Technical note: CDOM algorithm development for global inland waters

Reference: CCI-LAKES2-0006-TN

Issue 1.2 – 08/12/2022

11 rue Hermès, Parc Technologique du Canal — 31520 Ramonville Saint-Agne, France
Tel +33 (0)5 61 39 47 00 Fax +33(0)5 61 75 10 14
www.cls.fr

CHRONOLOGY ISSUES

Issue	Date	Object	Written by
1.0	15/11/2022	Initial version	Dalin Jiang, Evangelos Spyrakos – University of Stirling Stefan Simis, Xiaohan Liu – PML
1.1	22/11/2022	Revised version	Dalin Jiang, Evangelos Spyrakos – University of Stirling Stefan Simis, Xiaohan Liu – PML
1.2	08/12/2022	Revised version after ESA comments	Dalin Jiang, Evangelos Spyrakos – University of Stirling Stefan Simis, Xiaohan Liu – PML, A.Andral-CLS

Checked by	Science PIs : Stefan Simis – PML Jean-François Cretaux -LEGOS-CNES	S. Simis
Approved by	Project manager Alice Andral - CLS	Alice Andral
Authorized by	ESA Technical officer Clément Albergel - ESA	Albergel C.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

DISTRIBUTION

Company	Names	Email
ESA	Clément Albergel	clement.albergel@esa.int
BC	Carsten Brockman	carsten.brockmann@brockmann-consult.de
BC	Kerstin Stelzer	kerstin.stelzer@brockmann-consult.de
CLS	Alice Andral	aandral@groupcls.com
CLS	Anna Mangilli	amangilli@groupcls.com
CLS	Beatriz Calmettes	bcalmettes@groupcls.com
CLS	Christophe Fatras	cfatras@groupcls.com
CLS	Pierre Thibault	pthibaut@groupcls.com
CLS	Yann Bernard	ybernard@groupcls.com
CNR	Claudia Giardino	giardino.c@irea.cnr.it
CNR	Gary Free	free.g@irea.cnr.it
CNR	Mariano Bresciani	bresciani.m@irea.cnr.it
CNR	Monica Pinardi	pinardi.m@irea.cnr.it
CNR	Marina Amadori	amadori.m@irea.cnr.it
H2O Geo	Claude Duguay	claude.duguay@h2ogeomatics.com
H2O Geo	Yuhao Wu	mark.wu@h2ogeomatics.com
H2O Geo	Jaya Sree Mugunthan	jayasree.mugunthan@h2ogeomatics.com
LEGOS	Jean-François Cretaux	Jean-Francois.Cretaux@legos.obs-mip.fr
PML	Stefan Simis	stsi@pml.ac.uk
PML	Xiaohan Liu	liux@pml.ac.uk
Sertit	Hervé Yésou	Herve.yesou@unistra.fr
Sertit	Jérôme Maxant	maxant@unistra.fr
Sertit	Rémi Braun	remi.braun@unistra.fr
UoR	Chris Merchant	c.j.merchant@reading.ac.uk
UoB	Iestyn Woolway	Iestyn.woolway@bangor.ac.uk
UoR	Laura Carrea	l.carrea@reading.ac.uk
UoS	Dalin Jiang	dalin.jiang@stir.ac.uk
UoS	Evangelos Spyrakos	evangelos.spyrakos@stir.ac.uk
UoS	Ian Jones	ian.jones@stir.ac.uk

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

LIST OF CONTENTS/SOMMAIRE

1 Introduction 5
1.1 Objectives of CDOM option 5
1.2 Purpose of this document 5
2 Data collection 7
2.1 In situ data collection 7
2.2 Data processing 10
3 Assessment of published algorithm 11
3.1 Algorithm selection 11
3.2 Accuracy assessment 16
3.3 Algorithm assessment results 16
3.4 Algorithm comparison 17
4 Algorithm recalibration 20
4.1 Algorithm selection 20
4.2 Algorithm recalibration results 21
4.3 Algorithm comparison 23
5 Conclusions and recommendation 25
6 Reference 27

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

1 Introduction

1.1 Objectives of CDOM option

Lakes play an important role as regulators of the carbon cycle. Lakes can act as both a sink (sediment storage through flocculation from dissolved to particulate organic carbon) and a source for carbon (degradation and resulting mineralisation to CH4, CO and CO2). Approximately 90-95% of the Total Organic Carbon (TOC) in lakes consists of Dissolved Organic Carbon (DOC) making it one of the most effective estimates of carbon in lakes. Coloured dissolved organic matter (CDOM) is a, potentially lake or region-specific, proxy for DOC. Thus, observing CDOM at the global scale could mark a step change in studying global lake carbon.

Sentinel-3 provides the continuity (ENVISAT MERIS) of satellite observations with appropriate spectral, spatial and radiometric resolutions to retrieve CDOM time-series. The Lakes_cci project baseline activities provide chlorophyll-a and turbidity as products derived from Lake water-leaving reflectance (LWLR). However, it does not currently address CDOM because globally validated algorithms are lacking. Research is needed to bring CDOM into the adaptive algorithm selection methodology of LWLR. The development of retrieval algorithms for CDOM has thus far mainly focused on marine systems and particularly open ocean waters (e.g., Carder et al., 1999; Shanmugam, 2011). However, there has been an effort over the last decades to develop CDOM algorithms for optically complex inland waters, where suspended sediments, phytoplankton and related particles vary independently with CDOM (Kirk, 2011; Kutser et al., 2005; Olmanson et al., 2020). Although some progress has been made in the collection of in situ data and development of algorithms, there is no algorithm ready to be applied to lakes globally. The dynamic algorithm selection approach (Neil et al., 2019; Liu et al., 2021) based on the 13 distinct Optical Water Types (OWTs) defined by Spyrakos et al. (2018) provides a robust framework to introduce algorithm concepts that may not have global validity but perform well for optically distinct sets of lakes, which could be adopted for wider CDOM retrieval.

The Lakes_cci Phase 2 introduces R&D on CDOM retrieval at the global scale. The work includes a CDOM algorithm round-robin and tuning in an OWT framework, reporting to the inland water quality remote sensing community if it is possible to retrieve CDOM in global waters, for which optical cases reliable CDOM retrieval is feasible and where further work is needed. In line with existing efforts in the baseline activities for chlorophyll-a and turbidity, this work will characterise the uncertainty of the CDOM Lakes_cci product. This could ultimately lead to a global, long-term CDOM Lakes_cci product with characterised uncertainty per pixel. Activities foreseen in this work include the preparation of the in situ and satellite data to be analysed, the match up of in situ data with MERIS and OLCI, the benchmarking and tuning of CDOM algorithms for different OWTs, the validation of suggested algorithms with time series of available in situ data and the establishment of relationships between CDOM and DOC for specific lakes, regions or conditions.

1.2 Purpose of this document

This document, CCN-D-1 is the technical note of CDOM algorithm assessment for a global distribution of lakes, providing details on the in-situ data collection, selection, and assessment of published CDOM algorithms, algorithm recalibration and testing, algorithm comparison and recommendations for different water types. It also reports in which optical types the published CDOM algorithms can retrieve CDOM and where more work is needed to improve accuracy of CDOM estimates. This information will feed into deliverable CCN-D-2 “CDOM production validation” and will provide input to the final version of the

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

Algorithm Theoretical Basis Document (ATBD) and user requirement documents (URD) of the Lakes_cci project.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

2 Data collection

2.1 In situ data collection

In situ remote sensing reflectance (R_{rs}), with corresponding absorption coefficient of CDOM at 440 nm ($a_{CDOM}(440)$) data, were mainly compiled from LIMNADES database and GLORIA database. We also added previously unpublished data from the NERC project GloboLakes, and H2020 projects MONOCLE and CERTO. The latter includes in situ data from transitional aquatic systems around Europe. In total, 5082 in situ R_{rs}–$a_{CDOM}(440)$ data pairs were collected, which cover global inland and coastal waters (Figure 1), with $a_{CDOM}(440)$ ranging from 0.0025 m^{-1} to 42.47 m^{-1}, with a mean value of 1.01 m^{-1}.

R_{rs} was collected using radiometers deployed under or above the water surface. Not all in situ R_{rs} have the same spectral range: 2719 data points cover the range 400-850 nm, while 4414 data points cover the range 400-800 nm. $a_{CDOM}(440)$ from GloboLakes, MONOCLE and CERTO projects in LIMNADES was spectrophotometrically determined from 0.2 µm filtrates. Methodological detail for the other LIMNADES data sets is provided in Table 2 in Spyrakos et al. (2018), and GLORIA database details can be found in Lehmann et al. (2022).

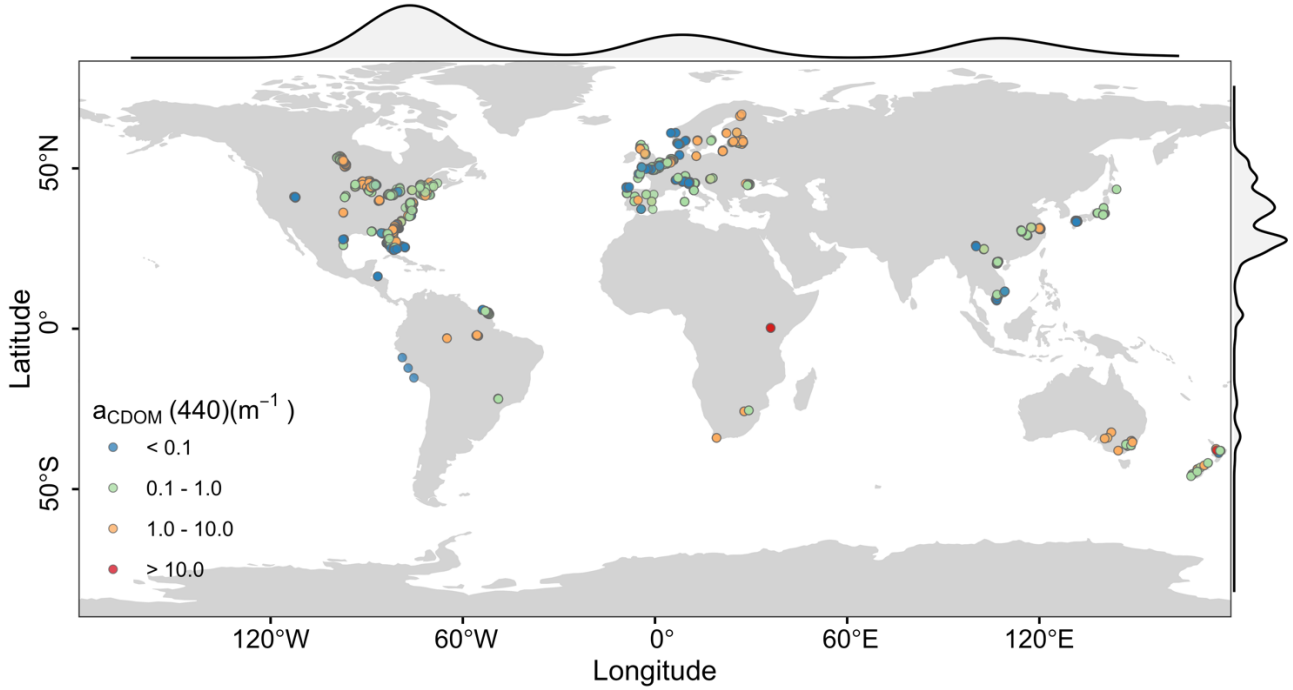

Figure 1. Spatial distribution of collected in situ R_{rs}–$a_{CDOM}(440)$ data, with colour indicating ranges of $a_{CDOM}(440)$ values, and density plots at the top and right sides of the graph are the distribution of data collected along longitudinal and latitudinal axes.

When classifying the in situ R_{rs} into the 13 lake optical water types from Spyrakos et al. (2018), the most commonly OWTs with highest similarity (based on spectral angle) were OWT 2 (16.9%), OWT 3 (18.5%), OWT 4 (11.0%) and OWT 9 (15.0%). Relatively few samples showed highest similarity to OWT 1 (0.7%), OWT 10 (0.7%) and OWT 7 (1.2%) (Table 1). A summary of characteristics of the 13 lake optical water types from Spyrakos et al. (2018) is provided in Table 2. Figure 2 shows histograms of $a_{CDOM}(440)$ associated with the samples separated by most similar OWTs. Values of $a_{CDOM}(440)$ corresponding most closely with OWT 6, 11 and 12 are generally higher than the other OWTs, and the $a_{CDOM}(440)$ of OWT 13 is the lowest among the 13 groups.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

Table 1. List of number of in situ R_{rs}-$a_{CDOM}(440)$ data for each of the OWT.

OWT	1	2	3	4	5	6	7	8	9	10	11	12	13
Number of data	37	857	941	558	128	403	63	244	764	35	394	475	183

Table 2. Summary of the 13 inland water optical water types defined in Spyrakos et al. (2018)

OWT	Dominant characteristics
1	Hypereutrophic waters with scum of cyanobacterial bloom and vegetation-like R_{rs}
2	Common case waters with diverse reflectance shape and marginal dominance of pigments and CDOM over inorganic suspended particles
3	Clear waters
4	Turbid waters with high organic content
5	Sediment-laden waters
6	Balanced effects of optically active constituents at shorter wavelength
7	Highly productive waters with high cyanobacteria abundance and elevated reflectance at red/near-infrared spectral region
8	Productive waters with cyanobacteria presence and with R_{rs} peak close to 700 nm
9	Optically neighbouring to OWT2 waters but with higher R_{rs} at shorter wavelengths
10	CDOM-rich waters
11	Waters high in CDOM with cyanobacteria presence and high absorption efficiency by NAP
12	Turbid, moderately productive waters with cyanobacteria presence
13	Very clear blue waters

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

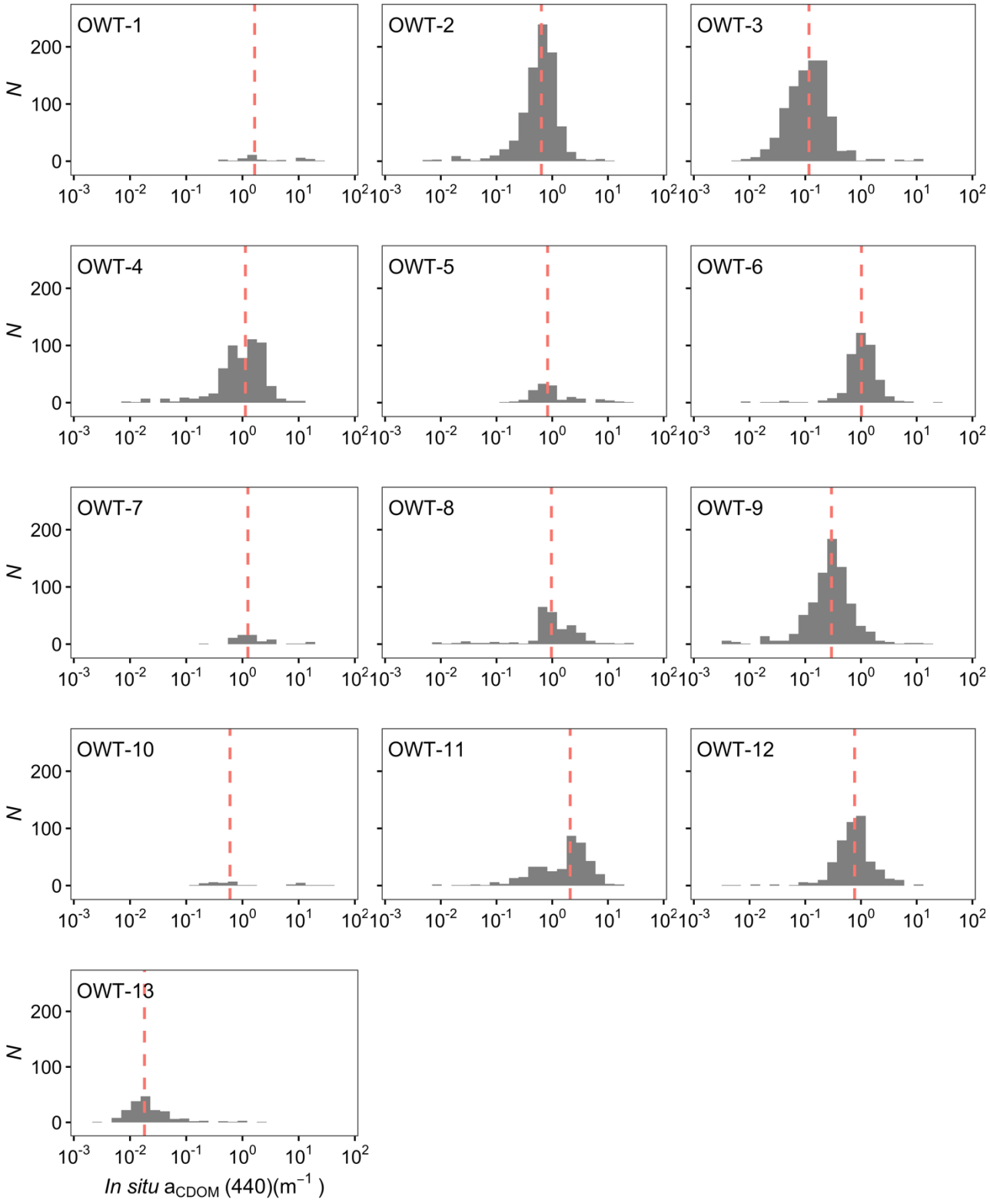

Figure 2. Histogram of in situ measured $a_{CDOM}(440)$ according to the most similar OWT. Red dashed lines are the median values.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

2.2 Data processing

Due to different instrumentations in collecting reflectance data (e.g., TriOS, ASD), the spectral intervals of original collected R_{rs} are different. In this report, all in situ R_{rs} data were firstly linearly interpolated into 1 nm intervals. Some spectra data showed clear offsets because of either residual reflected skylight or incorrect band selection for null-point correction at near-infrared wavelength, so all R_{rs} spectra were further processed using the method from either Ruddick et al. (2005) or Jiang et al. (2020) according to band availability, to address the R_{rs} magnitude offset problem. Finally, all interpolated R_{rs} spectra were further convolved to Sentinel-3 Ocean and Land Colour Instrument (OLCI) bands using the OLCI spectral response function (SRF). R_{rs} corresponding to OLCI wavebands were then used in a_{CDOM} algorithm assessment and recalibration. It should be noted that the R_{rs} in this report is defined as the remote sensing reflectance just above water surface with a units of sr^{-1}, it is the ratio between water-leaving radiance (L_w) and the irradiance (E_d) above water surface.

Some of the published algorithms we assessed are developed to estimate a_{CDOM} at 443 nm. To reduce the uncertainties caused by this band difference in algorithm assessment, 620 data pairs of in situ $a_{CDOM}(440)$ and $a_{CDOM}(443)$ were used to model and convert the relationship between the two wavebands as follows:

$$a_{\mathrm{CDOM}}(440) = 1.0495 \times a_{\mathrm{CDOM}}(443)^{1.0012} \quad (1)$$

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

3 Assessment of published algorithm

3.1 Algorithm selection

Based on comprehensive literature review, 16 algorithms were selected for the assessment of $a_{CDOM}(440)$ estimation. The selected algorithms include empirical, semi-analytical and machine learning approaches. A summary of the reference, the wavelength of estimated a_{CDOM}, the waveband(s) used for a_{CDOM} estimation, and the sensor(s) for which the algorithm was developed, are listed in Table 3. The last column in Table 3 indicates the abbreviation of the algorithm used in this report, where "org" specifies the original formulation of the algorithm prior to any adaptation of algorithm coefficients.

Table 3. Summary of algorithms for estimating $a_{CDOM}(440)$, selected from literature.

	Type	Algorithm	CDOM	Band used, sensor	Abbreviation
1	Empirical	Mannino et al., 2008	$a_{CDOM}(443)$	$R_{rs}490/R_{rs}560$, MODIS	M08-M-org
2		Mannino et al., 2008	$a_{CDOM}(443)$	$R_{rs}490/R_{rs}560$, SeaWiFS	M08-S-org
3		Mannino et al., 2014	$a_{CDOM}(443)$	$R_{rs}413/R_{rs}560$, MODIS (band ratio method)	M14-BM-org
4		Mannino et al., 2014	$a_{CDOM}(443)$	$R_{rs}413/R_{rs}560$, SeaWiFS (band ratio method)	M14-BS-org
5		Mannino et al., 2014	$a_{CDOM}(443)$	$R_{rs}443$, $R_{rs}560$, MODIS (MLR method)	M14-MM-org
6		Mannino et al., 2014	$a_{CDOM}(443)$	$R_{rs}443$, $R_{rs}560$, SeaWiFS (MLR method)	M14-MS-org
7		Ficek et al., 2011	$a_{CDOM}(440)$	$R_{rs}560/R_{rs}665$, in situ data	F11-org
8		Shanmugam, 2011	$a_{CDOM}(440)$	$R_{rs}443/R_{rs}560$, SeaWiFS	S11-org
9		Brezonik et al., 2015	$a_{CDOM}(440)$	B5/B12, Sentinel-3 OLCI	B15-org
10		Mabit et al., 2022	$a_{CDOM}(440)$	$R_{rs}665/R_{rs}560$, Landsat-8 OLI and Sentinel-2 MSI	M22-org
11	Semi-analytical	Zhu & Yu, 2013, original	$a_{CDOM}(443)$	Multiple bands, EO-Hyperion	Z13-org
12		Zhu & Yu, 2013, QAA_v6	$a_{CDOM}(443)$	--	Z13-v6
13		Zhu & Yu, 2013, QAA_Mishra	$a_{CDOM}(443)$	--	Z13-Ms
14		Zhu & Yu, 2013, QAA_hybrid	$a_{CDOM}(443)$	--	Z13-hy
15		Wang et al., 2017	$a_{CDOM}(443)$	Multiple bands, GOCI	W17-cj
16	Machine learning	Pahlevan et al., 2022	$a_{CDOM}(440)$	Landsat-8 OLI, Sentinel-2 MSI, Sentinel-3 OLCI	MDN

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

The original algorithm from Zhu & Yu (2013) is based on the Quasi-Analytical Algorithm (QAA), and some of the QAA steps are retuned in that paper. To test the feasibility of the Zhu & Yu (2013) a_{CDOM} algorithm, some other versions of QAA were selected to replace the QAA in the original Zhu & Yu (2013) algorithm, which include the QAA_v6 from IOCCG (2014), QAA_Mishra from Mishra et al. (2014), and QAA_hybrid from Jiang et al. (2019).

The equations of each of the algorithm are provided as follows, where the bands in original equations have been replaced by the nearest OLCI wavebands:

(1) Mannino et al. (2008)

$$a_{\mathrm{CDOM}}(443) = \ln\left(\frac{R_{\mathrm{rs}}(490)/R_{\mathrm{rs}}(560)-0.4363}{2.221}\right)/(-13.126) \quad (2)$$

$$a_{\mathrm{CDOM}}(443) = \ln\left(\frac{R_{\mathrm{rs}}(490)/R_{\mathrm{rs}}(560)-0.4247}{2.453}\right)/(-13.586) \quad (3)$$

Where Eq. 2 is developed for MODIS, and Eq. 3 is developed for SeaWiFS.

(2) Mannino et al. (2014)

$$a_{\mathrm{CDOM}}(443) = \ln\left(\frac{R_{\mathrm{rs}}(413)/R_{\mathrm{rs}}(560)-0.2678}{3.406}\right)/(-23.28) \quad (4)$$

$$a_{\mathrm{CDOM}}(443) = \ln\left(\frac{R_{\mathrm{rs}}(413)/R_{\mathrm{rs}}(665)-0.7857}{56.59}\right)/(-31.79) \quad (5)$$

$$a_{\mathrm{CDOM}}(443) = \exp[-3.664 - 1.291 \cdot \ln(R_{\mathrm{rs}}(443)) + 1.105 \cdot \ln(R_{\mathrm{rs}}(560))] \quad (6)$$

$$a_{\mathrm{CDOM}}(443) = \exp[-3.379 - 1.1513 \cdot \ln(R_{\mathrm{rs}}(443)) + 1.006 \cdot \ln(R_{\mathrm{rs}}(560))] \quad (7)$$

Where Eq. 4 is a band-ratio algorithm developed for MODIS, Eq. 5 a is band-ratio algorithm developed for SeaWiFS, Eq. 6 is a multiple linear regression (MLR) algorithm developed for MODIS, Eq. 7 is a multiple linear regression (MLR) algorithm developed for SeaWiFS.

(3) Ficek et al. (2011)

$$a_{\mathrm{CDOM}}(440) = 3.65[R_{\mathrm{rs}}(560)/R_{\mathrm{rs}}(665)]^{-1.93} \quad (8)$$

(4) Shanmugam (2011)

$$a_{\mathrm{CDOM}}(440) = a_{\mathrm{CDOM}}(350) \cdot e^{(-S(440-350)-\gamma^0)} \quad (9)$$

$$\gamma^0 = \frac{a_{\mathrm{CDOM}}(350)-(1/\gamma)}{a_{\mathrm{CDOM}}(350)+(1/\gamma)} \quad (10)$$

$$S = 0.0058 \times \left(\frac{a_{\mathrm{CDOM}}(412)}{a_{\mathrm{CDOM}}(350)}\right)^{-0.9677} \quad (11)$$

$$\gamma = 2.9332 \times \left(\frac{a_{\mathrm{CDOM}}(412)}{a_{\mathrm{CDOM}}(350)}\right)^{-0.7506} \quad (12)$$

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

$$a_{\mathrm{CDOM}}(350) = 0.5567 \times \left(\frac{R_{\mathrm{rs}}(443)}{R_{\mathrm{rs}}(560)}\right)^{-2.0421} \quad (13)$$

$$a_{\mathrm{CDOM}}(412) = 0.1866 \times \left(\frac{R_{\mathrm{rs}}(443)}{R_{\mathrm{rs}}(560)}\right)^{-1.9668} \quad (14)$$

(5) Brezonik et al. (2015)

$$a_{\mathrm{CDOM}}(440) = \exp[2.038 - 0.832 \times \ln\left(\frac{R_{\mathrm{rs}}(510)}{R_{\mathrm{rs}}(754)}\right)] \quad (15)$$

(6) Mabit et al. (2022)

$$a_{\mathrm{CDOM}}(440) = 20 \times \log_{10}\left(\frac{R_{\mathrm{rs}}(665)}{R_{\mathrm{rs}}(560)} + 1\right)^{1.8} \quad (16)$$

(7) Zhu & Yu (2013)

$$r_{\mathrm{rs}}(\lambda) = \frac{R_{\mathrm{rs}}(\lambda)}{0.52+2.1R_{\mathrm{rs}}(\lambda)} \quad (17)$$

$$u(\lambda) = 1 - \exp\left(\frac{-6.807 r_{\mathrm{rs}}(\lambda)^{1.186}}{0.31 - r_{\mathrm{rs}}(\lambda)}\right) \quad (18)$$

$$a(560) = a_{\mathrm{w}}(560) + 10^{-1.169-1.468x+0.274x^2} \quad (19)$$

$$x = \log_{10}\left(\frac{R_{\mathrm{rs}}(443)+R_{\mathrm{rs}}(490)}{R_{\mathrm{rs}}(560)+2\frac{R_{\mathrm{rs}}(665)}{R_{\mathrm{rs}}(490)}R_{\mathrm{rs}}(665)}\right) \quad (20)$$

$$b_{\mathrm{bp}}(560) = \frac{u(560)\times a(560)}{1-u(560)} - b_{\mathrm{bw}}(560) \quad (21)$$

$$b_{\mathrm{bp}}(443) = b_{\mathrm{bp}}(560)\left(\frac{560}{443}\right)^{Y} \quad (22)$$

$$Y = 2.2(1 - 1.2\exp\left(-0.9\frac{r_{\mathrm{rs}}(443)}{r_{\mathrm{rs}}(560)}\right)) \quad (23)$$

$$a(443) = \frac{(1-u(443))(b_{\mathrm{bw}}(443)+b_{\mathrm{bp}}(443))}{u(443)} \quad (24)$$

$$a_{\mathrm{p}}(443) = 0.63 b_{\mathrm{bp}}(560)^{0.88} \quad (25)$$

$$a_{\mathrm{CDOM}}(443) = a(443) - a_{\mathrm{w}}(443) - a_{\mathrm{p}}(443) \quad (26)$$

Where, r_{rs} is the remote sensing reflectance just below water surface, a_w is the absorption coefficient of pure water (Pope & Fry, 1997; Kou et al., 1993), b_{bw} is the backscattering coefficient of pure water (Zhang et al., 2009). In this report, we used $a_w(560)=0.062$ m^{-1}, $b_{bw}(560)=0.000779$ m^{-1}.

(8) Zhu & Yu (2013) with QAA_v6

The QAA_v6 uses $R_{rs}(665)$ to distinguish clear and turbid waters (IOCCG, 2014).

$$r_{\mathrm{rs}}(\lambda) = \frac{R_{\mathrm{rs}}(\lambda)}{0.52+1.7R_{\mathrm{rs}}(\lambda)} \quad (27)$$

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

$$u(\lambda) = \frac{-0.089+\sqrt{0.089^2+4\times0.1245\times r_{rs}(\lambda)}}{2\times0.1245} \quad (28)$$

If $R_{rs}(665) < 0.0015$ sr^{-1}, 560 nm is selected as the reference band (λ_0):

$$a(\lambda_0) = a(560) = a_w(560) + 10^{-1.146-1.366x-0.469x^2} \quad (29)$$

$$x = \log_{10}\left(\frac{r_{rs}(443)+r_{rs}(490)}{r_{rs}(560)+5\frac{r_{rs}(665)}{r_{rs}(490)}r_{rs}(665)}\right) \quad (30)$$

$$b_{bp}(\lambda_0) = b_{bp}(560) = \frac{u(560)\times a(560)}{1-u(560)} - b_{bw}(560) \quad (31)$$

If $R_{rs}(665) >= 0.0015$ sr^{-1}, 665 nm is selected as the reference band:

$$a(\lambda_0) = a(665) = a_w(665) + 0.39\left(\frac{r_{rs}(665)}{r_{rs}(443)+r_{rs}(490)}\right)^{1.14} \quad (32)$$

$$b_{bp}(\lambda_0) = b_{bp}(665) = \frac{u(665)\times a(665)}{1-u(665)} - b_{bw}(665) \quad (33)$$

Where $a_w(665)$=0.427 m^{-1}, $b_{bw}(665)$=0.000372 m^{-1}. After obtaining the $a(\lambda_0)$ and $b_{bp}(\lambda_0)$, $a(443)$ is estimated using the following equations:

$$Y = 2.0(1 - 1.2\exp(-0.9\frac{r_{rs}(443)}{r_{rs}(560)})) \quad (34)$$

$$b_{bp}(\lambda) = b_{bp}(\lambda_0)(\frac{\lambda_0}{\lambda})^Y \quad (35)$$

$$a(443) = \frac{(1-u(443))(b_{bw}(443)+b_{bp}(443))}{u(443)} \quad (36)$$

Finally, $a_{CDOM}(443)$ is estimated using Eqs. 25 – 26.

(9) Zhu & Yu (2013) with QAA_Mishra

r_{rs} and u are calculated using Eqs. 27 – 28, 709 nm is selected as the reference band (Mishra et al., 2014):

$$a(\lambda_0) = a(709) = a_w(709) + 10^{-0.7153-2.054x-1.047x^2} \quad (37)$$

$$x = \log_{10}\left(\frac{0.01\times r_{rs}(443)+r_{rs}(620)}{r_{rs}(709)+0.005\frac{r_{rs}(620)}{r_{rs}(443)}r_{rs}(620)}\right) \quad (38)$$

$$b_{bp}(\lambda_0) = b_{bp}(709) = \frac{u(709)\times a(709)}{1-u(709)} - b_{bw}(709) \quad (39)$$

Where $a_w(709)$=0.816 m^{-1}, $b_{bw}(709)$=0.000283 m^{-1}. After obtaining the $a(\lambda_0)$ and $b_{bp}(\lambda_0)$, the $a(443)$ and $b_{bp}(560)$ are calculated using Eqs. 34 – 36, and finally $a_{CDOM}(443)$ is estimated using Eqs. 25 – 26.

(10) Zhu & Yu (2013) with QAA_hybrid

The QAA_hybrid consists of QAA_v5 for clear waters and QAA_turbid for turbid waters, and the MCI is used to distinguish clear and turbid waters (Jiang et al., 2019):

$$\mathrm{MCI} = R_{rs}(709) - R_{rs}(665) - \left[\frac{(709-665)}{(754-665)}\left(R_{rs}(754) - R_{rs}(665)\right)\right] \quad (40)$$

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

If MCI <= 0.0010 sr^{-1}, 560 nm is selected as reference (QAA_v5), Eqs. 27 – 31, 34 – 36 are used to estimate $b_{bp}(560)$ and $a(443)$.

If MCI > 0.0010 sr^{-1}, 754 nm is selected as reference band (QAA_turbid):

$$a(754) \approx a_w(754) \tag{41}$$

$$b_{bp}(754) = \frac{u(754) \times a(754)}{1 - u(754)} - b_{bw}(754) \tag{42}$$

$$Y = -372.99\beta^2 + 37.286\beta + 0.84 \tag{43}$$

$$\beta = \log[u(754)/u(779)] \tag{44}$$

Where $a_w(754)$=2.868 m^{-1}, $b_{bw}(754)$=0.000217 m^{-1}. Eqs. 35 – 36 are used to estimate $b_{bp}(560)$ and $a(443)$.

Finally, $a_{CDOM}(443)$ can be estimated using Eqs. 25 – 26.

(11) Wang et al. (2017)

$$r_{rs}(\lambda) = \frac{R_{rs}(\lambda)}{\alpha(\lambda) + \beta(\lambda) R_{rs}(\lambda)} \tag{45}$$

$$\alpha(\lambda) = 0.3638 + 8.776 \times 10^{-4}\lambda - 9.193 \times 10^{-7}\lambda^2 + 3.17 \times 10^{-10}\lambda^3 \tag{46}$$

$$\beta(\lambda) = 1.357 + 8.608 \times 10^{-4}\lambda - 6.347 \times 10^{-7}\lambda^2 \tag{47}$$

u is calculated using Eq. 28, and 681 nm is selected as reference band:

$$a(\lambda_0) = a(681) = a_w(681) + 0.9398x^2 + 0.865x - 0.0852 \tag{48}$$

$$x = \frac{R_{rs}(681)}{R_{rs}(490)} \tag{49}$$

$$b_{bp}(\lambda_0) = b_{bp}(681) = \frac{u(681) \times a(681)}{1 - u(681)} - b_{bw}(681) \tag{50}$$

$$Y = 1.75 b_{bp}(681)^{-0.05} \tag{51}$$

Where $a_w(681)$=0.472 m^{-1}, $b_{bw}(681)$=0.000336 m^{-1}. $b_{bp}(560)$ and $a(443)$ are estimated using Eqs. 35 – 36. $a_p(443)$ is estimated as follows:

$$a_p(443) = 4.8024 b_{bp}(681)^{0.8055} \tag{52}$$

Finally, $a_{CDOM}(443)$ can be estimated using Eq. 26.

(12) MDN

The mixture density network (MDN) is a machine learning method proposed by Pahlevan et al. (2022), which is able to retrieve chlorophyll-a, total suspended matters (TSM) and $a_{CDOM}(440)$ from R_{rs} simultaneously. It is trained using in-situ data from global inland and coastal waters, which has a large overlap with the in-situ data used in this report. In the algorithm assessment in this report, the public available MDN model is directly used without any re-training.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

3.2 Accuracy assessment

The median absolute percentage difference (MAPD), the root mean square difference (RMSD), and Bias were used to evaluate the a_{CDOM} estimation results, with equations as follows:

$$\text{MAPD} = median\left(\left|\frac{X_{estimated,i}-X_{measured,i}}{X_{measured,i}}\right| \cdot 100\%\right) \quad (53)$$

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{N}[log_{10}(X_{estimated,i})-log_{10}(X_{measured,i})]^2}{N}} \quad (54)$$

$$\text{Bias} = 10^{Y} - 1, Y = \frac{\sum_{i=1}^{N}[log_{10}(X_{estimated,i})-log_{10}(X_{measured,i})]}{N} \quad (55)$$

where, $X_{measured}$ is in situ $a_{CDOM}(440)$ value, $X_{estimated}$ is the corresponding estimated $a_{CDOM}(440)$ value, N is the number of data points. The slope and R^2 of regression in log10 scale were also used for the assessment. It should be noted that negative estimates or estimates which higher than 500 m^{-1} are considered as invalid estimates and excluded in the error calculations.

3.3 Algorithm assessment results

Most empirical algorithms underestimate a_{CDOM} when applied to the global dataset (Figure 3a – 3j), especially in the $a_{CDOM}(440) > 1$ m^{-1} range. The two algorithms from Mannino et al. (2008) overestimate a_{CDOM} in the $a_{CDOM}(440) < 0.1$ m^{-1} range and underestimate in high $a_{CDOM}(440)$ range, with fewer than 4000 valid estimates (Figure 3a, 3b). The band ratio (BR) algorithms from Mannino et al. (2014) show similar results to Mannino et al. (2008), but with fewer valid estimates (N < 3000) (Figure 3c- 3d). The multi linear regression (MLR)-based algorithms from Mannino et al. (2014) obtain more valid estimates (> 5000) and show good results in the low $a_{CDOM}(440)$ range, but they clearly underestimate in the high $a_{CDOM}(440)$ range (Figure 3e – 3f). The results from Ficek et al. (2011) and Mabit et al. (2022) show similar results, and overall, they perform best among the tested empirical algorithms, with slopes higher than 0.5, R^2 higher than 0.4 and MAPD lower than 55% (Figure 3g, 3j). The results of those two algorithms generally align with the 1:1 line, but they show large uncertainties especially in low $a_{CDOM}(440)$ range. Results from the algorithm by Shanmugam (2011) also show good estimations in the low $a_{CDOM}(440)$ range, but underestimate in the high $a_{CDOM}(440)$ range (Figure 3h). Results from Brezonik et al. (2015) shows clear overestimations with a Bias of 2.4 (Figure 3i).

All semi-analytical algorithms show overestimations of $a_{CDOM}(440)$, especially in the high $a_{CDOM}(440)$ range (Figure 3k – 3o). Among the five tested semi-analytical algorithms, the Zhu & Yu (2013) with QAA_v6 shows fair results with R^2=0.52 and MAPD=73.0%, but with systematic bias of 0.8 (Figure 3l). The original Zhu & Yu (2013) algorithm, the Zhu & Yu (2013) with QAA_Mishra, the Zhu & Yu (2013) with QAA_hybrid, and the Wang et al. (2017) show clear overestimations with bias higher than 1 (Figure 3k, 3m – 3o), especially in high $a_{CDOM}(440)$ range.

The machine learning method (i.e., MDN) shows the best $a_{CDOM}(440)$ estimates compared to the other tested algorithms, with a MAPD of 31.9% and bias of -0.1 (Figure 3p). However, we note that there is a large overlap in the training dataset of the MDN model and this evaluation.

Generally, estimation accuracy is lower for high $a_{CDOM}(440)$ than in the low $a_{CDOM}(440)$ range based on results from the published algorithms.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

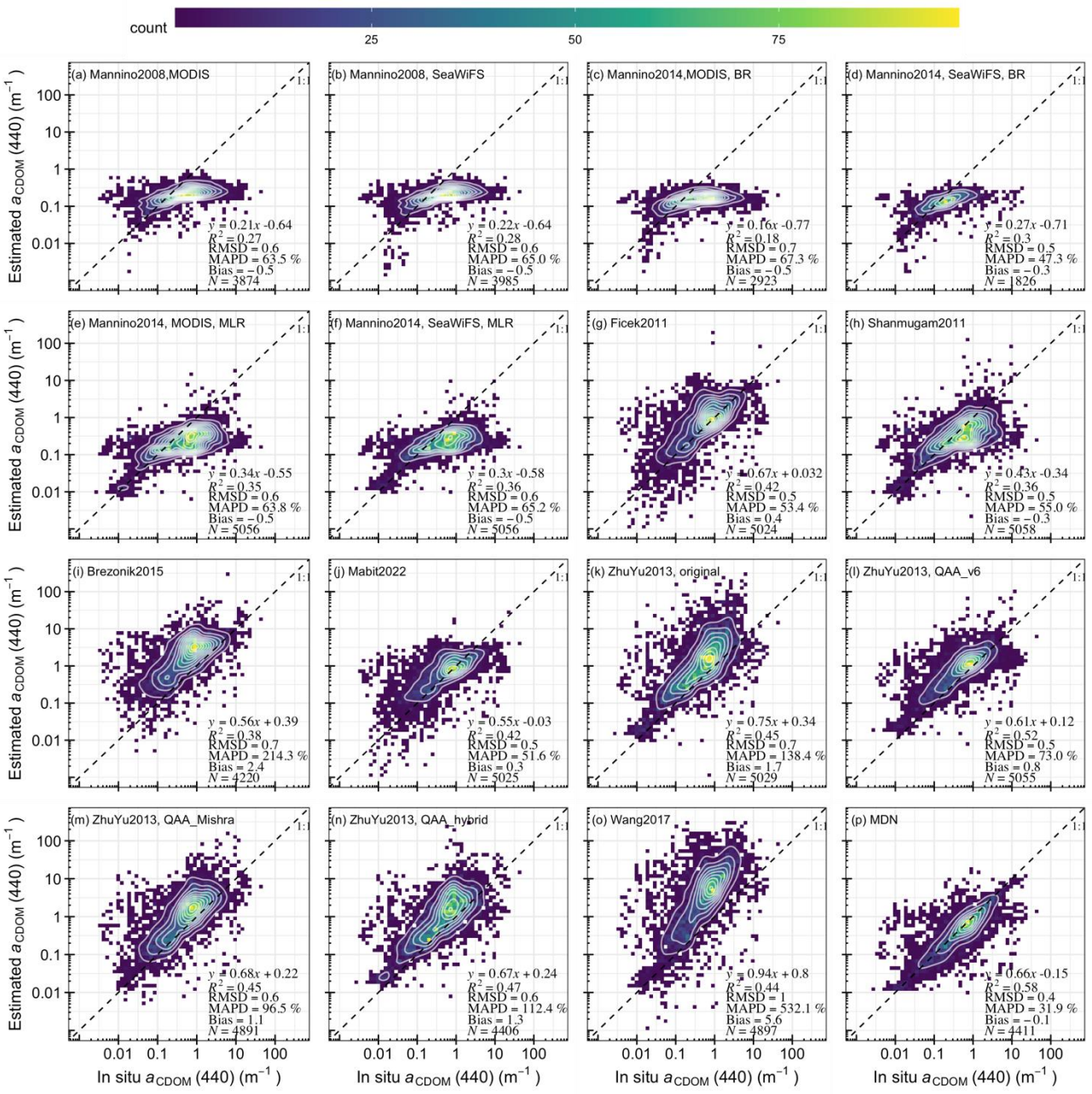

Figure 3. Comparison between in situ and estimated $a_{CDOM}(440)$ using previously published algorithms. Colour and white lines indicate the density of data points.

3.4 Algorithm comparison

Figure 4 shows a comparison of algorithm performance separated by the most similar OWT for each sample, where colour indicates the rank of the error index, having lower numbers for better performance (Slope and R^2 closer to 1, RMSD, MAPD and Bias closer to 0). MDN performed best for OWT2, 4, 5, 6, 7, 9, 11 and 12. Again, it should be noted that many of the compiled data have been used in MDN training.

When comparing the rest of the algorithms, it is found that:

- Ficek et al. (2011) is ranked high for OWT 1, 2, 4, 6 and 8.
- Shanmugam (2011) is ranked high for OWT 1, 9, 10, and 13.
- Mabit et al. (2022) is ranked high for OWT 2, 4, 6, 8, and 11.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

- The original Zhu & Yu (2013) is ranked high for OWT 5 and 12.
- The Zhu & Yu (2013) with QAA_v6 is ranked high for OWT 1, 5, 7 and 12.
- The Maninno et al. (2008, 2014) algorithms are ranked high for OWT 3 and 10.

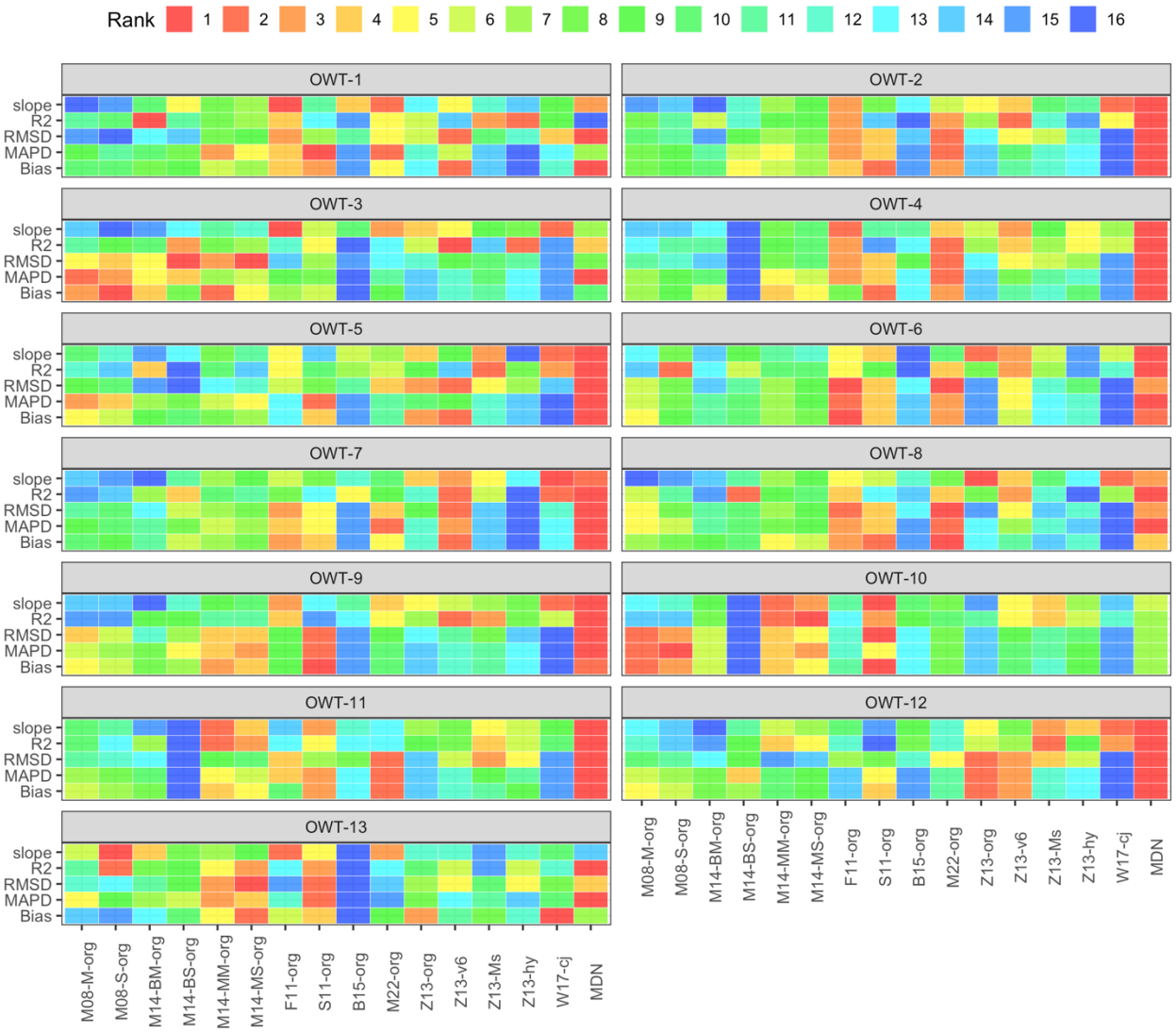

Figure 4. Rank of each error index for all the assessed algorithms for each OWT.

Some of the OWTs, e.g., OWT 4, 5, 8 and 11, even the algorithm performed the best still shows large uncertainties in $a_{CDOM}(440)$ estimations (Figure 5). The most challenging estimation is for OWT 11 (Figure 5d), which includes CDOM-rich waters (Figure 2 in this report) in the optical water type definition in Spyrakos et al. (2018).

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

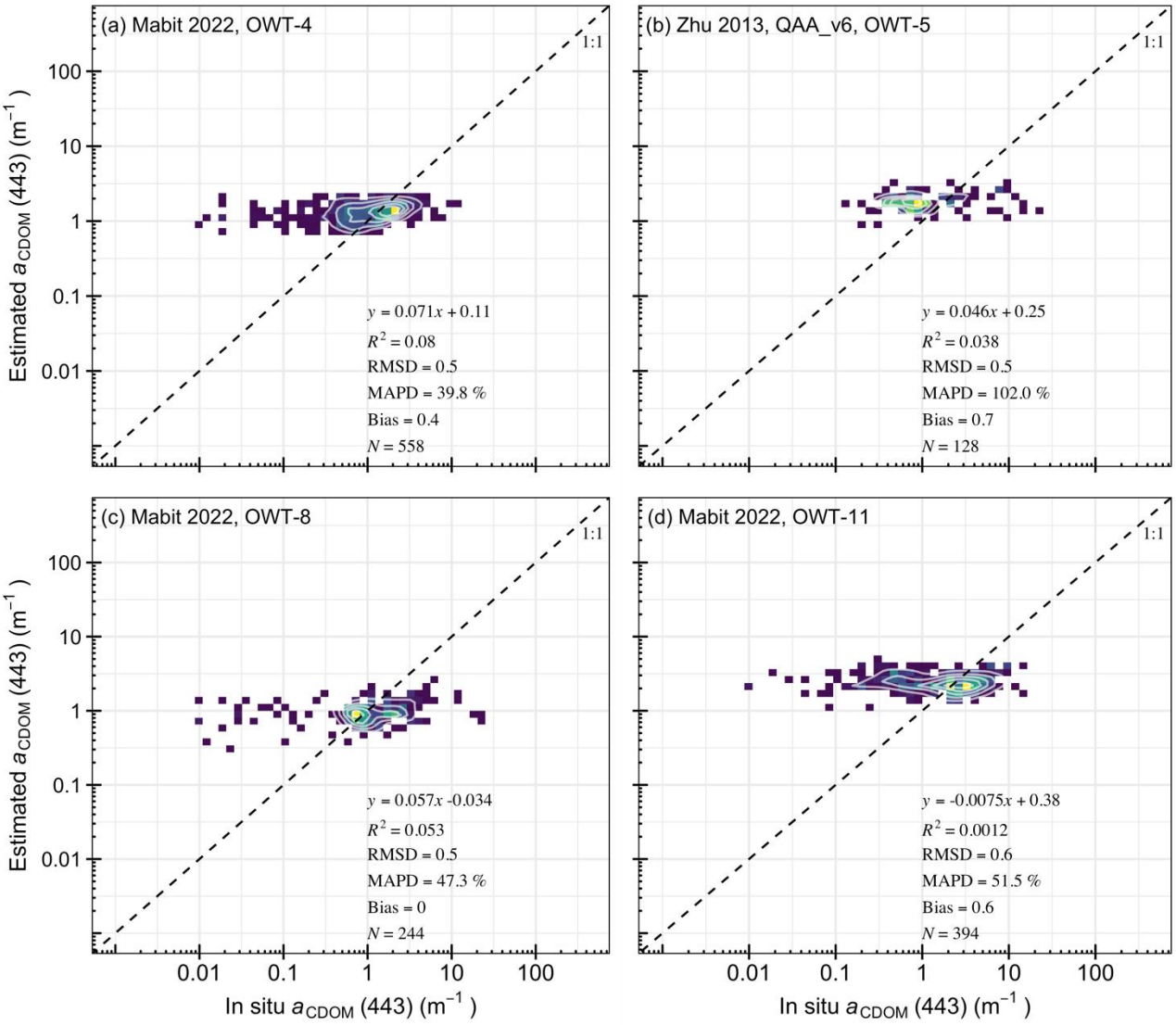

Figure 5. The estimated $a_{CDOM}(440)$ results from Mabit et al. (2022) for OWT 4, 8 and 11, from Zhu & Yu (2013) with QAA_v6 for OWT 5.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

4 Algorithm recalibration

4.1 Algorithm selection

In total, 15 published algorithms were included in recalibration (Table 4). These include single band algorithms, band ratio algorithms, and combinations thereof. Some of the algorithm (1, 3, 4, 5, 9, 10, 11 in Table 4) were already assessed in section 3 of this report, while others (2, 6, 7, 8, 12, 13, 14, 15 in Table 4) were included in addition to the earlier set. This is because the original algorithms target a_{CDOM} at other wavebands, for which fewer in situ data are available. For example, the algorithms from D'sa & Miller (2003) (6, 7, 8 in Table 4) estimate $a_{CDOM}(412)$, while Griffin et al. (2011) (15 in Table 4) estimate $a_{CDOM}(400)$. Their formulas are used in the recalibration to estimate $a_{CDOM}(440)$ in this report. The coefficients *a*, *b* and *c* in Table 4 were optimised using the non-linear square fitting in R with no pre-defined coefficients bounds. The last column of Table 4 indicates the abbreviation of each model, and they are used in the graphs in this report.

Table 4. List of selected algorithms for recalibration and corresponding references.

No.	Formula	Reference	Abbreviation
1	$a_{\mathrm{CDOM}}(440) = \exp(a \cdot \ln(R_{\mathrm{rs}}(443)) + \mathrm{b} \cdot \ln(R_{\mathrm{rs}}(560)) + c)$	Mannino et al., 2014	M14-MLR-cal
2	$a_{\mathrm{CDOM}}(440) = a(R_{\mathrm{rs}}(510)/R_{\mathrm{rs}}(665)) + b$	Castillo et al., 2008	C08-cal
3	$a_{\mathrm{CDOM}}(440) = a(R_{\mathrm{rs}}(443)/R_{\mathrm{rs}}(560))^{b}$	Shanmugam, 2011	S11-cal
4	$a_{\mathrm{CDOM}}(440) = a(R_{\mathrm{rs}}(560)/R_{\mathrm{rs}}(665))^{b}$	Ficek et al., 2011, Kutser et al., 2005, Kutser et al., 2009	F11-cal
5	$a_{\mathrm{CDOM}}(440) = \mathrm{a} \times \log_{10}(\frac{R_{\mathrm{rs}}(665)}{R_{\mathrm{rs}}(560)} + 1)^{b}$	Mabit et al., 2022, Menken et al., 2006	M22-cal
6	$a_{\mathrm{CDOM}}(440) = 10^{(a \times \log(\frac{R_{\mathrm{rs}}(413)}{R_{\mathrm{rs}}(510)}) + b)}$	D'sa & Miller, 2003	D03-413-cal
7	$a_{\mathrm{CDOM}}(440) = 10^{(a \times \log(\frac{R_{\mathrm{rs}}(443)}{R_{\mathrm{rs}}(510)}) + b)}$	D'sa & Miller, 2003	D03-443-cal
8	$a_{\mathrm{CDOM}}(440) = 10^{(a \times \log(\frac{R_{\mathrm{rs}}(510)}{R_{\mathrm{rs}}(560)}) + b)}$	D'sa & Miller, 2003	D03-510-cal
9	$a_{\mathrm{CDOM}}(440) = \exp(a \times \ln(\frac{R_{\mathrm{rs}}(510)}{R_{\mathrm{rs}}(754)}) + b)$	Brezonik et al., 2015	B15-cal
10	$a_{\mathrm{CDOM}}(440) = \ln(\frac{\frac{R_{\mathrm{rs}}(490)}{R_{\mathrm{rs}}(560)} + a}{b}) / c$	Mannino et al., 2008	M08-cal
11	$a_{\mathrm{CDOM}}(440) = \ln(\frac{\frac{R_{\mathrm{rs}}(413)}{R_{\mathrm{rs}}(560)} + a}{b}) / c$	Mannino et al., 2014	M14-BR-cal
12	$a_{\mathrm{CDOM}}(440) = a(R_{\mathrm{rs}}(665)/R_{\mathrm{rs}}(490)) + b(R_{\mathrm{rs}}(560)/R_{\mathrm{rs}}(490)) + c$	Liu et al., 2021	L21-cal

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

13	$a_{\mathrm{CDOM}}(440) = \exp(a\frac{R_{\mathrm{rs}}(443)}{R_{\mathrm{rs}}(665)} + b\frac{R_{\mathrm{rs}}(490)}{R_{\mathrm{rs}}(665)} + c)$	Olmanson et al., 2016	O16-cal
14	$a_{\mathrm{CDOM}}(440) = \exp(a\frac{R_{\mathrm{rs}}(665)}{R_{\mathrm{rs}}(560)} + b\frac{R_{\mathrm{rs}}(865)}{R_{\mathrm{rs}}(560)} + c)$	Olmanson et al., 2020	O20-cal
15	$a_{\mathrm{CDOM}}(440) = \exp(a \times \frac{R_{rs}(560)}{R_{rs}(490)} + b \times R_{rs}(665) + c)$	Griffin et al., 2011	G11-cal

4.2 Algorithm recalibration results

Figure 6 shows the results of algorithm recalibration.

The recalibrated algorithm of Mannino et al. (2008) shows slight improvement compared to the original formulation (MAPD: 63.8% vs 65.0%), while all show clear underestimations, especially in the high $a_{\mathrm{CDOM}}(440)$ range (Figure 6a).

Both the recalibrated band ratio and multiple linear regression models of Mannino et al. (2014) show slightly worse results compared to the original algorithms: the recalibrated band ratio algorithm shows clear underestimations (Figure 6b), and the multiple linear regression algorithm shows clear overestimations (Figure 6c).

The calibrated models of Ficek et al. (2011), Shammugam (2011) and Brezonik et al. (2015) show better estimations compared to the original formulations, with bias forced to zero, but remaining large uncertainty in the high $a_{\mathrm{CDOM}}(440)$ range (Figure 6d – 6f).

The recalibration of Mabit et al. (2022) shows slightly worse results compared to the original model, and overestimated $a_{\mathrm{CDOM}}(440)$, especially in the low $a_{\mathrm{CDOM}}(440)$ range (Figure 6g).

The recalibration of D'sa & Miller (2003) shows similar accuracy with the above-mentioned algorithms, with MAPD ranging 55.6% to 65.5% (Figure 6i – 6k).

The recalibration of Griffin et al. (2011) shows clear overestimations in the low $a_{\mathrm{CDOM}}(440)$ range.

The rest of the recalibrated models show either larger errors (Figure 6h, 6l, 6m) or small number of valid estimates (Figure 6n), which may due to the failure of model recalibration as the coefficients are not bounded in the recalibration, or the small number of available $R_{\mathrm{rs}}(865)$.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

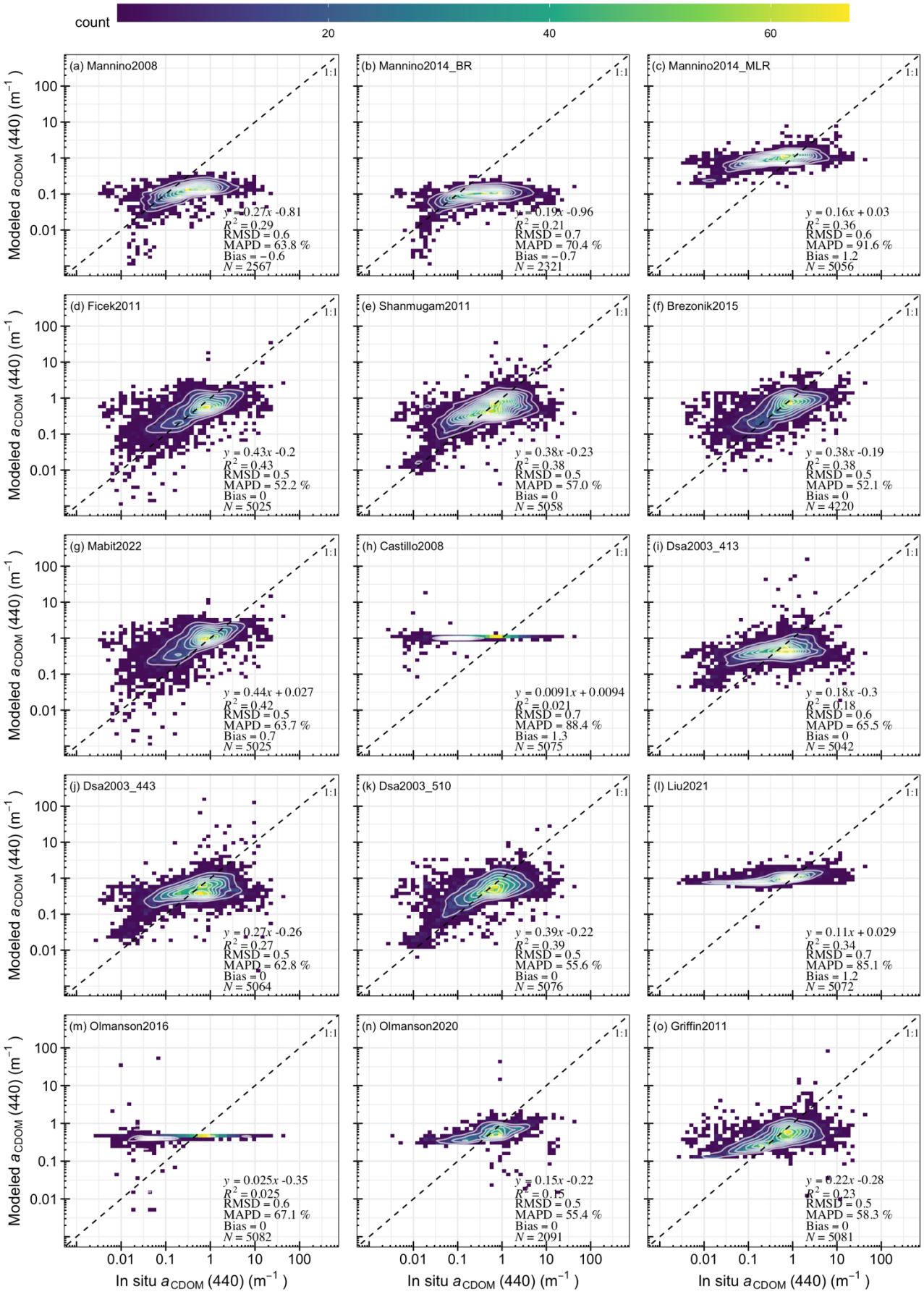

Figure 6. Results of recalibrated algorithms, where the formula of models are listed in Table 4.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

4.3 Algorithm comparison

Figure 7 shows the comparison of all original (section 3) and recalibrated (section 4) algorithms, where colour indicates the rank of each error index, i.e., slope and R^2 closer to 1, RMSD, MAPD and Bias closer to 0, the higher of the rank (red with lower number). It should be noted that the number of valid estimates is also considered in the algorithm comparison.

In terms of each OWT, it is found that:

OWT 1: the original and calibrated Shanmugam (2011), MDN, the Zhu & Yu (2013) with QAA_v6, and the calibrated Mannino et al. (2014) MLR algorithms perform better than the others.

OWT 2: the MDN, the original and calibrated Ficek et al. (2011), and the original Mabit et al. (2022), the calibrated D'sa & Miller (2003) algorithms performed better than others. The calibrated Olmanson et al. (2020) is ranked high as well, but with less valid estimates.

OWT 3: the MDN, the two original Mannino et al. (2008) and four original Mannino et al. (2014) and the original Shanmugam (2011) algorithms perform better than others.

OWT 4: The MDN, the original and calibrated Ficek et al. (2011), the original and calibrated Mabit et al. (2022) perform better than others.

OWT 5: The MDN, the calibrated Mannino et al. (2014) MLR, the calibrated Brezonik et al. (2015) and the calibrated Ficek et al. (2011) perform better than others. The calibrated Olmanson et al. (2020) is ranked high as well, but with less valid estimates.

OWT 6: The MDN, the original and calibrated Mabit et al. (2022), the original Ficek et al. (2011) perform better than others.

OWT 7: The MDN, the calibrated Mannino et al. (2014) MLR, the calibrated Liu et al. (2021), the Zhu & Yu (2013) with QAA_v6 and original Mabit et al. (2022) algorithms perform better than others.

OWT 8: The MDN, the original Mabit et al. (2022), the original Ficek et al. (2011), and the calibrated Griffin et al. (2011) algorithms perform better than others.

OWT 9: The MDN, the calibrated Ficek et al. (2011), the original Shanmugam (2011) and the calibrated Griffin et al. (2011) algorithms perform better than others.

OWT 10: The calibrated and original Shanmugam (2011), and the calibrated D'sa & Miller (2003) algorithms perform better than the others.

OWT 11: The MDN, the original and calibrated Mabit et al. (2022), the calibrated Mannino et al. (2014) MLR, the calibrated Ficek et al. (2011) algorithms perform better than others.

OWT 12: The MDN, the Mannino et al. (2014) MLR, the calibrated Ficek et al. (2011), and the calibrated Brezonik et al. (2015) algorithms perform better than others. The calibrated Olmanson et al. (2020) is ranked high as well, but with less valid estimates.

OWT 13: The original and calibrated Shanmugam (2011), the two original Mannino et al. (2014) MLR, and the MDN algorithms perform better than others.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

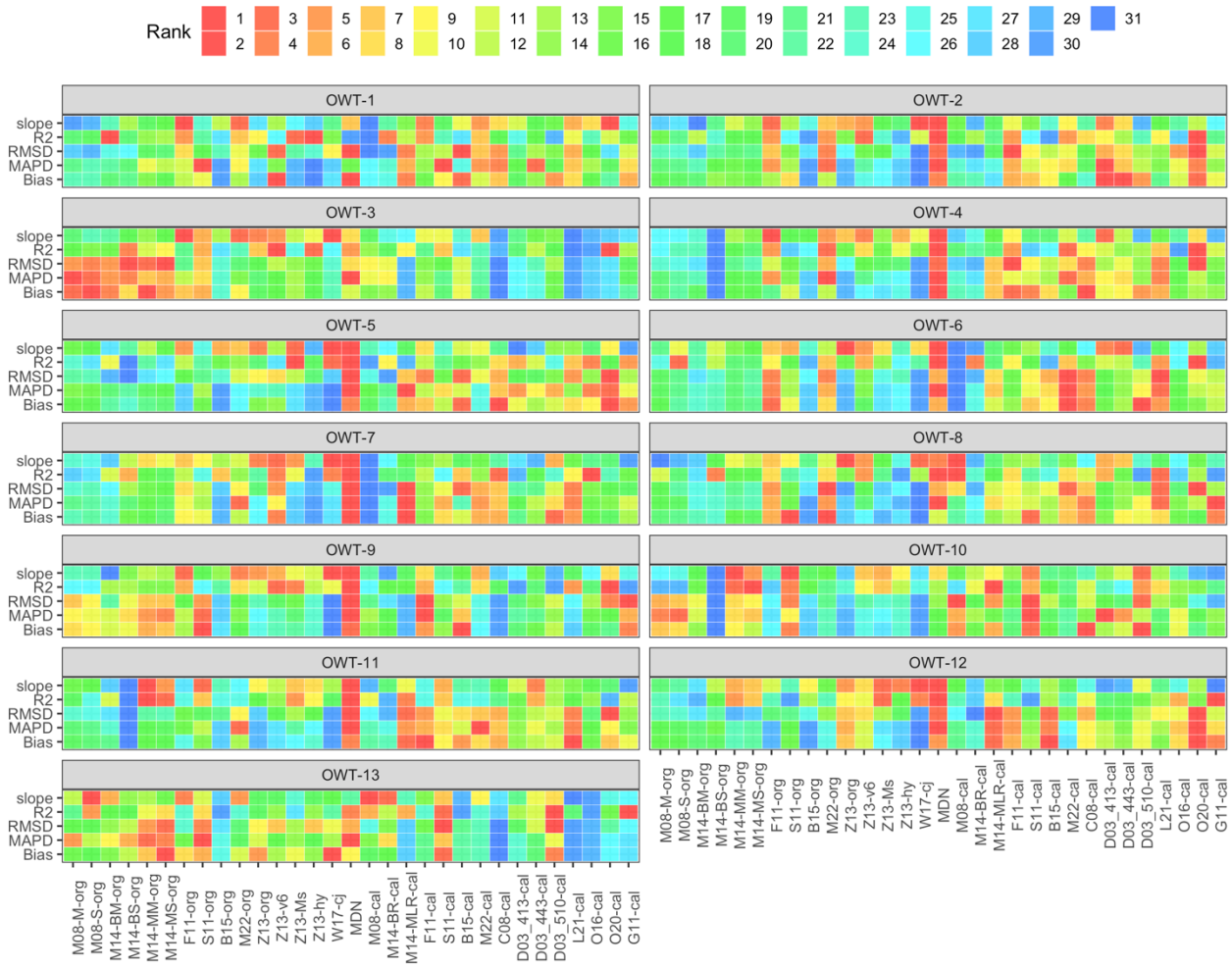

Figure 7. Comparison of performance for all original and recalibrated algorithms.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

5 Conclusions and recommendation

16 previously published algorithms and 15 recalibrated algorithms for estimating $a_{CDOM}(440)$ are evaluated in this report. Based on these first results, it is possible to estimate $a_{CDOM}(440)$ in medium-low value ranges (e.g., < 1 m^{-1}) from several published or recalibrated algorithms (e.g., Shanmugam (2011), Ficek et al. (2011)), but results show that high uncertainties in estimating $a_{CDOM}(440)$ in high value ranges (e.g., > 1 m^{-1}) remains from the published or recalibrated algorithms, especially for OWT 8, 11, and 12.

Model recalibrations showed slightly improvement in $a_{CDOM}(440)$ estimations for some OWTs (e.g., OWT 5, 6, 10, 11 and 12, see Figure 7), but they cannot address the problem of high uncertainties in estimating $a_{CDOM}(440)$ in high value ranges. The difficulty of accurately estimating $a_{CDOM}(440)$ in high value ranges is probably due to: (1) the limited remote sensing signal from water because of strong absorption of CDOM, which are not able to detect the variations of $a_{CDOM}(440)$ in the water, (2) the fact that phytoplankton pigment and suspended matter may mask the a_{CDOM} spectral signature, thus decreasing algorithm specificity and increasing algorithmic uncertainty.

From the comparison of different algorithms in section 4.3, it is found that the original or the calibrated Shanmugam (2011), Ficek et al. (2011), Mabit et al. (2022), Mannino (2014) MLR, and the calibrated Brezonik et al. (2015) algorithms are frequently ranked high, which means they have higher accuracy in $a_{CDOM}(440)$ estimation than the other algorithms. Further analysis reveals that:

- For OWT 4, 6, 8 and 11, either the Ficek et al. (2011) or Mabit et al. (2022) are ranked high, and the common feature of these two algorithms is that both use the band ratio $R_{rs}(560)/R_{rs}(665)$. Therefore, the green-red band ratio algorithm is more likely to diagnostically capture the $a_{CDOM}(440)$ variations for those water types.
- For OWT 1, 3, 10 and 13, algorithms which use short wavelengths (blue to green) are ranked high (e.g., Shanmugam (2011), Maninno et al. (2014) MLR, D'sa & Miller (2003)). Therefore, blue to green wavelengths are useful in $a_{CDOM}(440)$ estimation for those water types where phytoplankton abundance is low. Although OWT-1 has high reflectance in NIR, the blue to green wavelengths still show ability for $a_{CDOM}(440)$ estimation.
- NIR wavebands shows potential in $a_{CDOM}(440)$ estimations for OWT 5 and OWT 12, for which the algorithms from Brezonik et al. (2015) and Olmanson et al. (2020) are ranked high.

To inform next steps, there are several possible ways to improve the accuracy of $a_{CDOM}(440)$ estimations, which include:

- Apply a linear calibration to the results of published algorithms, for example, the Zhu & Yu (2013) with QAA_v6 shows systematic overestimation of $a_{CDOM}(440)$, a linear function to shift the estimations and make it more accurate would be useful, especially in medium-low $a_{CDOM}(440)$ ranges (will be done within the CDOM option).
- Select the best algorithm for each OWT and use weighted blending of results to adjust in ranges where an observation bears similarity to more than one OWT (will be carried out within the CDOM option).
- Recalibrate the algorithms using satellite-derived reflectance, which is needed to eliminate systematic bias from atmospheric correction in the $a_{CDOM}(440)$ retrieval accuracy (will be done within the CDOM option).
- Bound the range of coefficients to be recalibrated in model recalibration process to prevent losing sensitivity of original algorithms to $a_{CDOM}(440)$ variations (will be done within the previous activity using satellite-derived reflectance).
- Quantify the uncertainties from each algorithm (ideally per OWT), then flag or exclude estimates where uncertainties exceed acceptable thresholds (will be carried out within the CDOM option).

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

- Develop new $a_{CDOM}(440)$ algorithms, making use of increased in situ datasets, especially for CDOM-rich waters (depending on available resources).

Based on the current results of assessment of published and recalibrated algorithms using in situ observations, the current recommendation is to consider:

- the original algorithm of Shanmugam (2011) for OWT 1, 3, 10 and 13.
- the original algorithm of Ficek et al., (2011) for OWT 2, 4, and 9.
- the calibrated MLR algorithm of Mannino et al. (2014) for OWT 5, 7 and 12.
- the calibrated Mabit et al. (2022) algorithm for OWT 6, 8 and 11.

The MDN should be considered as well, after carefully interpreting independent validation results. The results of the modelled $a_{CDOM}(440)$ when switching between the recommended algorithms based on the most similar OWT for each sample is shown in Figure 8, with lower MAPD (42.5%) and RMSD (0.4) than published and recalibrated algorithms.

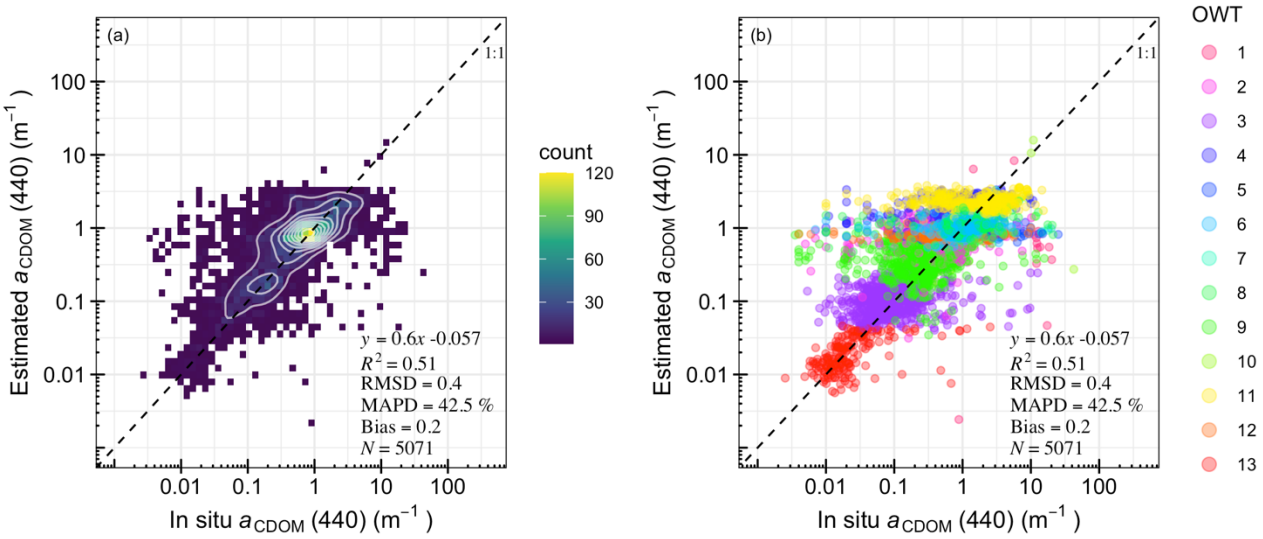

Figure 8. In situ vs modelled $a_{CDOM}(440)$ when combing the recommended algorithms. (a) Data shown as density plot, (b) data shown for each OWT.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

6 Reference

Brezonik, P. L., Olmanson, L. G., Finlay, J. C., & Bauer, M. E. (2015). Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. Remote Sensing of Environment, 157, 199-215.

Carder, K. L., Chen, F. R., Lee, Z. P., Hawes, S. K., & Kamykowski, D. (1999). Semianalytic Moderate-Resolution Imaging Spectrometer algorithms for chlorophyll a and absorption with bio-optical domains based on nitrate-depletion temperatures. *Journal of Geophysical Research: Oceans*, *104*(C3), 5403-5421.

D'Sa, E. J., & Miller, R. L. (2003). Bio-optical properties in waters influenced by the Mississippi River during low flow conditions. Remote sensing of environment, 84(4), 538-549.

Del Castillo, C. E., & Miller, R. L. (2008). On the use of ocean color remote sensing to measure the transport of dissolved organic carbon by the Mississippi River Plume. Remote Sensing of Environment, 112(3), 836-844.

Ficek, D., Zapadka, T., & Dera, J. (2011). Remote sensing reflectance of Pomeranian lakes and the Baltic. Oceanologia, 53(4), 959-970.

Griffin, C. G., Frey, K. E., Rogan, J., & Holmes, R. M. (2011). Spatial and interannual variability of dissolved organic matter in the Kolyma River, East Siberia, observed using satellite imagery. Journal of geophysical research: Biogeosciences, 116(G3).

IOCCG, 2014. Update of the Quasi-Analytical Algorithm (QAA_v6). Available online. http://www.ioccg.org/groups/Software_OCA/QAA_v6_2014209.pdf.

Jiang, D., Matsushita, B., Setiawan, F., Vundo, A., 2019. An improved algorithm for estimating the Secchi disk depth from remote sensing data based on the new underwater visibility theory. ISPRS J. Photogramm. Remote Sens. 152, 13–23.

Jiang, D., Matsushita, B., Yang, W., 2020. A simple and effective method for removing residual reflected skylight in above-water remote sensing reflectance measurements. ISPRS J. Photogramm. Remote Sens. 165, 16–27.

Kirk, J. (2011). Light and photosynthesis in aquatic ecosystems. Cambridge: Cambridge university press.

Kou, L., Labrie, D., Chylek, P., 1993. Refractive indices of water and ice in the 0.65- to 2.5-μm spectral range. Appl. Opt. 32 (19), 3531–3540.

Kutser, T., Pierson, D. C., Kallio, K. Y., Reinart, A., & Sobek, S. (2005). Mapping lake CDOM by satellite remote sensing. Remote Sensing of Environment, 94(4), 535-540.

Kutser, T., Tranvik, L., & Pierson, D. C. (2009). Variations in colored dissolved organic matter between boreal lakes studied by satellite remote sensing. Journal of Applied Remote Sensing, 3(1), 033538.

Lehmann, M., et al., 2022. GLORIA - A globally representative hyperspectral in situ dataset for optical sensing of water quality. Scientific Data (submitted).

Liu, G., Li, S., Song, K., Wang, X., Wen, Z., Kutser, T., ... & Hou, J. (2021). Remote sensing of CDOM and DOC in alpine lakes across the Qinghai-Tibet Plateau using Sentinel-2A imagery data. Journal of Environmental Management, 286, 112231.

Mabit, R., Bélanger, S., Araújo, C. A., & Singh, R. K. (2022). Empirical remote sensing algorithms to retrieve SPM and CDOM in Québec coastal waters. Frontiers in Remote Sensing, 15.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.

Zoho Sign Document ID: 4AC07D10E-STKMD030NUQSBWTQ2TOLFY9DNQJH9DNBP4ICMJGAYGU

Mannino, A., Novak, M. G., Hooker, S. B., Hyde, K., & Aurin, D. (2014). Algorithm development and validation of CDOM properties for estuarine and continental shelf waters along the northeastern US coast. Remote Sensing of Environment, 152, 576-602.

Mannino, A., Russ, M. E., & Hooker, S. B. (2008). Algorithm development and validation for satellite-derived distributions of DOC and CDOM in the US Middle Atlantic Bight. Journal of Geophysical Research: Oceans, 113(C7).

Menken, K. D., Brezonik, P. L., & Bauer, M. E. (2006). Influence of chlorophyll and colored dissolved organic matter (CDOM) on lake reflectance spectra: Implications for measuring lake properties by remote sensing. Lake and Reservoir Management, 22(3), 179-190.

Mishra, S., Mishra, D. R., & Lee, Z. (2013). Bio-optical inversion in highly turbid and cyanobacteria-dominated waters. IEEE transactions on geoscience and remote sensing, 52(1), 375-388.

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.

Olmanson, L. G., Brezonik, P. L., Finlay, J. C., & Bauer, M. E. (2016). Comparison of Landsat 8 and Landsat 7 for regional measurements of CDOM and water clarity in lakes. Remote Sensing of Environment, 185, 119-128.

Olmanson, L. G., Page, B. P., Finlay, J. C., Brezonik, P. L., Bauer, M. E., Griffin, C. G., & Hozalski, R. M. (2020). Regional measurements and spatial/temporal analysis of CDOM in 10,000+ optically variable Minnesota lakes using Landsat 8 imagery. Science of the Total Environment, 724, 138141.

Pahlevan, N., Smith, B., Alikas, K., Anstee, J., Barbosa, C., Binding, C., ... & Ruiz-Verdù, A. (2022). Simultaneous retrieval of selected optical water quality indicators from Landsat-8, Sentinel-2, and Sentinel-3. Remote Sensing of Environment, 270, 112860.

Pope, R.M., Fry, E.S., 1997. Absorption spectrum (380–700 nm) of pure water. II. Integrating cavity measurements. Appl. Opt. 36 (33), 8710–8723.

Ruddick, K., De Cauwer, V., Van Mol, B., 2005. Use of the near infrared similarity reflectance spectrum for the quality control of remote sensing data. In: Remote Sensing of the Coastal Oceanic Environment. International Society for Optics and Photonics.

Shanmugam, P. (2011). New models for retrieving and partitioning the colored dissolved organic matter in the global ocean: Implications for remote sensing. Remote Sensing of Environment, 115(6), 1501-1521.

Spyrakos, E., O'donnell, R., Hunter, P. D., Miller, C., Scott, M., Simis, S. G., ... & Tyler, A. N. (2018). Optical types of inland and coastal waters. Limnology and Oceanography, 63(2), 846-870.

Wang, Y., Shen, F., Sokoletsky, L., & Sun, X. (2017). Validation and calibration of QAA algorithm for CDOM absorption retrieval in the Changjiang (Yangtze) estuarine and coastal waters. Remote Sensing, 9(11), 1192.

Zhang, X., Hu, L., He, M.X., 2009. Scattering by pure seawater: effect of salinity. Opt. Express 17 (7), 5698–5710.

Zhu, W., Yu, Q., & Tian, Y. Q. (2013). Uncertainty analysis of remote sensing of colored dissolved organic matter: Evaluations and comparisons for three rivers in North America. ISPRS journal of photogrammetry and remote sensing, 84, 12-22.

Open/Public/Público © 2019 CLS. All rights reserved. Proprietary and Confidential.