration. TSM and the organic fraction of TSM (OSM) were measured gravimetrically from the dried and combusted residue on pre-combusted and pre-weighed filter pads (APHA 2540D and APHA 2540E).

Laboratory triplicates of $a_{\mathrm{CDOM}}(\lambda)$ were transferred to a 0.1 m cuvette and optical densities of the filtrates were measured for a wavelength range from 350 nm to 800 nm and a bandwidth of 2 nm with the UV WinLab software package and a PerkinElmer LAMBDA 35 UV/Vis spectrophotometer. The signal from a de-ionized (DI) water blank was subtracted automatically and $a_{\mathrm{CDOM}}(\lambda)$ was calculated as:

$$a_{\mathrm{CDOM}}(\lambda) = \frac{\ln(10)}{l}[OD_{\mathrm{s}}(\lambda) - OD_{\mathrm{null}}], \tag{A.2}$$

where $OD_S(\lambda)$ is the optical density of the sample, $OD_{null}(\lambda)$ is the average optical density of the sample from 700 to 720 nm (if $OD_S(440) < 0.05$) or 780 to 800 nm (if $OD_S(440) \geq 0.05$) for the null point correction, and l is the pathlength of the cuvette in m. The results represent the averaged laboratory triplicates. The particulate and NAP absorption coefficients were measured through the quantitative filter technique (Mitchell et al., 2000). Laboratory triplicates were transferred to an optical glass slide mounted against the transmittance port of a Labsphere RSA-PE-20 accessory and the optical densities of the particles retained on the filters were measured over the 400 to 800 nm wavelength range and a bandwidth of 2 nm with the same spectrophotometer. The signal from a previously measured blank filter was subtracted and $a_p(\lambda)$ was calculated as:

$$a_{\mathrm{p}}(\lambda) = \frac{\ln(10)}{\frac{V_{\mathrm{f}}}{A}}\left[0.3907 \times \left[OD_{\mathrm{fp}}(\lambda) - OD_{\mathrm{null}}\right] + 0.3201 \times \left[OD_{\mathrm{fp}}(\lambda) - OD_{\mathrm{null}}\right]^2\right], \tag{A.3}$$

where $OD_{fp}(\lambda)$ is the optical density of the sample and $OD_{fp}(\lambda)$ is the average optical density of the sample from 780 to 800 nm required for the null point correction of the measurements. V and A are the volume of water filtered in m^3 and the area of the filter in m^2 used to calculate the pathlength in m. The quadratic function for the pathlength amplification correction (Cleveland and Weidemann, 1993) included in the equation was derived for a small set of samples collected in 2015 with the same instrument. The effects of the absorption by pigments were removed through reaction with a sodium hypochlorite solution (Ferrari and Tassan, 1999) and $a_{\mathrm{NAP}}(\lambda)$ was measured similarly to $a_p(\lambda)$ after a 20-minute reaction time for the rinsed samples. The subtraction of $a_{\mathrm{NAP}}(\lambda)$ from $a_p(\lambda)$ resulted in $a_\phi(\lambda)$, where the results represent the averaged laboratory triplicates.

Appendix B. Description of University of Stirling (UoStirling) *in situ* data

Radiometric measurements were made with a ship-mounted set of three hyperspectral sensors (HyperSAS, Seabird Scientific Inc.) installed just above the water surface. The first sensor pointed at the water surface and collected upwelling radiance $L_u(\lambda)$, which comprises both water-leaving radiance $L_w(\lambda)$ and the reflected sky irradiance $\rho_s L_s(\lambda)$. The sky irradiance $L_s(\lambda)$ was measured with a second sensor, while a third sensor measured downwelling irradiance $E_d(\lambda)$. $R_{rs}(\lambda)$ was then derived through Simis and Olsson, (2013):

$$\begin{aligned} R_{rs}(\lambda) &= L_w(\lambda, 0^+)/E_d(\lambda), \\ L_w(\lambda, 0^+) &= L_t(\lambda) - \rho_s L_s(\lambda). \end{aligned} \tag{B.1}$$

The radiometric measurements were taken under the consideration of the two largest challenges to reduce measurement errors: maintenance of optimal viewing geometry and an accurate value determination of ρ_s under changing illumination conditions (Aas, 2010). Measurements were kept in this dataset if they were taken under cloud free conditions and on calm waters to ensure a correct alignment of the ship at a viewing azimuth angle (φ_v) $> 90°$ (ideally at 135°) (Hooker and Morel, 2003).

Water constituents were measured with the same method and protocols as described in Appendix A for the U.S. Wisconsin DNR *in situ* data. Chla concentration was derived via spectrophotometric determination (ISO 12060:1992). $a_p(\lambda)$ was measured at Plymouth Marine Laboratory (PML, England). Before 2014 $a_p(\lambda)$ was derived through the method by Tassan and Ferrari (1995) and after 2013 via (Röttgers and Gehnke, 2012; Stramski et al., 2015). CDOM absorption for these samples was calculated through Stedmon et al. (2000). Water transparency (Z_{SD}) was measured through a standard Secchi Disk lowered into the water from nadir until invisible to the viewer's eye.

Appendix C. Uncertainty models

In this paper, we restricted the modelling of the RVDs to non-linear RF models. Here we compare the RF to a linear model, the so-called Least Absolute Shrinkage and Selection Operator (LASSO). LASSO shrinks parameter coefficients to zero for variables that did not contribute significantly to an uncertainty model (Tibshirani, 1996). LASSO is an extension to the standard linear regression model:

$$y_i = \beta_0 + \sum\nolimits_{j=1}^{n} \beta_i c_{ij}, \tag{C.1}$$

where y_i is the target variable (here a residual value), c_{ij} an explanatory variable and beta (β_i) the learned coefficient (or weight) of the i-th observation for the j-th of all explanatory variables n. β_0 is the intercept. The coefficients $\beta_0, \beta_1, \cdots, \beta_n$ are estimated through minimising the residual sum of squares:

$$\widehat{\beta} = \underset{\beta}{\operatorname{argmin}} \sum\nolimits_{i=1}^{m} \left(y_i - \beta_0 - \sum\nolimits_{j=1}^{n} \beta_j c_{ij}\right)^2. \tag{C.2}$$

To estimate the beta-terms for LASSO, a penalty term $\alpha\sum_{j=1}^{n}|\beta_j|$, also called L1-norm or L1 penalty, is added:

$$\widehat{\beta}_{LASSO} = \underset{\beta}{\operatorname{argmin}} \sum\nolimits_{i=1}^{m} \left(y_i - \beta_0 - \sum\nolimits_{j=1}^{n} \beta_j c_{ij}\right)^2 + \alpha \sum\nolimits_{j=1}^{n} |\beta_j|. \tag{C.3}$$

The L1 penalty has the effect of forcing some of the coefficient estimates to be exactly zero when α is sufficiently large. The larger the penalisation, the fewer explanatory variables are present in the model because their coefficients are zero. The L1 penalty performs variable selection, as relevant variables to the model receive non-zero coefficients. In LASSO, α is a model hyper-parameter which we optimised with the same procedure as for the RFs. LASSO is restricted to model the residual values linearly and would thus require explicit formulations of non-linearities and interaction terms. Since we did not have knowledge about these, we opted for use of RFs. In linear models, such as LASSO, the β terms are sensitive to value magnitude and range, which differ between variables (e.g., TSM in g m^{-3} and $a_{CDOM}(443)$ in m^{-1}). The variables therefore need to be normalised for LASSO to fall in the range between 0 and 1:

$$\mathrm{c}_{\mathrm{norm}} = \frac{c - \min(c)}{\max(c) - \min(\mathrm{c})}. \tag{C.4}$$

Conversely, RF models are independent of unit scales and do not require prior normalisation. Table 6 shows that differences between RF and LASSO started to become significant with increasing chla retrieval uncertainty. A likely reason for this performance difference was the ability of the RF to estimate outliers of the RVDs better through its underlying non-linear model. An exhaustive dataset with many explanatory variables is highly labour and cost intensive. Hence, we required an uncertainty model that can be constructed with a limited set of explanatory variables (here n = 89). Unlike popular neural networks, LASSO and RFs do not require high amounts of measurements. Other methods, such as support vector regression machines (Boser et al., 1992) or more advanced gradient boosting decision trees (Chen and Guestrin, 2016) may provide similar performance such as the RFs

Table 6
LASSO and RF uncertainty quantification accuracies (linear scale) for each chla retrieval algorithm included in the uncertainty analysis. RF outperformed LASSO across all metrics for all algorithms.

Chla algorithm	MdAD [mg m^{-3}]		MAD [mg m^{-3}]		RMSE [mg m^{-3}]	
	LASSO	RF	LASSO	RF	LASSO	RF
OC4 -opt	0.78	0.50	1.15	0.79	1.84	1.08
OC6 -opt	0.73	0.58	1.08	0.84	1.72	1.13
G11 -opt	0.85	0.68	1.34	1.05	2.16	1.51
MDN HICO	0.83	0.64	1.25	0.96	1.79	1.26
QAAv6 (443)	1.09	0.62	1.68	0.96	2.61	1.39
GSM HICO	2.28	1.56	4.21	3.08	7.03	6.02
3SAA	1.50	0.92	2.42	1.56	3.77	2.42
Gons05	2.62	2.17	3.61	3.47	5.01	4.84

used in this study and can be considered as model alternatives.

Appendix D. SHAP and standard variable importance

ML libraries such as Scikit-learn (Pedregosa et al., 2011) provide measures to estimate variable impact on a model (thereby defining its importance). To enable the explanation of the quantified uncertainty by the RFs, the measure to infer variable impact must be consistently defined and available for each individual observation. During the construction process of a RF, each node in a sub-decision tree represents a split using some selection of the input variables. For RFs specifically, the aim is to gain information about the variable impact on the RF that resulted in a reduction of variance (or a decrease in error). However, the impact of a variable cannot be consistently derived for ML models such as RFs using standard variable importance measures (Altmann et al., 2010; Grömping, 2009; Strobl et al., 2008). In practice, the implementations of the variance reduction criterion vary between the libraries (Zhou and Hooker, 2021). As an example, the popular eXtreme Gradient Boosting (XGBoost) library (Chen and Guestrin, 2016) has five metrics (total gain, gain, total cover, cover, weight) to measure variable importance. The use of these metrics results in different variable rankings when applied to the same dataset. Besides inconsistency drawbacks, standard importance measures share a limitation: the contribution of a variable to the model for a single observation is not discernible from the overall variable impact. For these reasons standard importance measures are insufficient to model the RVD of a chla algorithm. We therefore used SHAP that ensures a consistent measure of variable impact (Lundberg et al., 2017) and enables to obtain the impact of a variable for an individual observation (Du et al., 2020; Lundberg et al., 2020). A recent review on the practices of variable importance measures can be found in Belle and Papantonis (2021).

References

Aas, E., 2010. Estimates of radiance reflected towards the Zenith at the surface of the sea. Ocean Sci. 6, 861–876. https://doi.org/10.5194/os-6-861-2010.

Altmann, A., Toloşi, L., Sander, O., Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. Bioinformatics 26, 1340–1347. https://doi.org/10.1093/bioinformatics/btq134.

Aulló-Maestro, M.E., 2019. Bio-geo-optical properties and remote sensing of CDOM in optically complex inland waters. PhD Dissertation. University of Stirling.

Belle, V., Papantonis, I., 2021. Principles and Practice of Explainable Machine Learning. Front. Big Data 4, 39. https://doi.org/10.3389/fdata.2021.688969.

Bergstra, J., Ca, J.B., Ca, Y.B., 2012. Random Search for Hyper-Parameter Optimization Yoshua Bengio. J. Mach. Learn. Res.

Binding, C., Jerome, J., Bukata, R., Booty, W., 2008. Spectral absorption properties of dissolved and particulate matter in Lake Erie. Remote Sens. Environ. 112, 1702–1711. https://doi.org/10.1016/j.rse.2007.08.017.

Binding, C.E., Greenberg, T.A., Bukata, R.P., 2013. The MERIS Maximum Chlorophyll Index; its merits and limitations for inland water algal bloom monitoring. J. Great Lakes Res. 39, 100–107. https://doi.org/10.1016/j.jglr.2013.04.005.

Binding, C.E., Greenberg, T.A., Jerome, J.H., Bukata, R.P., Letourneau, G., 2011. An assessment of MERIS algal products during an intense bloom in Lake of the Woods. J. Plankton Res. 33, 793–806. https://doi.org/10.1093/plankt/fbq133.

Binding, C.E., Jerome, J.H., Bukata, R.P., Booty, W.G., 2010. Suspended particulate matter in Lake Erie derived from MODIS aquatic colour imagery. Int. J. Remote Sens. 31, 5239–5255. https://doi.org/10.1080/01431160903302973.

Boser, B.E., Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A Training Algorithm for Optimal Margin Classifiers. In: Proceedings of the 5TH Annual ACM Workshop on Computational Learning Theory, pp. 144–152.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Breiman, L., 1996. Bagging Predictors. Mach. Learn. 24, 123–140. https://doi.org/10.1023/A:1018054314350.

Bresciani, M., Cazzaniga, I., Austoni, M., Sforzi, T., Buzzi, F., Morabito, G., Giardino, C., 2018. Mapping phytoplankton blooms in deep subalpine lakes from Sentinel-2A and Landsat-8. Hydrobiologia 824, 197–214. https://doi.org/10.1007/s10750-017-3462-2.

Bresciani, M., Stroppiana, D., Odermatt, D., Morabito, G., Giardino, C., 2011. Assessing remotely sensed chlorophyll-a for the implementation of the Water Framework Directive in European perialpine lakes. Sci. Total Environ. 409, 3083–3091. https://doi.org/10.1016/j.scitotenv.2011.05.001.

Brewin, R.J.W., Sathyendranath, S., Müller, D., Brockmann, C., Deschamps, P.Y., Devred, E., Doerffer, R., Fomferra, N., Franz, B., Grant, M., Groom, S., Horseman, A., Hu, C., Krasemann, H., Lee, Z.P., Maritorena, S., Mélin, F., Peters, M., Platt, T., Regner, P., Smyth, T., Steinmetz, F., Swinton, J., Werdell, J., White, G.N., 2015. The Ocean Colour Climate Change Initiative: III. A round-robin comparison on in-water bio-optical algorithms. Remote Sens. Environ. 162, 271–294. https://doi.org/10.1016/j.rse.2013.09.016.

Bricaud, A., Claustre, H., Ras, J., Oubelkheir, K., 2004. Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations. J. Geophys. Res. Oceans 109. https://doi.org/10.1029/2004JC002419.

Bricaud, A., Mejia, C., Blondeau-Patissier, D., Claustre, H., Crepon, M., Thiria, S., 2007. Retrieval of pigment concentrations and size structure of algal populations from their absorption spectra using multilayered perceptrons. Appl. Opt. 46, 1251–1260. https://doi.org/10.1364/AO.46.001251.

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 Abstract. Spectral absorption coefficients of total particulate matter mg m-3). As pre. J. Geophys. Res. 103, 31033–31044.

Carlson, R.E., 1977. A trophic state index for lakes. Limnol. Oceanogr. 22, 361–369. https://doi.org/10.4319/lo.1977.22.2.0361.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Claustre, H., Hooker, S.B., van Heukelem, L., Berthon, J.F., Barlow, R., Ras, J., Sessions, H., Targa, C., Thomas, C.S., van der Linde, D., Marty, J.C., 2004. An intercomparison of HPLC phytoplankton pigment methods using in situ samples: Application to remote sensing and database activities. Mar. Chem. 85, 41–61. https://doi.org/10.1016/j.marchem.2003.09.002.

Cleveland, J.S., Weidemann, A.D., 1993. Quantifying absorption by aquatic particles: A multiple scattering correction for glass-fiber filters. Limnol. Oceanogr. 38, 1321–1327. https://doi.org/10.4319/lo.1993.38.6.1321.

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 https://doi.org/10.1029/2003GL018065.

Dierssen, H.M., Ackleson, S.G., Joyce, K.E., Hestir, E.L., Castagna, A., Lavender, S., McManus, M.A., 2021. Living up to the Hype of Hyperspectral Aquatic Remote Sensing: Science, Resources and Outlook. Front. Environ. Sci. 9.

Du, M., Liu, N., Hu, X., 2020. Techniques for interpretable machine learning. Commun. ACM 63, 68–77. https://doi.org/10.1145/3359786.

Ferrari, G.M., Tassan, S., 1999. A method using chemical oxidation to remove light absorption by phytoplankton pigments. J. Phycol. 35, 1090–1098. https://doi.org/10.1046/j.1529-8817.1999.3551090.x.

Filazzola, A., Mahdiyan, O., Shuvo, A., Ewins, C., Moslenko, L., Sadid, T., Blagrave, K., Imrit, M.A., Gray, D.K., Quinlan, R., O'Reilly, C.M., Sharma, S., 2020. A database of chlorophyll and water chemistry in freshwater lakes. Sci. Data 7, 1–10. https://doi.org/10.1038/s41597-020-00648-2.

Garver, S.A., Siegel, D.A., 1997. Inherent optical property inversion of ocean color spectra and its biogeochemical interpretation: 1. Time series from the Sargasso Sea. J. Geophys. Res. Oceans 102, 18607–18625. https://doi.org/10.1029/96JC03243.

Giardino, C., Bresciani, M., Cazzaniga, I., Schenk, K., Rieger, P., Braga, F., Matta, E., Brando, V.E., 2014a. Evaluation of Multi-Resolution Satellite Sensors for Assessing Water Quality and Bottom Depth of Lake Garda. Sensors 14, 24116–24131. https://doi.org/10.3390/s141224116.

Giardino, C., Bresciani, M., Stroppiana, D., Oggioni, A., Morabito, G., 2014b. Optical remote sensing of lakes: An overview on Lake Maggiore. J. Limnol. 73, 201–214. https://doi.org/10.4081/jlimnol.2014.817.

Giardino, C., Bresciani, M., Valentini, E., Gasperini, L., Bolpagni, R., Brando, V.E., 2015. Airborne hyperspectral data to assess suspended particulate matter and aquatic vegetation in a shallow and turbid lake. Remote Sens. Environ. 157, 48–57. https://doi.org/10.1016/j.rse.2014.04.034.

Giardino, C., Candiani, G., Zilioli, E., 2005. Detecting chlorophyll-a in Lake Garda using TOA MERIS radiances. Photogramm. Eng. Remote Sens. 71, 1045–1051. https://doi.org/10.14358/PERS.71.9.1045.

Gilerson, A., Zhou, J., Hlaing, S., Ioannou, I., Schalles, J., Gross, B., Moshary, F., Ahmed, S., 2007. Fluorescence component in the reflectance spectra from coastal waters. Dependence on water composition. Opt. Express 15, 15702. https://doi.org/10.1364/oe.15.015702.

Giardino, C., Pepe, M., Brivio, P.A., Ghezzi, P., Zilioli, E., 2001. Detecting chlorophyll, Secchi disk depth and surface temperature in a sub-alpine lake using Landsat imagery. Sci. Total Environ. 268, 19–29.

Gilerson, A., Gitelson, A.A., Zhou, J., Gurlin, D., Moses, W., Ioannou, I., 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. https://doi.org/10.1364/oe.18.024109.

Gitelson, A., Garbuzov, G., Szilagyi, F., Mittenzwey, K.H., Karnieli, A., Kaiser, A., 1993. Quantitative remote sensing methods for real-time monitoring of inland waters quality. Int. J. Remote Sens. 14, 1269–1295. https://doi.org/10.1080/01431169308953956.

Gitelson, A.A., 1992. The peak near 700 nm on radiance spectra of algae and water: Relationships of its magnitude and position with chlorophyll. Int. J. Remote Sens. 13, 3367–3373. https://doi.org/10.1080/01431169208904125.

Gitelson, A.A., Dall'Olmo, G., Moses, W., Rundquist, D.C., Barrow, T., Fisher, T.R., Gurlin, D., Holz, J., 2008. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. Remote Sens. Environ. 112, 3582–3593. https://doi.org/10.1016/j.rse.2008.04.015.

Gitelson, A.A., Gurlin, D., Moses, W.J., Barrow, T., 2009. A bio-optical algorithm for the remote estimation of the chlorophyll-a concentration in case 2 waters. Environ. Res. Lett. 4, 5. https://doi.org/10.1088/1748-9326/4/4/045003.

Gons, H., Auer, M., Effler, S., 2008. MERIS satellite chlorophyll mapping of oligotrophic and eutrophic waters in the Laurentian Great Lakes. Remote Sens. Environ. 112, 4098–4106. https://doi.org/10.1016/j.rse.2007.06.029.

Gons, H., Rijkeboer, M., Ruddick, K., 2005. Effect of a waveband shift on chlorophyll retrieval from MERIS imagery of inland and coastal waters. J. Plankton Res. 27.

Gons, H., Rijkeboer, M., Ruddick, K., 2002. A chlorophyll-retrieval algorithm for satellite imagery (Medium Resolution Imaging Spectrometer) of inland and coastal waters. J. Plankton Res. 24 https://doi.org/10.1093/plankt/24.9.947.

Gordon, H.R., Brown, O.B., Evans, R.H., Brown, J.W., Smith, R.C., Baker, K.S., Clark, D. K., 1988. A semianalytic radiance model of ocean color. J. Geophys. Res. 93, 10909. https://doi.org/10.1029/JD093iD09p10909.

Gower, J.F.R., 1980. Observations of in situ fluorescence of chlorophyll-a in Saanich Inlet. Bound.-Layer Meteorol. 18, 235–245. https://doi.org/10.1007/BF00122022.

Gower, J.F.R., Brown, L., Borstad, G.A., 2004. Observation of chlorophyll fluorescence in west coast waters of Canada using the MODIS satellite sensor. Can. J. Remote Sens. 30, 17–25. https://doi.org/10.5589/m03-048.

Gower, J.F.R., Doerffer, R., Borstad, G.A., 1999. Interpretation of the 685nm peak in water-leaving radiance spectra in terms of fluorescence, absorption and scattering, and its observation by MERIS. Int. J. Remote Sens. 20, 1771–1786. https://doi.org/10.1080/014311699212470.

Guanter, L., Ruiz-Verdú, A., Odermatt, D., Giardino, C., Simis, S., Estellés, V., Heege, T., Domínguez-Gómez, J.A., Moreno, J., 2010. Atmospheric correction of ENVISAT/MERIS data over inland waters: Validation for European lakes. Remote Sens. Environ. 114, 467–480. https://doi.org/10.1016/j.rse.2009.10.004.

Gupana, R.S., Odermatt, D., Cesana, I., Giardino, C., Nedbal, L., Damm, A., 2021. Remote sensing of sun-induced chlorophyll-a fluorescence in inland and coastal waters: Current state and future prospects. Remote Sens. Environ. https://doi.org/10.1016/j.rse.2021.112482.

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 Sens. Environ. 115, 3479–3490. https://doi.org/10.1016/j.rse.2011.08.011.

Grömping, U., 2009. Variable importance assessment in regression: Linear regression versus random forest. Am. Stat. 63, 308–319. https://doi.org/10.1198/tast.2009.08199.

Hieronymi, M., Müller, D., Doerffer, R., 2017. The OLCI Neural Network Swarm (ONNS): A Bio-Geo-Optical Algorithm for Open Ocean and Coastal Waters. Front. Mar. Sci. 4, 140. https://doi.org/10.3389/fmars.2017.00140.

Hooker, S.B., Morel, A., 2003. Platform and environmental effects on above-water determinations of water-leaving radiances. J. Atmos. Oceanic Technol. 20, 187–205. https://doi.org/10.1175/1520-0426(2003)020<0187:PAEEOA>2.0.CO;2.

Hooker, S.B., van Heukelem, L., Thomas, C.S., Claustre, H., Ras, J., Barlow, R., Sessions, H., Schlüter, L., Perl, J., Trees, C., Stuart, V., Head, E., Clementson, L., Fishwick, J., Llewellyn, C., Aiken, J., 2005. The Second SeaWiFS HPLC Analysis Round-Robin Experiment (SeaHARRE-2). NASA Technical Memorandum.

Huot, Y., Babin, M., Bruyant, F., Grob, C., Twardowski, M.S., Claustre, H., Claustre, H., 2007. Does chlorophyll a provide the best index of phytoplankton biomass for primary productivity studies? European Geosciences Union.

IOCCG, 2019. Uncertainties in Ocean Colour Remote Sensing. International Ocean-Colour Coordinating Group (IOCCG), Dartmouth, Canada.

Jamet, C., Loisel, H., Dessailly, D., 2012. Retrieval of the spectral diffuse attenuation coefficient $K_d(\lambda)$ in open and coastal ocean waters using a neural network inversion. J. Geophys. Res. Oceans 117, n/a-n/a. https://doi.org/10.1029/2012JC008076.

Jorge, D.S.F., Loisel, H., Jamet, C., Dessailly, D., Demaria, J., Bricaud, A., Maritorena, S., Zhang, X., Antoine, D., Kutser, T., Bélanger, S., Brando, V.O., Werdell, J., Kwiatkowska, E., Mangin, A., d'Andon, O.F., 2021. A three-step semi analytical algorithm (3SAA) for estimating inherent optical properties over oceanic, coastal, and inland waters from remote sensing reflectance. Remote Sens. Environ. 263, 112537 https://doi.org/10.1016/j.rse.2021.112537.

Kallio, K., Koponen, S., Ylöstalo, P., Kervinen, M., Pyhälahti, T., Attila, J., 2015. Validation of MERIS spectral inversion processors using reflectance, IOP and water quality measurements in boreal lakes. Remote Sens. Environ. 157, 147–157. https://doi.org/10.1016/j.rse.2014.06.016.

Kutser, T., Vahtmäe, E., Paavel, B., Kauer, T., 2013. Removing glint effects from field radiometry data measured in optically complex coastal and inland waters. Remote Sens. Environ. 133, 85–89. https://doi.org/10.1016/j.rse.2013.02.011.

Lee, Z., Carder, K.L., Arnone, R.A., 2002. Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. Appl. Opt. 41, 5755. https://doi.org/10.1364/ao.41.005755.

Liu, X., Steele, C., Simis, S., Warren, M., Tyler, A., Spyrakos, E., Selmes, N., Hunter, P., 2021. Retrieval of Chlorophyll-a concentration and associated product uncertainty in optically diverse lakes and reservoirs. Remote Sens. Environ. 267, 112710 https://doi.org/10.1016/j.rse.2021.112710.

Loisel, H., Stramski, D., 2000. Estimation of the inherent optical properties of natural waters from the irradiance attenuation coefficient and reflectance in the presence of Raman scattering. Appl. Opt. 39, 3001–3011.

Loisel, H., Stramski, D., Dessailly, D., Jamet, C., Li, L., Reynolds, R.A., 2018. An Inverse Model for Estimating the Optical Absorption and Backscattering Coefficients of Seawater From Remote-Sensing Reflectance Over a Broad Range of Oceanic and Coastal Marine Environments. J. Geophys. Res. Oceans 123, 2141–2171. https://doi.org/10.1002/2017JC013632.

Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems 2017-December, pp. 4766–4775.

Lundberg, S., Lee, S.-I., 2016. An unexpected unity among methods for interpreting model predictions. arXiv.

Lundberg, S.M., Allen, P.G., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. https://doi.org/10.5555/3295222.

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., Lee, S.-I., 2020. From local explanations to global understanding with explainable AI for trees. Nature Mach. Intell. 2, 56–67. https://doi.org/10.1038/s42256-019-0138-9.

Lundberg, S.M., Erion, G.G., Lee, S.-I., 2018. Consistent Individualized Feature Attribution for Tree Ensembles. arXiv.

Manzo, C., Bresciani, M., Giardino, C., Braga, F., Bassani, C., 2015. Sensitivity analysis of a bio-optical model for Italian lakes focused on Landsat-8, Sentinel-2 and Sentinel-3. Eur. J. Remote Sens. 48, 17–32. https://doi.org/10.5721/EuJRS20154802.

Maritorena, S., Siegel, D.A., Peterson, A.R., 2002. Optimization of a semianalytical ocean color model for global-scale applications. Appl. Opt. 41, 2705. https://doi.org/10.1364/ao.41.002705.

Matthews, M.W., 2014. Eutrophication and cyanobacterial blooms in South African inland waters: 10years of MERIS observations. Remote Sens. Environ. 155, 161–177. https://doi.org/10.1016/j.rse.2014.08.010.

Matthews, M.W., Bernard, S., 2013. Characterizing the absorption properties for remote sensing of three small optically-diverse south african reservoirs. Remote Sens. 5, 4370–4404. https://doi.org/10.3390/rs5094370.

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. https://doi.org/10.1016/j.rse.2012.05.032.

Matthews, M.W., Odermatt, D., 2015. Improved algorithm for routine monitoring of cyanobacteria and eutrophication in inland and near-coastal waters. Remote Sens. Environ. 156, 374–382. https://doi.org/10.1016/j.rse.2014.10.010.

McKee, D., Röttgers, R., Neukermans, G., Calzado, V.S., Trees, C., Ampolo-Rella, M., Neil, C., Cunningham, A., 2014. Impact of measurement uncertainties on determination of chlorophyll-specific absorption coefficient for marine phytoplankton. J. Geophys. Res. Oceans 119, 9013–9025. https://doi.org/10.1002/2014JC009909.

McKinna, L.I.W., Cetinić, I., Werdell, P.J., 2021. Development and Validation of an Empirical Ocean Color Algorithm with Uncertainties: A Case Study with the Particulate Backscattering Coefficient. e2021JC017231 J. Geophys Res. Oceans 126. https://doi.org/10.1029/2021JC017231.

Mélin, F., Franz, B.A., 2014. Assessment of satellite ocean colour radiometry and derived geophysical products. In: Experimental Methods in the Physical Sciences. Academic Press, pp. 609–638. https://doi.org/10.1016/B978-0-12-417011-7.00020-9.

Mélin, F., Sclep, G., Jackson, T., Sathyendranath, S., 2016. Uncertainty estimates of remote sensing reflectance derived from comparison of ocean color satellite data sets. Remote Sens. Environ. 177, 107–124. https://doi.org/10.1016/j.rse.2016.02.014.

Merchant, C.J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A.C., Reuter, M., Sathyendranath, S., Sandven, S., Sofieva, V.F., Wagner, W., 2017. Uncertainty information in climate data records from Earth observation. Earth Syst. Sci. Data 9, 511–527. https://doi.org/10.5194/essd-9-511-2017.

Miles, J., 2014. Tolerance and Variance Inflation Factor. Wiley StatsRef: Statistics Reference Online, Major Reference Works. https://doi.org/10.1002/9781118445112.stat06593.

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 Sens. Environ. 117, 394–406. https://doi.org/10.1016/j.rse.2011.10.016.

Mitchell, B.G., Bricaud, A., Carder, K., Cleveland, J., Ferrari, G., Gould, R., Kahru, M., Kishino, M., Maske, H., Moisan, T., Moore, L., Nelson, N., Phinney, D., Reynolds, R., Sosik, H., Stramski, D., Tassan, S., Trees, C., Weidemann, A., Wieland, J., Vodacek, A., 2000. Determination of spectral absorption coefficients of particles, dissolved material and phytoplankton for discrete water samples.

Mittenzwey, K.-H., Ullrich, S., Gitelson, A.A., Kondratiev, K.Y., 1992. Determination of chlorophyll a of inland waters on the basis of spectral reflectance. Limnol. Oceanogr. 37, 147–149. https://doi.org/10.4319/lo.1992.37.1.0147.

Mobley, C.D., 1999. Estimation of the remote-sensing reflectance from above-surface measurements. Appl. Opt. 38, 7442. https://doi.org/10.1364/ao.38.007442.

Morel, A., Prieur, L., 1977. Analysis of Variations in Ocean Color. Limnology and Oceanography 22, 709–722.

Morley, S.K., Brito, T.V., Welling, D.T., 2018. Measures of Model Performance Based On the Log Accuracy Ratio. Sp. Weather 16, 69–88. https://doi.org/10.1002/2017SW001669.

Moses, W.J., Gitelson, A.A., Berdnikov, S., Povazhnyy, V., 2009. Estimation of chlorophyll-a concentration in case II waters using MODIS and MERIS data - Successes and challenges. Environ. Res. Lett. 4, 45005. https://doi.org/10.1088/1748-9326/4/4/045005.

Moses, W.J., Sterckx, S., Montes, M.J., De Keukelaere, L., Knaeps, E., 2017. Atmospheric Correction for Inland Waters, in: Bio-Optical Modeling and Remote Sensing of Inland Waters. Elsevier Inc., pp. 69–100. https://doi.org/10.1016/B978-0-12-804644-9.00003-3.

Mouw, C.B., Chen, H., McKinley, G.A., Effler, S., O'Donnell, D., Perkins, M.G., Strait, C., 2013. Evaluation and optimization of bio-optical inversion algorithms for remote sensing of Lake Superior's optical properties. J. Geophys. Res. Oceans 118, 1696–1714. https://doi.org/10.1002/jgrc.20139.

Neil, C., Spyrakos, E., Hunter, P.D., Tyler, A.N., 2019. A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. Remote Sens. Environ. 229, 159–178. https://doi.org/10.1016/j.rse.2019.04.027.

Neville, R.A., Gower, J.F.R., 1977. Passive remote sensing of phytoplankton via chlorophyll α fluorescence. J. Geophys. Res. 82, 3487–3493. https://doi.org/10.1029/JC082i024p03487.

Odermatt, D., Giardino, C., Heege, T., 2010. Chlorophyll retrieval with MERIS Case-2-Regional in perialpine lakes. Remote Sens. Environ. 114, 607–617. https://doi.org/10.1016/j.rse.2009.10.016.

Odermatt, D., Heege, T., Nieke, J., Kneubühler, M., Itten, K., 2008. Water Quality Monitoring for Lake Constance with a Physically Based Algorithm for MERIS Data. Sensors 8, 4582–4599. https://doi.org/10.3390/s8084582.

Ohde, T., Siegel, H., 2003. Derivation of immersion factors for the hyperspectral TriOS radiance sensor. J. Opt. A: Pure Appl. Opt. https://doi.org/10.1088/1464-4258/5/3/103.

O'Reilly, 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 103, 24937–24953. https://doi.org/10.1029/98JC02160.

O'Reilly, Werdell, P.J., 2019. Chlorophyll algorithms for ocean color sensors - OC4, OC5 & OC6. Remote Sens. Environ. 229, 32–47. https://doi.org/10.1016/j.rse.2019.04.021.

Pahlevan, N., Mangin, A., Balasubramanian, S.V., Smith, B., Alikas, K., Arai, K., Barbosa, C., Bélanger, S., Binding, C., Bresciani, M., Giardino, C., Gurlin, D., Fan, Y., Harmel, T., Hunter, P., Ishikaza, J., Kratzer, S., Lehmann, M.K., Ligi, M., Ma, R., Martin-Lauzer, F.R., Olmanson, L., Oppelt, N., Pan, Y., Peters, S., Reynaud, N., Sander de Carvalho, L.A., Simis, S., Spyrakos, E., Steinmetz, F., Stelzer, K., Sterckx, S., Tormos, T., Tyler, A., Vanhellemont, Q., Warren, M., 2021a. ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. Remote Sens. Environ. 258 https://doi.org/10.1016/j.rse.2021.112366.

Pahlevan, N., Smith, B., Binding, C., Gurlin, D., Li, L., Bresciani, M., Giardino, C., 2021b. Hyperspectral retrievals of phytoplankton absorption and chlorophyll-a in inland and nearshore coastal waters. Remote Sens. Environ. 253, 112200 https://doi.org/10.1016/j.rse.2020.112200.

Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M.K., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. Remote Sens. Environ. 240, 111604 https://doi.org/10.1016/j.rse.2019.111604.

Pedregosa, F., Michel, V., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot, Édouardand, M., Duchesnay, Édouard, Duchesnay, E., Fré, 2011. Scikit-learn: Machine Learning in Python. JMLR 12, 2825–2830.

Pepe, M., Giardino, C., Borsani, G., Cardoso, A.C., Chiaudani, G., Premazzi, G., Rodari, E., Zilioli, E., 2001. Relationship between apparent optical properties and photosynthetic pigments in the sub-alpine Lake Iseo. Sci. Total Environ. 268, 31–45. https://doi.org/10.1016/S0048-9697(00)00691-4.

Poikane, S., Alves, M.H., Argillier, C., van den Berg, M., Buzzi, F., Hoehn, E., de Hoyos, C., Karottki, I., Laplace-Treyture, C., Solheim, A.L., Ortiz-Casas, J., Ott, I., Phillips, G., Pilke, A., Pádua, J., Remec-Rekar, S., Riedmüller, U., Schaumburg, J., Serrano, M.L., Soszka, H., Tierney, D., Urbanic, G., Wolfram, G., 2010. Defining chlorophyll-a reference conditions in European Lakes. Environ. Manage. 45, 1286–1298. https://doi.org/10.1007/s00267-010-9484-4.

Röttgers, R., Gehnke, S., 2012. Measurement of light absorption by aquatic particles: Improvement of the quantitative filter technique by use of an integrating sphere approach. Appl. Opt. 51, 1336–1351. https://doi.org/10.1364/AO.51.001336.

Ruiz-Verdú, A., Dominguez-Gomez, J.-A., Pena-Martinez, R., 2005. Use of CHRIS for Monitoring Water Quality in Rosarito Reservoir. In: ESA Special Publication, ESA Special Publication, pp. 26.

Ruiz-Verdú, A., Simis, S.G.H., de Hoyos, C., Gons, H.J., Peña-Martínez, R., 2008. An evaluation of algorithms for the remote sensing of cyanobacterial biomass. Remote Sens. Environ. 112, 3996–4008. https://doi.org/10.1016/j.rse.2007.11.019.

Salama, M.S., Stein, A., 2009. Error decomposition and estimation of inherent optical properties. Appl. Opt. 48, 4947–4962. https://doi.org/10.1364/AO.48.004947.

Schaeffer, B., Salls, W., Coffer, M., Lebreton, C., Werther, M., Stelzer, K., Urquhart, E., Gurlin, D., 2022. Merging of the Case 2 Regional Coast Colour and Maximum-Peak Height chlorophyll-a algorithms: validation and demonstration of satellite-derived retrievals across US lakes. Environ. Monit. Assess. 194, 1–25. https://doi.org/10.1007/s10661-021-09684-w.

Schalles, J.F., 2006. Optical remote sensing techniques to estimate phytoplankton chlorophyll a concentrations in coastal waters with varying suspended matter and cdom concentrations. In: Remote Sensing and Digital Image Processing, Remote Sensing and Digital Image Processing. Springer International Publishing, pp. 27–79.

Seegers, B.N., Stumpf, R.P., Schaeffer, B.A., Loftin, K.A., Werdell, P.J., 2018. Performance metrics for the assessment of satellite data products: an ocean color case study. Opt. Express 26, 7404. https://doi.org/10.1364/oe.26.007404.

Shi, K., Li, Y., Li, L., Lu, H., Song, K., Liu, Z., Xu, Y., Li, Z., 2013. Remote chlorophyll-a estimates for inland waters based on a cluster-based classification. Sci. Total Environ. 444, 1–15. https://doi.org/10.1016/j.scitotenv.2012.11.058.

Simis, S.G.H., Olsson, J., 2013. Unattended processing of shipborne hyperspectral reflectance measurements. Remote Sens. Environ. 135, 202–212. https://doi.org/10.1016/j.rse.2013.04.001.

Smith, M.E., Robertson Lain, L., Bernard, S., 2018. An optimized Chlorophyll a switching algorithm for MERIS and OLCI in phytoplankton-dominated waters. Remote Sens. Environ. 215, 217–227. https://doi.org/10.1016/j.rse.2018.06.002.

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. https://doi.org/10.1080/01431160600815533.

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-Verdú, A., Schalles, J.F., Tebbs, E., Zhang, Y., Tyler, A.N., 2018. Optical types of inland and coastal waters. Limnol. Oceanogr. 63, 846–870. https://doi.org/10.1002/lno.10674.

Stedmon, C.A., Markager, S., Kaas, H., 2000. Optical properties and signatures of chromophoric dissolved organic matter (CDOM) in Danish coastal waters. Estuar. Coast. Shelf Sci. 51, 267–278. https://doi.org/10.1006/ecss.2000.0645.

Stramski, D., Reynolds, R.A., Kaczmarek, S., Uitz, J., Zheng, G., 2015. Correction of pathlength amplification in the filter-pad technique for measurements of particulate absorption coefficient in the visible spectral region. Appl. Opt. 54, 6763. https://doi.org/10.1364/ao.54.006763.

Strobl, C., Boulesteix, A.L., Kneib, T., Augustin, T., Zeileis, A., 2008. Conditional variable importance for random forests. BMC Bioinf. 9, 1–11. https://doi.org/10.1186/1471-2105-9-307.

Tassan, S., Ferrari, G.M., 1995. An alternative approach to absorption measurements of aquatic particles retained on filters. Limnol. Oceanogr. 40, 1358–1368. https://doi.org/10.4319/lo.1995.40.8.1358.

Tibshirani, R., 1996. Regression Shrinkage and Selection via the Lasso. Journal of the Royal Statistical Society. Series B (Methodol.) 58, 267–288.

Warren, M.A., Simis, S.G.H., Martinez-Vicente, V., Poser, K., Bresciani, M., Alikas, K., Spyrakos, E., Giardino, C., Ansper, A., 2019. Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. Remote Sens. Environ. 225, 267–289. https://doi.org/10.1016/j.rse.2019.03.018.

Welschmeyer, N.A., 1994. Fluorometric analysis of chlorophyll a in the presence of chlorophyll b and pheopigments. Limnol. Oceanogr. 39, 1985–1992. https://doi.org/10.4319/lo.1994.39.8.1985.

Werdell, P.J., McKinna, L.I.W., Boss, E., Ackleson, S.G., Craig, S.E., Gregg, W.W., Lee, Z., Maritorena, S., Roesler, C.S., Rousseaux, C.S., Stramski, D., Sullivan, J.M., Twardowski, M.S., Tzortziou, M., Zhang, X., 2018. An overview of approaches and challenges for retrieving marine inherent optical properties from ocean color remote sensing. Prog. Oceanogr. https://doi.org/10.1016/j.pocean.2018.01.001.

Werther, M., Spyrakos, E., Simis, S.G.H., Odermatt, D., Stelzer, K., Krawczyk, H., Berlage, O., Hunter, P., Tyler, A., 2021. Meta-classification of remote sensing reflectance to estimate trophic status of inland and nearshore waters. ISPRS J. Photogramm. Remote Sens. 176, 109–126. https://doi.org/10.1016/j.isprsjprs.2021.04.003.

Zhou, Z., Hooker, G., 2021. Unbiased measurement of feature importance in tree-based methods. ACM Trans. Knowl. Discov. Data 15, 1–26.

Zibordi, G., Voss, K.J., 2014. In situ optical radiometry in the visible and near infrared. In: Experimental Methods in the Physical Sciences. Academic Press, pp. 247–304. https://doi.org/10.1016/B978-0-12-417011-7.00010-6.