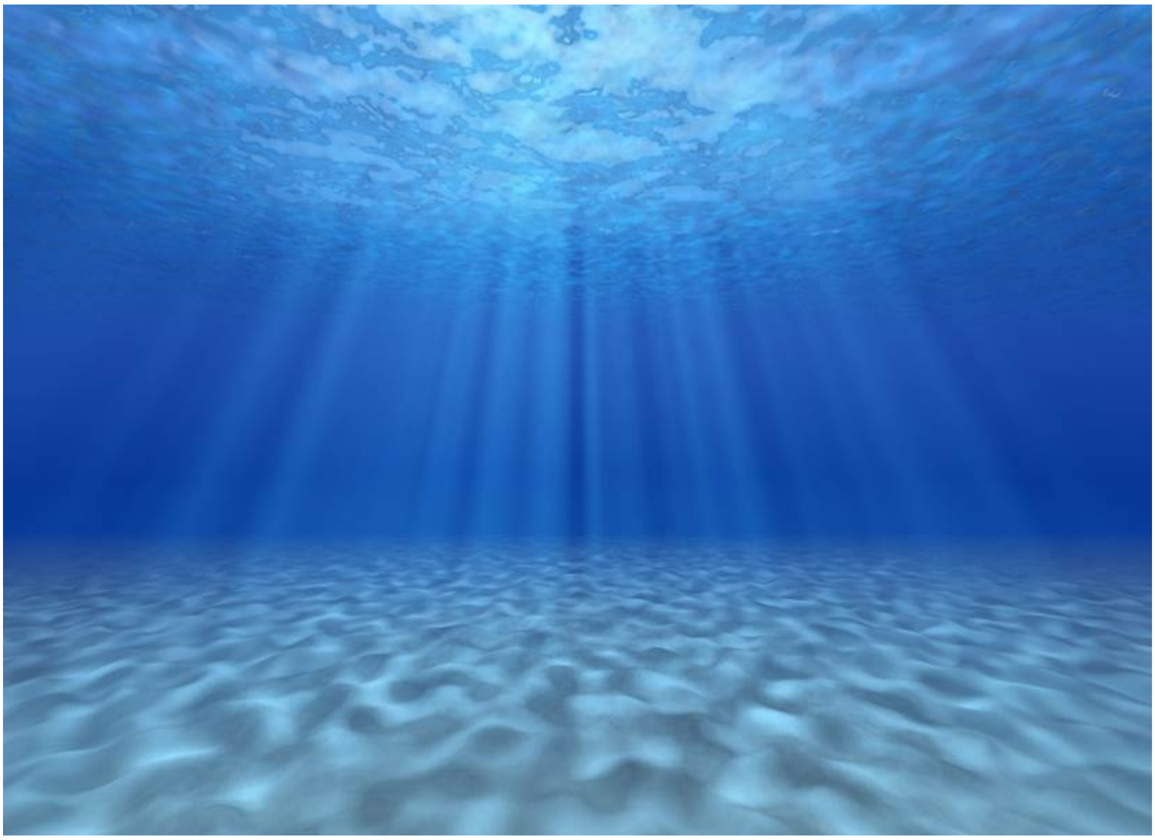
Sentinel-3 OLCI Inherent Optical Properties

Product Validation Report

PVR



Date: 31 July 2019

EUMETSAT contract: 17/214791

EUMETSAT Ref.: EUM/RSP/REP/20/1160645

Acronyms

PVR	Product Validation Report
ESA	European Space Agency
OLCI	Ocean and Land Colour Imager

Symbol	definition	Dimension
λ	Wavelength	nm
θ_s	Sun Zenith Angle	Degrees
μ_w	Cosine of the angle of refraction of the solar beam just beneath the sea surface	rad
η	Ratio of molecular scattering to total scattering	

Radiometry and Apparent Optical Properties (AOPs)

R_{rs}	Remote Sensing Reflectance (above water)	sr^{-1}
r_{rs}	Remote Sensing Reflectance (bellow water)	sr^{-1}
K_d	Attenuation coefficient for downwelling irradiance	m^{-1}

Inherent Optical Properties (IOPs)

$a(\lambda)$	Total absorption coefficient	m^{-1}
$a_w(\lambda)$	Pure sea water absorption coefficient	m^{-1}
$a_{nw}(\lambda)$	Non water absorption coefficient	m^{-1}
$a_{phy}(\lambda)$	Phytoplankton absorption coefficient	m^{-1}
$a^*_{phy}(\lambda)$	Phytoplankton specific absorption coefficient	$m^2\ mg^{-1}$
$a_{cdm}(\lambda)$	Colored Dissolved Matter absorption coefficient	m^{-1}
$a_{CDOM}(\lambda)$	Colored Dissolved Organic Matter absorption coefficient	m^{-1}
$b_b(\lambda)$	Total backscattering coefficient	m^{-1}
$b_{bp}(\lambda)$	Particulate backscattering coefficient	m^{-1}
$b_w(\lambda)$	Pure sea water backscattering coefficient	m^{-1}

S	CDOM spectral slope	m^{-1}
Y	Particle backscattering slope	m^{-1}

Table of Content

Page

List of Figures

Page

List of Tables

Page

1 Introductions

The purpose of this document is to show the performance of the 2SAA IOPs-algorithm (described in the ATBD) using different synthetic, in situ, and match-up data sets. While the final products will all be provided at 443 nm, this report also shows the performance of the model at all available bands.

The operational IOPs are the following ones:

- The non-water absorption coefficient, a_{nw} (443), and particulate backscattering coefficient, b_{bp} (443)
- The absorption coefficient by colored dissolved organic matter, a_{cdm} (443)
- The absorption coefficient by phytoplankton, a_{phy} (443)
- The absorption coefficient by colored dissolved organic matter, a_{cdom} (443).

The other IOPs that will be delivered through SNAP as demonstrative products are:

- The non-water absorption coefficient and particulate backscattering coefficient at 400, 412.5, 490, 510, 560, 620, and 665
- The vertical attenuation coefficient for the diffuse attenuation coefficient for downward irradiance at 490 nm.
- The absorption coefficient by phytoplankton at 400, 412.5, 490, 510, 560, 620, and 665 nm.
- The spectral slope of $b_{bp}(\lambda)$

Another set of parameters will be provided as measurements of uncertainty:

- The percentage of error during the Rrs reconstruction
- The percentage of uncertainty based of the given water class (fixed)
- A flag addressing the possible failures during the inversion.

Three data sets are used to evaluate the performance of the two-step algorithm (2SAA) to retrieve IOPs from Rrs (see ATBD). The first step used Loisel et al., 2018) algorithm (LS2) to obtain the total absorption and backscattering coefficients from Rrs. The second step focusses on the estimation of the absorption coefficient of phytoplankton (a_{phy}) and colored detrital matter (a_{cdm}) from non-water absorption coefficient (a_{nw}) using a slightly modified version of Zhang et al., 2015. a_{cdom} is estimated from a modified version of Loisel et al.,2014.

The data sets are divided in three groups, i) synthetic, which were generated through radiative transfer simulations with input of synthetic IOP data, ii) *in situ* with IOPs measurements which are subject to measurements uncertainties, and iii) matchup, which were obtained from remote sensing images from past and current ocean color sensors and in situ measurements collected from oceanographic cruises or from bio-argo floats (only for the particulate backscattering coefficient, b_{bp}).

2 Description of the synthetic data sets

The synthetic group is composed of two data sets, IOCCG and CCRR. IOCCG was created as part of the Ocean Colour Coordinating Group (IOCCG) Working Group (IOCCG, 2006) project on inverse bio-optical algorithms, where IOPs are mainly driven by the chlorophyll concentration. CCRR was developed in the frame of CoastColour Round Robin for coastal waters (Nechad et al., 2015), with a broader range of IOPs, in which, the IOPs may or not be driven by chlorophyll. As synthetic data set are free of measurement errors, it allows evaluating the uncertainties of the tested models, which are associated solely with the algorithmic formulation.

2.1 The IOCCG data set

The IOCCG data set covers a large range of bio-optical properties of natural waters (**Figure 1**). For example, the a_{nw} values at 443 nm, which represents a sum of phytoplankton, non-algal particulate and CDOM contributions, ranges between 0.01 and 3.17 m^{-1} with a median value of 0.216 m^{-1}. The a_{phy} (443) values range between 0.006 and 0.42 m^{-1} with a median value of 0.057 m^{-1}. The a_{cdm} (443) values range between 0.005 and 2.75 m^{-1} with a median value of 0.159 m^{-1}. The particulate backscattering coefficient b_{bp} (443) ranges between 0.0006 and 0.127 m^{-1} with a median value of 0.012 m^{-1}. Overall 500 IOP scenarios were used as input to generate the IOCCG data set. For each set of input IOPs the simulations were made for three sun zenith angles of 0, 30, and 60°.

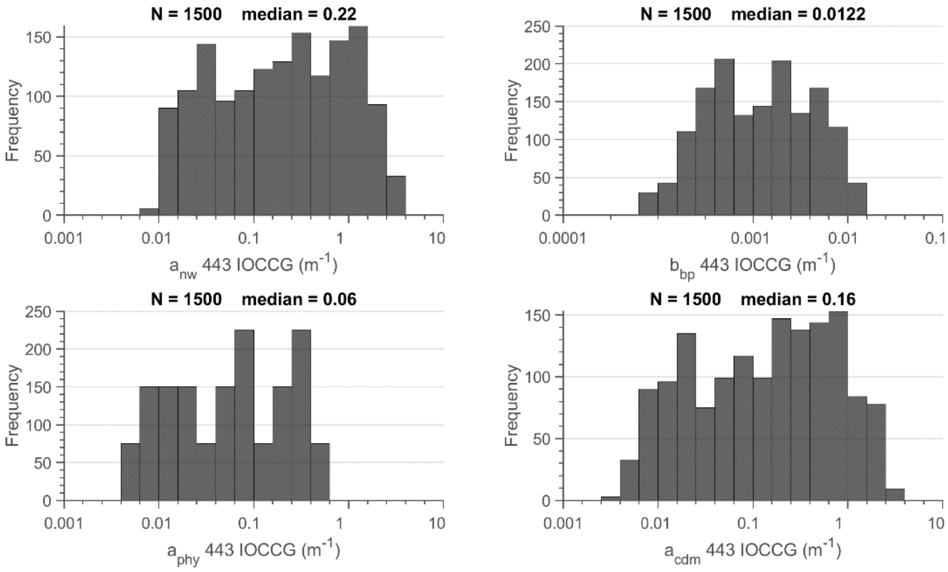


Figure 1: Histogram of the IOPs distribution for the IOOCG synthetic data set.

2.2 The CCRR data set

The data set CCRR covers an even broader range, including coastal waters and extremely complex waters (**Figure 2**). The a_{nw} values at 443 nm range between 0.01 and 22.89 m^{-1} with a median value of 0.32 m^{-1}. The a_{phy} (443) values range between 0.005 and 1.49 m^{-1} with a median value of 0.1 m^{-1}. The a_{cdm} (443) values range between 0.005 and 2.75 m^{-1} with a median value of 0.2 m^{-1}. The particulate backscattering coefficient b_{bp} (550) values range between 0.00016 and 4.64 m^{-1}with a median value of 0.0186 m^{-1}. Overall 5000 IOP scenarios were used as input to generate CCRR data set. For each set of input IOPs the simulations were made for three sun zenith angles of 0, 40, and 60°.

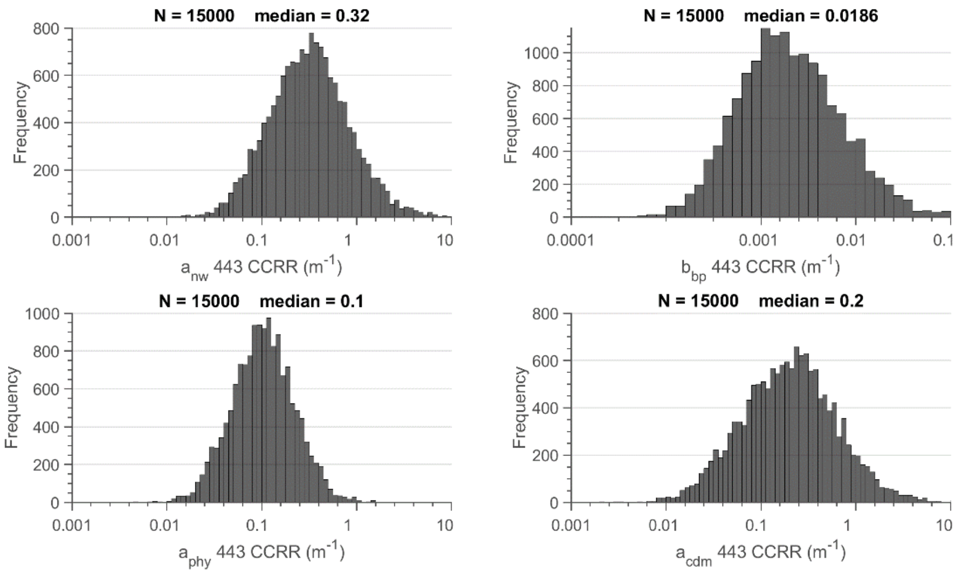


Figure 2: Histogram of the IOPs distribution for the CCRR synthetic data set.

3 Description of the *in situ* data sets

3.1 Open and coastal waters data set

The in situ data set is composed of twelve different data sets gathering measurements collected in open, coastal, and inland waters (Table 1). The first seven data sets comprise in situ measurements collected in different oceanic and coastal environments, and span a broad range of trophic and environmental conditions, while the last five data sets comprise in situ measurements collected in inland waters (lakes and Saint Laurent river). These data sets are used to evaluate the model for the scenario in which in situ measurements of R_{rs} provide input to the model. In this study, the model-derived IOPs are compared with in situ measurements of IOPs. Thus, this assessment is subject to uncertainties in both the algorithmic formulation of the model and measurements of R_{rs} and IOPs. For the evaluation we separated the in situ group in 2 data sets: coastal and ocean waters, and inland waters. The global distribution of IOPs for each in situ data sets is provided in **Figure 3-7**.

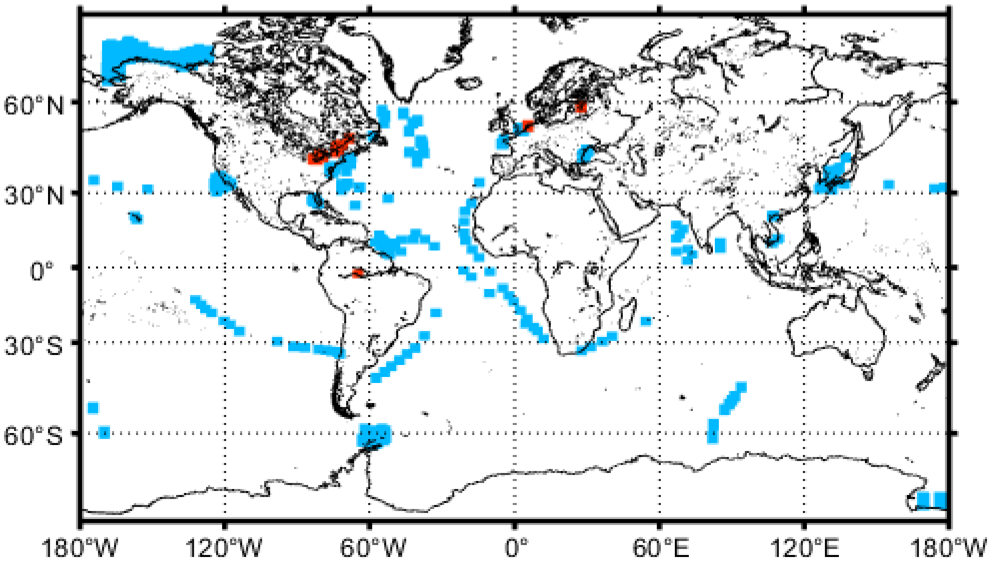


Figure 3: Distribution of the in situ stations with a_{nw} measurements at 443 nm. Coastal and open and coastal waters in blue and lakes in red.

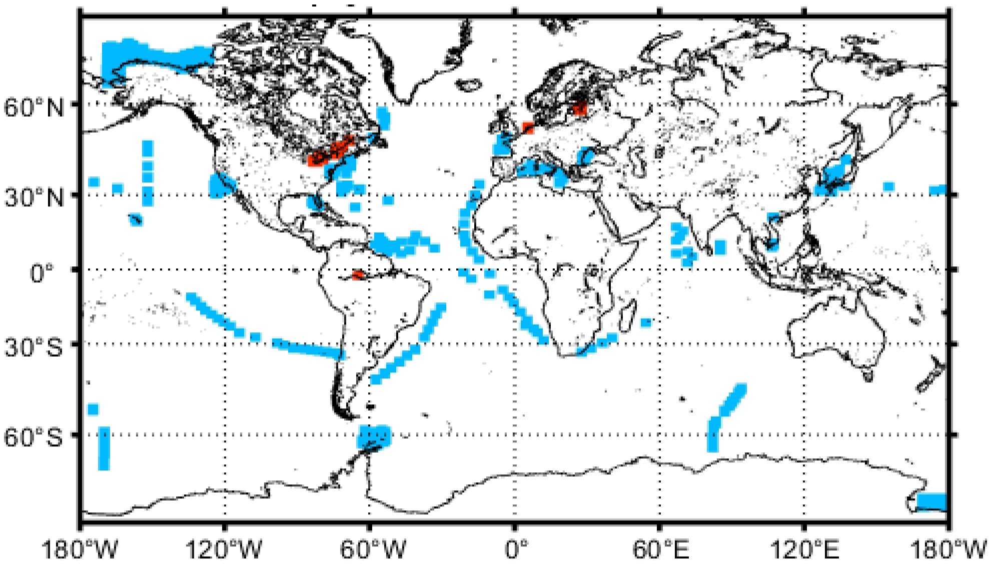


Figure 4: Distribution of the in situ stations with a_{phy} measurements at 443 nm. Coastal and open and coastal waters in blue and lakes in red.

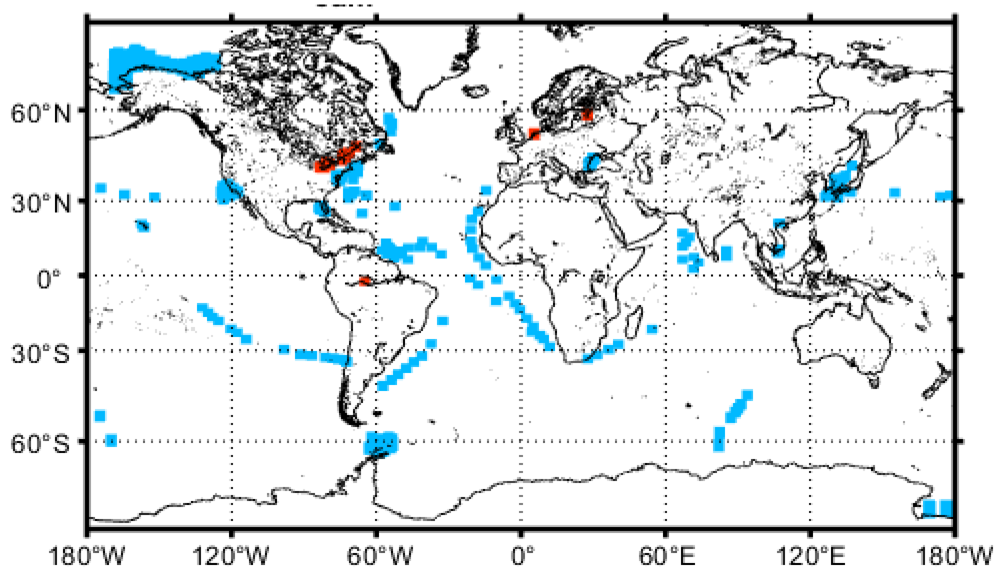


Figure 5: Distribution of the in situ stations with a_{cdm} measurements at 443 nm. Coastal and open and coastal waters in blue and lakes in red.

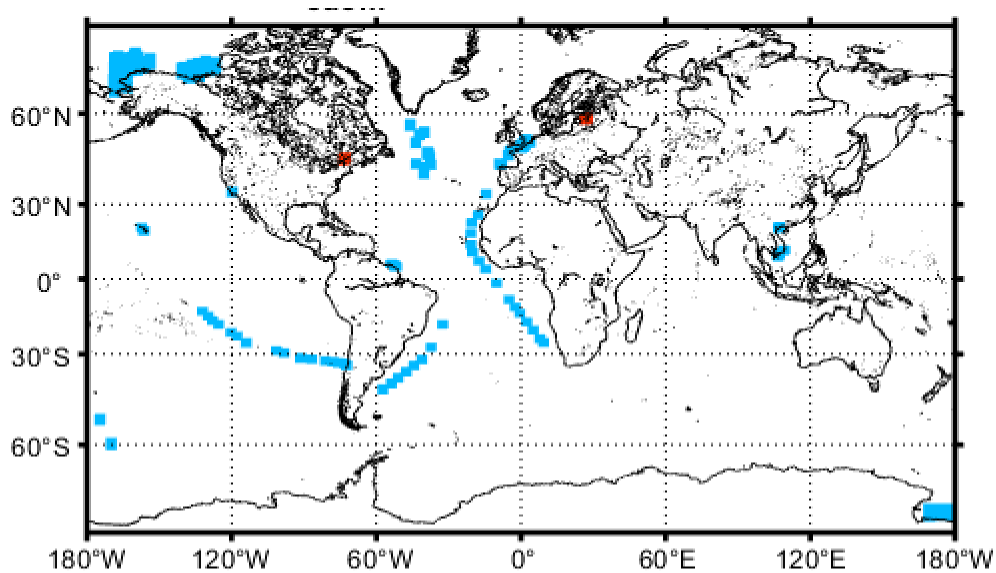


Figure 6: Distribution of the in situ stations with a_{cdom} measurements at 412 nm. Coastal and open and coastal waters in blue and lakes in red.

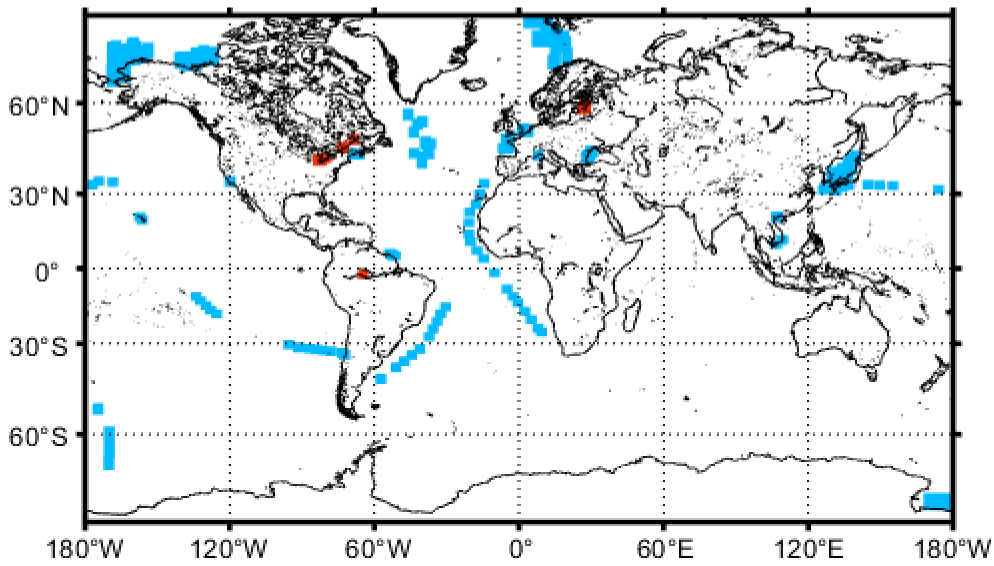


Figure 7: Distribution of the in situ stations with b_{bp} measurements between 443 – 510 nm. Coastal and open and coastal waters in blue and lakes in red.

The **LOG data set** (**Table 1**) is mainly dominated by case 2 waters. The spatial distribution of this data set is described on Loisel et al., (2018) (as DS2 data set) and the IOPs and Rrs protocols are described in Loisel et al., (2007), Lubac et al., (2008) and Neukermans et al., (2012). The a_{nw} values at 443 nm range between 0.095 and 5.27 m^{-1} with a median value of 0.67 m^{-1}. The a_{phy} (443) values range between 0.003 and 10.94 m^{-1} with a median value of 0.25 m^{-1}. The a_{cdm} (443) values range between 0.04 and 1.76 m^{-1} with a median value of 0.51 m^{-1}. The b_{bp} (490) values range between 0.002 and 0.85 m^{-1} with a median value of 0.07 m^{-1}.

The **Scripps data set** (**Table 1**) comprises globally distributed data including polar and lower latitude regions and encompassing contrasting bio-optical environments that range from the clearest waters in ocean subtropical gyres to extremely turbid coastal waters. The spatial distribution of this data set is described on Loisel et al., (2018) (as DS3 data set) and the IOPs and Rrs protocols are described in Stramski et al., (2008) and Zheng et al., (2014). The a_{nw} values at 443 nm ranges between 0.002 and 2.910 m^{-1} with a median value of 0.058 m^{-1}. The a_{phy} (443) values range between 0.001 and 0.275 m^{-1} with a median value of 0.012 m^{-1}. The a_{cdm} (443) range between 0.00045 and 2.720 m^{-1} with a mean value of 0.042 m^{-1}. The b_{bp} (560) range between 0.0004 and 0.05 m^{-1} with a median value of 0.0013 m^{-1}.

The **NAAMES data set** (**Table 1**) was collected in the scope of the NASA EV-S North Atlantic Aerosols and Marine Ecosystems Study (NAAMES), with 2 missions in the subarctic Atlantic. The project collected profiles of b_{bp}, a_{nw} and Rrs and the full description of the protocols and the studies associated with this project is listed in https://naames.larc.nasa.gov/science-publications.html. The a_{nw} values at 443 nm ranges between 0.012 and 0.149 m^{-1} with a median value of 0.034 m^{-1}. The b_{bp} (490) values range between 0.0005 and 0.0053 m^{-1} with a median value of 0.0013 m^{-1}.

The **BOUSSOLE data set** (**Table 1**) was obtained from the "BOUée pour l'acquiSition d'une Série Optique à Long termE" (Boussole) project. The measurements were collected from the buoy deployed in the Ligurian Sea, and all the information pertaining it is described in Antoine et al., (2006). The b_{bp} (490) values range between 0.0005 and 0.0053 m^{-1} with a median value of 0.0013 m^{-1}.

The **PnB data set** (**Table 1**) is composed by samples collected at eight sampling stations in California's (USA) optically complex waters. The a_{nw} values at 440 nm ranges between 0.009 and 2.599 m^{-1} with a mean value of 0.45 m^{-1}. The a_{phy} (443) values range between 0.003 and 1.8 m^{-1} with a mean value of 0.152 m^{-1}. The a_{cdm} (443) values range between 0.004 and 1.8 m^{-1} with a median value of 0.29 m^{-1}. The b_{bp} (560) values range between 0.001 and 0.007 m-1 with a median value of 0.002 m^{-1}.

The **Peacetime data set** (**Table 1**) was collected as part of the Peacetime cruise in the Mediterranean Sea, in which the aim of the project is to study the impact of the processes induced by atmospheric deposition in the air-sea interface. The a_{phy} (443) values range between 0.004 and 0.008 m^{-1} with a median value of 0.008 m^{-1}.

The **Valente data set** (**Table 1**) (Valente et al., 2015 and Valente et al., 2016) gathers data which were acquired from several sources (MOBY, BOUSSOLE, AERONET-OC, SeaBASS, NOMAD, MERMAID, AMT, ICES, HOT, GeP&CO), between 1997 and 2012. Note that the data which are present in the previously described data set have been removed. The a_{nw} at 440 nm ranges between 0.005 and 2.60 m^{-1} with a median value of 0.115 m^{-1}. The a_{phy} (443) range between 0.002 and 1.48 m^{-1} with a median value of 0.043 m-1. The a_{cdm} (443) range between 0.003 and 1.8 m^{-1} with a median value of 0.06 m^{-1}. The b_{bp} (490) range between 0.005 and 2.6 m^{-1} with a median value of 0.115 m^{-1}.

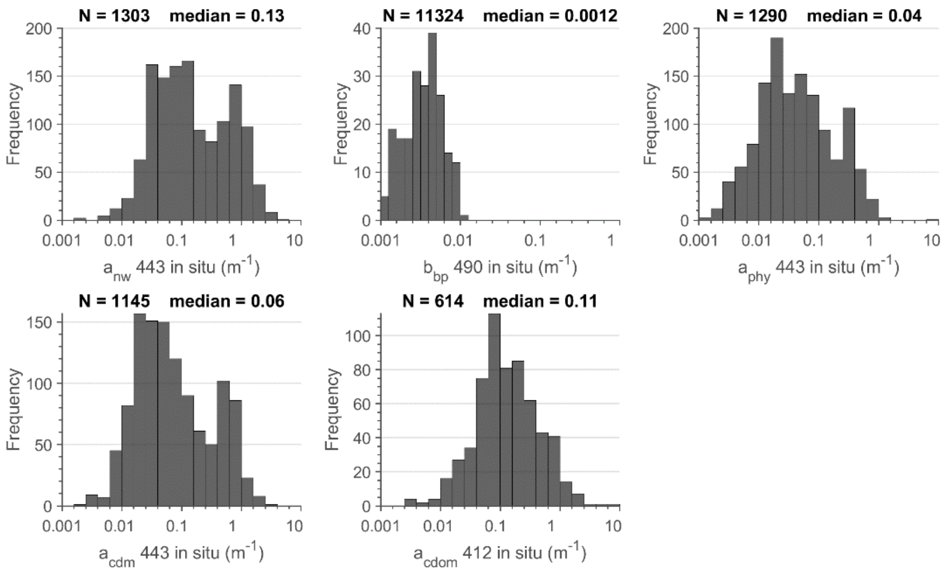


Figure 8: Histogram of the IOPs distribution for the open and coastal waters *in situ* data set.

Table 1: General overview of the IOPs distribution that was used to compose the open and coastal waters data set. a_{nw}, a_{phy}, a_{cdm} and a_{cdom} were set at 443 nm, and b_{bp} was set at 665 nm. The terms min, median, max, std and N were used to address the minimum, median, maximum, standard deviation and number of samples for each subset of the data set.

	LOG	Scripps	NAAMES	BOUSSOLE	P&B	Peacetime	Valente
a_{nw}_min	0.095	0.002	0.012	-	0.023	-	0.005
a_{nw}_median	0.670	0.058	0.034	-	0.103	-	0.115
a_{nw}_max	5.268	2.910	0.149	-	0.552	-	2.598
a_{nw}_std	0.706	0.370	0.037	-	0.129	-	0.456
N	175	151	14	0	115	0	848
b_{bp}_min	0.0022	-	0.0005	0.0003	-	-	0.0005
b_{bp}_median	0.0709	-	0.0013	0.0012	-	-	0.0033
b_{bp}_max	0.8482	-	0.0053	0.0036	-	-	0.0479
b_{bp}_std	0.1743	-	0.0014	0.0005	-	-	0.0065
N	186	0	15	10888	0	0	235
a_{phy}_min	0.003	0.001	-	-	0.014	0.004	0.002
a_{phy}_median	0.249	0.012	-	-	0.050	0.005	0.043
a_{phy}_max	10.942	0.275	-	-	0.459	0.008	1.480
a_{phy}_std	1.766	0.049	-	-	0.108	0.001	0.174
N	49	181	0	0	111	12	937
a_{cdm} _min	0.042	0.000	-	-	0.012	-	0.003
a_{cdm} _median	0.511	0.042	-	-	0.050	-	0.064
a_{cdm} _max	1.762	2.720	-	-	0.152	-	1.802
a_{cdm} _std	0.458	0.387	-	-	0.027	-	0.295
N	34	153	0	0	110	0	848
a_{cdom} _min	0.014	0.000	0.004	-	0.009	-	-
a_{cdom} _median	0.241	0.051	0.029	-	0.071	-	-

a_{cdom} _max	9.235	1.858	0.116	-	0.195	-	-
a_{cdom} _std	0.650	0.330	0.028	-	0.032	-	-
N	327	156	14	0	117	0	0

3.2 Inland water data set

The **St-Laurence data set** was collected in the St. Lawrence during two field missions in 2013 and 2014. The a_{nw} (443) values range between 0.29 and 4.94 m^{-1} with a median value of 1.41 m^{-1}. The a_{phy} (443) values range between 0.017 and 0.84 m^{-1} with a median value of 0.24 m^{-1}. The a_{cdm} (443) values range between 0.21 and 4.53 m^{-1} with a median value of 1.11 m^{-1}. The b_{bp} (665) values range between 0.0044 and 0.1987 m^{-1} with a median value of 0.06 m^{-1}.

The **Great-Lakes data set** (**Table 2**) was collected in the Great Lakes between 2011 and 2017 with optical properties mainly driven by phytoplankton. The a_{nw} values at 443 nm ranges between 0.047 and 2.49 m^{-1} with a median value of 0.37 m^{-1}. The a_{phy} (443) values range between 0.009 and 0.76 m^{-1} with a median value of 0.15 m^{-1}. The a_{cdm} (443) values range between 0.02 and 1.73 m^{-1} with a mean value of 0.3 m^{-1}. The b_{bp} (665) values range between 0.0051 and 0.5347 m^{-1} with a median value of 0.0486 m^{-1}.

The **Estonian-Lakes data set** (**Table 2**) was collected in Estonian lakes in 2013, 2016 and 2017. The full data acquisition procedure is described in Kutser et al., (2016). The a_{nw} at 440 nm ranges between 1.89 and 5.46 m^{-1} with a mean value of 3.63 m^{-1}. The a_{phy} (443) range between 0.13 and 3.59 m^{-1} with a mean value of 1.94 m^{-1}. The a_{cdm} (443) range between 0.93 and 2.02 m^{-1} with a mean value of 1.49 m^{-1}. The b_{bp} (665) range between 0.0286 and 0.093 m^{-1} with a median value of 0.064 m^{-1}.

The **Netherlands-Lakes data set** (**Table 2**) was collected in Ijsselmeer and Markermeer (Netherlands) optically complex lakes, with 10 samplings stations visited during one field mission performed in 2015. The optical properties are driven by both a_{phy} and a_{cdm}. The a_{nw} at 440 nm ranges between 1.89 and 5.46 m^{-1} with a mean value of 3.63 m^{-1}. The a_{phy} (443) range between 0.13 and 3.59 m^{-1} with a mean value of 1.94 m^{-1}. The a_{cdm} (443) range between 0.93 and 2.02 m^{-1} with a mean value of 1.49 m^{-1}. The are no measurements of b_{bp}.

Lastly, **the Mamiraua-Lakes (Brazil)** data set (**Table 2**) was collected in four different lakes inside the Mamirauá Sustainable Development Reserve (MSDR) during the FAPESP project number 2014/23903-9, with a total of 102 sampling stations over the course of two years (2015 and 2016). The optical properties are driven by both b_{bp} and a_{cdm}. The a_{nw} at 440 nm ranges between 3.15 and 8.81

m^{-1} with a median value of 4.72 m^{-1}. The a_{phy} (443) range between 0.11 and 1.33 m^{-1} with a median value of 0.66 m^{-1}. The a_{cdm} (443) range between 2.48 and 8.18 m^{-1} with a median value of 4.03 m^{-1}. The b_{bp} (510) range between 0.0438 and 6.7 m^{-1} with a median value of 0.15 m^{-1}.

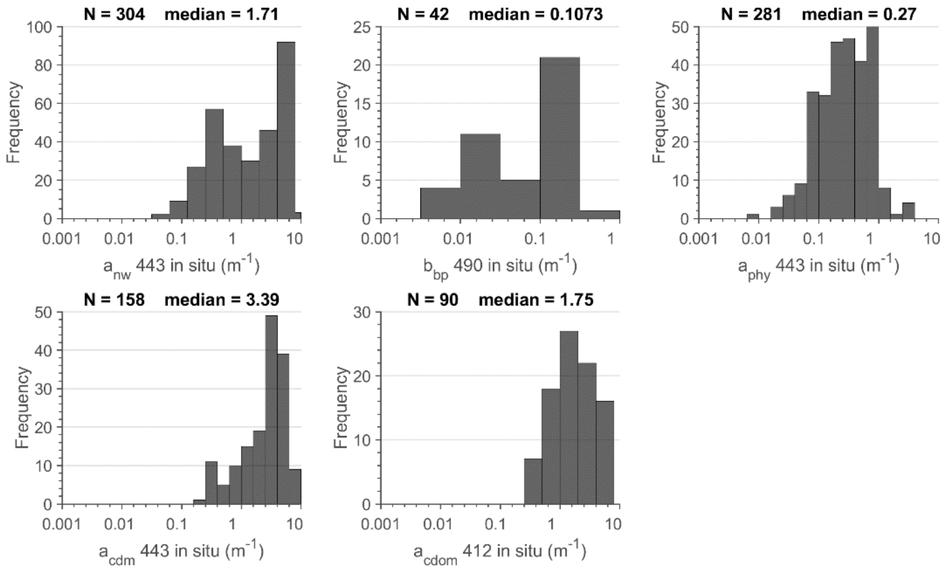


Figure 9: Histogram of the IOPs distribution for the inland waters in situ data set.

Table 2: General overview of the IOPs distribution that was used to compose the inland water data set. a_{nw}, a_{phy}, a_{cdm} and a_{cdom} were set at 443 nm, and b_bp was set at 665 nm. The terms min, median, max, std and N were used to address the minimum, median, max, std and N were used to address the minimum, median, maximum, standard deviation and number of samples for each subset of the data set.

	St-Laurence	Great-Lakes	Estonian-Lakes	Netherlands-Lakes	Mamiraua-Lakes
a_{nw}_min	0.295	0.047	1.904	1.896	3.159
a_{nw}_median	1.416	0.372	3.160	3.978	4.726
a_{nw}_max	4.945	2.498	6.743	5.470	8.817
a_{nw}_std	1.085	0.398	1.388	1.261	1.161
N	56	121	25	9	93
b_{bp}_min	0.004	0.005	0.029	-	-
b_{bp}_median	0.065	0.049	0.064	-	-
b_{bp}_max	0.199	0.535	0.094	-	-
b_{bp}_std	0.052	0.093	0.023	-	-
N	42	130	25	0	0
a_{phy}_min	0.017	0.009	-	0.127	0.111
a_{phy}_median	0.243	0.146	-	1.705	0.664
a_{phy}_max	0.836	0.762	-	3.594	1.334
a_{phy}_std	0.177	0.132	-	1.191	0.250
N	57	121	0	10	93
a_{cdm} _min	0.211	-	-	0.931	2.484

a_{cdm} _median	1.113	-	-	1.569	4.033
a_{cdm} _max	4.531	-	-	2.015	8.189
a_{cdm} _std	0.967	-	-	0.379	1.196
N	56	0	0	9	93
a_{cdom} _min	0.390	-	2.393	1.188	-
a_{cdom} _median	1.121	-	4.443	1.271	-
a_{cdom} _max	4.901	-	7.382	1.965	-
a_{cdom} _std	1.016	-	1.419	0.238	-
N	56	0	25	9	0

4 Description of the match-up data set

The match-up group is composed of three data sets, **GlobColour (GC), MERIS-ESA, and Bio Argo data sets.** GlobColour data set was created combining the in situ data set presented in the last section with the merged product from GlobColour (http://hermes.acri.fr). For this processing, all the concurrently images with in situ data from either MEdium Resolution Imaging Spectrometer (MERIS), Moderate-Resolution Imaging Spectroradiometer (Modis), Sea-viewing Wide Field-of-view Sensor (SeaWiFS) or Visible Infrared Imaging Radiometer Suite (VIIRS) were selected, and the Rrs was extracted. For the MERIS-ESA data set, the Full resolution (FR - 260 x 290 m) and Reduced Resolution (RR - 1040 x 1160 m) were used. RR is usually the product available for the end user, meanwhile FR are provided only under specific conditions. Bio Argo data set, was developed combining the b_{bp} data set collected from the Bio Argo floats over the globe (Claustre et al. 2010) with the OLCI concurrent images. This product was also developed by GlobColour.

4.1 The Globcolour-merged data set

GC IOPs are mainly representative of open ocean waters (**Figure 11**). The a_{nw} values at 443 nm ranges between 0.0047 and 1.4 m^{-1} with a median value of 0.12 m^{-1}. The a_{phy} (443) values range between 0.0018 and 0.475 m^{-1} with a median value of 0.05 m^{-1}. The a_{cdm} (443) values range between 0.0027 and 0.822 m^{-1} with a median value of 0.06 m^{-1}. The b_{bp} (490) values range between 0.0003 and 0.19 m^{-1} with a median value of 0.0019 m^{-1}. The full histogram of the IOPs can be observed in **Figure 10**.

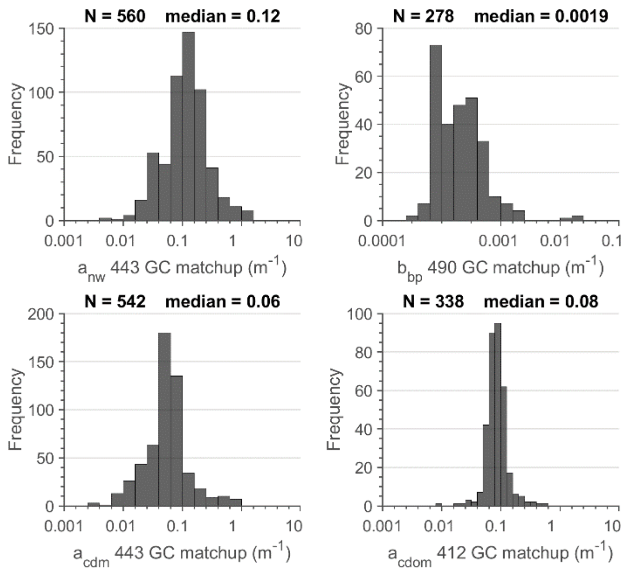


Figure 10: Histogram of the IOPs distribution for the matchup data set developed by Glob Colour.

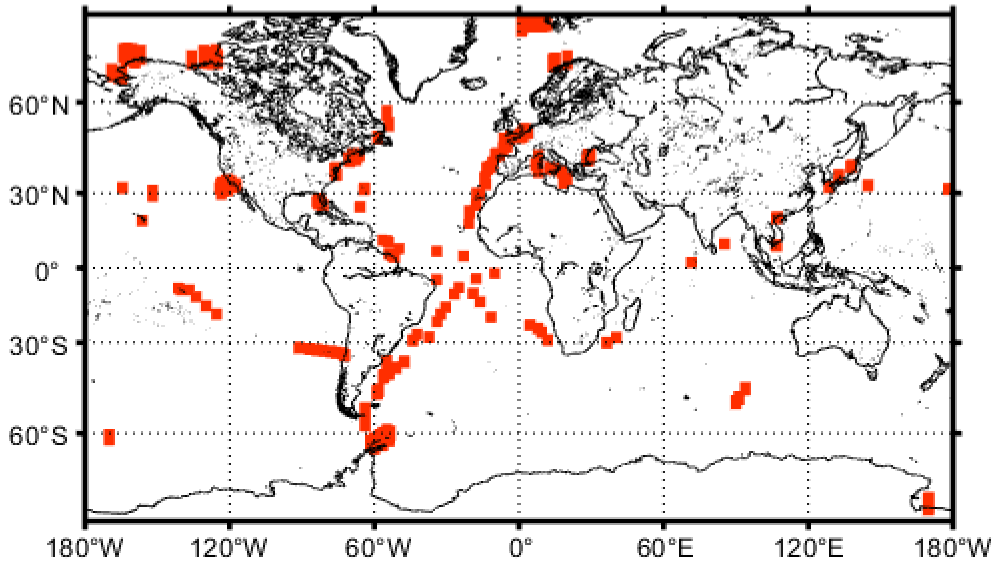


Figure 11: Distribution of the data sets containing concomitant IOP at 443 nm and Rrs from the GLobColour Merge product.

4.2 The MERIS (ESA process) data set (FR and RR)

This data set was subdivided in two groups Full Resolution (FR) and Reduced Resolution (RR). The full histogram of the IOPs for the two subsets can be observed in **Figure 12** and **Figure 13** and the spatial distribution can be observed in **Figure 14**. For RR, the a_{nw} value at 443 nm ranges between 0.0153 and 3.15 m^{-1} with a median value of 0.16 m^{-1}. The a_{phy} (443) value range between 0.0034 and 0.718 m^{-1} with a median value of 0.07 m^{-1}. The a_{cdm} (443) value range between 0.0085 and 1.817 m^{-1} with a median value of 0.06 m^{-1}. The b_{bp} (490) value range between 0.0001 and 0.172 m-1 with a median value of 0.002 m^{-1}.

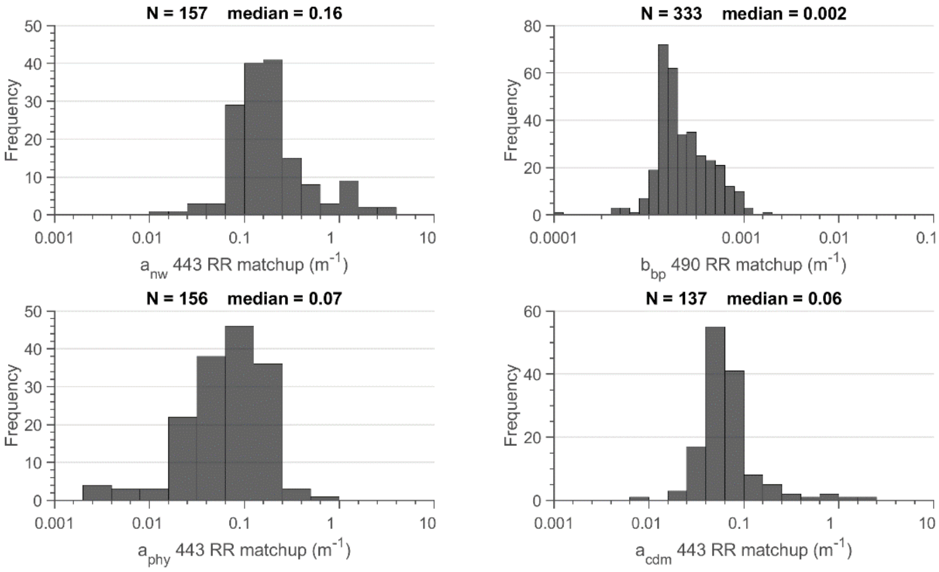


Figure 12: Histogram of the IOPs distribution for the MERIS Reduced resolution data set.

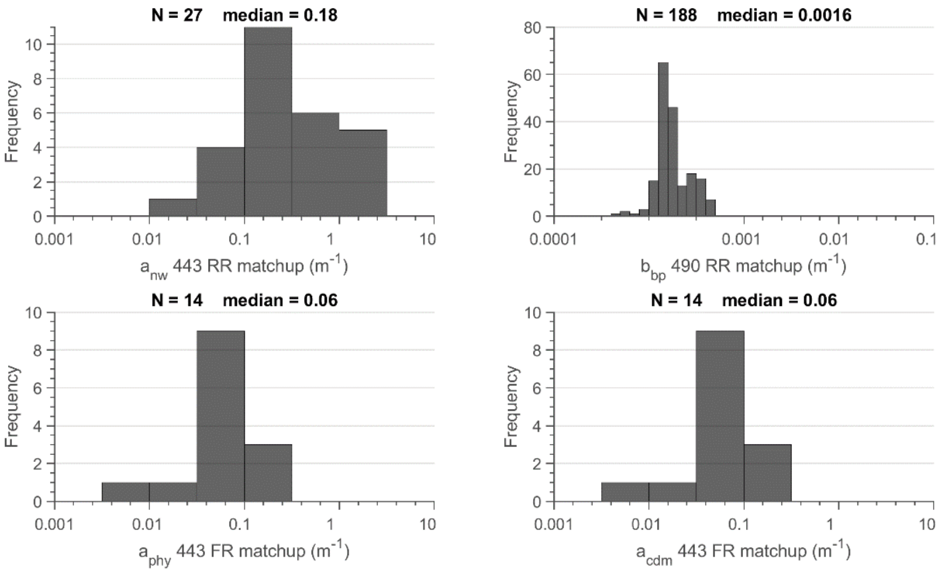


Figure 13: Histogram of the IOPs distribution for the MERIS Full resolution data set.

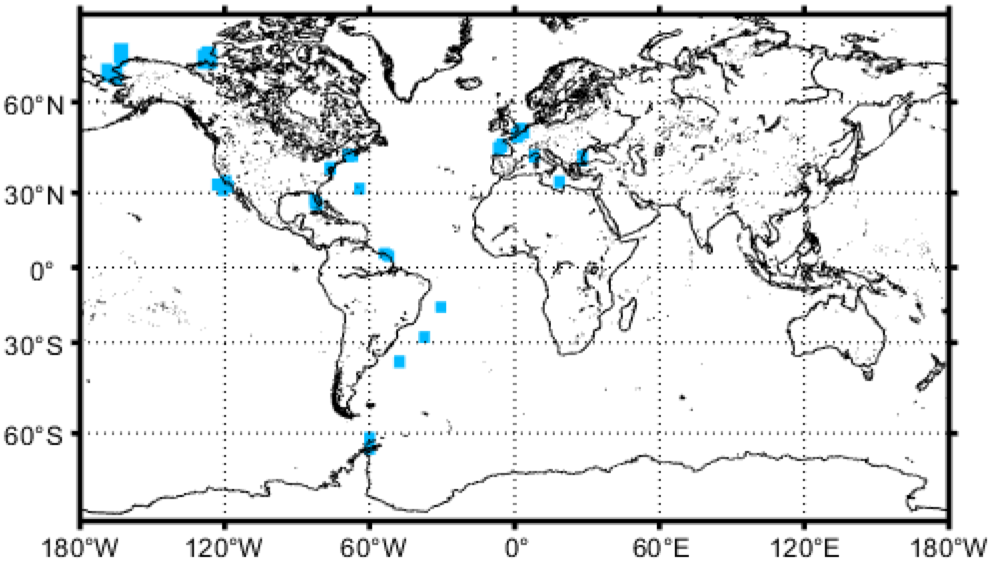


Figure 14: Distribution of the data sets containing concomitant any in situ IOP at 443 nm and Rrs from the MERIS reduced resolution data set.

4.3 The bio-argo data set

The Bio- geochemical Argo data set will be used to considerably increase the number of match-up data set for the validation of b_{bp} (bio-geochemical argo measures b_{bp} at 700 nm). These data have been provided by the International Argo Program and the national programs that contribute to it: (http://www.argo.ucsd.edu, http://argo.jcommops.org). The b_{bp} ranged between 0.0001 and 0.0169 m^{-1}, with a median of 0.0006. The full histogram of the IOPs can be observed in **Figure 15** and the spatial distribution can be observed in **Figure 16**.

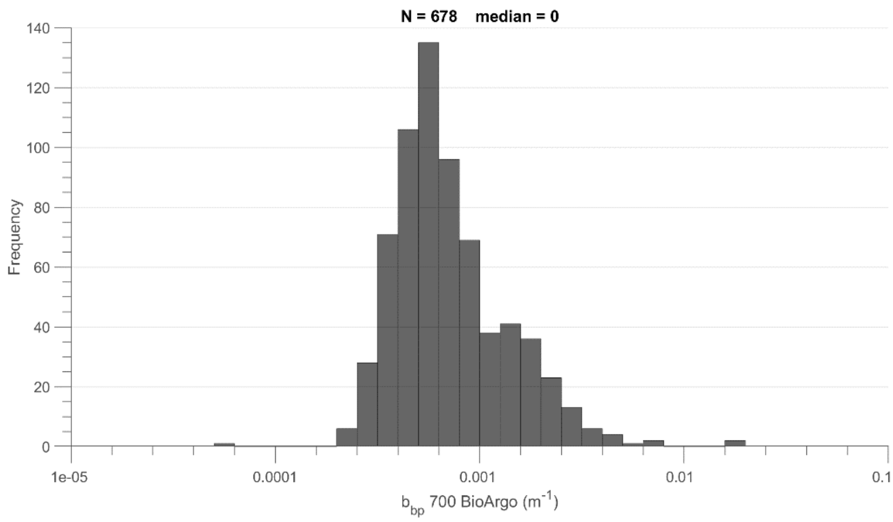


Figure 15: Histogram of the b_{bp} distribution at 700 nm for the Bio Argos data set.

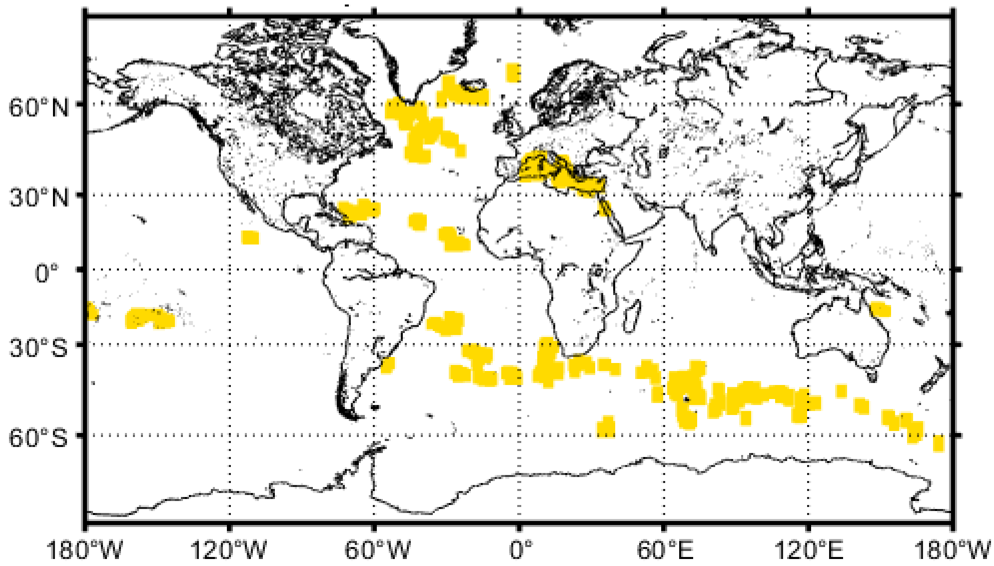


Figure 16: Distribution of the data sets containing b_{bp} at 700 nm from Bio Argos and valid OLCI image.

5 Statistical metric used for the validation

5.1 Statistical Indicators of model performance

To assess model performance, we use scatterplots of model predictions and observations as well as quantitative statistical metrics of differences between the corresponding model predictions and observations. We calculated several statistical indicators that are typically utilized in the assessment of invers models. These indicators include the root-mean-square deviation, $RMSD_{log}$ and RMSD, calculated in the logarithmic and linear space, respectively:

$$RMSD_{log} = \left(\frac{\sum_{i=1}^{N} (\log10(IOP_i^{mod}) - (\log10(IOP_i^{obs}))^2}{N - DF} \right)^{0.5}$$

$$RMSD = \left(\frac{\sum_{i=1}^{N} ((IOP_i^{mod}) - ((IOP_i^{obs}))^2}{N - DF} \right)^{0.5}$$

where N is the number of data points, DF is the degrees of freedom, IOP_i^{mod} is the model-derived value of IOP, and IOP_i^{obs} is the known IOP value that was either measured in situ (e.g. open and coastal waters data set, lakes or matchup).

We also report the mean bias value, MB, representing the difference between the means of the two data sets, i.e., untransformed model-derived data and corresponding untransformed measured data. MB is a component of total RMSD. Other indicators reported for untransformed data sets include the Pearson correlation coefficient, r, the median ratio of model-derived to measured values, MR, which provides a non-dimensional measure of bias including its sign, the median absolute percent difference, MAPD, calculated as the median of the individual absolute percent differences between the modeled and measured data and the MPD, calculated as the median of the individual percent differences between the modeled and measured data, maintaining the sign.

5.2 Classification

A class based analysis was also included for the in situ data sets. For this case, the same statistical methodology described in section 5.1 was applied. The data sets were separated in 17 distinct radiometric classes, based on the Mélin & Vantrepotte (2015) classification (see ATBD) and the statistical parameters were calculated for each class. This additional step was done during the validation process as an additional parameter used to assess the IOPs uncertainties. For any given pixel on the OLCI image, the uncertainty of the IOP retrieval will be calculated from this process.

6 Validation of IOPs

This section will be divided based on the data sets used. The data sets were divided based on the sources (Synthetic, in situ or matchup). The synthetic and in situ data sets were used during the development and calibration of the algorithm, meanwhile the matchup data sets (Remote Sensing images and Floats) were used to assess the performance using the available platforms. From the 2SAA different outputs, first, the outputs from LS2 will be provided (a_{nw} and b_{bp}), followed by the outputs from Zhang et al., 2015 (a_{phy} and a_{cdm}) and Loisel et al., 2014 (a_{cdom}).

6.1 Validation over the synthetic data sets

For both the IOCCG and CCRR data sets the model performance is consistent over the entire dynamic range and spectral range of the particulate backscattering values, with however a better performance over the IOCCG than CCRR data set (**Figure 19**, **Figure 20**, and **Table 3**, **Table 4**). The retrieval accuracy is slightly deteriorated for points belonging to Class 1 (**Figure 21**). $a_{nw}(\lambda)$ (as well as a_{phy} and a_{cdm}) is also well estimated over the whole range of variability, and especially in the blue part of the spectrum (**Figure 17**, **Figure 18**). The model performance for $a_{nw}(\lambda)$ deteriorates significantly for wavelengths of 555 and 670 nm. As already discussed in Loisel et al. (2018), the retrieval of a_{nw} in the green and mainly in the red part of the spectrum is very challenging. This result is associated primarily with relatively poor performance of the model in clear ocean waters where molecular water is the dominant absorbing component and $a_{nw}(\lambda)$ has small contribution to $a(\lambda)$ within the long-wavelength portion of the spectrum. Except for this part of the spectrum the retrieval of a_{nw} at other bands, and especially at 443 nm, is excellent. One may note a slight over-estimation of a_{phy} (443) for complex optical classes 1 and 2 (**Figure 22**, **Figure 25** and **Table 3**, **Table 4**). In contrast to a_{phy} (443), the retrieval of a_{cdm} (443) is excellent for both IOCCG and CCRR data sets (**Figure 23**, **Figure 26** and **Table 3**, **Table 4**). The retrieval of a_{cdom} (412) is performed with a relatively good accuracy (MAPD of 23%) with a slight under (over)-estimation at higher(lower) values. A shift can also be observed for classes >=15, in which a_{cdm} is underestimated and a_{phy} is overestimated.

Table 3: Statistic parameters obtained during the validation exercise performed on the IOCCG data set. RMSD is Root Mean Square Deviation, RMSD_log is Root Mean Square Deviation from two logarithmic distributions, MB is the Mean Bias, MR is the Median Ratio, MAPD is the Mean Absolute Percentage Difference and r is the coefficient of correlation from two logarithmic distributions.

	RMSD	RMSD_log	MB	MR	MAPD	r
a_{nw} (443)	0.235	0.082	0.112	1.074	9.899	0.998
b_{bp} (443)	0.010	0.074	0.004	1.095	12.427	0.994
a_{phy} (443)	0.118	0.206	0.028	1.32	37.049	0.955
a_{cdm} (443)	0.236	0.235	0.105	1.119	21.416	0.974
a_{cdom} (443)	0.155	0.153	-0.045	1.145	23.079	0.98

Table 4: Statistic parameters obtained during the validation exercise performed on the CCRR data set. RMSD is Root Mean Square Deviation, RMSD_log is Root Mean Square Deviation from two logarithmic distributions, MB is the Mean Bias, MR is the Median Ratio, MAPD is the Mean Absolute Percentage Difference and r is the coefficient of correlation from two logarithmic distributions.

	RMSD	RMSD_log	MB	MR	MAPD	r
a_{nw} (443)	0.358	0.099	0.097	1.243	24.663	0.995
b_{bp} (443)	0.063	0.102	0.009	1.244	24.665	0.995
a_{phy} (443)	0.327	0.224	0.028	0.992	29.509	0.775
a_{cdm} (443)	0.504	0.193	0.07	1.399	38.684	0.966

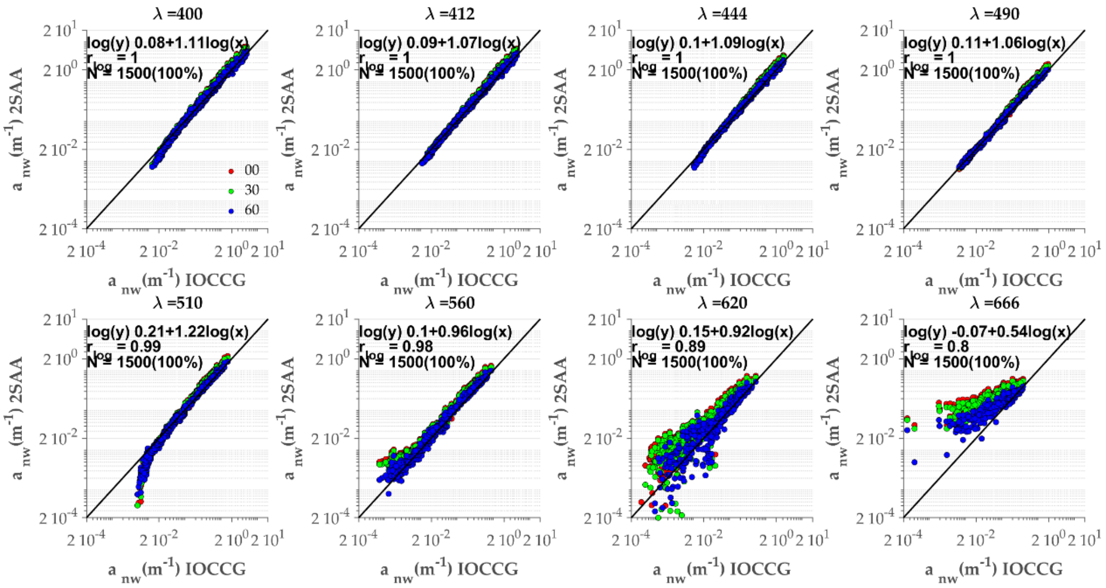


Figure 17: Distribution of the measured a_{nw} vs the estimated a_{nw} using the proposed algorithm for the IOCCG data set. The different colors stand for different sun zenith angles used as input during the simulation.

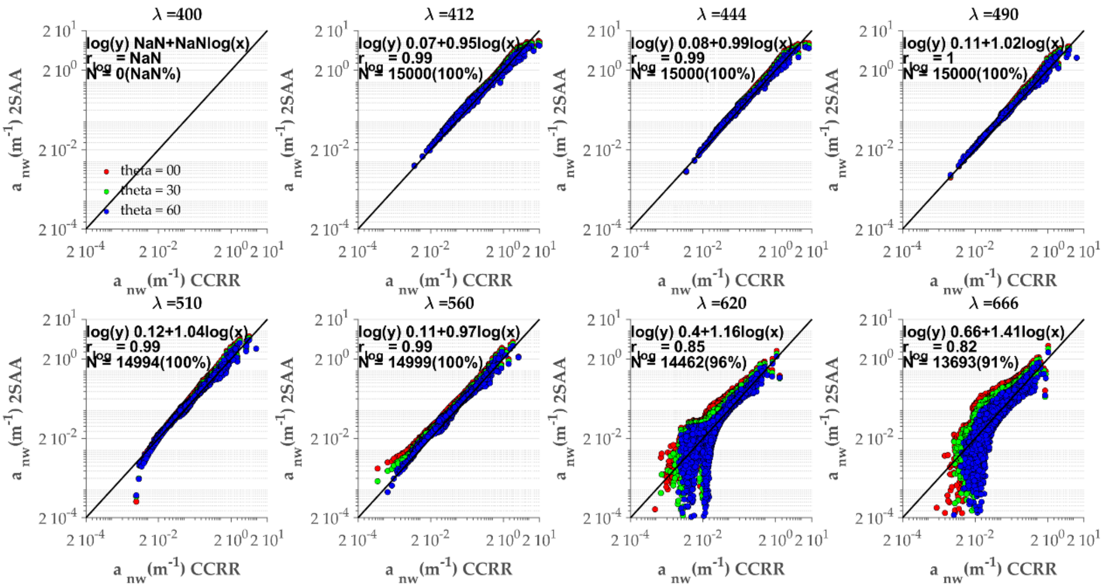


Figure 18: Distribution of the measured a_{nw} vs the estimated a_{nw} using the proposed algorithm for the CCRR data set. The different colors stand for different sun zenith angles used as input during the simulation.

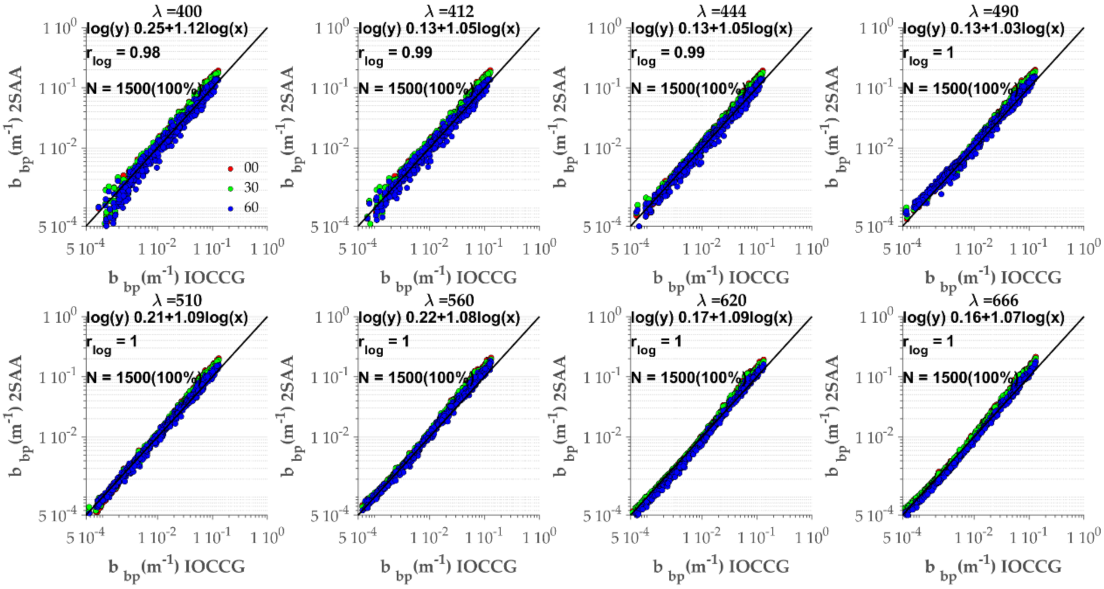


Figure 19: Distribution of the measured b_{bp} vs the estimated b_{bp} using the proposed algorithm for the IOCCG data set. The different colors stand for different sun zenith angles used as input during the simulation.

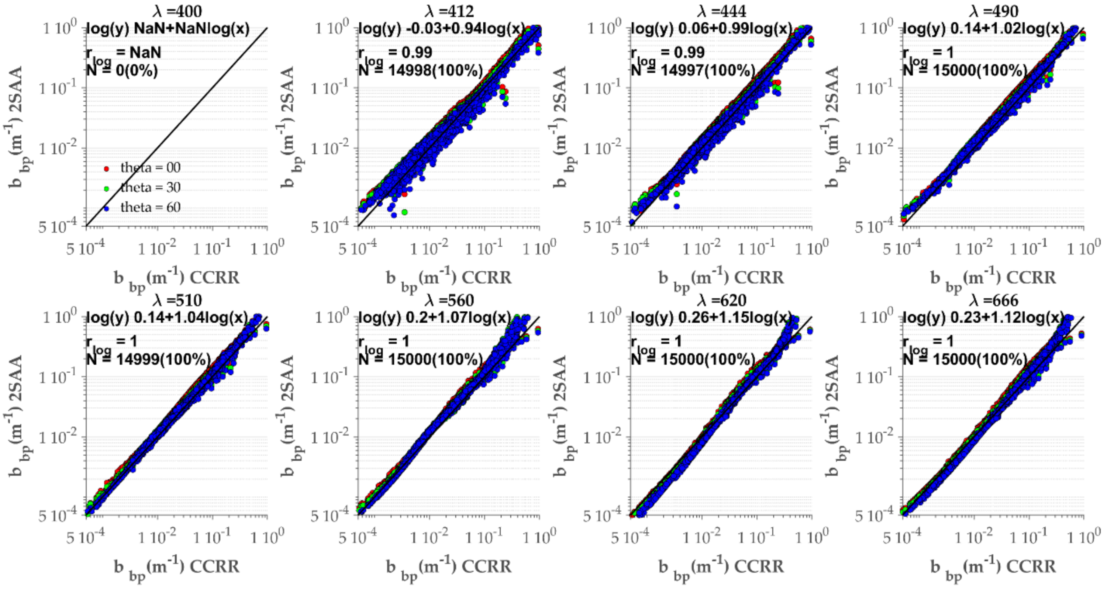


Figure 20: Distribution of the measured b_{bp} vs the estimated b_{bp} using the proposed algorithm for the CCRR data set. The different colors stand for different sun zenith angles used as input during the simulation.

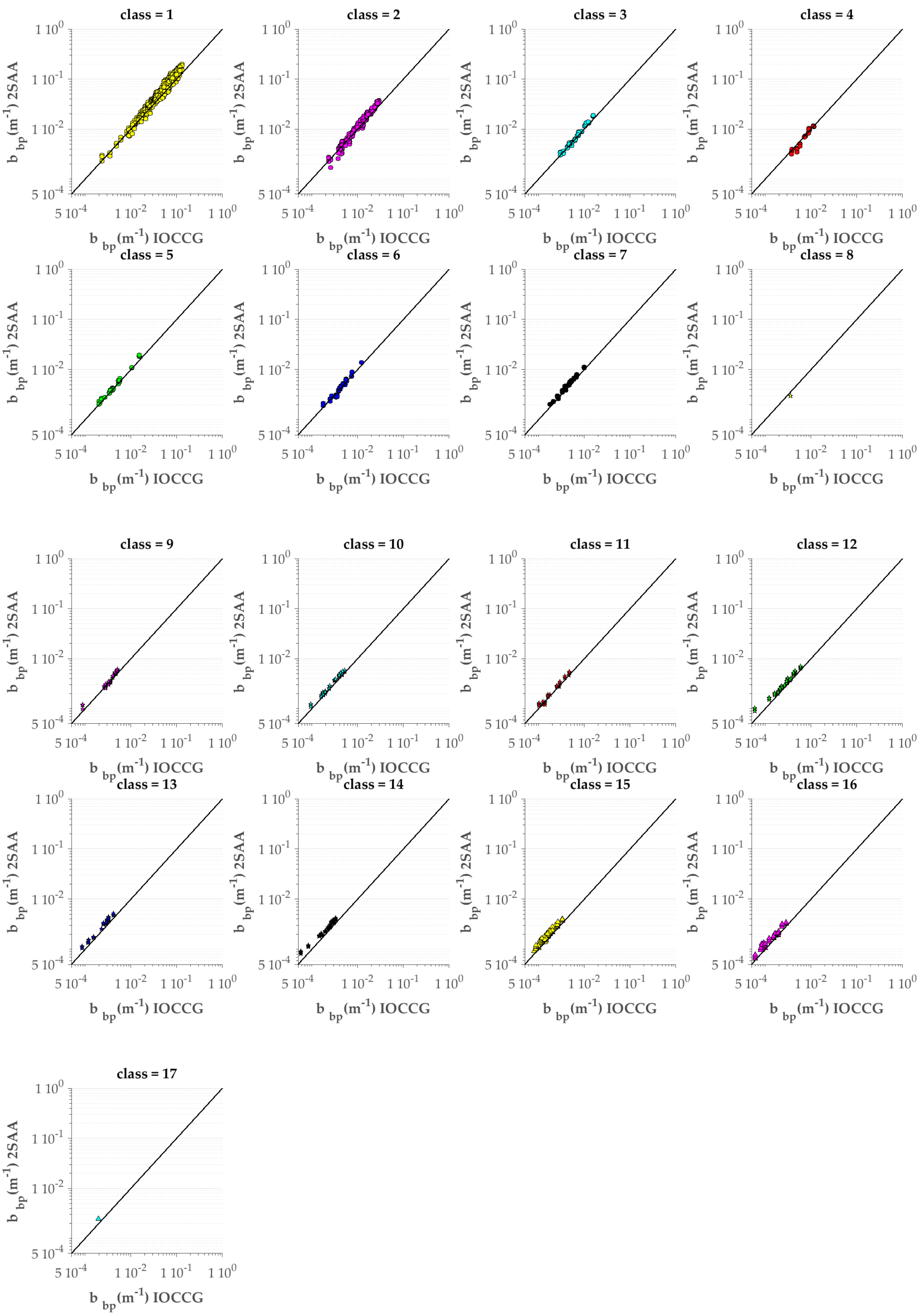


Figure 21: Distribution of the measured b_{bp} vs the estimated b_{bp} using the proposed algorithm for the IOCCG data set. The colors (7 different) and symbols (3 different) are due to the different water classes obtained during the inversion. The spectra from classes 1 to 17 can be observed in the ATBD.

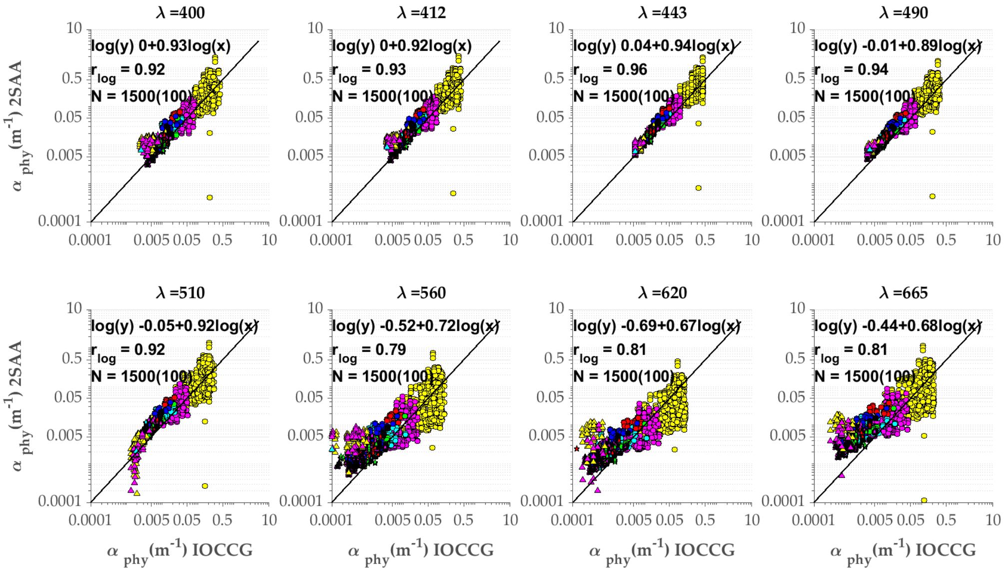


Figure 22: Distribution of the measured a_{phy} vs the estimated a_{phy} using the proposed algorithm for the IOCCG data set. The different colors are due to the different water classes obtained during the inversion.

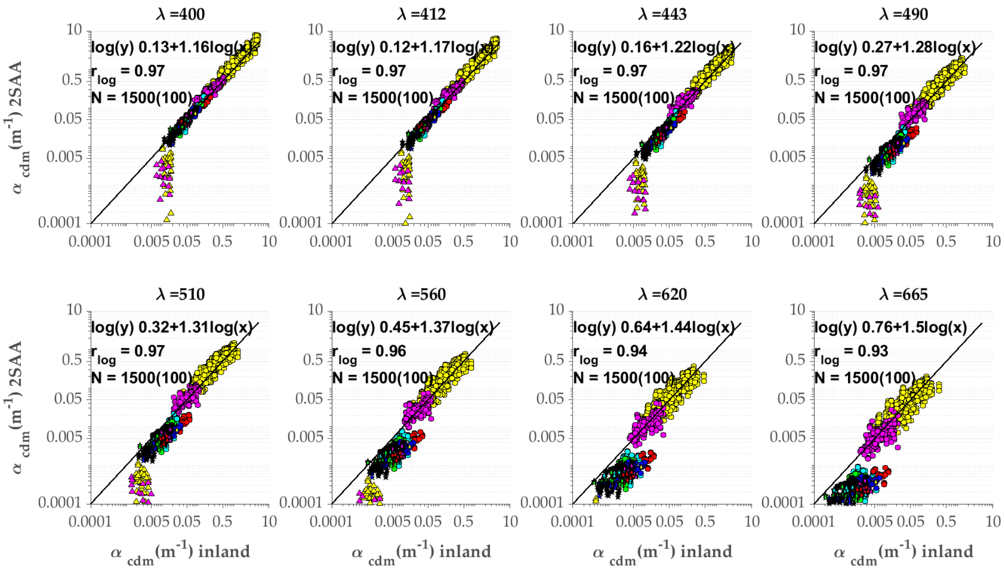


Figure 23: Distribution of the measured a_{cdm} vs the estimated a_{cdm}, using the proposed algorithm for the IOCCG data set. The different colors and symbols are due to the different water classes obtained during the inversion.

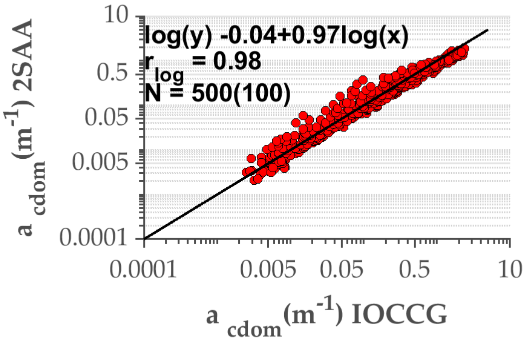


Figure 24: Distribution of the measured a_{cdom} vs the estimated a_{cdom} using the proposed algorithm for the IOCCG data set.

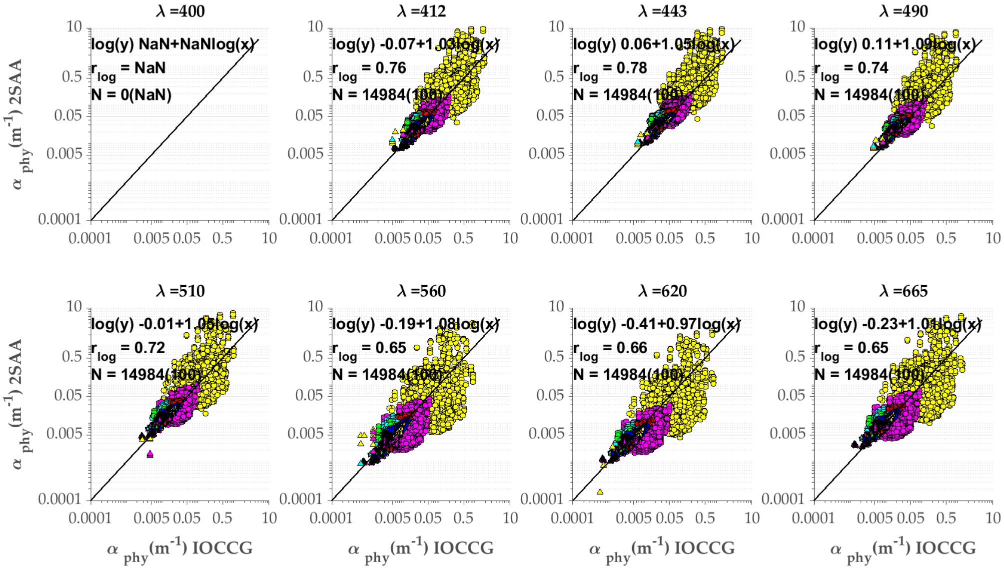


Figure 25: Distribution of the measured a_{phy} vs the estimated a_{phy} using the proposed algorithm for the CCRR data set. The different colors and symbols are due to the different water classes obtained during the inversion.

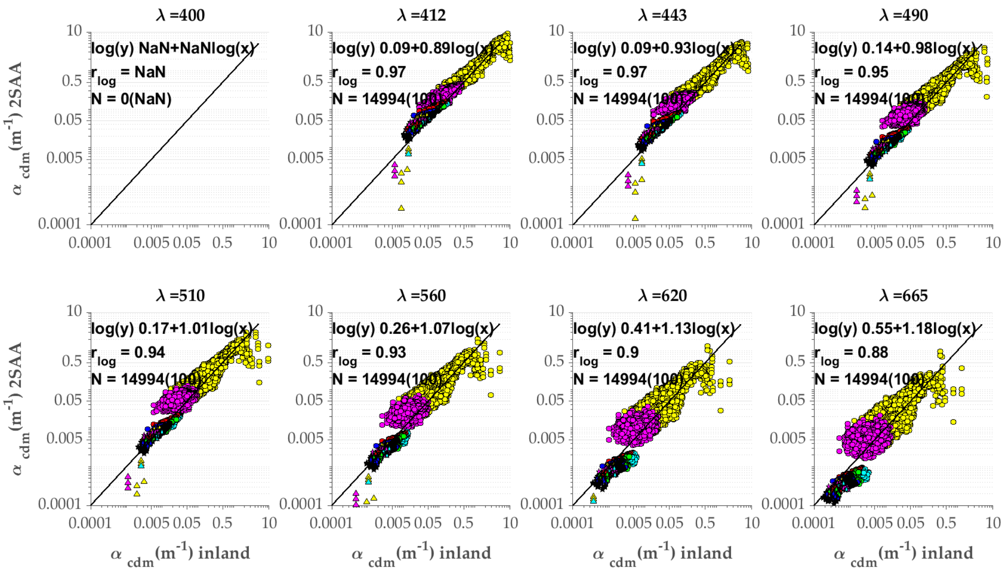


Figure 26: Distribution of the measured a_{cdm} vs the estimated a_{cdm} using the proposed algorithm for the CCRR data set. The different colors and symbols are due to the different water classes obtained during the inversion.

6.2 Validation over the *in situ* data sets

6.2.1 Open and coastal waters

For this section, another parameter will be included, the class based statistics. It is important to note that not all subsets have measurements at all wavelengths, so we can expect a higher difference in N for each wavelength. As can be observed in **Figure 27** and **Figure 28**, the accuracy is high for both a_{nw} and b_{bp}, with a higher dispersion for the wavelengths in the red (r>0.85 at 443). Another aspect is the higher failure rate at 620 and 665 nm, in which the algorithm retrieved negative values for a_{nw}, especially for clear waters, and was subsequently flagged as a failure. This issue, due to the strong contribution of aw into a at these bands, was also identified in Loisel et al., (2018). The statistic parameters obtained for each IOP at the reference wavelength are provided in **Table 5**.

Table 5: Statistic parameters obtained during the validation exercise performed on the open and coastal waters data set. RMSD is Root Mean Square Deviation, RMSD_log is Root Mean Square Deviation from two logarithmic distributions, MB is the Mean Bias, MR is the Median Ratio, MAPD is the Mean Absolute Percentage Difference and r is the coefficient of correlation from two logarithmic distributions.

	RMSD	RMSD_log	MB	MR	MAPD	r	Slope
a_{nw} (443)	0.26	0.20	-0.0612	0.823	22.70	0.96	1.05
b_{bp} (443)	0.001	0.09	0.0002	1.09	13.26	0.85	1.11

a_{phy} (443)	0.44	0.27	-0.024	0.89	34.17	0.89	1.05
a_{cdm} (443)	0.235	0.266	-0.28	0.65	34.53	0.89	0.91
a_{cdom} (443)	0.853	0.491	0.256	1.33	54.56	0.79	1.18

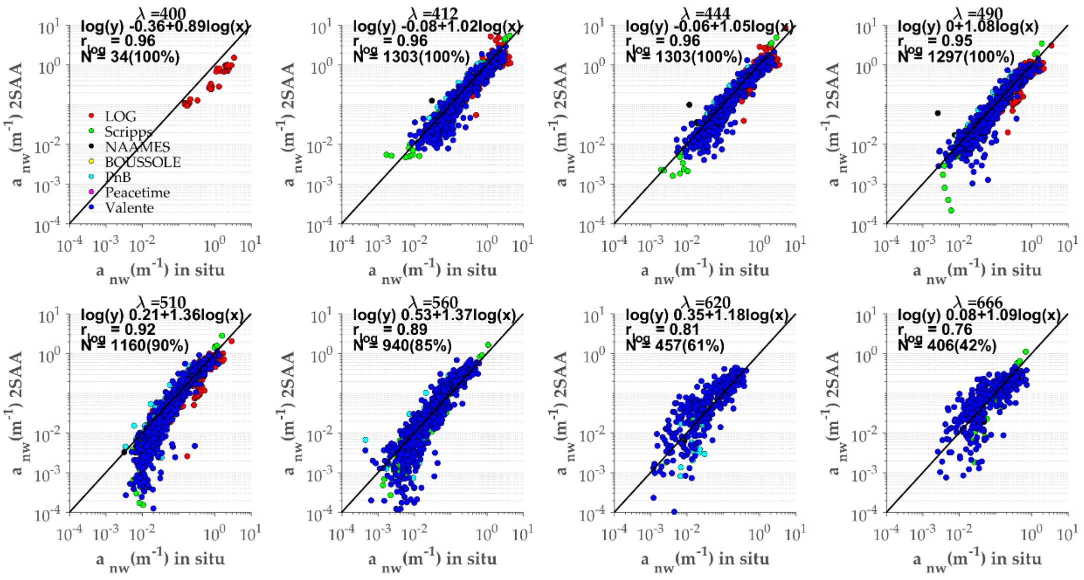


Figure 27: Distribution of the measured a_{nw} vs the estimated a_{nw} using the proposed algorithm for the open and coastal waters data set. The different colors are due to the subsets used to compose the data set.

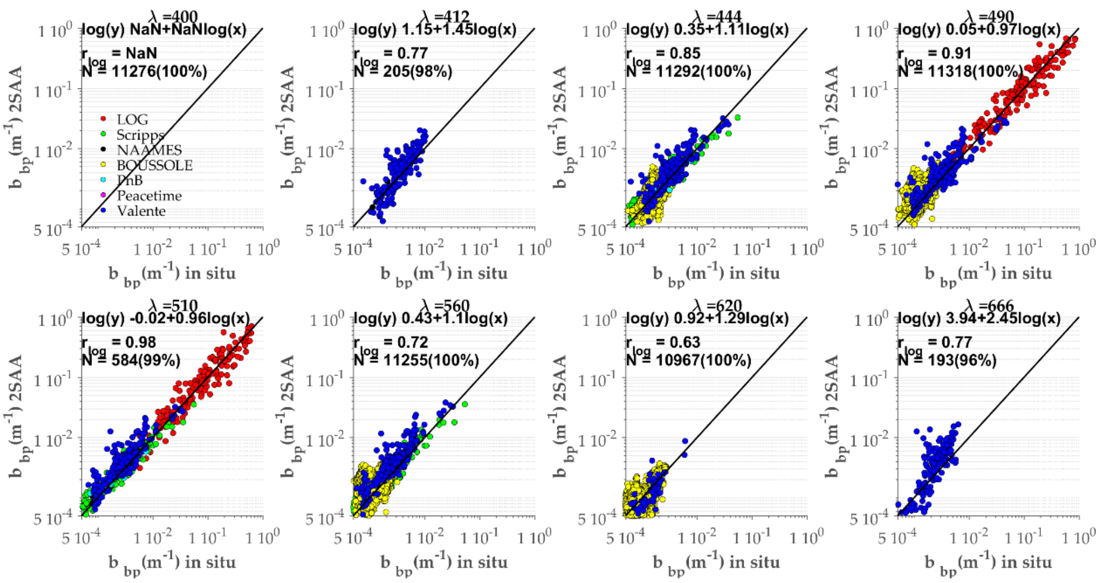


Figure 28: Distribution of the measured b_{bp} vs the estimated b_{bp} using the proposed algorithm for the open and coastal waters data set. The different colors are due to the subsets used to compose the data set.

For the second step of the IOPs retrieval an emphasis will be given to the results at the nominal wavelength (443 nm) For the three IOPs, we observed high correlation between the estimated and measured in situ data (**Figure 29**, **Figure 30** and **Figure 31**; **Table 5**). b_{bp} is the best estimated parameters (MAPD=13%), and a_{cdm} (MAPD=35%), a_{phy} (MAPD=36%), and a_{cdom} (MAPD=38%) are retrieved with nearly the same accuracy.

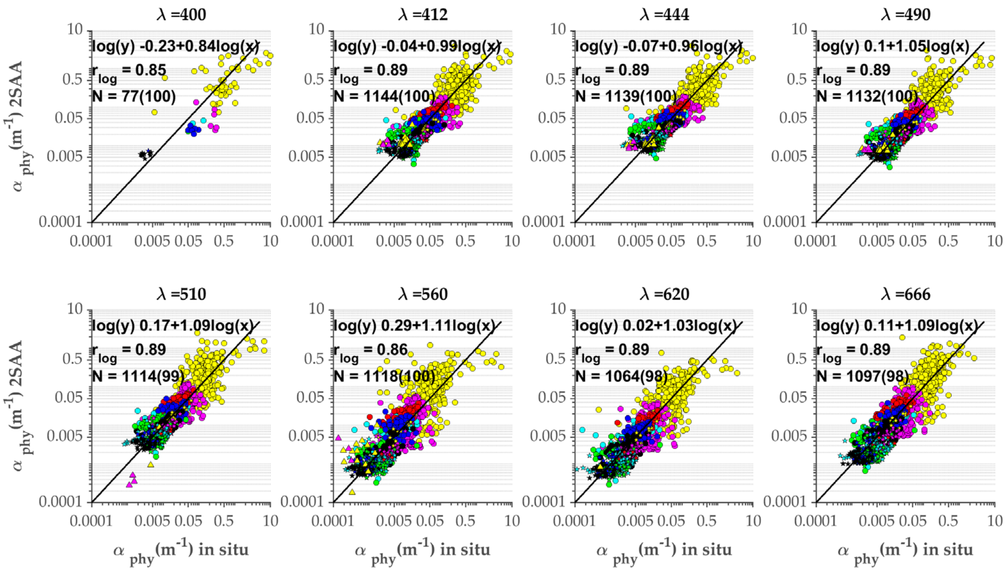


Figure 29: Distribution of the measured a_{phy} vs the estimated a_{phy} using the proposed algorithm for the open and coastal waters data set. The different colors and symbols are due to the different water classes obtained during the inversion.

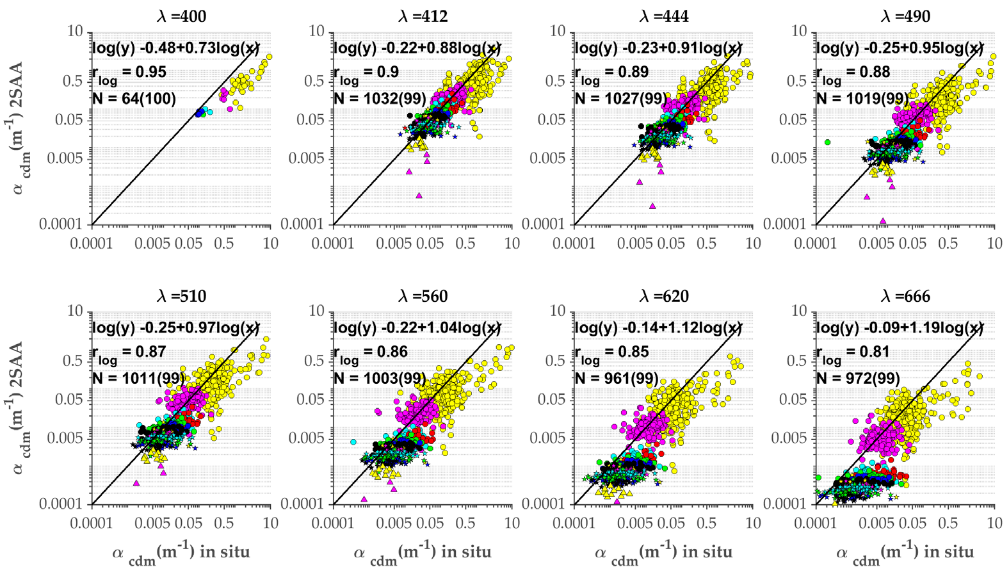


Figure 30: Distribution of the measured a_{cdm} vs the estimated a_{cdm} using the proposed algorithm for the open and coastal waters data set. The different colors and symbols are due to the different water classes obtained during the inversion.

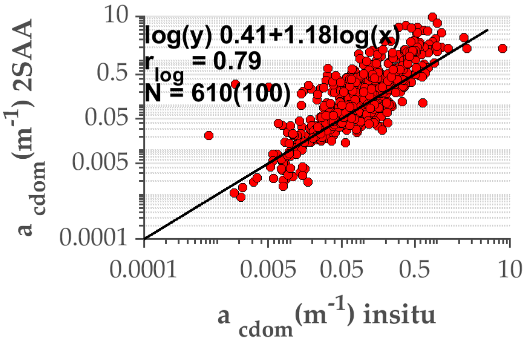


Figure 31: Distribution of the measured a_{cdom} vs the estimated a_{cdom} using the proposed algorithm for the open and coastal waters data set.

To give an overview of the optical properties of the in situ data for each class used during the validation, **Table 6-10**, shows the minimum, maximum, median, standard deviation and number of samples for each class and IOP.

Table 6: Non water absorption coefficient minimum, maximum, median, standard deviation and number of sampling points for each class at 443 nm.

443 nm	a_{nw} min	a_{nw} median	a_{nw} max	a_{nw} std	N
Class 1	0,095	0,832	5,268	0,565	416
Class 2	0,066	0,210	0,711	0,112	150
Class 3	0,012	0,120	0,368	0,042	77
Class 4	0,103	0,142	0,400	0,063	52
Class 5	0,035	0,076	0,178	0,026	71
Class 6	0,075	0,107	0,180	0,023	44
Class 7	0,022	0,059	0,188	0,026	86
Class 8	0,054	0,096	0,148	0,028	12
Class 9	0,045	0,058	0,117	0,017	35
Class 10	0,008	0,039	0,211	0,028	101
Class 11	0,023	0,045	0,074	0,019	6
Class 12	0,012	0,034	0,125	0,019	83
Class 13	0,021	0,032	0,182	0,033	24
Class 14	0,009	0,027	0,086	0,016	44
Class 15	0,014	0,029	0,055	0,009	32
Class 16	0,005	0,016	0,057	0,010	40
Class 17	0,002	0,007	0,092	0,028	10

Table 7: Particle backscattering coefficient minimum, maximum, median, standard deviation and number of sampling points for each class at 443 nm.

443 nm	b_{bp} min	b_{bp} median	b_{bp} max	b_{bp} std	N
Class 1	0,0017	0,0078	0,0546	0,0125	41
Class 2	0,0018	0,0059	0,0148	0,0040	23
Class 3	0,0018	0,0029	0,0078	0,0014	47
Class 4	0,0021	0,0045	0,0147	0,0024	103
Class 5	0,0010	0,0023	0,0047	0,0007	80
Class 6	0,0006	0,0037	0,0047	0,0006	365
Class 7	0,0009	0,0021	0,0040	0,0008	636
Class 8	0,0011	0,0021	0,0036	0,0005	394
Class 9	0,0010	0,0014	0,0034	0,0006	2069
Class 10	0,0006	0,0014	0,0030	0,0005	1120
Class 11	0,0009	0,0012	0,0026	0,0004	281
Class 12	0,0006	0,0015	0,0027	0,0004	3091
Class 13	0,0008	0,0016	0,0025	0,0003	2327
Class 14	0,0006	0,0015	0,0021	0,0003	693
Class 15	0,0007	0,0011	0,0015	0,0003	7
Class 16	0,0006	0,0007	0,0011	0,0002	19
Class 17	0,0006	0,0006	0,0006	0,0000	2

Table 8: Phytoplankton absorption coefficient minimum, maximum, median, standard deviation and number of sampling points for each class at 443 nm.

443 nm	a_{phy} min	a_{phy} median	a_{phy} max	a_{phy} std	N
Class 1	0,012	0,287	10,942	0,730	306
Class 2	0,006	0,093	0,381	0,059	153
Class 3	0,004	0,047	0,128	0,029	75
Class 4	0,019	0,053	0,144	0,033	50
Class 5	0,004	0,027	0,094	0,019	79

Class 6	0,022	0,050	0,108	0,021	55
Class 7	0,004	0,027	0,092	0,014	94
Class 8	0,036	0,044	0,121	0,031	11
Class 9	0,012	0,029	0,068	0,013	42
Class 10	0,002	0,018	0,043	0,008	107
Class 11	0,015	0,032	0,049	0,017	4
Class 12	0,006	0,015	0,042	0,006	97
Class 13	0,005	0,012	0,021	0,004	27
Class 14	0,003	0,009	0,021	0,004	68
Class 15	0,003	0,009	0,022	0,005	43
Class 16	0,002	0,005	0,010	0,002	45
Class 17	0,001	0,002	0,006	0,001	13

Table 9: Colored detritus matter absorption coefficient minimum, maximum, median, standard deviation and number of sampling points for each class at 443 nm.

443 nm	a_{cdm} min	a_{cdm} median	a_{cdm} max	a_{cdm} std	N
Class 1	0,041	0,561	2,720	0,368	299
Class 2	0,031	0,126	0,441	0,077	141
Class 3	0,021	0,066	0,153	0,031	67
Class 4	0,028	0,103	0,257	0,041	47
Class 5	0,012	0,051	0,120	0,020	70
Class 6	0,028	0,050	0,108	0,015	45
Class 7	0,007	0,031	0,096	0,019	83
Class 8	0,019	0,032	0,080	0,017	10
Class 9	0,012	0,030	0,102	0,016	35
Class 10	0,003	0,024	0,179	0,026	93
Class 11	0,018	0,024	0,031	0,006	4
Class 12	0,003	0,018	0,112	0,017	81
Class 13	0,005	0,021	0,172	0,034	24
Class 14	0,003	0,017	0,078	0,016	44
Class 15	0,007	0,018	0,040	0,008	32

Class 16	0,003	0,011	0,052	0,010	40
Class 17	0,000	0,005	0,086	0,026	10

Table 10: Colored dissolved organic matter absorption coefficient minimum, maximum, median, standard deviation and number of sampling points for each class at 412 nm.

412 nm	a_{cdom} min	a_{cdom} median	a_{cdom} max	a_{cdom} std	N
Class 1	0,002	0,188	7,844	0,547	275
Class 2	0,014	0,075	0,334	0,061	70
Class 3	0,019	0,057	0,178	0,035	48
Class 4	0,031	0,058	0,144	0,027	19
Class 5	0,010	0,045	0,080	0,014	42
Class 6	0,014	0,040	0,212	0,046	24
Class 7	0,007	0,024	0,082	0,015	23
Class 8	0,013	0,070	0,171	0,048	12
Class 9	0,009	0,027	0,035	0,007	10
Class 10	0,010	0,022	0,039	0,007	34
Class 11	0,001	0,015	0,027	0,013	3
Class 12	0,012	0,015	0,056	0,012	13
Class 13	0,009	0,011	0,012	0,001	4
Class 14	0,006	0,008	0,010	0,002	5
Class 15	0,006	0,008	0,016	0,004	5
Class 16	0,002	0,009	0,015	0,003	15
Class 17	0,000	0,002	0,008	0,003	9

6.2.2 Inland waters

A higher uncertainty is expected for the inland waters data set, due to their expected higher optical complexity and to the fact that the algorithms (especially the second step, as well as the retrieval of Kd) were initially not developed for these waters (**Table 11**). However, the results for a_{nw} (**Figure 32**) and b_{bp} (**Figure 33**) are in agreement with the ones obtained for the coastal and open ocean in situ data set. A higher dispersion was observed for b_{bp} at 443 and 510 nm for the Mamiraua-Lakes data set, which most likely due to the extremely high attenuation of those lakes, which complicates the in situ measurement of b_{bp} (related to the pathlength correction).

Table 11: Statistic parameters obtained during the validation exercise performed on the inland waters data set. RMSD is Root Mean Square Deviation, RMSD_log is Root Mean Square Deviation from two logarithmic distributions, MB is the Mean Bias, MR is the Median Ratio, MAPD is the Mean Absolute Percentage Difference and r is the coefficient of correlation from two logarithmic distributions.

	RMSD	RMSD_log	MB	MR	MAPD	r	Slope
a_{nw} (443)	1.43	0.29	-0.55	0.835	30.27	0.88	0.9
b_{bp} (620)	0.09	0.24	-0.02	0.813	26.32	0.89	0.92
a_{phy} (443)	1.22	0.49	0.008	0.90	51.89	0.54	1.15
a_{cdm} (443)	0.49	0.26	-0.379	0.66	42.27	0.85	0.88
a_{cdom} (443)	1.03	0.34	-0.66	0.59	41.17	0.78	0.84

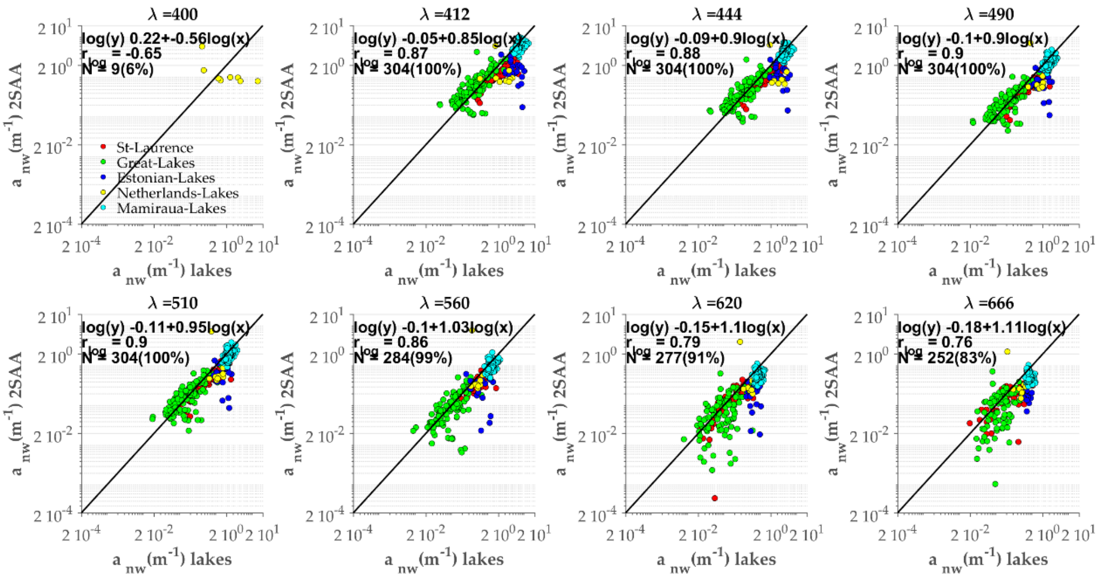


Figure 32: Distribution of the measured a_{nw} vs the estimated a_{nw} using the proposed algorithm for the lakes data set. The different colors are due to the subsets used to compose the data set.

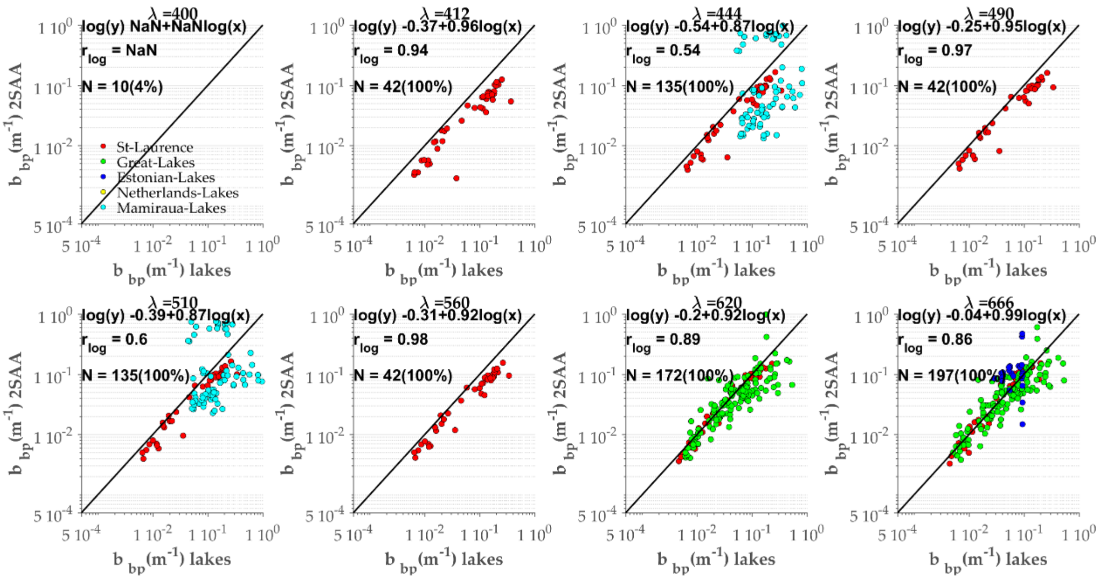


Figure 33: Distribution of the measured b_{bp} vs the estimated b_{bp} using the proposed algorithm for the lakes data set. The different colors are due to the subsets used to compose the data set.

For a_{phy} (**Figure 34**) and a_{cdm} (**Figure 35**) at 443 nm, the correlation coefficient was slightly lower (r>0.74), with an overestimation of a_{phy} for high absorption values and an underestimation of a_{cdm} under the same conditions. The highest MAPD values are observed for a_{phy}. a_{cdom}, is estimated with a MAPD of 41%, and is slightly under-estimated, especially for high a_{cdom} values, a pattern which is not observed over the synthetic data set. (**Figure 36**).

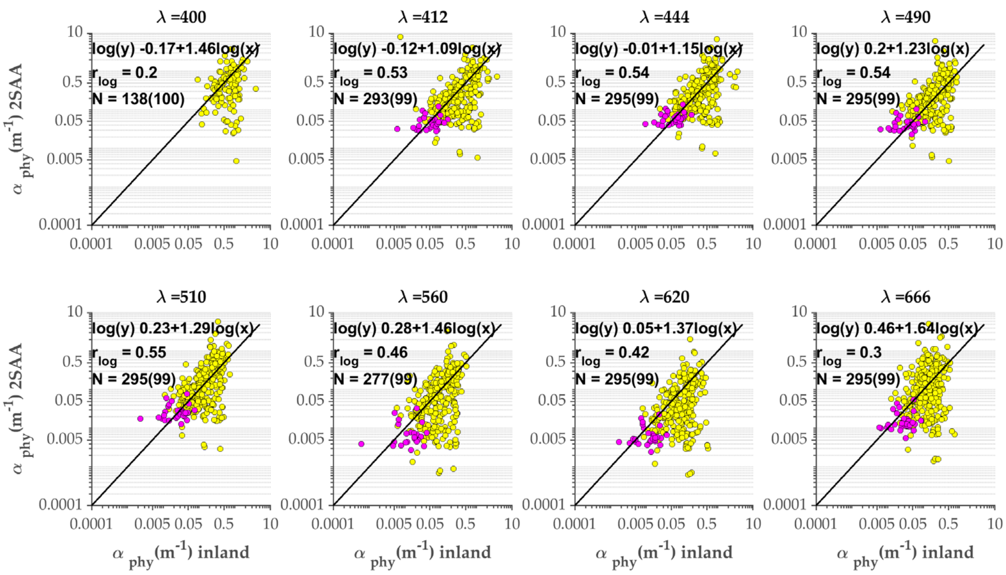


Figure 34: Distribution of the measured a_{phy} vs the estimated a_{phy} using the proposed algorithm for the lakes data set. The different colors are due to the different water classes obtained during the inversion.

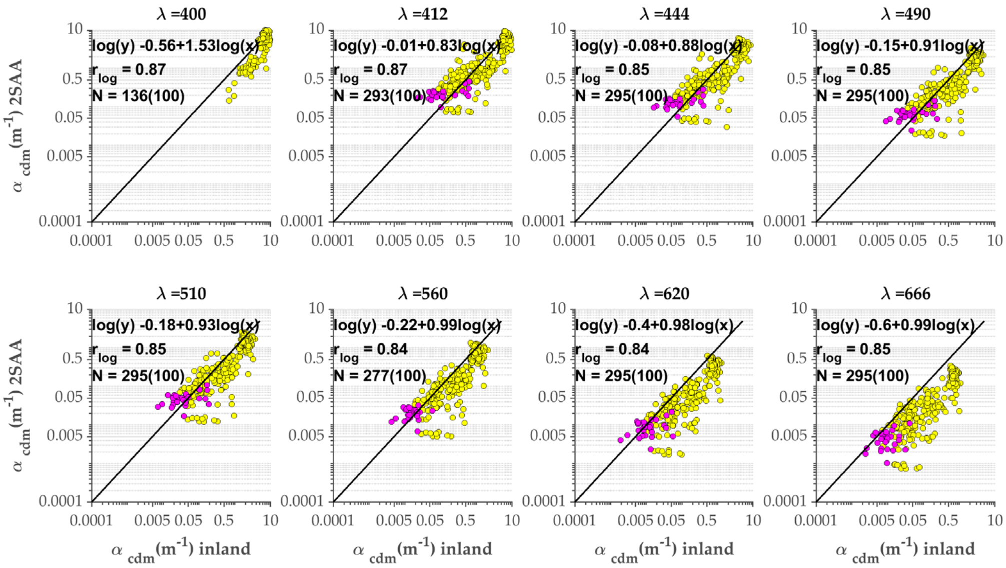


Figure 35: Distribution of the measured a_{cdm} vs the estimated a_{cdm} using the proposed algorithm for the lakes data set. The different colors and symbols are due to the different water classes obtained during the inversion.

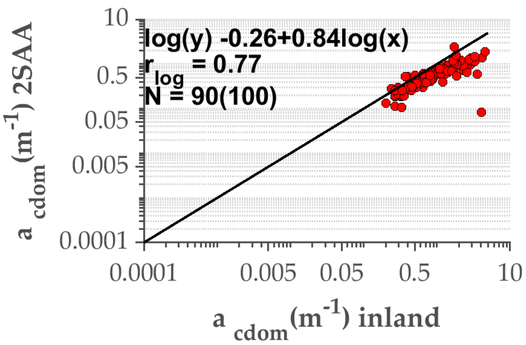


Figure 36: Distribution of the measured a_{cdom} vs the estimated a_{cdom} using the proposed algorithm for the lakes data set.

An overview of the in situ IOPs values for each class used during the validation of the algorithms for Inland waters is provided in **Table 12-16**.

Table 12: Non water absorption coefficient minimum, maximum, median, standard deviation and number of sampling points for each class.

443 nm	a_{nw} min	a_{nw} median	a_{nw} max	a_{nw} std	N
Class 1	0,14	2,09	8,81	2,16	275
Class 2	0,04	0,15	0,49	0,12	29

Table 13: Particle backscattering coefficient minimum, maximum, median, standard deviation and number of sampling points for each class.

443 nm	b_{bp} min	b_{bp} median	b_{bp} max	b_{bp} std	N
Class 1	0,005	0,069	0,535	0,089	144
Class 2	0,005	0,009	0,092	0,017	28

Table 14: Phytoplankton absorption coefficient minimum, maximum, median, standard deviation and number of sampling points for each class.

443 nm	a_{phy} min	a_{phy}	a_{phy} max	a_{phy} std	N
Class 1	0,02	0,32	3,59	0,47	252
Class 2	0,01	0,07	0,19	0,04	29

Table 15: Colored detritus matter absorption coefficient minimum, maximum, median, standard deviation and number of sampling points for each class.

443 nm	a_{cdm} min	a_{cdm}	a_{cdm} max	a_{cdm} std	N
Class 1	0,30	3,46	8,19	1,78	152
Class 2	0,21	0,36	0,40	0,08	6

Table 16: Colored dissolved organic matter absorption coefficient minimum, maximum, median, standard deviation and number of sampling points for each class.

443 nm	a_{cdom} min	a_{cdom} median	a_{cdom} max	a_{cdom} std	N
Class 1	0,24	1,14	4,58	1,02	84
Class 2	0,20	0,29	0,36	0,05	6

6.2.3 Class based analysis performance for each IOP

A combined class based analysis was performed for the open and coastal waters and inland data sets (**Figure 37**). The 17 initial OWC were divided in four groups, class 1 and 2, class 3 to 9, class 10 to 14 and class 15 to 17. In the case of inland waters, a single group was formed since all points belong to either class 1 or 2. As can be observed, there are significant differences in the number of data points for each IOP and class, so, for classes with a lower amount of sampling points, or with a lower dynamic range, the statistical parameters may be biased. In general, the highest variability in RMSD and RMSD_log can be observed for the classes 1 to 2 (optically complex waters), due to the higher magnitude of all the IOPs. For the MR the uncertainty is relatively similar for the classes between 3 and 9, getting higher for the class 1 and 2 and 15 to 17 (around +- 0.5), but that may be due to the higher absolute value of the IOPS for classes 1 and 2 and lower number of sampling points for this class. The MAPD showed a distribution similar to the MR, with an uncertainty under 50% in most cases. Although we can observe a spike in MAPD for a_{cdom} and a_{cdm} of clear waters, that is most likely due to the low amount of sampling points for this class. The determination coefficient showed a broad range of variability, having the lowest precision for the classes 10 to 14.

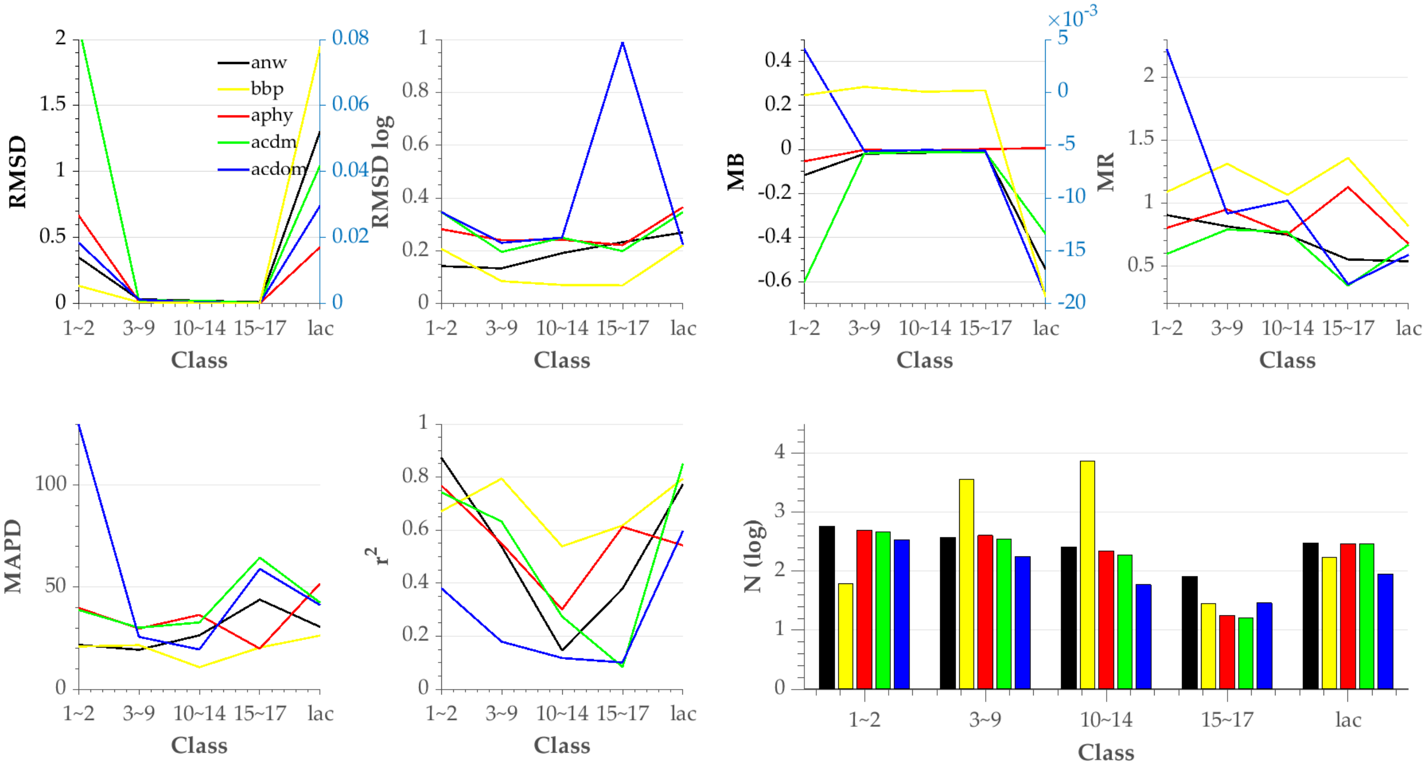


Figure 37: Class based analysis for each IOP (colours), class (-1 to 17) and statistical parameter selected for the lakes data set. a_{nw} is the non-water absorption coefficient (m-1), b_{bp} is the particulate backscattering coefficient, a_{phy} is the phytoplankton absorption coefficient m-1), a_{cdm} is the colored dissolved matter absorption coefficient (m-1), a_{cdom} is the colored dissolved organic matter absorption coefficient (m-1). RMSD is the root mean squared deviation, RMSD log is the root mean squared deviation from the distribution using logarithmic scale, MB is the mean bias, MR is the median ratio, MAPD is the mean absolute percentage difference and N (log) is the number of sampling points for each class (in log10 scale). For the plots with 2 axis, the second axis in blue was used to show the b_{bp} values for that statistical parameter.

6.3 Validation over the match-up data sets

6.3.1 The Globcolour-merged data set

The first matchup data set was derived from MERIS, SeaWiFS, MODIS and VIIRS images, in which a combined product was generated. A good agreement was found for a_{nw} (**Figure 38**), b_{bp} (**Figure 39**) and a_{phy} (**Figure 40**) at the reference wavelength (**Table 17**). The highest uncertainty is found for a_{cdm}, (**Table 17**), with an overall underestimation of a_{cdm} (**Figure 41**). For a_{cdom} a lower correlation coefficient was observed (r=0.312), but that may be due to the low dynamic range of the *in situ* measurements (**Figure 42** and **Table 17**).

Table 17: Statistic parameters obtained during the validation exercise performed on the Globcolour-merged data set. RMSD is Root Mean Square Deviation, RMSD_log is Root Mean Square Deviation from two logarithmic distributions, MB is the Mean Bias, MR is the Median Ratio, MAPD is the Mean Absolute Percentage Difference and r is the coefficient of correlation from two logarithmic distributions.

	RMSD	RMSD_log	MB	MR	MAPD	r	Slope
a_{nw} (443)	0.165	0.209	0.008	0.921	27.85	0.888	1.14
b_{bp} (443)	0.001	0.204	0.000	1.122	26.142	0.736	0.83
a_{phy} (443)	0.068	0.256	-0.012	0.627	38.354	0.788	0.98
a_{cdm} (443)	0.109	0.451	-0.008	1.046	52.778	0.598	1.3
a_{cdom} (443)	0.360	0.342	0.055	1.168	41.510	0.312	1.57

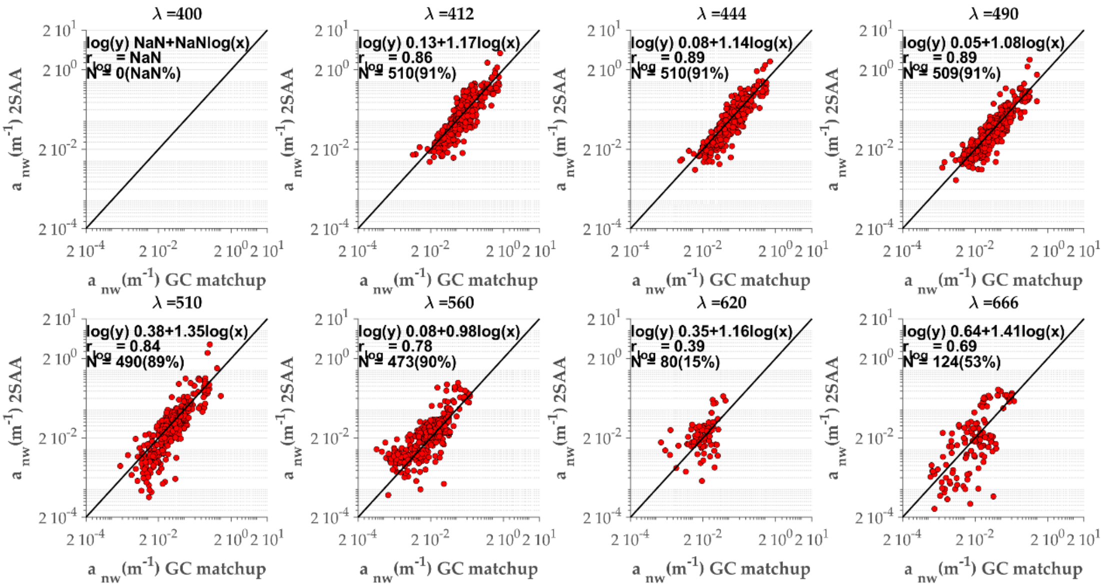


Figure 38: Distribution of the measured a_{nw} vs the estimated a_{nw} using the proposed algorithm for the GlobColour matchup data set.

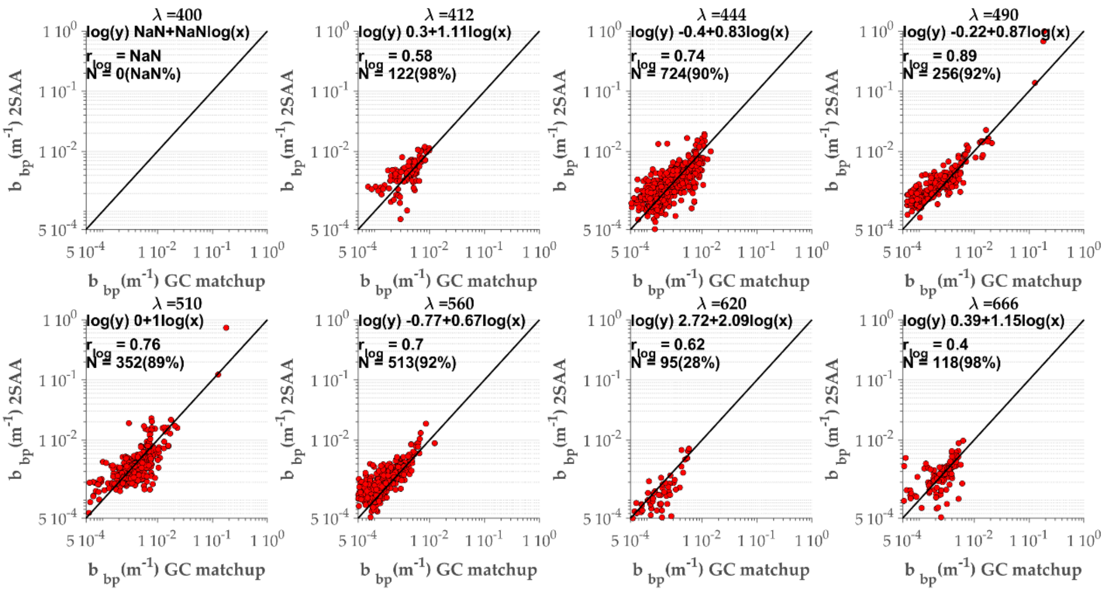


Figure 39: Distribution of the measured b_{bp} vs the estimated b_{bp} using the proposed algorithm for the GlobColour matchup data set.

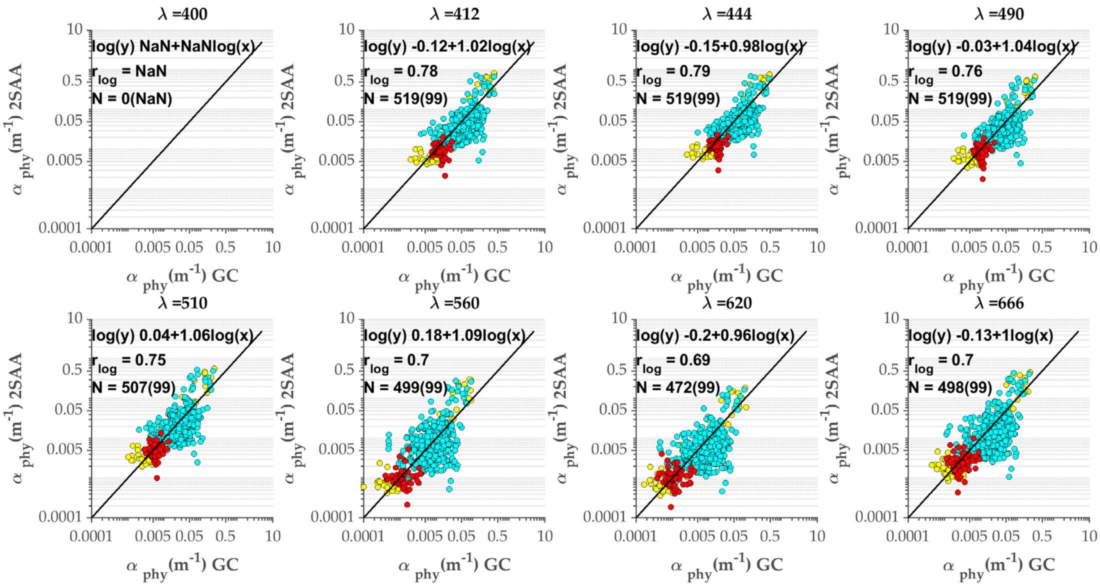


Figure 40: Distribution of the measured a_{phy} vs the estimated a_{phy} using the proposed algorithm for the GlobColour matchup data set.

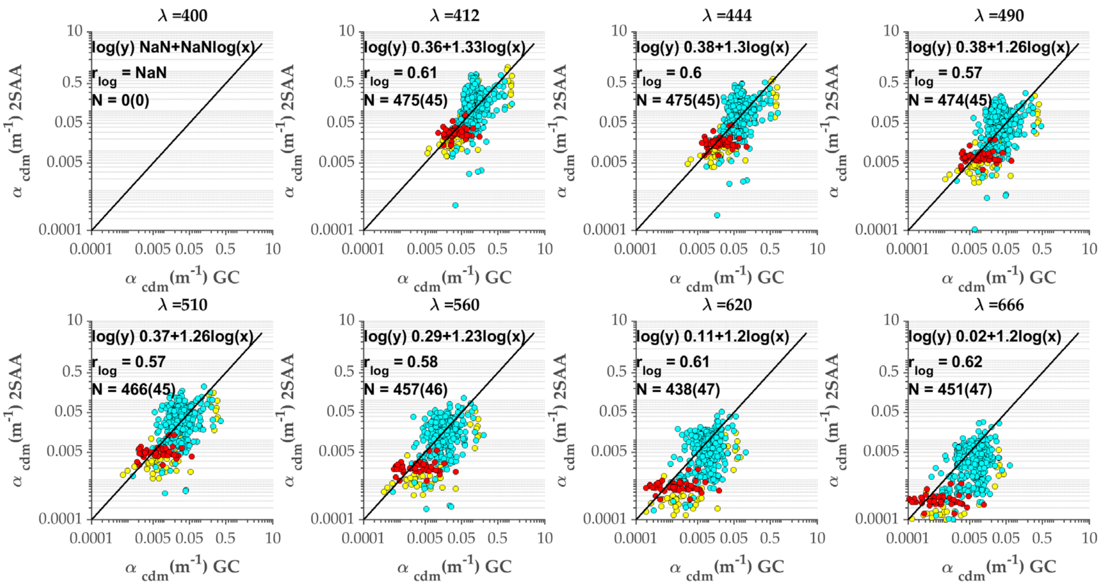


Figure 41: Distribution of the measured a_{cdm} vs the estimated a_{cdm} using the proposed algorithm for the GlobColour matchup data set.

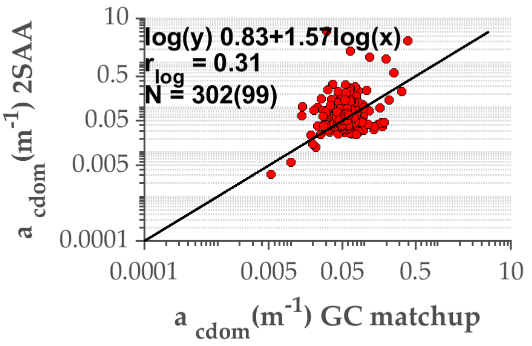


Figure 42: Distribution of the measured a_{cdom} vs the estimated a_{cdom} using the proposed algorithm for the GlobColour matchup data set.

6.3.2 The MERIS (ESA process) data set (FR and RR)

The second matchup data set was build using only the MERIS FR and RR images provided by ESA. The validation results using the FR and RR are provided in **Figure 43-46** and **Figures 47-50, Table 18.**

Table 18: Statistic parameters obtained during the validation exercise performed on the MERIS (ESA process) data set. RMSD is Root Mean Square Deviation, RMSD_log is Root Mean Square Deviation from two logarithmic distributions, MB is the Mean Bias, MR is the Median Ratio, MAPD is the Mean Absolute Percentage Difference and r is the coefficient of correlation from two logarithmic distributions.

	RMSD	RMSD_log	MB	MR	MAPD	r	Slope
a_{nw} (443)	0.104	0.209	-0.091	0.81	36.44	0.84	1.24
b_{bp} (443)	0.0016	0.225	-0.0005	1.04	22.38	0.66	0.99
a_{phy} (443)	0.059	0.359	-0.045	0.74	40.70	0.69	0.74
a_{cdm} (443)	0.146	0.48	0.057	2.18	117.63	0.50	2.76

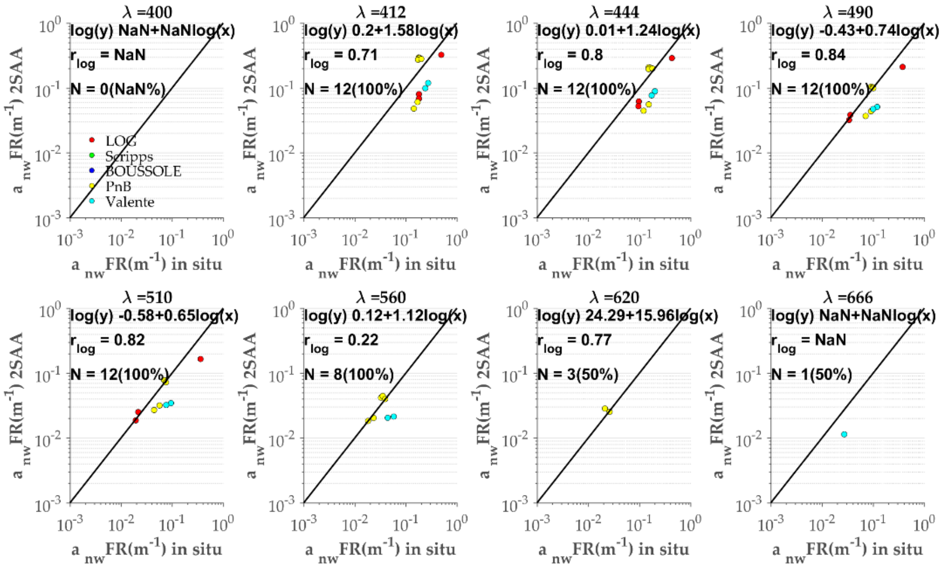


Figure 43: Distribution of the measured a_{nw} vs the estimated a_{nw} using the proposed algorithm for the MERIS full resolution data set. The different colors are due to the subsets used to compose the data set.

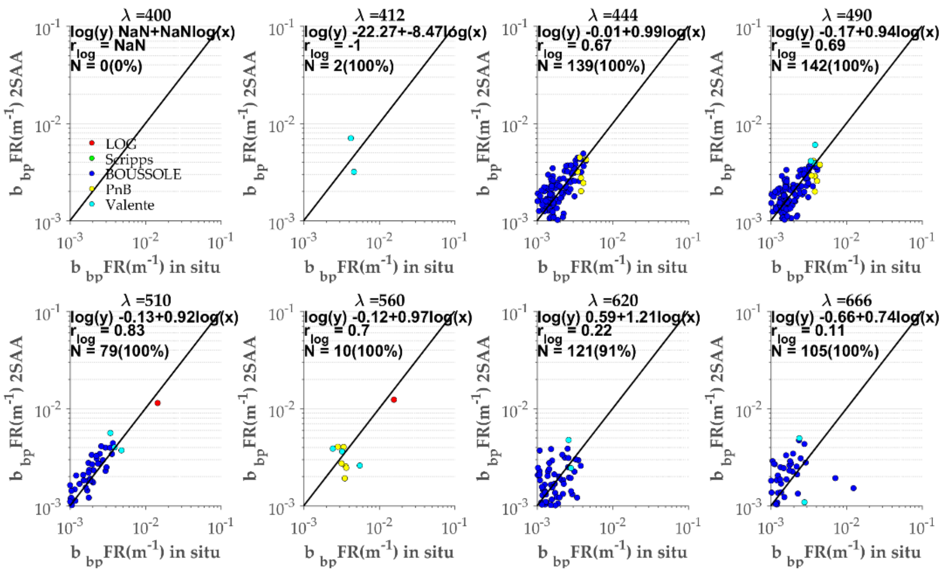


Figure 44: Distribution of the measured b_{bp} vs the estimated b_{bp} using the proposed algorithm for the MERIS full resolution data set. The different colors are due to the subsets used to compose the data set.

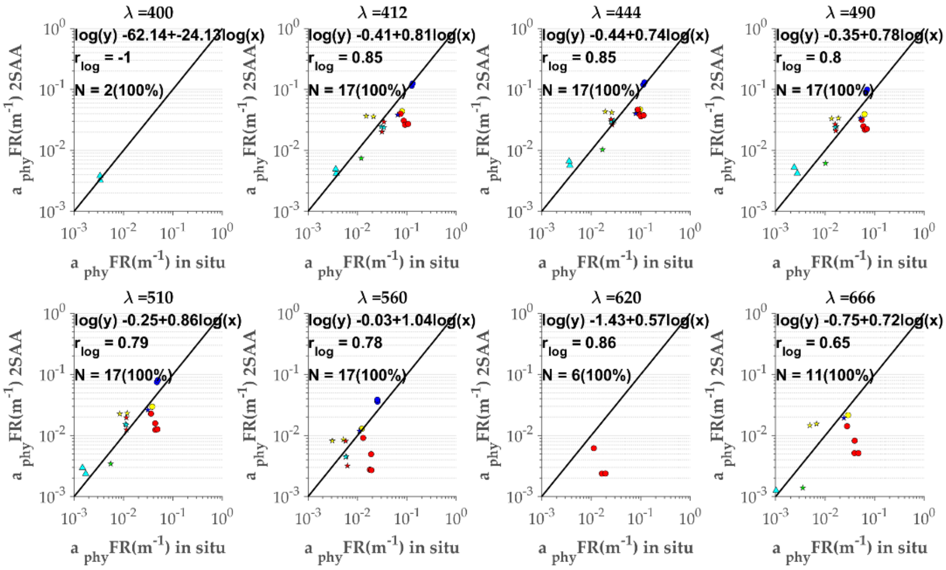


Figure 45: Distribution of the measured a_{phy} vs the estimated a_{phy} using the proposed algorithm for the MERIS full resolution data set. The different colors are due to the subsets used to compose the data set.

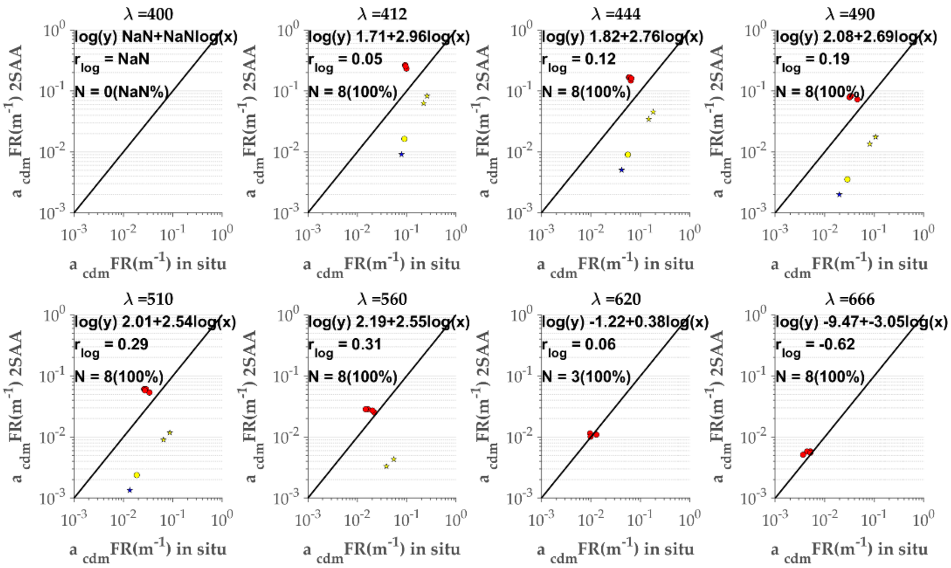


Figure 46: Distribution of the measured a_{cdm} vs the estimated a_{cdm} using the proposed algorithm for the MERIS full resolution data set. The different colors are due to the subsets used to compose the data set.

The performance of the algorithm for the retrieval of IOPs is difficult to evaluate based on this data set, due to the relatively low number of match-up MERIS-FR data points. The number of data point is however sufficient for b_{bp}, showing an excellent agreement between estimated and measured b_{bp} values. For instance, the MAPD, MR, and MB values are 22.38%, 1.0038, and -0.0005, respectively.

In contrast to MERIS-FR, the MERIS-RR match-up data points is composed by a much higher number of data points allowing the performance of the model to be evaluated. The non-water absorption coefficient, a_{nw} at 443 nm, is estimated with an excellent accuracy (**Figure 47**,**Table 18**). An excellent retrieval accuracy is also observed for b_{bp}, except for some data points belonging to P&B (Yellow dots in **Figure 48**). Meanwhile, a bias was identified for a_{cdm} (**Figure 50**), with an overestimation for high absorption and an underestimation for lower absorption. Overall, a_{phy} was underestimated in most cases (**Figure 49**).

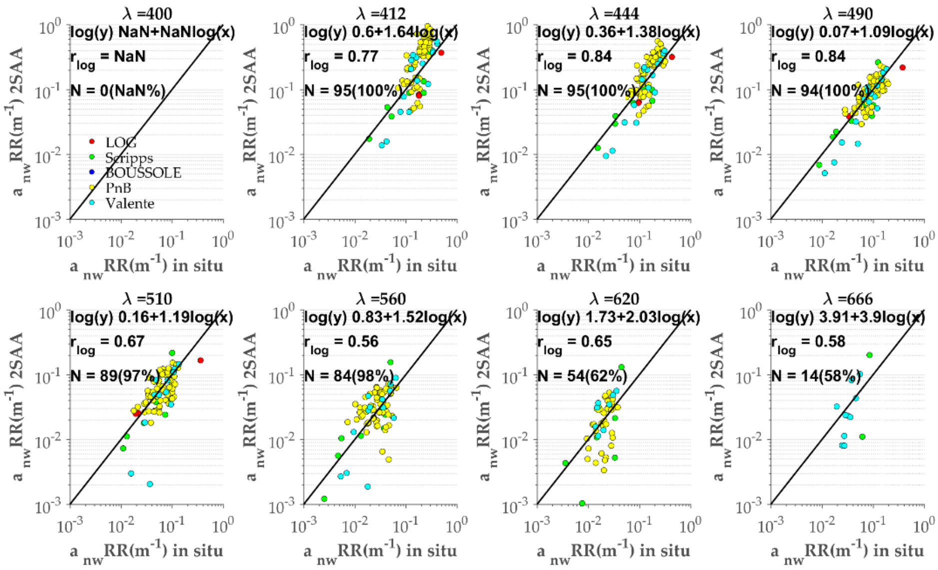


Figure 47: Distribution of the measured a_{nw} vs the estimated a_{nw} using the proposed algorithm for the MERIS reduced resolution data set. The different colors are due to the subsets used to compose the data set.

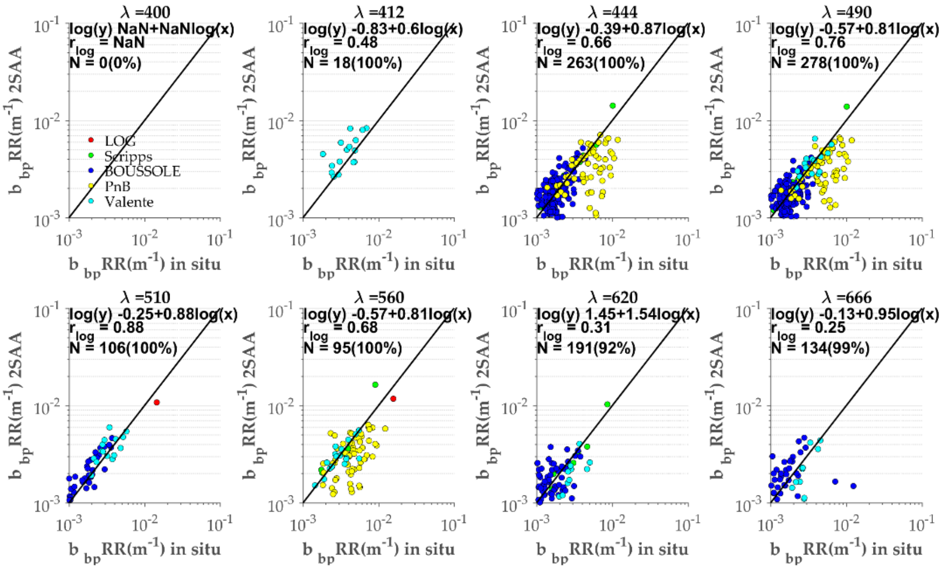


Figure 48: Distribution of the measured b_{bp} vs the estimated b_{bp} using the proposed algorithm for the MERIS reduced resolution data set. The different colors are due to the subsets used to compose the data set.

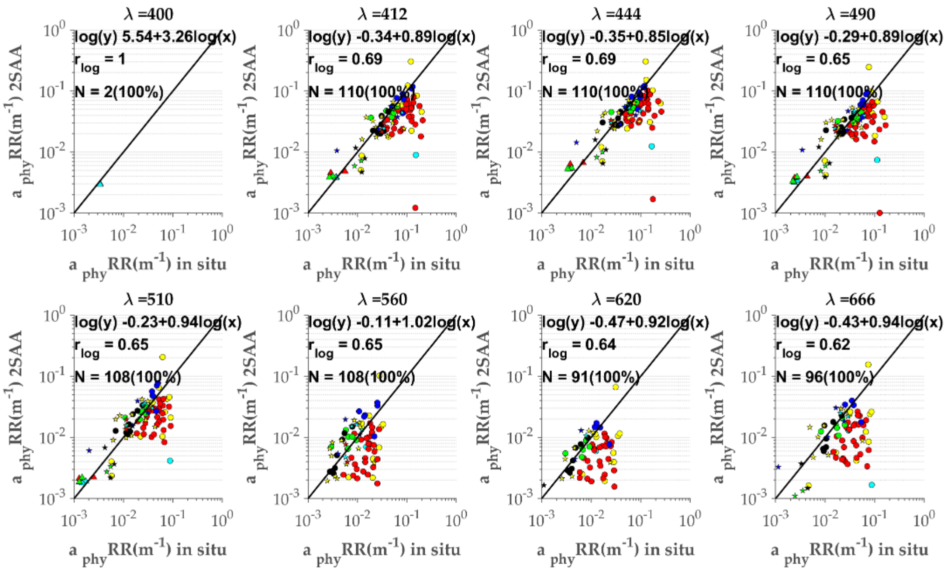


Figure 49: Distribution of the measured a_{phy} vs the estimated a_{phy} using the proposed algorithm for the MERIS reduced resolution data set. The different colors are due to the subsets used to compose the data set.

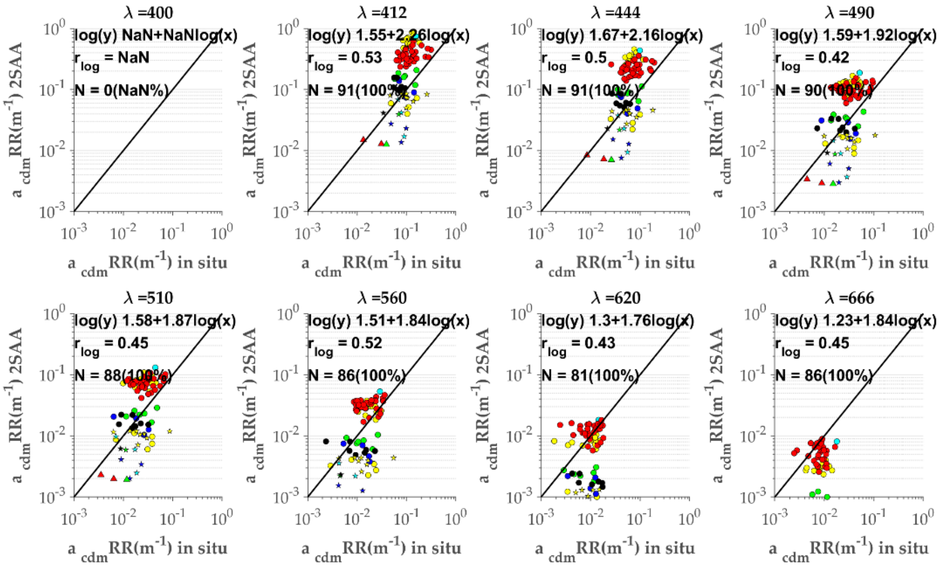


Figure 50: Distribution of the measured a_{cdm} vs the estimated a_{cdm} using the proposed algorithm for the MERIS reduced resolution data set. The different colors are due to the subsets used to compose the data set.

6.3.3 The Bio-geochemical argo data set

The Bioargo data set is composed of b_{bp} in situ measurements performed at 700 nm. Meanwhile the b_{bp} was estimated at 4 OLCI bands to calculate the b_{bp} spectral slope, Y (the 400 and 412 bands are not taken into account to limit the effect of absorption on the b_{bp} inversion, and the red band is not considered because of the large uncertainty in the b_{bp} retrieval at this band in oligotrophic waters). To extrapolate the satellite b_{bp} values estimated at 443 nm to 700 nm, a linear regression was made between $b_{bp}(443)$ estimated at 443 nm using the calculated b_{bp} spectral slope. The exact wavelengths and structure of the variable used during this linear regression can be seen bellow:

$$x = log10\left[\, b_{bp}(443),\ b_{bp}(490),\ b_{bp}(510),\ b_{bp}(560)\right]$$

$$y = log10\left[\left(\frac{443}{443}\right), \left(\frac{490}{443}\right), \left(\frac{510}{443}\right), \left(\frac{560}{443}\right)\right]$$

Then, the b_{bp} at 700 nm was calculated as follows:

$$b_{bp}(700) = b_{bp}\left(\frac{443}{700}\right)^{Y}$$

Where Y is the slope of the linear regression between x and y where x represents the b_{bp} estimated from the satellite R_{rs}.

As can be observed in the figure bellow (**Figure 51**), there is a good agreement between the measured and modelled data, with a more significant scatter around the 1:1 line in oligotrophic waters.

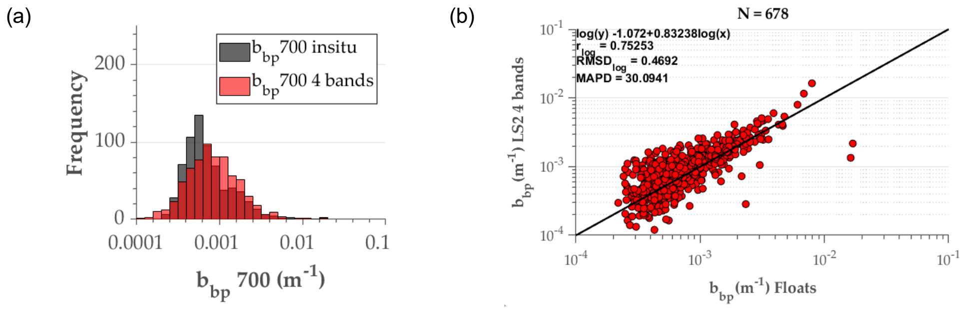


Figure 51: (a) Histograms of the distribution of the b_{bp} (700 nm) measured by the Bio Argos floats and estimated by the 2SAA inversion. (b) Distribution of the measured b_{bp} vs the estimated b_{bp} using the proposed algorithm for the Bio-geochemical Argos.

Another aspect tested for this data set is the magnitude of Y in comparison to the standard Y product (Lee et al., 2002) used for most ocean colour algorithms (**Figure 52**). As can be seen, although most of the values are saturated at 2 for the standard algorithm, this behavior was not observed for 2SAA.

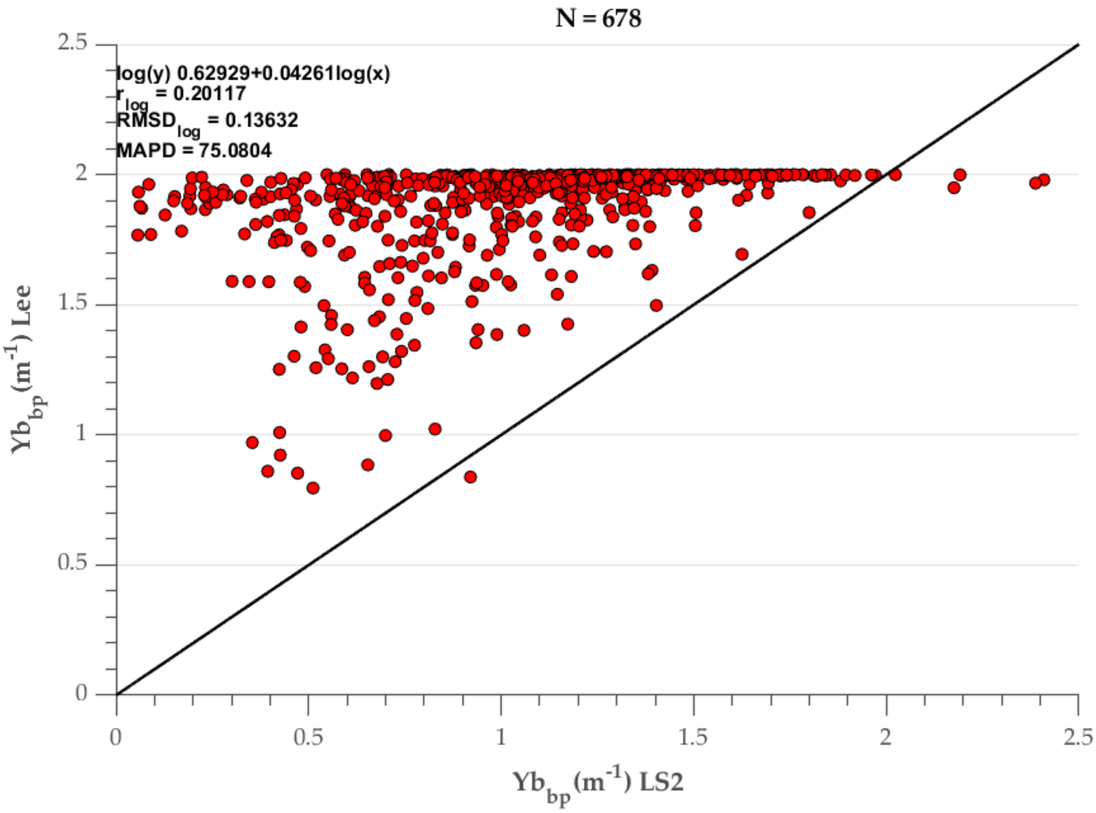


Figure 52: Distribution of the Y b_{bp} estimated using the proposed algorithm by 2SAA vs the Y b_{bp} estimated using Lee et al., (2002) for the BioArgos data set.

7 Application to remote sensing images

7.1 Level 3 Images - OLCI & VIIRS

As an example of the algorithm performance on remote sensing images, we selected one monthly L3 R_{rs} image (May 2018). We also used 3 different products to evaluate the discrepancies between them. The products were: 1) Generalized Inherent Optical Property (GIOP) for the VIIRS image (provided by NASA); 2) Garver-Siegel-Maritorena (GSM) for the OLCI image (provided by Glob Colour); 3) GIOP for the OLCI image (provided by LOG).

For a_{nw} we can see a high agreement with all the products, except for extremely clear waters (MAPD<40% in general). The highest correlation was found between 2SAA and GIOP provided by LOG (Figure 55). The difference when you compare 2SAA results against GIOP provided by NASA (**Figure 53**) and LOG **(Figure 55)** are probably due to the differences in sensor (OLCI vs VIIRS) and atmospheric correction protocols. When comparing the density plots of 2SAA against GSM OLCI (**Figure 54**) and GIOP VIIRS (**Figure 55**), it is possible to identify a plateau at 0.01 m^{-1}, in which the algorithm can't successfully retrieve any absorption bellow this magnitude. Meanwhile for 2SAA, a saturation wasn't observed at any given magnitude. This differences can be clearly observed near the gyres, in which MAPD can identify a difference of over 100%

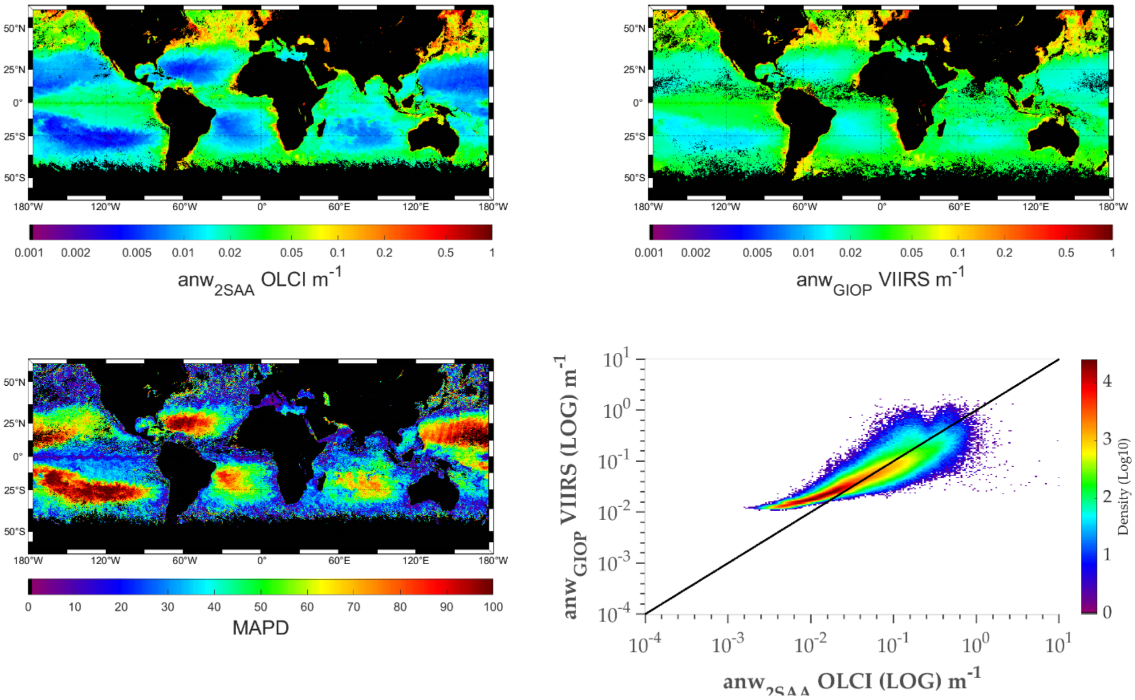


Figure 53: Comparison of the a_{nw} (443) estimated using the proposed algorithm for OLCI (top left) for one monthly image against the a_{nw} (443) estimated by GIOP for VIIRS and provided by NASA (top right). On the bottom left the MAPD between the two products. On the bottom right, the density plot of the two images.

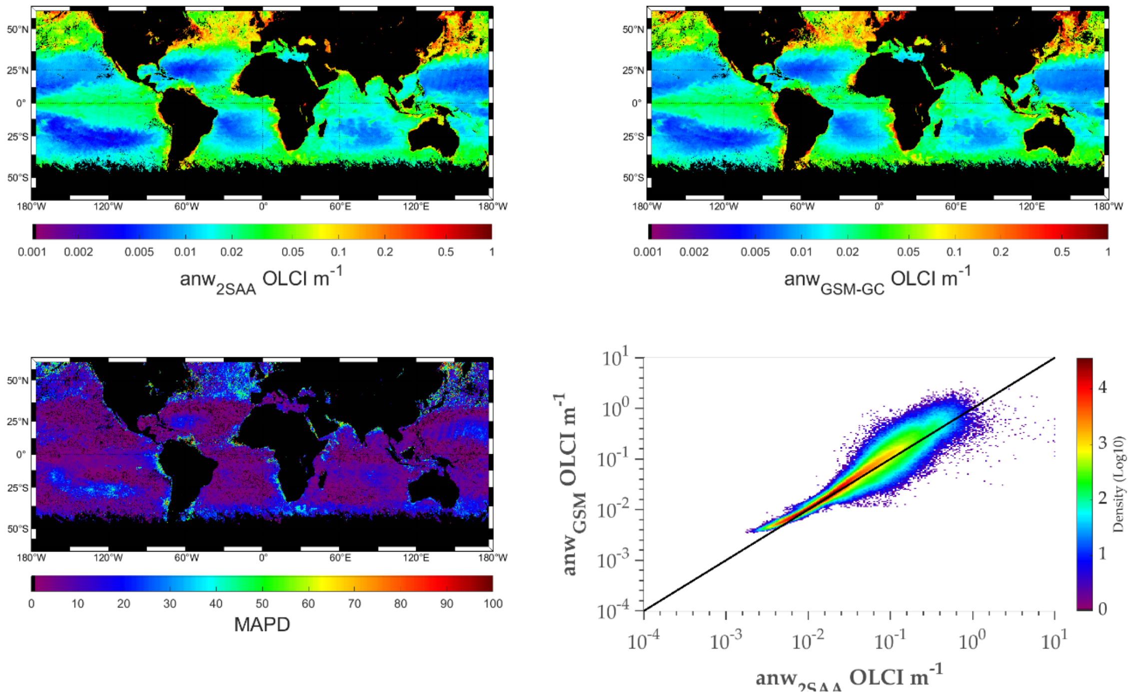


Figure 54: Comparison of the a_{nw} (443) estimated using the proposed algorithm for OLCI (top left) for one monthly image against the a_{nw} (443) estimated by GSM for

OLCI and provided by GlobColour (top right). On the bottom left the MAPD between the two products. On the bottom right, the density plot of the two images.

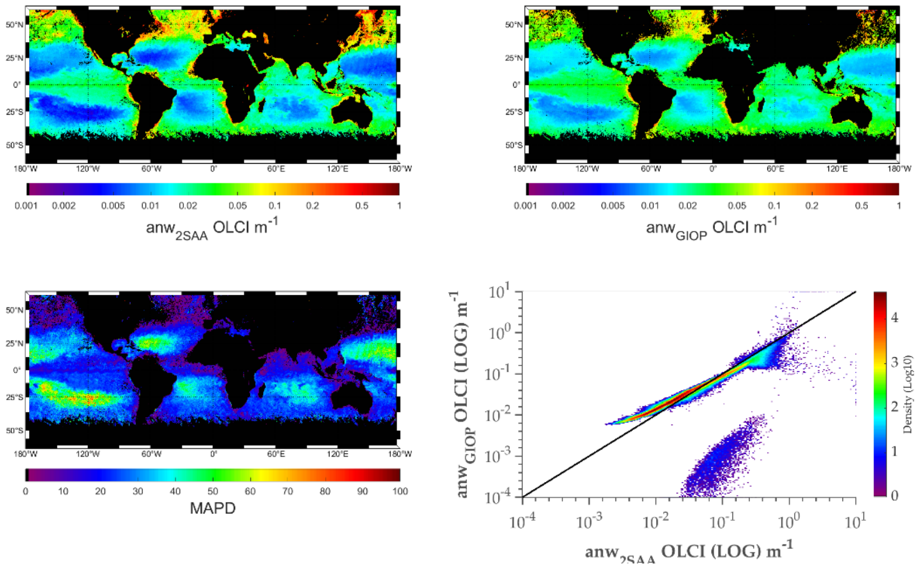


Figure 55: Comparison of the a_{nw} (443) estimated using the proposed algorithm for OLCI (top left) for one monthly image against the a_{nw} (443) estimated by GIOP for OLCI and provided by LOG (top right). On the bottom left the MAPD between the two products. On the bottom right, the density plot of the two images.

For b_{bp}, a higher difference was observed among the different products, especially in the gyres, an area in which b_{bp} uncertainty is higher. The highest agreement was observed between 2SAA and GSM-GC (**Figure 57**), followed by GIOP provided by LOG (**Figure 58**) and GIOP provided by NASA (**Figure 56**). This over-estimation of b_{bp} observed over clear waters by GIOP and GSM (also from QAA, not shown) is certainly due to the formulation used in these different model for the b_{bp} spectral slope, which quickly saturates, in contrast to the one calculated by 2SAA (see **Figure 59**).

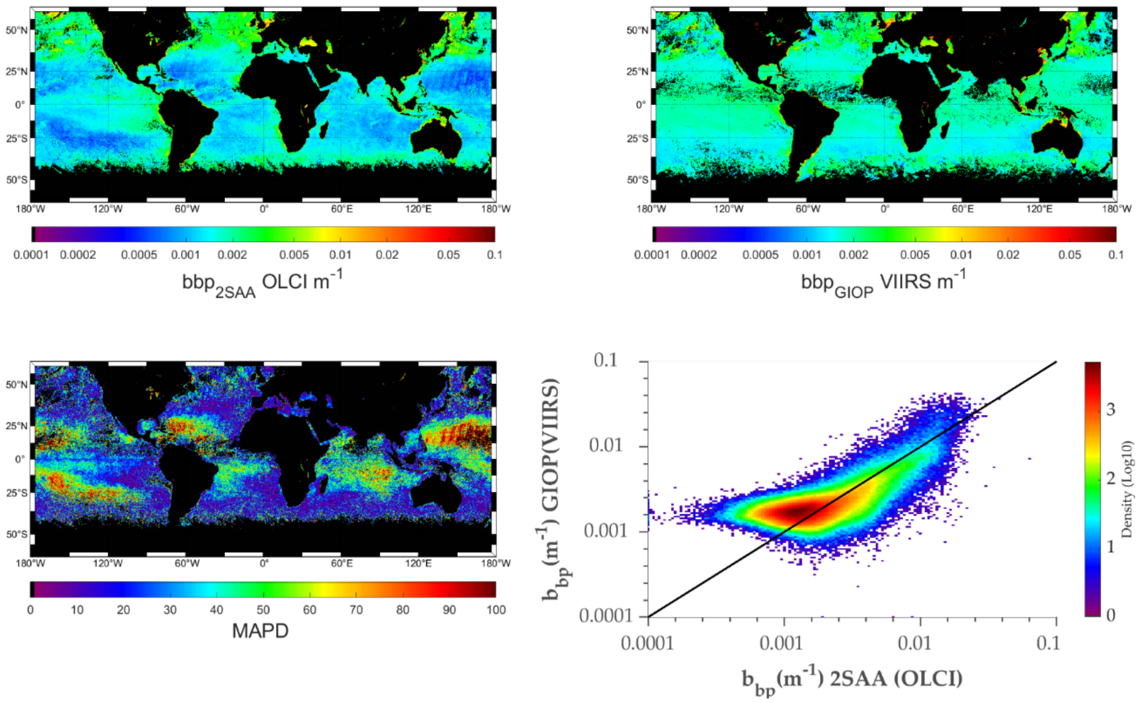


Figure 56: Comparison of the b_{bp} (443) estimated using the proposed algorithm for OLCI (top left) for one monthly image against the b_{bp} (443) estimated by GIOP for VIIRS and provided by NASA (top right). On the bottom left the MAPD between the two products. On the bottom right, the density plot of the two images.

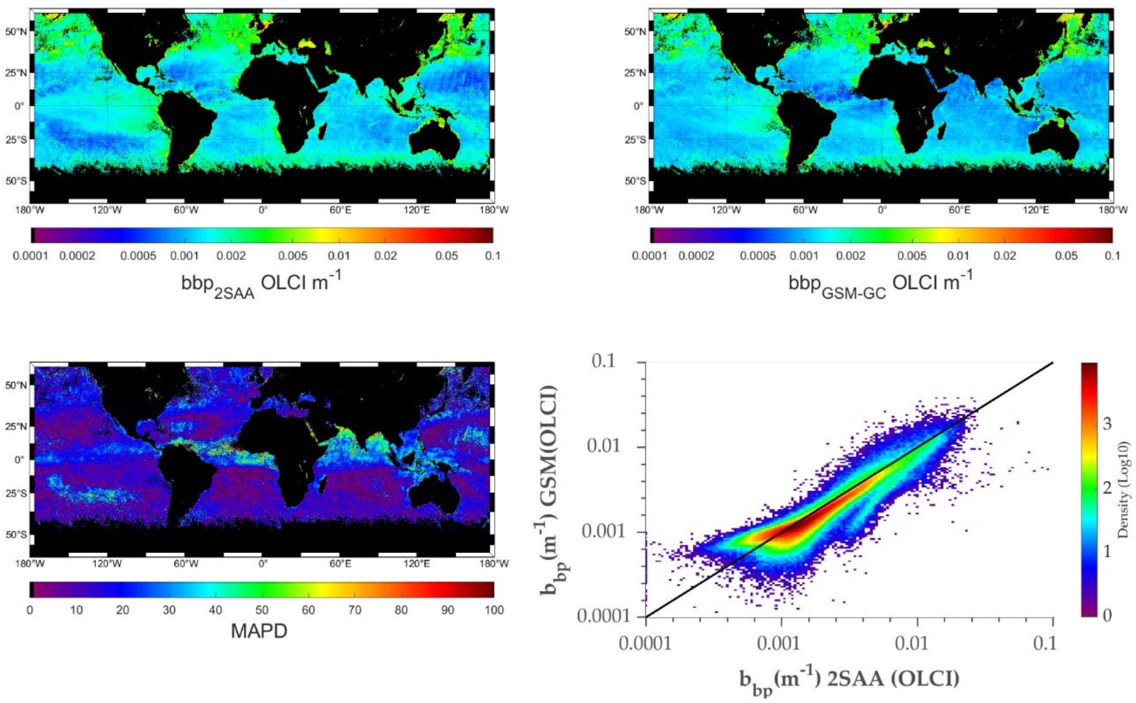


Figure 57: Comparison of the b_{bp} (443) estimated using the proposed algorithm for OLCI (top left) for one monthly image against the b_{bp} (443) estimated by GSM for OLCI and provided by GlobColour (top right). On the bottom left the MAPD between the two products. On the bottom right, the density plot of the two images.

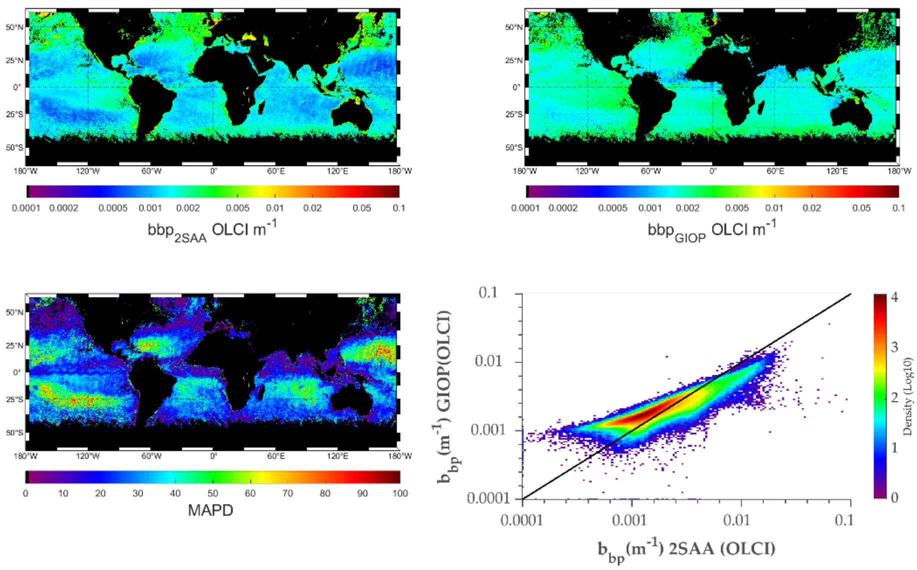


Figure 58: Comparison of the b_{bp} (443) estimated using the proposed algorithm for OLCI (top left) for one monthly image against the b_{bp} (443) estimated by GIOP for OLCI and provided by LOG (top right). On the bottom left the MAPD between the two products. On the bottom right, the density plot of the two images.

To show the differences in Yb_{bp} estimated by different methodologies, **Figure 59** show the Yb_{bp} using 2SAA, Lee et al., (2002) and Morel & Maritorena (2001). The first clear difference is the dynamic range for each algorithm. For 2SAA, the magnitude is distributed between 0 and 3, with a shape similar to a normal distribution peaking at 1 and 1.5. Meanwhile, for the other 2 algorithms, there is a clear saturation. For Lee et al., (2002) the distribution follow an exponential curve peaking at 2, while for Morel & Maritorena (2001) the algorithm saturate at 0.8. Considering the differences in the optical properties of the particles over the globe, a difference in Yb_{bp} is expected over different water masses.

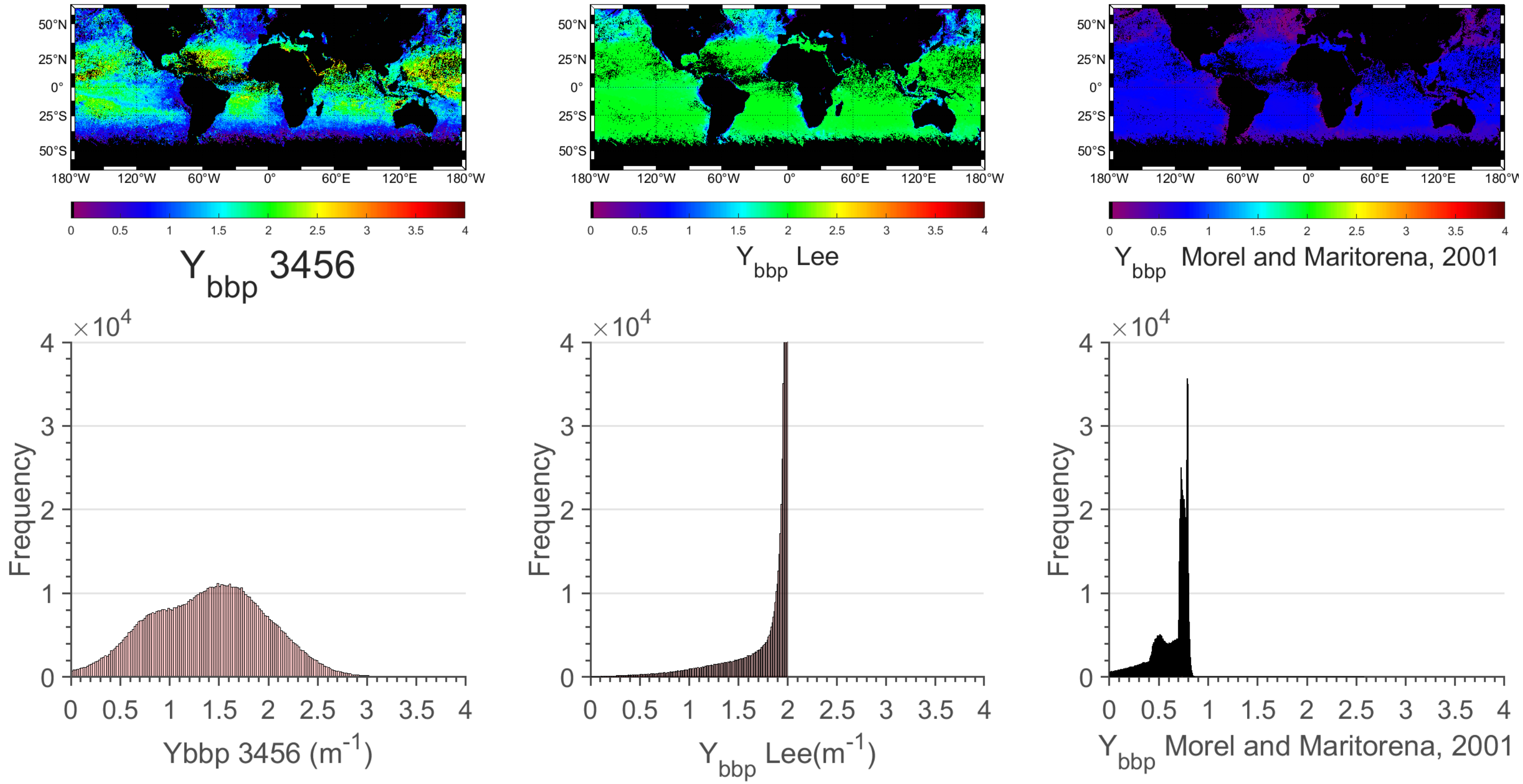


Figure 59: Comparison of the Yb_{bp}estimated using three different methodologies for one OLCI monthly image. In the first line the global Ybbp is displayed, and in the second line the correspondent histogram is displayed. Yb_{bp} 3456 was calculated using the bands between 443 and 560nm. Yb_{bp} QAA was calculated following Let et al., 2002. Yb_{bp} Morel and Maritorena, 2001 was calculated following the corresponding study and used the Chl-a calculated from Zhang et al., 2015.

For a_{phy} (**Figure 60**, **Figure 61** and **Figure 62**) and a_{cdm} (**Figure 63**, **Figure 64**, and **Figure 65**) the biggest difference between the algorithms is in the Pacific (near the gyres), in which it is possible to see the impact of a_{nw} in the retrieval of a_{phy} and a_{cdm}. Zhang et al., 2015 retrieved lower values for a_{phy} for all the Pacific and near the gyres, meanwhile for a_{cdm} the behavior is the also similar, with a lower retrieval value close to the gyres. This difference is expected, due to the lowera_{nw} retrieved values by 2SAA.

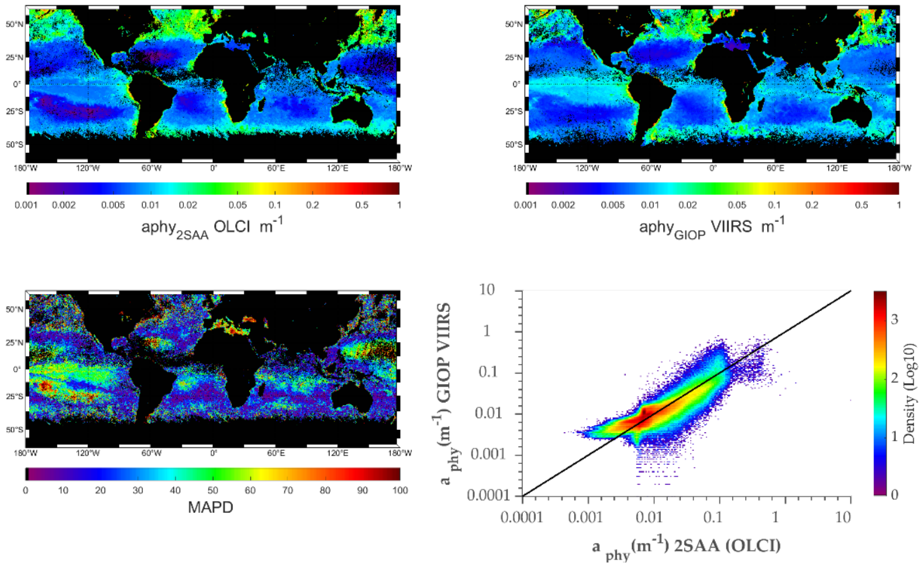


Figure 60: Comparison of the a_{phy} (443) estimated using the proposed algorithm for OLCI (top left) for one monthly image against the a_{phy} (443) estimated by GIOP for VIIRS and provided by NASA (top right). On the bottom left the MAPD between the two products. On the bottom right, the density plot of the two images.

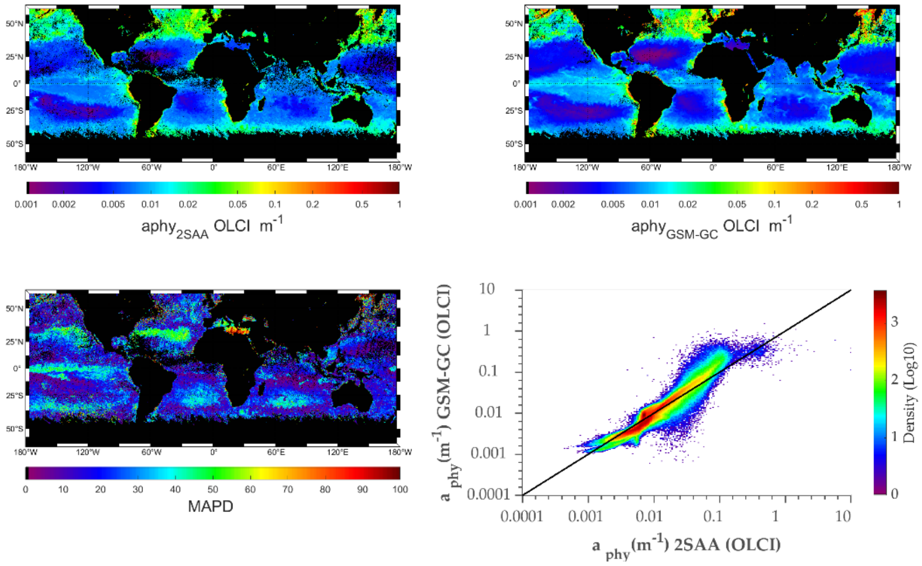


Figure 61: Comparison of the a_{phy} (443) estimated using the proposed algorithm for OLCI (top left) for one monthly image against the a_{phy} (443) estimated by GSM for OLCI and provided by GlobColour (top right). On the bottom left the MAPD between the two products. On the bottom right, the density plot of the two images.

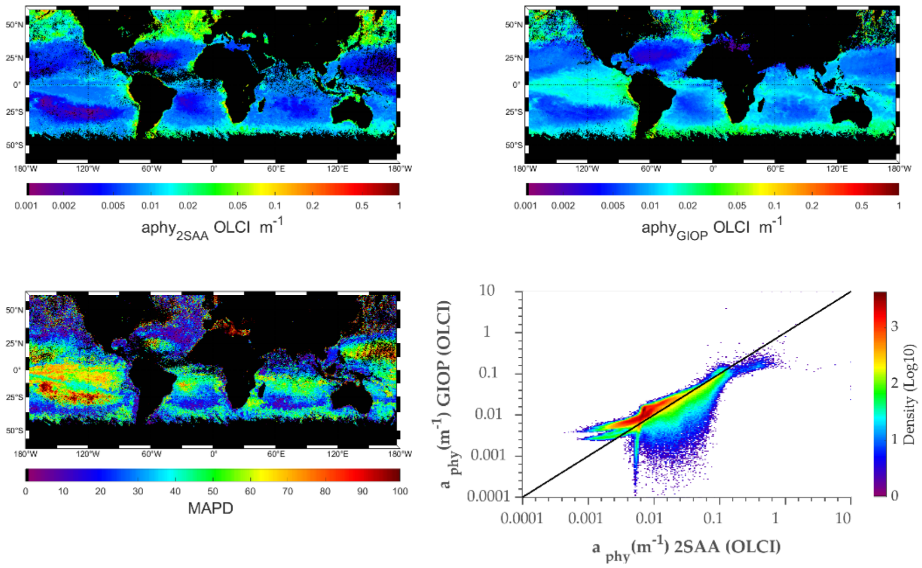


Figure 62: Comparison of the a_{phy} (443) estimated using the proposed algorithm for OLCI (top left) for one monthly image against the a_{phy} (443) estimated by GIOP for OLCI and provided by LOG (top right). On the bottom left the MAPD between the two products. On the bottom right, the density plot of the two images.

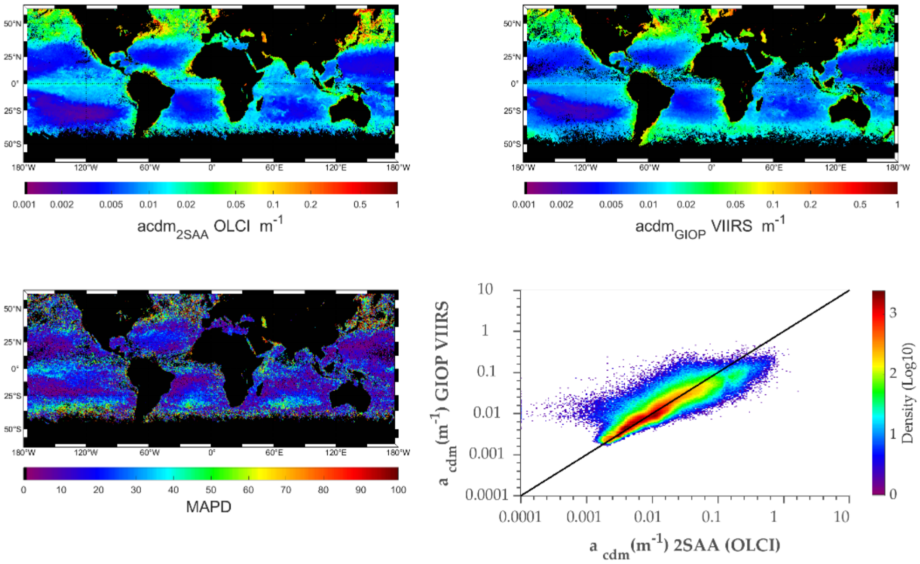


Figure 63: Comparison of the a_{cdm} (443) estimated using the proposed algorithm for OLCI (top left) for one monthly image against the a_{cdm} (443) estimated by GIOP for VIIRS and provided by NASA (top right). On the bottom left the MAPD between the two products. On the bottom right, the density plot of the two images.

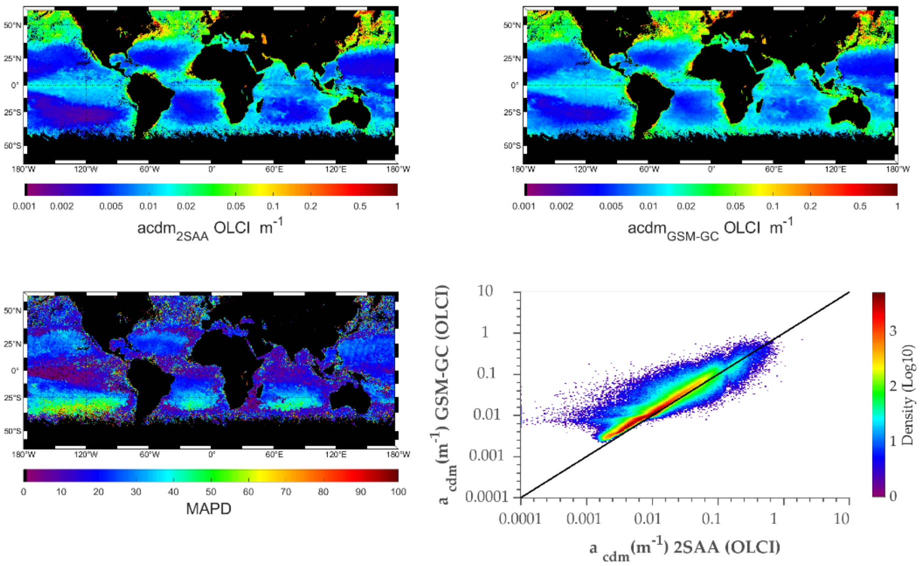


Figure 64: Comparison of the a_{cdm} (443) estimated using the proposed algorithm for OLCI (top left) for one monthly image against the a_{cdm} (443) estimated by GSM for OLCI and provided by GlobColour (top right). On the bottom left the MAPD between the two products. On the bottom right, the density plot of the two images.

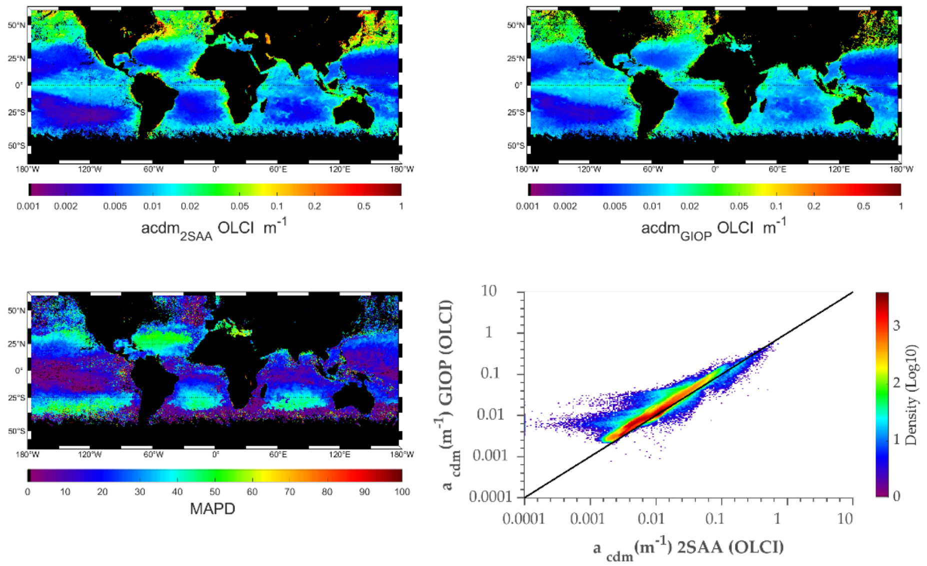


Figure 65: Comparison of the a_{cdm} (443) estimated using the proposed algorithm for OLCI (top left) for one monthly image against the a_{cdm} (443) estimated by GIOP for OLCI and provided by LOG (top right). On the bottom left the MAPD between the two products. On the bottom right, the density plot of the two images.

Lastly, in **Figure 66** we can observe the retrieval flag used in this algorithm. It is clear that the algorithm has a high retrieval percentage, and could retrieve all the IOPs for this image.

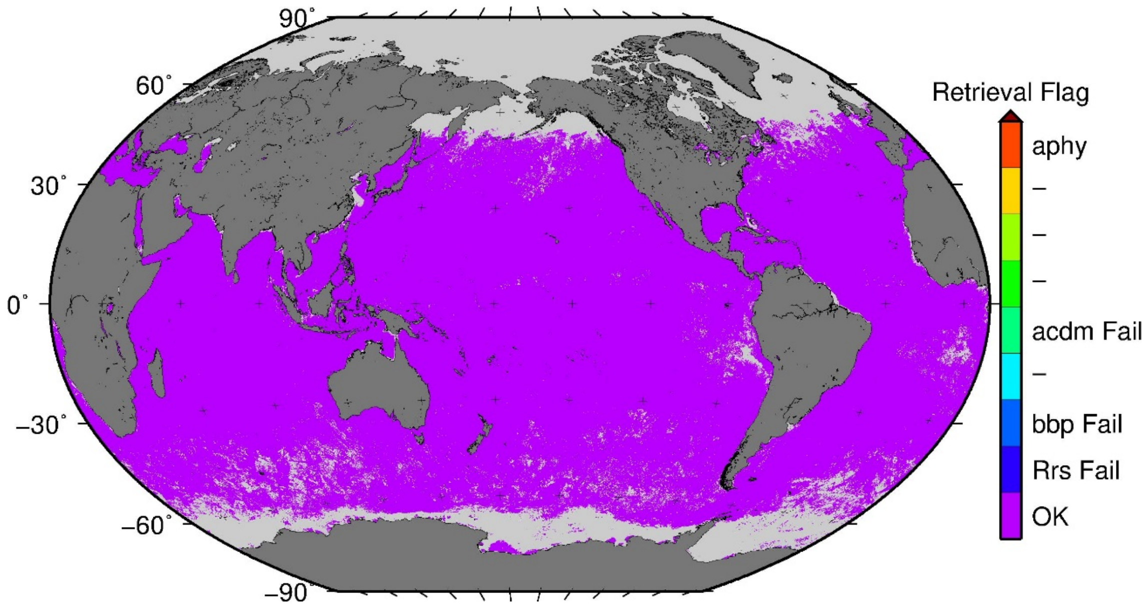


Figure 66 Retrieved flag for each IOP parameter estimated during the inversion

8 Sensitivity analysis in the algorithm due to uncertainty in a_{nw}

We examine here from the in situ data set how the uncertainty in a_{nw} is propagated throughout the second step of the algorithm.

For the uncertainty propagation, the noise added to the a_{nw} was estimate using equation 1.

$$a_{nw}^{*}(B_n) = a_{nw}(B_n) \pm a_{nw}(B_n)\, rnd(0{,}1)\, Noise$$

Where a_{nw}^{*} is a_{nw} with the addition of the noise, B_n is the band number, $rnd(0{,}1)$ is a standard normal random distribution with mean 0 and standard deviation 1, and noise is the uncertainty in a_{nw} (between 0 and 1). For each OLCI band, one hundred random values were generated using a standard normal distribution (Total of 800 per spectrum). The same values in $rnd(0{,}1)$ were used for all the spectra, so that all the variability in a_{nw}^{*} is due to the changes in $rnd(0{,}1)$. For the $Noise$ values, we assumed different scenarios between 0 and 100%. In **Figure 67** we can see the expected impact of the uncertainty in a_{nw} for a_{phy} and a_{cdm} estimation with an uncertainty of 0%,1%, 5%, 10% and 20%.

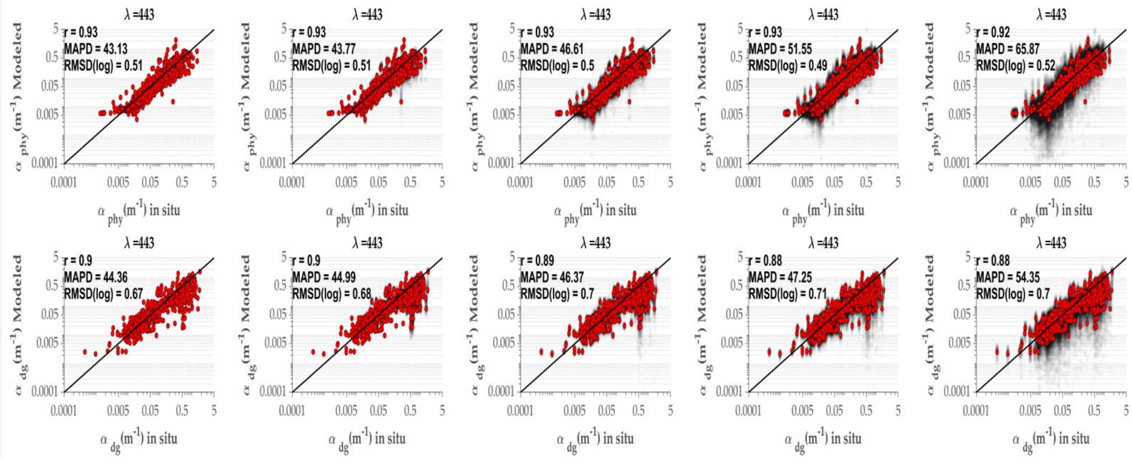


Figure 67: Noise propagation during the Zhang et al., 2015 algorithm under different conditions for a_{nw} uncertainty. The uncertainties are 0%,1%, 5%, 10% and 20%.

It is clear that as we increase the uncertainty in a_{nw}, the uncertainty is propagated throughout the inversion algorithm, resulting in a higher dispersion. This can be observed by comparing the black dots dispersion between each figure, and by the increase in the MAPD value as we increase the uncertainty.

The direct impact on the MAPD was also calculated by setting an uncertainty of 0%,5%, 10%, 20%, 40%, 60%, 80% and 100%, and normalizing the MAPD by the MAPD on the 0% scenario. In **Figure 68** we can see the results of the simulation, in which we identified a higher impact of the a_{nw} uncertainty in the a_{phy} retrieval than for a_{cdm}. Another aspect is that the MAPD doesn't increase linearly with the uncertainty in a_{nw}. While the impact on a_{cdm} was relatively close to the 1:1 line, the impact on a_{phy} was almost 2 times higher. So, the propagation of the uncertainty in Rrs estimatives during the first step of the 2SAA inversion may induce have a negative impact in the retrieval of a_{cdm} and a_{phy}.

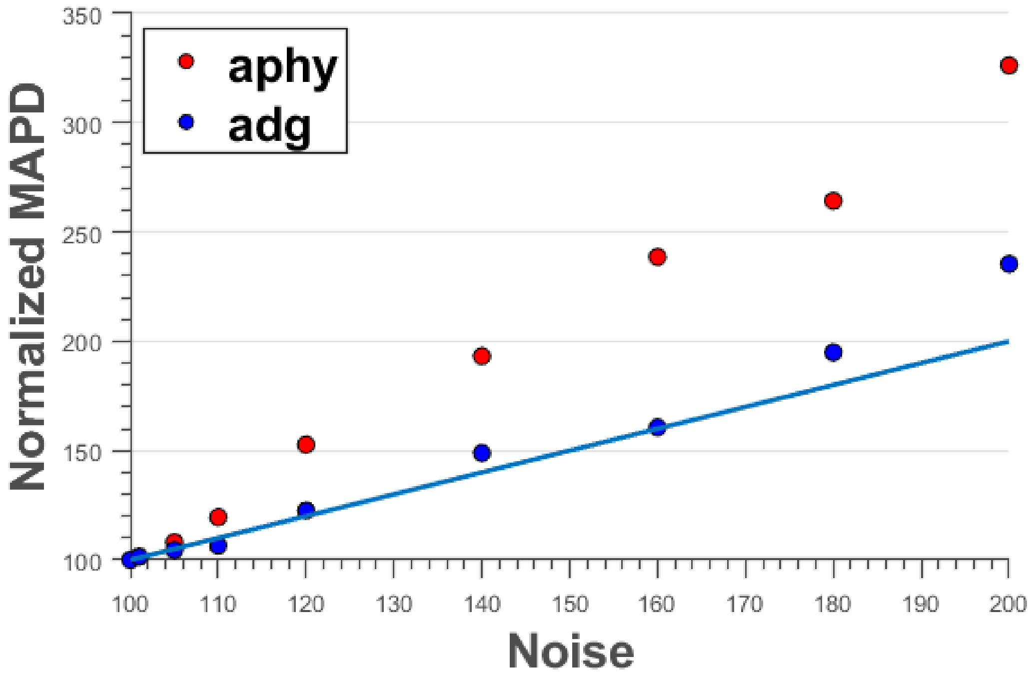


Figure 68: Normalized MAPD for a_{phy} and a_{cdm} under different uncertainties (Noise) for a_{nw}. The uncertainties described are: 0%,5%, 10%, 20%, 40%, 60%, 80% and 100%.

9 Conclusions and perspectives

In general, the IOPs are estimated with a relatively good accuracy over inland, open and coastal waters. For instance, using the in situ data set which is impacted by R_{rs} and IOPs errors measurements, $b_{bp}(443)$, $a_{phy}(443)$, $a_{cdm}(443)$, and $a_{cdom}(443)$ are estimated with a MAPD of 13.26%, 34.17%, 34.53% and 54.56% over open and coastal waters, respectively. Over inland waters, $b_{bp}(620)$, $a_{phy}(443)$, $a_{cdm}(443)$, and $a_{cdom}(443)$ are estimated with a MAPD of 26.32%, 51.89%, 42.27% and 41.17%, respectively. Note that b_{bp} MAPD was estimated at 620 nm due to the abnormal behavior of the 2SAA model over Amazon lakes in Brazil (most likely due to some measurement issues due to pathlength correction). More in situ data should be collected in inland waters to increase the IOPs validation range. Match-up exercise for inland waters has not been provided due to the rough spatial resolution of the sensors. Using classification in the inversion process allowed to improve the inversion of a_{phy} and a_{cdm} in clear oceanic waters. From the classification method adopted in this study, it is important to note that a precise estimation of the class membership probability associated with an unknown spectrum (fuzzy logic approach described in Moore et al., 2001, Mélin and Vantrepotte 2015) would provide more precise information, particularly useful in the objective to use OWCs for deriving class based inversion algorithms.

The inter-comparison between OLCI and VIIRS IOPs products are, by essence, impacted by the different Rrs values. It should therefore be necessary to apply 2SAA to Rrs-VIIRS data to complete this inter-comparison exercise.

In general, the 2SAA was robust when compared to the two available products tested during the validation (GIOP and GSM). It displayed a higher dynamic range, especially for extremely clear waters, where those 2 algorithms show saturation patterns for all IOPs, most likely due to the b_{bp} spectral behavior adopted in these two later models. Sensitivity analysis should be performed to implement the calculation for b_{bp}, and consequently Yb_{bp}, by 2SAA in the GIOP/GSM algorithm to estimate the two absorption parameters. Further performance validation should be made to check whether this change is significant or not.

10 Acknowledgment

We would like to thank the collaborative effort of different research groups to build up the dataset used during this study. Specifically, Julia Uitz (Peacetime), Herve Claustre (Bio-Argo), Vincenzo Vellucci (BOUSSOLE), Stephane Maritorena (Plume & Blooms), Simon Belanger (St-Laurence), Caren Binding (Great-Lakes), Tiit Kutser (Estonian lakes), Annelies Hommerson (Netherlands lakes) and Claudio Barbosa (Mamiraua lakes).

11 References

Antoine, D., Chami, M., Claustre, H., d'Ortenzio, F., Morel, A., Bécu, G., Gentili, B., Louis, F., Ras, J., Roussier, E. and Scott, A.J., 2006. BOUSSOLE: a joint CNRS-INSU, ESA, CNES, and NASA ocean color calibration and validation activity.

Claustre, H., 2009. Bio-optical profiling floats as new observational tools for biogeochemical and ecosystem studies: Potential synergies with ocean color remote sensing.

Kutser, T., Paavel, B., Verpoorter, C., Ligi, M., Soomets, T., Toming, K. and Casal, G., 2016. Remote sensing of black lakes and using 810 nm reflectance peak for retrieving water quality parameters of optically complex waters. Remote Sensing, 8(6), p.497.

Lee, Z., Carder, K.L. and Arnone, R.A., 2002. Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. Applied optics, 41(27), pp.5755-5772.

Loisel, H., Mériaux, X., Berthon, J.F. and Poteau, A., 2007. Investigation of the optical backscattering to scattering ratio of marine particles in relation to their biogeochemical composition in the eastern English Channel and southern North Sea. Limnology and Oceanography, 52(2), pp.739-752.

Loisel, H., Vantrepotte, V., Dessailly, D. and Mériaux, X., 2014. Assessment of the colored dissolved organic matter in coastal waters from ocean color remote sensing. Optics Express, 22(11), pp.13109-13124.

Loisel, H., Stramski, D., Dessailly, D., Jamet, C., Li, L. and 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. Journal of Geophysical Research: Oceans, 123(3), pp.2141-2171.

Lubac, B., Loisel, H., Guiselin, N., Astoreca, R., Artigas, L.F. and Mériaux, X., 2008. Hyperspectral versus multispectral remote sensing approach to detect phytoplankton blooms in coastal waters: application to a Phaeocystis bloom. Journal of Geophysical Research, (113), pp.1-17.

Mélin, F. and Vantrepotte, V., 2015. How optically diverse is the coastal ocean?. Remote Sensing of Environment, 160, pp.235-251.

Moore, T.S., Campbell, J.W. and Feng, H., 2001. A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms. IEEE Transactions on Geoscience and Remote sensing, 39(8), pp.1764-1776.

Morel, A. and Maritorena, S., 2001. Bio-optical properties of oceanic waters: A reappraisal. Journal of Geophysical Research: Oceans, 106(C4), pp.7163-7180.

Nechad, B., Ruddick, K.G., Schroeder, T., Oubelkheir, K., Blondeau-Patissier, D., Cherukuru, R.C.N., Brando, V.E., Dekker, A.G., Clementson, L.A., Banks, A.C. and Maritorena, S., 2015. CoastColour Round Robin datasets: a database to evaluate the performance of algorithms for the retrieval of water quality parameters in coastal waters. Earth system science data, 7, pp.319-348.

Neukermans, G., Loisel, H., Mériaux, X., Astoreca, R. and McKee, D., 2012. In situ variability of mass-specific beam attenuation and backscattering of marine particles with respect to particle size, density, and composition. Limnology and Oceanography, 57(1), pp.124-144.

Stramski, D., Reynolds, R.A., Babin, M., Kaczmarek, S., Lewis, M.R., Röttgers, R., Sciandra, A., Stramska, M., Twardowski, M.S., Franz, B.A. and Claustre, H., 2008. Relationships between the surface concentration of particulate organic carbon and optical properties in the eastern South Pacific and eastern Atlantic Oceans. Biogeosciences, 5(1), pp.171-201.

Werdell, P.J., Franz, B.A., Bailey, S.W., Feldman, G.C., Boss, E., Brando, V.E., Dowell, M., Hirata, T., Lavender, S.J., Lee, Z. and Loisel, H., 2013. Generalized ocean color inversion model for retrieving marine inherent optical properties. Applied optics, 52(10), pp.2019-2037.

Zhang, X., Huot, Y., Bricaud, A. and Sosik, H.M., 2015. Inversion of spectral absorption coefficients to infer phytoplankton size classes, chlorophyll concentration, and detrital matter. Applied optics, 54(18), pp.5805-5816.

Zheng, G., Stramski, D. and Reynolds, R.A., 2014. Evaluation of the Quasi-Analytical Algorithm for estimating the inherent optical properties of seawater from ocean color: Comparison of Arctic and lower-latitude waters. Remote sensing of environment, 155, pp.194-209.