



Evaluation of four atmospheric correction algorithms for MODIS-Aqua images over contrasted coastal waters

C. Goyens^a, C. Jamet^{a,*}, T. Schroeder^b

^a CNRS, UMR 8187, Univ Lille Nord de France, ULCO, LOG, F-62930 Wimereux, France

^b CSIRO Land and Water, 41 Boggo Road, Brisbane, QLD 4102, Australia

ARTICLE INFO

Article history:

Received 25 July 2012

Received in revised form 30 October 2012

Accepted 4 December 2012

Available online January 10 2013

Keywords:

MODIS-Aqua

Ocean color

Atmospheric correction

Turbid water

Validation

Classification

ABSTRACT

The use of satellites to monitor the color of the ocean requires effective removal of the atmospheric signal. This can be performed by extrapolating the aerosol optical properties to the visible from the near-infra-red (NIR) spectral region assuming that seawater is totally absorbent in this latter part of the spectrum, the so-called black pixel assumption. While this assumption is verified for most phytoplankton dominated waters, it is invalid in turbid waters. Consequently, for the past ten years, several algorithms have been developed on alternative assumptions. Studies comparing these algorithms are of great interest for further improvement in water leaving radiance ($L_w(\lambda)$) retrievals from satellite images explaining the focus of the present research. Four published atmospheric correction algorithms for MODIS-Aqua are compared: (1) the standard NIR algorithm of NASA, (2) the NIR similarity spectrum algorithm, (3) the NIR-SWIR algorithm and (4) an Artificial Neural Network algorithm. The MODIS-Aqua estimated normalized $L_w(\lambda)$ are validated with AERONET-Ocean Color data and cruise measurements presenting moderately to highly turbid waters. Based on a match-up exercise, the former three algorithms show the best results in the green region of the spectrum (relative error, RE, between 11 and 20%) and the largest errors in the blue and red region of the spectrum (RE exceeding 30%). In contrast, the Artificial Neural Network algorithm performs better in the red band (RE of 22%). The latter tends to overestimate the normalized $L_w(\lambda)$ at all wavelengths while the NIR similarity spectrum algorithm tends to underestimate it. Retrievals of aerosol products, such as the Ångström coefficient, $\alpha(531,869)$, and the optical thickness, $\tau(869)$, present RE above 44% and 72%, respectively.

The performance of the algorithms is also investigated as a function of water types. For water masses mainly dominated by phytoplankton, the standard NIR algorithm performs better. In contrast, for water masses mainly dominated by detrital and mineral material, the neural network-based algorithm shows the best results. The largest errors are encountered above water masses dominated by high phytoplankton and CDOM concentrations. This work conducted to a number of perspectives for improving the atmospheric correction algorithms.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

With their broad spatial coverage and fine temporal resolutions, satellite remote sensing data represent a crucial tool for monitoring large areas of oceans. The remote sensing of ocean color radiometry, defined as the spectrum of the up-welling light from the sea (Gordon, 1987), allows to investigate the bio-optical properties of the seawater. These are then used as proxies for various geophysical parameters (e.g., concentrations of chlorophyll-a, Chl-a, and suspended particulate matter, SPM). The first satellite dedicated for ocean color research, the Coastal Zone Color Scanner (CZCS), was launched in 1978 and operated till 1986. Eleven years later the Sea-Viewing Wide Field of View Sensor (SeaWiFS) was launched followed by the Moderate Resolution Imaging Spectroradiometer (MODIS-Aqua) and the Medium Resolution Imaging Spectroradiometer (MERIS), both launched in 2002. SeaWiFS and

MERIS have been operable until December 2010 and April 2012, respectively.

Since the launch of the CZCS, improvements in radiometric and spectral performances and in the reliability of ocean color products have been impressive. Nevertheless, processing satellite ocean color data remains a challenging task. This includes, among others, the effective removal of the contribution of the atmosphere to the signal measured by the satellite, the so-called atmospheric correction process. In clear open ocean waters, for which the optical properties are mainly driven by phytoplankton and co-varying biological constituents, it is assumed that seawater absorbs all the light in the red and near-infrared (NIR) region of the spectrum, referred as the black-pixel assumption (Gordon & Wang, 1994, thereafter named GW94). This allows to estimate the contribution of the atmosphere and subsequently the light scattered by the water column and measured by the satellite, known as the water-leaving reflectance, $\rho_w(\lambda)$. While the black-pixel assumption is verified in clear ocean waters, it induces significant errors when applied in turbid waters. The latter are indeed optically more complex due to the backscattering

* Corresponding author. Tel.: +33 3 21 99 64 22.

E-mail address: cedric.jamet@univ-littoral.fr (C. Jamet).

and absorption of Colored Dissolved Organic Matter (CDOM) and non-algal particles (Siegel et al., 2000). Moreover these water masses present high spatial and temporal variability. Consequently, global parametrizations are less applicable (e.g., Loisel et al., 2010; Werdell et al., 2007) and significant errors remain in retrieving geophysical parameters and water properties from satellite remote sensing (e.g., Zibordi et al., 2009a, 2009b). Numerous algorithms have been developed with alternative hypotheses taking into account non-negligible NIR ocean contribution to the measured signal. A global validation of these algorithms is required in order to gain insight in their performances and limitations and to aid movement towards improvements. To our knowledge, several studies already compared the performances of new developed algorithms against the standard atmospheric correction algorithms (e.g., Kuchinke et al., 2009; Ruddick et al., 2000; Schroeder et al., 2007; Wang & Shi, 2007; Wang et al., 2009; Werdell et al., 2010). However only few studies (Jamet et al., 2011) have tackled a round-robin evaluation of all available atmospheric correction algorithms. This is the focus of the present research. It is a complementary work to the study of Jamet et al. (2011) who evaluated, based on a similar match-up exercise, three SeaWiFS atmospheric correction algorithms.

MODIS-Aqua images processed with four commonly used atmospheric correction algorithms are compared with coastal in-situ data. The four algorithms are (1) the NASA official algorithm based on an iterative procedure and a bio-optical model (Bailey et al., 2010; Stumpf et al., 2003) (referred later as 'STD'), (2) a GW94-based algorithm assuming spatial homogeneity of the atmosphere and ocean over the region of interest in the NIR (Ruddick et al., 2000) (referred later as 'SIMIL'), (3) a NIR-based algorithm using the Short-Wave-Infra-Red (SWIR) bands in very turbid waters (Wang et al., 2009) (referred later as 'NIR-SWIR') and (4) a direct inversion approach using a neural network (Schroeder et al., 2007) (referred later as 'NN'). Besides a global evaluation of the algorithms, the performances of the algorithms are also investigated as a function of water type. This is particularly relevant to gain insight in their applicability and limitations. Therefore, in-situ data are assigned to different classes of water types (Vantrepotte et al., 2012) and an evaluation is made per water type.

The next section describes the different algorithms used in the present research. Data and methods are outlined in Section 3. This section includes an overview of the in-situ and satellite data used for the evaluation, an explanation of the match-up selection and algorithm evaluation criteria, and a brief description of the classification method of Vantrepotte et al. (2012). In Section 4 the evaluation of the algorithms over the entire validation dataset as well as their sensitivities to different water types are outlined and discussed. These results conducted to the definition of some challenges and perspectives for further improvements.

2. Algorithms

The signal measured by the satellite sensor can be decomposed into several terms (Gordon, 1997):

$$\rho_t(\lambda) = \rho_r(\lambda) + \rho_a(\lambda) + \rho_{ra}(\lambda) + t_v(\lambda)\rho_{wc}(\lambda) + T_v(\lambda)\rho_g(\lambda) + t_v(\lambda)\rho_w(\lambda) \quad (1)$$

where $\rho_t(\lambda)$ represents the reflectance measured by the satellite and $\rho_r(\lambda)$, $\rho_a(\lambda)$ and $\rho_{ra}(\lambda)$ the contribution of the atmosphere, notably, the scattered sunlight by air molecules, aerosols and the coupling between both air and aerosol molecules, respectively. $\rho_{wc}(\lambda)$ is the reflectance related to the whitecaps and $\rho_g(\lambda)$ to the glitter. t_v and T_v are the diffuse and direct atmospheric transmittance from sea to sensor, respectively. Ultimately, $\rho_w(\lambda)$ is the water leaving reflectance. Together, $\rho_r(\lambda)$, $\rho_a(\lambda)$ and $\rho_{ra}(\lambda)$ represent about 90% of $\rho_t(\lambda)$. $\rho_g(\lambda)$ and $\rho_{wc}(\lambda)$ can be calculated for a given sun and sensor zenith angle (Cox & Munk, 1954) and an empirical wind speed formulation (Frouin et al.,

1996), respectively. During pre-processing $\rho_t(\lambda)$ is corrected for gas absorption, Rayleigh scattering, white-caps and glitter (Gordon, 1997). This term is referred as the Rayleigh corrected reflectance, $\rho_{rc}(\lambda)$:

$$\rho_{rc} = \rho_t(\lambda) - \rho_r(\lambda) - t_v(\lambda)\rho_{wc}(\lambda) - T_v(\lambda)\rho_g(\lambda). \quad (2)$$

In order to ease the comparison of our results with other studies, we focussed on the normalized water-leaving radiance, $L_{wn}(\lambda)$ which is related to $\rho_w(\lambda)$ as follow:

$$L_{wn}(\lambda) = \frac{\rho_w(\lambda)F_0(\lambda)\cos\theta_0}{\pi} \quad (3)$$

with F_0 being the extra-terrestrial solar irradiance at a given wavelength λ and θ_0 the sun zenith angle.

2.1. GW94-based algorithm with iterative procedure ('STD')

The STD algorithm, initially developed by Stumpf et al. (2003) and revised by Bailey et al. (2010), uses an iterative method including a bio-optical model allowing to account for the water contribution in the NIR region of the spectrum. The algorithm uses 80 aerosol models developed with observations from the global AERONET network and the radiative transfer code described in Ahmad et al. (2010). First the algorithm assumes $L_{wn}(\lambda)$ being null in the NIR (i.e. black pixel assumption, GW94) allowing to retrieve $L_{wn}(\lambda)$ in the blue and green region of the spectrum. Estimations of L_{wn} at 443, 488 and 547 nm are then used as inputs into the bio-optical model. The latter allows to achieve a first guess in the concentration of Chl-a (MODIS Chl-a OC3 algorithm, O'Reilly et al. (2000)) which in turn allows to determine the particulate backscattering in the NIR, $b_{bp}(NIR)$. Knowing $b_{bp}(NIR)$, it is possible to determine the contribution of the particles to $L_{wn}(NIR)$. The process is then re-iterated with $L_{wn}(NIR)$ equals zero after the modeled reflectance has been removed from $L_{rc}(NIR)$. The iteration process continues until $L_{wn}(NIR)$ converges. The bio-optical model of the STD algorithm contains assumptions that have been derived empirically from the global in-situ NASA bio-Optical Marine Algorithm Data set, NOMAD (Werdell & Bailey, 2005).

2.2. GW94-based algorithm assuming spatial homogeneity of $\rho_w(NIR)$ and $\rho_a(NIR)$ ratios ('SIMIL')

The SIMIL algorithm includes two alternative assumptions, one on the water optical properties, the NIR similarity spectrum assumption, and one on the atmosphere (Ruddick et al., 2000). The first assumption arises from the fact that the sea-water NIR spectrum shape is largely determined by pure water absorption, and hence invariant, while the magnitude of the signal is approximately proportional to the backscatter coefficient. Consequently, the ratio of any two NIR water leaving reflectances, α , is constant. For MODIS-Aqua images α is defined as:

$$\alpha = \frac{\rho_w(748)/t_v(748)}{\rho_w(869)/t_v(869)} = 1.945. \quad (4)$$

The second assumption is based on the fact that the atmosphere composition does not vary significantly in space and therefore the ratio of the aerosol reflectance $\rho_a(\lambda) + \rho_{ra}(\lambda)$, referred as $\epsilon(\lambda_1, \lambda_2)$, can be considered as spatially homogeneous over the area of interest. In clear waters, where $\rho_w(\lambda)$ in the NIR region of the spectrum is close to zero, $\rho_a(\lambda) + \rho_{ra}(\lambda)$ can be approximated by $\rho_{rc}(\lambda)$. Accordingly, over clear water pixels $\epsilon(748, 869)$ is calculated as follows:

$$\epsilon(748, 869) = \frac{\rho_{rc}(748)}{\rho_{rc}(869)}. \quad (5)$$

Assuming spatial homogeneity of the atmosphere over the area of interest allows to estimate $\epsilon(748,869)$ over a clear water sub-scene and to use it for the atmospheric correction of the entire image. A clear water sub-scene is defined as any pixel within the image for which $L_{rc}(\lambda)$ is non-negative and below $0.015 \text{ mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$ at 748 and 869 nm.

The SIMIL algorithm is a three-step process. First the atmospheric correction is done with the initial black pixel assumption (i.e. GW94) in order to retrieve the scatter-plot $\rho_{rc}(748)$ versus $\rho_{rc}(869)$ for the region of interest and subsequently to estimate $\epsilon(748,869)$ over the clear water sub-scene. Next, using Eqs. (4) and (5), $\rho_a(\lambda)$ is estimated at each wavelength taking into account the non-zero water leaving reflectance and allowing to determine the appropriate aerosol model for the region of interest. Ultimately, the initial atmospheric correction algorithm of GW94 is run again but forced with the previously selected aerosol model.

2.3. GW94-based algorithm using the SWIR bands in very turbid waters (NIR-SWIR)

The NIR-SWIR approach combines the STD algorithm and the GW94 approach using the SWIR bands depending on the turbidity (Wang & Shi, 2007; Wang et al., 2009). This approach assumes that ocean water absorbs strongly in the SWIR part of the spectrum and that the contribution of the in-water constituents remains null. Accordingly, in highly turbid waters, $\rho_w(\text{SWIR})$ can be safely considered as null and the SWIR bands at 1240 and 2130 nm are used instead of the NIR bands for aerosol model selection. Highly turbid waters are defined based on a turbidity index computed from $\rho_{rc}(\lambda)$ at 748 and 1240 nm (Shi & Wang, 2007). For any pixel presenting a turbidity index above 1.05, the SWIR atmospheric correction approach is applied. The refinement of Wang et al. (2009) includes a second turbidity criterion. If, after a first processing with the SWIR aerosol model selection, the estimated $L_{wn}(869)$ is below $0.08 \text{ mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$ and the estimated Chl-a value lower or equal to 1 mg l^{-1} , the atmospheric correction is reset to the STD atmospheric correction method. Improvements in the retrievals of $\rho_w(\lambda)$ have been demonstrated when using the SWIR algorithm in extremely turbid waters (e.g., Dogliotti et al., 2011; Wang & Shi, 2007; Wang et al., 2009).

2.4. Direct inversion approach by neural network ('NN')

The MODIS NN algorithm used in this study is based on an inverse modeling of extensive radiative transfer (RT) simulations within a coupled ocean-atmosphere system, similar to an approach previously implemented for MERIS (Schroeder et al., 2007). This method directly inverts $\rho_t(\lambda)$ into remote sensing reflectance, $R_{rs}(\lambda)$, defined as:

$$R_{rs}(\lambda) = \frac{\rho_w(\lambda) \cos \theta_0}{\pi} \quad (6)$$

The direct inversion takes into account the spectral information at each wavelength simultaneously and not only in the NIR region of the spectrum like the previously described algorithms. Consequently, the NN method does not attempt to decouple oceanic and atmospheric signals to perform the atmospheric correction (see Eq. (1)). Another distinct difference is that the forward model parameterization of the NN includes only 8 aerosol models — compared to 80 for the three other algorithms. The RT simulations consist of reflectance data at top of atmosphere (TOA) and at mean sea level (MSL) and cover a large range of oceanic and atmospheric conditions for a variety of different sun and observing geometries. Specifically, the optically active water constituents cover the following concentration ranges: $0.05\text{--}50 \text{ mg l}^{-1}$ for Chl-a, $0.05\text{--}50 \text{ g l}^{-1}$ for SPM and $0.001\text{--}1 \text{ m}^{-1}$ absorption of CDOM at 443 nm. The simulated data were used to adapt the inverse model during a supervised learning procedure. The NN inverse model consists

of a multilayer perceptron that learned the functional relationship between TOA and MSL reflectance. More details about the forward model parameterization can be found in (Schroeder et al., 2007). When applied to MODIS data the algorithm requires the TOA reflectance at all ocean color bands as input with further information about the sun and observing geometry as well as the surface pressure. In addition, MODIS Level-2 flags are used to mask pixels affected by land, high glint and clouds. The associated NN outputs consist of MSL R_{rs} at eight MODIS bands up to 748 nm including spectral aerosol optical thickness at four AERONET bands (440, 550, 670 and 870 nm). Additional NN flags are used for further quality control of the outputs that check for out-of-scope conditions compared to the simulated data ranges.

3. Data and methods

3.1. In-situ measurements

The validation is carried out with the AERONET-Ocean Color network (AERONET-OC) data (Zibordi et al., 2006a, 2009b) and field data from 4 cruises coordinated by the Laboratoire d'Océanographie et Géosciences (LOG). The latter includes a large diversity of water types from moderately turbid waters in the North East Atlantic to very turbid waters in the coastal area of French Guiana and North Sea-English Channel. A detailed description of the sampling protocols and instruments are described in Lubac and Loisel (2007) and Nechad et al. (2010) for the North Sea-English Channel dataset and in Loisel et al. (2009) and Vantrepotte et al. (2011) for French Guiana.

AERONET-OC data provides $L_{wn}(\lambda)$ data in the 412–1020 nm spectral range (Zibordi et al., 2009a, 2009b) and the aerosol optical thickness τ from which we retrieved the Ångström coefficient, $\alpha(531,869)$. The AERONET-OC data used in the present research were taken at one site located in the Adriatic Sea (AAOT), two in the Baltic Sea (HLT and GDLT) and two along the East Coast of the USA (MVCO and COVE). The deployment sites and the number of observations for each site are given in Table 1. More details about instruments, station properties and particularities can be found in, among others, Berthon and Zibordi (2004), Zibordi et al. (2006b), Mélin et al. (2007), Feng et al. (2008), Zibordi et al. (2009a, 2009b).

3.2. Satellite data

The MODIS-Aqua instrument has been designed with 9 bands in the visible and NIR from 412 to 869 nm as well as with 3 bands in the SWIR (1240, 1640 and 2130 nm). In the present paper, the data analysis focuses on L_{wn} at 412, 443, 488, 531, 547 and 667 nm. The SWIR bands at 1240 and 2130 nm are used for the SWIR aerosol model selection approach in the NIR-SWIR algorithm.

MODIS-Aqua images were processed from L1A to L1B and subsequently from L1B to L2 for the three GW94-based atmospheric correction algorithms using the SeaWiFS Data Analysis System version 6.1 (SeaDAS 6.1, OBPG <http://oceancolor.gsfc.nasa.gov/>). In contrast to the other algorithms, the NN algorithm does only provide the $R_{rs}(\lambda)$ as output product. Therefore $R_{rs}(\lambda)$ were converted to $L_{wn}(\lambda)$ as described previously in Section 2.4, with F0 being the full band pass extra-terrestrial solar irradiance. Since the $R_{rs}(\lambda)$ outputs from the NN algorithm are not corrected for bi-directional effects, the AERONET-OC data and $L_{wn}(\lambda)$ retrievals from the STD, SIMIL and SWIR algorithms used in this study were not corrected neither. For the GW94-based algorithms, pixels with excessive cloud cover, large solar ($> 70^\circ$) and sensor ($> 60^\circ$) zenith angles, sun glint contamination or where sensors are saturating, were masked out. For the NN algorithm, pixels flagged as cloud and high glint or contaminated by solar and sensor zenith angles above 75.71° and 52.84° , respectively, were not processed.

Table 1
In-situ data location, description and distance from land.

	Lat	Lon	Distance from land (km)	Number of obs.	Period of obs.
AAOT (Adriatic Sea)	43.31° N	12.51° E	15	4348	2002–2007
GDLT (Baltic Proper)	58.59° N	17.467° E	16.5	924	2005–2009
HLT (Gulf of Finland)	59.95° N	24.92° E	22.22	665	2006–2009
MVCO (U.S. Atlantic Coast)	41.32° N	70.57° W	3.7	973	2004–2005
COVE (U.S. Atlantic Coast)	36.90° N	75.71° W	24	450	2006–2009
Belcolor (2009)	51.17–51.92° N	1.36–3.73° W	–	30	2009/06/18–2009/09/15
Belgica (2009)	42.47–51.39° N	9.46° W–2.81° E	–	14	2009/05/30–2009/06/07
French Guiana (2009)	4.15–5.98° N	53.95–51.29° W	–	30	2009/10/12–2009/10/17

Ultimately, as the spectral channels between the MODIS-Aqua images and the in-situ data differ slightly, a linear interpolation is performed to adjust the in-situ channels to the MODIS-Aqua channels.

3.3. Match-up exercise

The match-up exercise was a similar version of the approach described in Bailey and Werdell (2006) and Jamet et al. (2011). For each match-up pair, the median and standard deviation of the MODIS estimated $L_{wn}(\lambda)$, $R_{rs}(\lambda)$, $\alpha(531,869)$ and $\tau(869)$ were calculated over a 3 by 3 pixel window around the station. The median value was preferred to the average value as the former was less sensitive to large outliers and the 3 by 3 pixel window was used since the MODIS-Aqua images have geolocation accuracy of about 150 m at satellite nadir (Wolfe et al., 2002). For the MVCO station, the pixel window size was reduced to 2 by 3 allowing to avoid the adjacency effects from nearby land lying only 3 km North of the station (Feng et al., 2008; Jamet et al., 2011). Prior to data analysis, five exclusion criteria were applied ensuring that the MODIS-Aqua data did effectively represent the in-situ measurements. These were the following:

1. When the time difference between the satellite overpass and the in-situ measurement exceeded 2 hours the match-up pair was excluded. If redundant data remained, the match-up with the smallest time difference was retained.
2. Match-up pairs presenting in-situ $L_{wn}(667)$ below $0.183 \text{ mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$ were not retained. This threshold coincides with the turbid flag within SeaDAS and thus ensures that turbid waters only were retained (Robinson et al., 2003).
3. Match-up pairs were excluded when in a 3 by 3 pixel window (2 by 3 for the MVCO station) more than 3 pixels were flagged.
4. Spatial homogeneity within the pixel window was imposed. Match-up pairs presenting a MODIS estimated standard deviation for $L_{wn}(547)$ above 20% of the mean value were excluded.
5. For the NN algorithm, match-up pairs were rejected for further processing when flagged as out of scope compared to the simulated data ranges. This flag is raised if either the TOA radiances, the geometry or the surface pressure become out of range.

A time difference of two hours between the satellite overpass and the in-situ measurement greatly reduced the number of match-up pairs. However, a small time difference was preferred since larger time differences do not ensure that the MODIS estimated $L_{wn}(\lambda)$ actually corresponds to the in-situ measurement. This is particularly true in coastal regions where the spatial and temporal variability in water masses is important due to the high tides and rapid currents

(e.g. North Sea and English Channel). Therefore, we also investigated the errors on the retrieved $L_{wn}(\lambda)$ when considering only match-ups with less than 1 hour time difference between in-situ and satellite estimated $L_{wn}(\lambda)$.

The homogeneity criterion was applied on $L_{wn}(547)$ (as done by Zibordi et al. (2009a)) rather than on $\tau(869)$ as the former is more restrictive. The spatial variability in $L_{wn}(\lambda)$ is indeed much shorter compared to the spatial variability in atmospheric composition, particularly in turbid waters (Feng et al., 2008; Schroeder et al., 2007). Moreover, Zibordi et al. (2009a) positively tested the appropriateness of the 20% spatial homogeneity criterion at $L_{wn}(547)$. Nevertheless, in order to verify if these observations were valid with our dataset we also calculated the number of excluded match-up pairs when using the homogeneity criterion on $\tau(869)$.

Besides a visual inspection on scatter-plots, the algorithm performance was compared based on six statistical values: the percent bias, the percent relative error (RE), the root mean square error (RMSE), the spearman correlation (R) and the slope and intercept of the linear regression between in-situ data and MODIS $L_{wn}(\lambda)$ estimates. The bias, RE and RMSE are calculated as follows:

$$\text{BIAS} = \frac{1}{N} \sum \left(100 \frac{L_{wn}^{est}(\lambda) - L_{wn}^{obs}(\lambda)}{L_{wn}^{obs}(\lambda)} \right), \quad (7)$$

$$\text{RE} = \frac{1}{N} \sum \left(100 \frac{|L_{wn}^{est}(\lambda) - L_{wn}^{obs}(\lambda)|}{L_{wn}^{obs}(\lambda)} \right), \quad (8)$$

$$\text{RMSE} = \sqrt{\left(\frac{\sum (L_{wn}^{est}(\lambda) - L_{wn}^{obs}(\lambda))^2}{N} \right)}, \quad (9)$$

where L_{wn}^{obs} and L_{wn}^{est} are the in-situ and the MODIS estimated $L_{wn}(\lambda)$, respectively, and N is the number of observations.

Table 2
Excluded and total number of match-up pairs per algorithm. Values in parentheses indicate percent values.

	STD	SIMIL	NN	SWIR
Non-turbid	1135 (58)	1135 (58)	1135 (58)	1135 (58)
<6 pixels	860 (44)	862 (44)	922 (47)	880 (45)
Non-uniform (nLw(547))	155 (8)	121 (6)	100 (5)	337 (17)
Non-uniform ($\tau(869)$)	111 (6)	107 (5)	149 (8)	111 (6)
Flagged NN	–	–	370 (19)	–
Total	364 (18)	386 (20)	282 (14)	324 (16)

Table 3
Number of match-up pairs per algorithm after data filtration with the exclusion criteria.

	Belcolor	Belgica	Cove	GDLT	French Guiana	HLT	MVCO	AAOT
STD	2	1	23	54	2	25	54	203
NN	4	0	16	30	1	22	45	164
SIMIL	3	1	23	55	2	27	55	220
SWIR	2	1	21	38	2	21	45	194

3.4. Data classification

In order to understand the behavior of the algorithms as a function of the water type, we used a classification scheme developed by Vantrepotte et al. (2012) including four homogeneous optical classes over contrasted coastal areas. Their classification approach is based on a previous study of Lubac and Loisel (2007) and consists to cluster

in-situ $L_{wn}(\lambda)$ spectra normalized by the area below the curve. This normalization allows to classify the spectra based on their shape rather than on their magnitude. The former is indeed related to the nature of the optically significant water constituents rather than on their relative concentrations (Lubac et al., 2008; Mélin et al., 2011; Vantrepotte et al., 2012). In summary, Vantrepotte et al. (2012) defined two optically mixed classes of water masses mainly determined by detrital and mineral material (Class 1) or phytoplankton (Class 2), one class presenting water masses optically influenced by high loads of particulate matter mainly from mineral origin (Class 3), and a last class optically impacted by high concentrations of CDOM and phytoplankton (Class 4). More details about the classification approach and classes properties can be found in Vantrepotte et al. (2012).

The in-situ $L_{wn}(\lambda)$ spectra were assigned to one of the four classes based on the novelty detection method as described in D'Alimonte et al. (2003); Mélin et al. (2011); Vantrepotte et al. (2012).

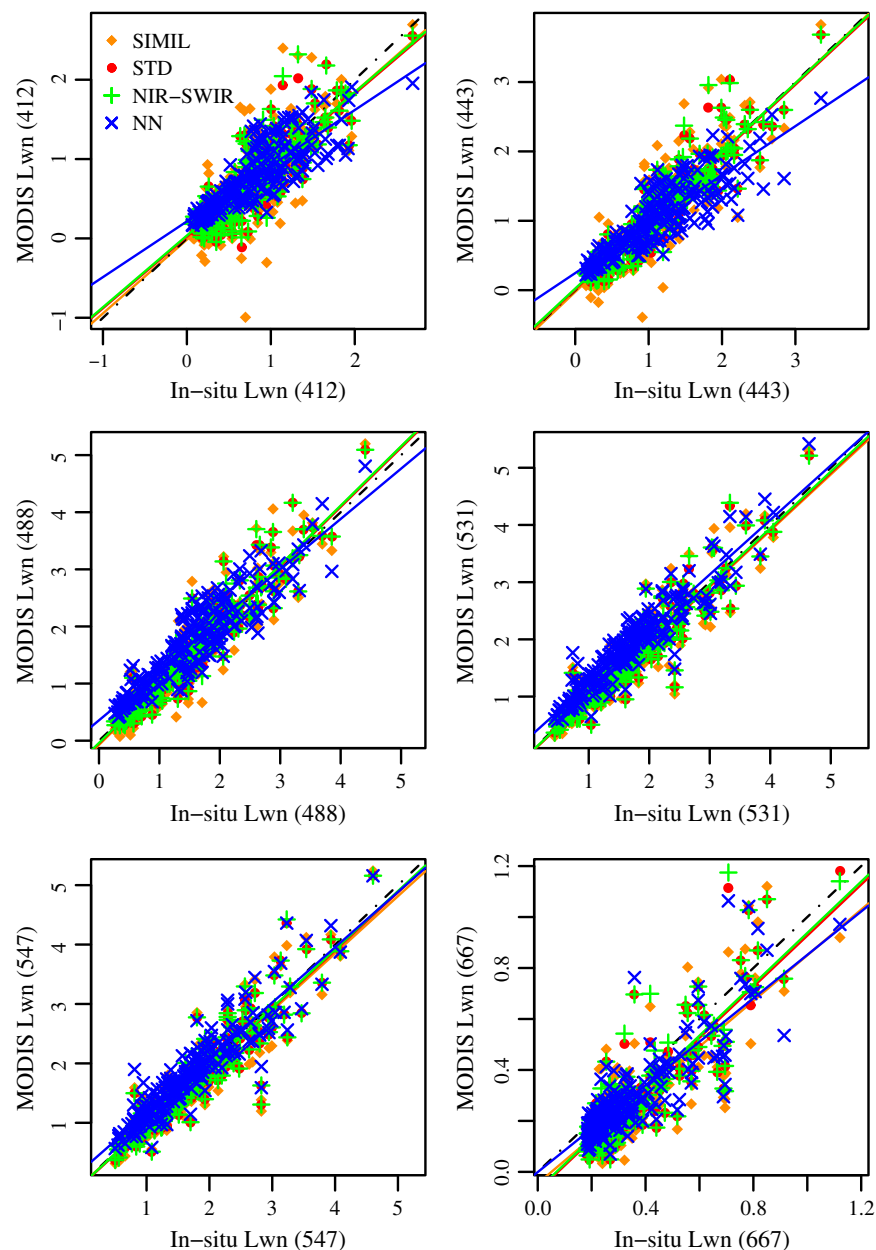


Fig. 1. MODIS-Aqua estimated $L_{wn}(\lambda)$ ($\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$) versus in-situ $L_{wn}(\lambda)$ ($\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$) at six different wavelengths, 412, 433, 488, 531, 547 and 667 nm. The black dot-dashed line represents the 1:1 line.

4. Results and discussion

4.1. Match-up selection

In total, 1973 match-up pairs, with ± 2 hour time difference between in-situ and satellite $L_{wn}(\lambda)$, have been retained. The match-up dataset was further filtered with the exclusion criteria described in Section 3.3. Table 2 shows the number of match-up pairs per algorithm that did not pass the exclusion criteria out of the 1973 match-ups. Note that there is an overlap between the different exclusion criteria. For instance, for the GW94-based and the NN algorithms about 40 and 48% of the 1135 non-turbid match-ups, respectively, were also excluded because their pixel window showed less than six valid pixels. The bottom line in Table 2 indicates the remaining number of match-up pairs per algorithm. Compared to the other algorithms, the NN algorithm generated more match-ups with less than 6 valid pixels while the NIR-SWIR algorithm was more affected by the spatial homogeneity criterion on $L_{wn}(547)$. A possible explanation for the latter is, as observed by Werdell et al. (2010), the increasing dynamic range in $L_{wn}(\lambda)$ due to the switch between the NIR and SWIR aerosol model selection approach within a single window. For comparison we also checked the number of match-up pairs that should be excluded if the homogeneity criterion was applied on $\tau(869)$. As explained previously, except for the NN algorithm, this homogeneity criterion was less severe (Table 2).

The turbidity criterion appeared to be the most restrictive with more than 50% of the initial number of match-ups excluded from the data analysis. Overall, the SIMIL algorithm showed the largest number of match-up pairs followed by the STD algorithm while the NN algorithm showed the lowest total number of match-ups. However, the latter is filtered by an additional exclusion criteria ('flagged NN' in Table 2). Indeed, out of the 370 flagged NN match-up pairs, 133 match-ups were not excluded by any other exclusion criteria. While the NN algorithm presents the advantage to retrieve $L_{wn}(\lambda)$ at all wavelengths simultaneously and to couple atmospheric and bio-optic model, it presents the disadvantage to be constructed by training and hence to be dependent on the training dataset. If a given water mass is not represented in the training dataset, the algorithm might fail. Moreover, it significantly reduces the number of match-up pairs when $L_t(\lambda)$ is out of the simulated range. Hence, the NN algorithm could be further improved by using a training dataset including more diverse water types. Without excluding the match-ups with inputs out-of-range, the NN algorithm presents the largest amount of match-up pairs (i.e. 415).

Table 3 shows the number of match-ups per station and per algorithm. A significant number of match-up pairs from the North Sea, English Channel and French Guiana were flagged as cloud or presented pixels where the sensor saturated explaining the low number of match-up pairs for these regions. For cross comparison, we imposed that the number of values for each algorithm matched. This means that the data presenting a valid match-up for the 4 algorithms only were retained reducing the number of total match-ups to 211 (2, 1, 14, 21, 18, 32, and 123 match-up pairs for BelColor, French Guiana, COVE, GDLT, HLT, MVCO and AAOT stations, respectively).

4.2. Global $L_{wn}(\lambda)$ validation

Fig. 1 displays the scatter-plots of the MODIS-Aqua estimated $L_{wn}(\lambda)$ versus the measured in-situ $L_{wn}(\lambda)$. Table 4 shows the corresponding statistical performance of each algorithm and per wavelength as well as the number of negative $L_{wn}(\lambda)$ retrievals. When compared to the 1:1 line, Fig. 1 shows good agreements between the MODIS-Aqua estimated $L_{wn}(\lambda)$ and the in-situ measured $L_{wn}(\lambda)$. This is particularly true for the GW94-based algorithms at 443, 488 and 547 nm. At 412 and 667 nm the band of data are more dispersed.

As indicated in Table 4, the SIMIL algorithm retrieved the largest number of negative $L_{wn}(\lambda)$ in the blue (15 and 4 at 412 and 443 nm,

Table 4

Statistical results for the retrieved values of $L_{wn}(\lambda)$ obtained with the SIMIL, STD, NIR-SWIR and NN algorithms and number of negative values for $L_{wn}(\lambda)$: R (dimensionless), Intercept (dimensionless), Slope (dimensionless), RE (%), RMSE ($\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$) and Bias (%). For each wavelength, the best statistical performance is rendered in bold.

	R	Intercept	Slope	RE	RMSE	Bias	$L_{wn} < 0$
<i>nLw(412)</i>							
SIMIL	0.75	−0.015	0.93	45	0.381	−11	15
STD	0.86	0.021	0.90	35	0.262	−2	2
NIR-SWIR	0.84	0.036	0.90	36	0.271	1	1
NN	0.81	0.214	0.70	31	0.270	11	0
<i>nLw(443)</i>							
SIMIL	0.86	−0.025	1.00	30	0.357	−4	4
STD	0.93	0.009	0.98	21	0.241	1	0
NIR-SWIR	0.92	0.023	0.99	21	0.254	3	0
NN	0.86	0.246	0.71	24	0.323	3	0
<i>nLw(488)</i>							
SIMIL	0.92	−0.075	1.04	20	0.351	−3	0
STD	0.95	−0.052	1.04	13	0.263	−0.2	0
NIR-SWIR	0.95	−0.04	1.04	13	0.274	1	0
NN	0.91	0.357	0.88	27	0.385	20	0
<i>nLw(531)</i>							
SIMIL	0.92	−0.022	0.98	14	0.32	−4	0
STD	0.95	−0.024	0.99	11	0.269	−3	0
NIR-SWIR	0.94	−0.015	0.99	11	0.273	−2	0
NN	0.94	0.273	0.954	19	0.338	16	0
<i>nLw(547)</i>							
SIMIL	0.91	−0.003	0.96	14	0.329	−4	0
STD	0.93	−0.02	0.98	11	0.286	−3	0
NIR-SWIR	0.93	−0.011	0.98	11	0.288	−2	0
NN	0.92	0.232	0.93	15	0.315	10	0
<i>nLw(667)</i>							
SIMIL	0.81	−0.046	0.90	33	0.136	−26	0
STD	0.87	−0.075	1.00	30	0.12	−25	0
NIR-SWIR	0.86	−0.07	1.00	30	0.121	−23	0
NN	0.85	−0.001	0.85	22	0.107	−15	0

respectively) followed by the STD algorithm and the NIR-SWIR algorithm (2 and 1 at 412 nm, respectively). The NN algorithm did not show any negative $L_{wn}(\lambda)$ retrievals as it outputs the reflectance on a logarithmic scale. The SIMIL algorithm tended to underestimate $L_{wn}(\lambda)$ at all wavelengths (negative bias ranging between -3 and -26%). At 412 nm, about 60% of the MODIS retrieved L_{wn} are underestimated with the SIMIL and STD algorithms, 57% with the NIR-SWIR algorithm and 45% with the NN algorithm.

For the STD and NIR-SWIR algorithms, the best results were obtained at 488, 531 and 547 nm with R coefficients above 0.9, biases below -3% and RE around 12%. The SIMIL algorithm showed RE ranging between 14 and 45% (with maxima at 412 and 667 nm) and biases between -26 and -4% . Such as observed by Jamet et al. (2011), who compared three SeaWiFS algorithms and among others

Table 5

Statistical results for the retrieved aerosol products obtained with the SIMIL, STD, NIR-SWIR and NN algorithms: RE (%), RMSE (dimensionless) and Bias (%).

	R	Intercept	Slope	RE	RMSE	Bias
<i>$\alpha(531, 869)$</i>						
SIMIL	0.319	0.939	0.289	52	0.586	6
STD	0.403	0.842	0.374	46	0.553	9
NIR-SWIR	0.434	0.764	0.421	44	0.550	7
<i>$\tau(869)$</i>						
SIMIL	0.607	0.036	0.680	75	0.042	51
STD	0.606	0.035	0.682	74	0.042	49
NIR-SWIR	0.613	0.035	0.674	72	0.040	47
NN	0.579	0.053	0.919	154	0.070	109

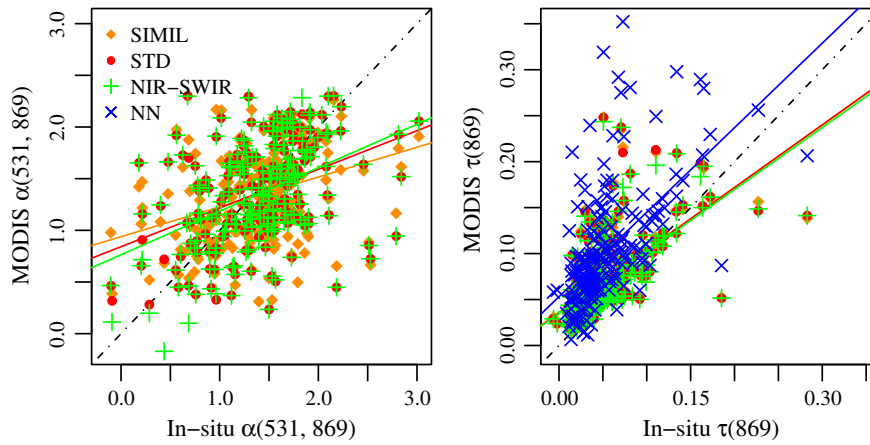


Fig. 2. MODIS-Aqua estimated $\alpha(531,869)$ (left) and $\tau(869)$ (right) versus in-situ aerosol products.

the STD and SIMIL algorithm, our results indicated least accurate $L_{wn}(\lambda)$ retrievals with the SIMIL algorithm compared to the STD algorithm. In contrast, [Jamet et al. \(2011\)](#) observed the largest errors at 670 nm (35 and 43% with the STD and SIMIL algorithms, respectively) while we observed larger errors at 412 nm (35 and 30% at 412 and 667 nm, respectively, for the STD, and 45 and 33% for the SIMIL algorithm).

The performances of the STD and NIR-SWIR algorithms remained very close. For 27% of the match-up pairs the NIR-SWIR algorithm used

the NIR approach for the aerosol model selection method and the difference between the STD and NIR-SWIR estimated $L_{wn}(\lambda)$ was below $0.01 \text{ mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$ for more than 82, 85 and 88% of the match-ups at 412, 547 and 667 nm, respectively. Accordingly, with our in-situ dataset, the NIR-SWIR algorithm did not show significant improvement compared to the STD algorithm. Modifying the turbidity index threshold, as defined by [Wang et al. \(2009\)](#) and/or the cloud albedo threshold or cloud scheme (e.g. using the MODIS 1240 nm band such as [Wang et al. \(2009\)](#)) may result in more pixels processed with the SWIR aerosol

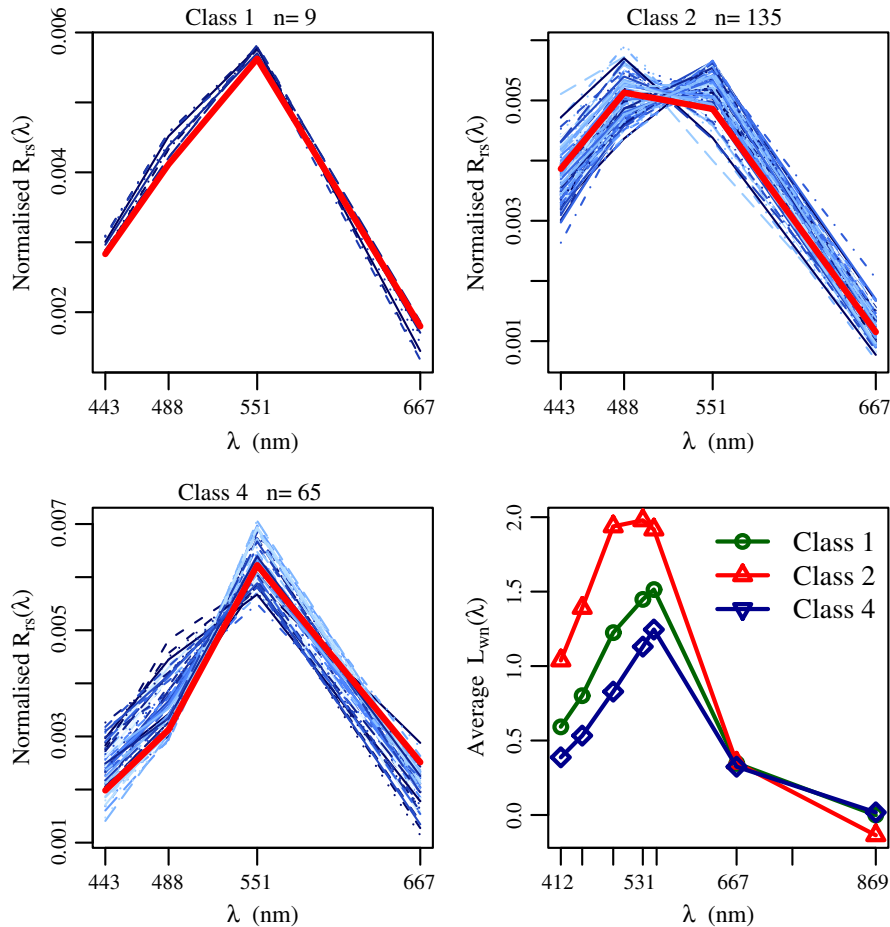


Fig. 3. Top row and bottom left plots: $R_{rs}(\lambda)$ normalized by the area below the curve assigned to each class together with the average normalized spectra given by [Vantrepotte et al. \(2012\)](#) (red plain line). Bottom right plot: average in-situ $L_{wn}(\lambda)$ ($\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$) spectra for the three classes.

Table 6Mean values and standard deviation of in-situ $L_{wn}(\lambda)$ per class ($\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$).

	$L_{wn}(412)$	$L_{wn}(443)$	$L_{wn}(488)$	$L_{wn}(531)$	$L_{wn}(547)$	$L_{wn}(667)$	$L_{wn}(748)$
Class 1	0.59 ± 0.22	0.80 ± 0.28	1.22 ± 0.43	1.45 ± 0.49	1.51 ± 0.53	0.35 ± 0.12	0.16 ± 0.05
Class 2	1.04 ± 0.39	1.39 ± 0.51	1.94 ± 0.67	1.98 ± 0.70	1.92 ± 0.70	0.35 ± 0.16	0.11 ± 0.47
Class 4	0.39 ± 0.25	0.53 ± 0.25	0.83 ± 0.50	1.13 ± 0.60	1.25 ± 0.65	0.32 ± 0.19	0.17 ± 0.08

model selection approach. However, further work should be done in order to verify if a more frequent switch towards the SWIR aerosol model selection will lead to improvement and not to additional noise in the aerosol model selection. Moreover, to ensure high quality match-ups for all algorithms the default cloud mask from MODIS-Aqua was preferred.

The NN algorithm revealed a degradation in accuracy in the green bands with an overestimation at 531 and 547 nm (+16 and +10% bias, respectively), a high RMSE (above $0.3 \text{ mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$) and a large RE (19 and 15% at 531 and 547 nm, respectively). However, at 412 and 667 nm, the NN algorithm performed the best with RE of 31 and 22% and biases of 11 and −15%, respectively. Improvement in $L_{wn}(\lambda)$ retrievals in the red with the NN algorithm was also observed by Schroeder et al. (2007) who validated MERIS $L_{wn}(\lambda)$ estimations. Similarly, Schroeder (2012, pers. comm., 28 Sep) reported a significant improvement in the blue and NIR spectral region with the NN algorithm from an unpublished match-up exercise with MODIS images similar to the present research.

According to the results above, it seems that higher $L_{wn}(\lambda)$ signals, observed mainly in the green region of the spectrum, tend to be better predicted by the GW94-based algorithms, while the lower $L_{wn}(\lambda)$ signals, observed in the blue and red region of the spectrum, are better predicted by the NN algorithm. This can be explained by the error function which minimizes the cost function on a logarithmic scale when training the NN algorithm. Hence, the cost function constrained the parameters of the neural network allowing larger errors for higher $L_{wn}(\lambda)$ signals and lower for lower $L_{wn}(\lambda)$ signals.

With 1 hour time threshold, the number of match-ups decreased from 211 to 179. The RE for all algorithms and at all wavelengths decreased on average by 3%. The largest decrease in RE was observed in the blue region of the spectrum (−4, −6, −9 and −9% for the SIMIL, STD, NIR-SWIR and NN algorithms, respectively). For the other wavelengths the differences in RE did not overpass 4%, except for the NN algorithm at 488, 531 and 547 nm (−6, 6 and 5%). Although better results were obtained with a shorter time interval between the satellite overpass and the in-situ measurements, the relative performances of the algorithms remained the same.

4.3. Global validation of aerosol products

The accuracy of the retrievals of $\alpha(531,869)$ and $\tau(869)$ are given in Table 5 and illustrated in Fig. 2. For $\alpha(531,869)$, RE varied between 44 and 52% and biases between 6 and 9%. For $\tau(869)$, the statistics were not as good with RE ranging between 72 and 154% and biases ranging from 47 to 109%. The aerosol product retrievals were the least accurate with the NN algorithm (RE and biases of 154 and 109% for $\tau(869)$ and 44 and 7% for $\alpha(531,869)$) and the most accurate with the NIR-SWIR algorithm (RE and biases of 72 and 47% for $\tau(869)$ and 52 and 6% for $\alpha(531,869)$). In order to investigate how errors on $\alpha(531,869)$ and $\tau(869)$ affected the $L_{wn}(\lambda)$ retrievals, we calculated the correlations

Table 7Mean values and standard deviation of $\tau(869)$ and $\alpha(531,869)$ per class (dimensionless).

	$\tau(869)$	$\alpha(531,869)$
Class 1	0.054 ± 0.050	1.39 ± 0.35
Class 2	0.055 ± 0.039	1.38 ± 0.57
Class 4	0.049 ± 0.037	1.41 ± 0.44

between $\tau(869)$ and $\alpha(531,869)$ and the absolute errors in $L_{wn}(412)$ (i.e. difference between MODIS estimations and in-situ observations). Such as Feng et al. (2008), we observed a systematic underestimation of $\alpha(531,869)$ with an overestimation of $\tau(869)$ for the three GW94-based algorithms (not shown here). However, the correlation coefficients between $\alpha(531,869)$ and $\tau(869)$ remained moderate (between −0.39 and −0.51 for the three algorithms). $\tau(869)$ and the absolute errors in $L_{wn}(412)$ were also negatively correlated for the SIMIL, STD and NIR-SWIR algorithms (−0.4, −0.35 and −0.27, respectively). Accordingly, with high aerosol loads, the GW94-based algorithms tended to underestimate the L_{wn} signal in the blue. For the NN algorithm high aerosol loads did not have any impact on the retrieval errors in $L_{wn}(\lambda)$. The aerosol products are indeed not used as input for the correction procedure and the NN algorithm is optimized only based on $R_{rs}(\lambda)$ data.

4.4. $L_{wn}(\lambda)$ validation per water type

Four wavelengths (443, 488, 547, and 667 nm) were used to classify the in-situ $L_{wn}(\lambda)$ spectra. The wavelengths at 412 and 531 nm were discarded for the same reasons mentioned by Vantrepotte et al. (2012), notably, because in coastal regions, disagreement is

Table 8Statistical results of the retrieved $L_{wn}(\lambda)$ for the SIMIL, STD, NIR-SWIR and NN algorithms: RE (%), RMSE ($\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$) and Bias (%). For each wavelength, the best statistical performance is rendered in bold.

	Class 1			Class 2			Class 4		
	RE	RMSE	Bias	RE	RMSE	Bias	RE	RMSE	Bias
<i>nLw(412)</i>									
SIMIL	44	0.291	−43	31	0.381	−5	73	0.397	−18
STD	51	0.290	−45	24	0.287	−8	53	0.201	14
NIR-SWIR	51	0.292	−44	24	0.297	−6	56	0.207	18
NN	27	0.236	5	24	0.309	−2	46	0.174	35
<i>nLw(443)</i>									
SIMIL	29	0.26	−26	21	0.367	1	50	0.355	−11
STD	34	0.286	−29	15	0.265	−1	31	0.177	9
NIR-SWIR	34	0.288	−28	15	0.279	1	32	0.191	11
NN	25	0.281	−3	20	0.374	−6	33	0.191	22
<i>nLw(488)</i>									
SIMIL	23	0.303	−22	14	0.357	2	30	0.349	−8
STD	25	0.313	−23	10	0.279	0	18	0.224	3
NIR-SWIR	25	0.310	−24	10	0.29	1	18	0.237	4
NN	21	0.348	5	19	0.416	11	44	0.32	40
<i>nLw(531)</i>									
SIMIL	20	0.352	−19	11	0.303	−2	20	0.352	−4
STD	22	0.361	−19	9	0.256	−3	14	0.283	1
NIR-SWIR	21	0.358	−19	9	0.262	−2	14	0.284	2
NN	18	0.291	−0.5	14	0.328	11	30	0.366	27
<i>nLw(547)</i>									
SIMIL	20	0.377	−19	11	0.303	−2	18	0.374	−5
STD	22	0.383	−19	10	0.265	−3	14	0.316	−1
NIR-SWIR	21	0.379	−19	9	0.269	−2	14	0.316	0
NN	16	0.303	−7	12	0.287	6	23	0.372	19
<i>nLw(667)</i>									
SIMIL	45	0.183	−45	32	0.132	−24	34	0.139	−26
STD	48	0.179	−44	32	0.12	−28	25	0.112	−16
NIR-SWIR	48	0.178	−43	31	0.121	−25	25	0.112	−15
NN	36	0.149	−36	22	0.108	−17	19	0.099	−7

often observed between in-situ and remote sensing $L_{wn}(412)$ and because the inclusion of $L_{wn}(531)$ did not improve the classification due to its low discrimination character. Out of the 211 spectra, two spectra remained unclassified, 9 spectra were assigned to the first class, 135 to the second and 65 to the fourth. None of our match-up pairs were assigned to Class 3.

The three first plots in Fig. 3 show the $R_{rs}(\lambda)$ of the in-situ data normalized by their area for each class. The class averaged normalized $R_{rs}(\lambda)$ calculated by Vantrepotte et al. (2012) are also indicated. The bottom right plot in Fig. 3 shows the average $L_{wn}(\lambda)$ spectra for the three classes obtained with our in-situ data. Table 6 shows the mean and standard deviation of the in-situ $L_{wn}(\lambda)$ for each class.

Class 1 contained six match-ups from the MVCO station, two match-ups from the AAOT station and one from the turbid waters of French Guiana. Class 2 contained almost all in-situ data from the AAOT station (112 out of 123) as well as some in-situ data from COVE and MVCO stations. The AAOT station presents large variabilities in bio-optical quantities as it is located between clear ocean waters and turbid coastal waters with a larger occurrence of open ocean water types (Berthon & Zibordi, 2004; Mélin et al., 2011; Zibordi et al., 2009b). Consequently, it is not surprising that most in-situ data from this station were allocated to Class 2 defined by Vantrepotte et al. (2012) as a mixed class encountered in transition zones between coastal turbid waters and open oceans.

Class 4 corresponds to water masses largely influenced by CDOM concentrations and phytoplankton blooms. All the classified in-situ data from the HLT and GDLT stations were assigned to Class 4. Indeed, as observed in previous studies (e.g., Darecki & Stramski, 2004; Zibordi et al., 2009a, 2009b), these two stations exhibit water masses fed by river discharges rich in humic substances. This results in high CDOM concentrations, an increase in primary production and subsequently in high Chl-a concentrations corresponding to the definition of Class 4. The domination of CDOM absorption can be noticed on the bottom right plot of Fig. 3 and in Table 6 where $L_{wn}(\lambda)$ in the blue is relatively lower ($0.39 \pm 0.25 \text{ mW} \cdot \text{cm}^{-2} \cdot \mu\text{m}^{-1} \cdot \text{sr}^{-1}$) compared to the other two classes. The 3 classes presented rather continental aerosol types with average values of $\alpha(531,869)$ around 1.4 (standard deviations ± 0.5) (Table 7).

Table 8 gives the RMSE, biases and RE per class and per algorithm. For a better comparison, the statistical parameters are also visually represented in Figs. 4 and 5.

The algorithm with the best performance for Class 1 was the NN algorithm with a RE ranging from 16 to 36% and biases from -36 to 5% (Table 8). As shown in Fig. 5, at all wavelengths, the NN algorithms showed lower RE for Class 1 compared to the three other algorithms. For this class, the NIR-SWIR algorithm only switched for one match-up pair towards the SWIR aerosol model selection approach. When comparing only the GW94-based algorithms, the SIMIL algorithm

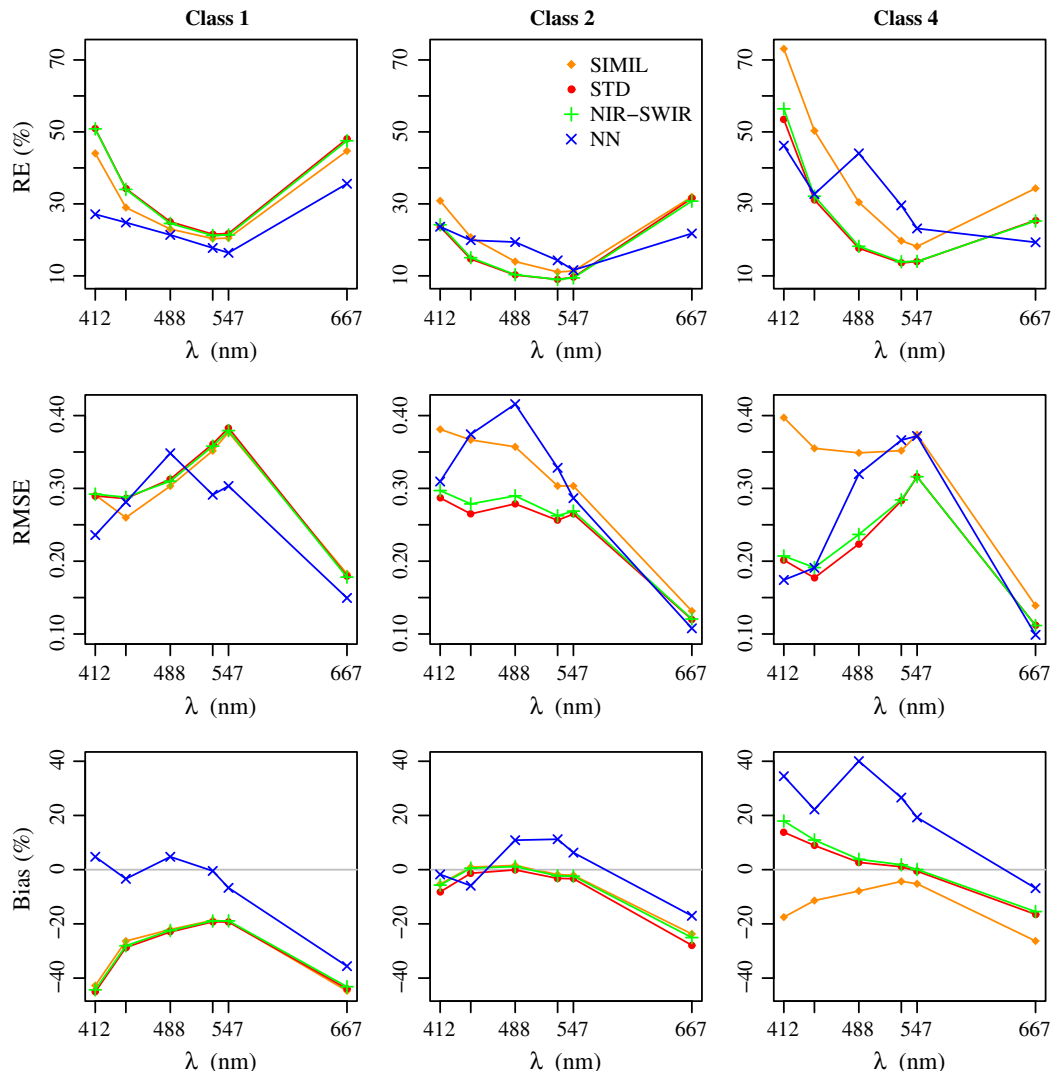


Fig. 4. RMSE ($\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$), biases (%) and RE (%) per class (from left to right column: Class 1, 2 and 4) and per algorithm as a function of the wavelength.

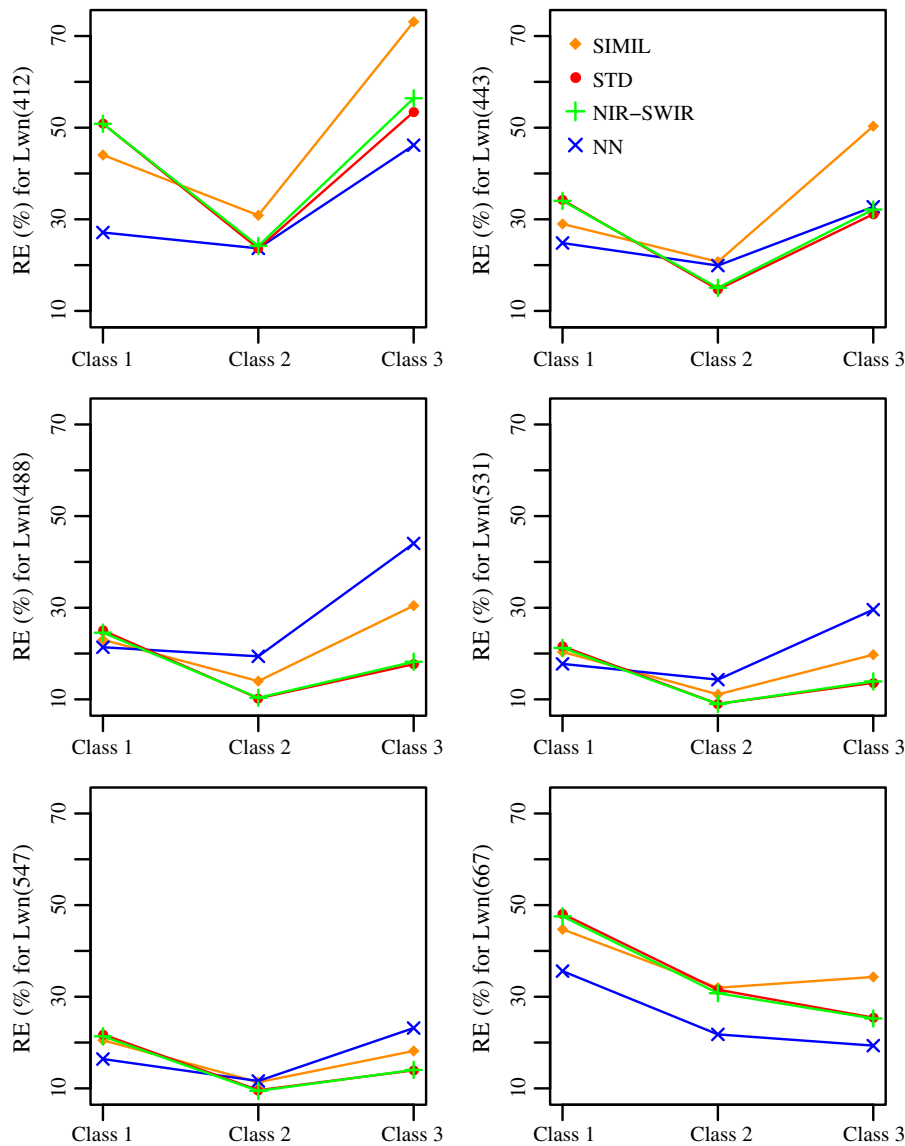


Fig. 5. RE (%) as a function of the water type (Classes 1, 2 and 4) per wavelength.

gave the most satisfying results for this water type. This can be explained by the validity range of the NIR similarity spectrum assumption. Indeed, as observed by several studies (e.g., Doron et al., 2011; Neukermans et al., 2009; Ruddick et al., 2006; Shi & Wang, 2009), the NIR similarity spectrum assumption is only valid for a given range in $L_{wn}(NIR)$. For very clear and extremely turbid waters, Ruddick et al. (2006) mentioned that this assumption should be adjusted. In the present research we found that the linear correlation coefficient between the in-situ $L_{wn}(748)$ and $L_{wn}(869)$ data was about 83% for Class 1, while it was only about 68 and 63% for Class 2 and 4, respectively.

For Class 2, the agreement with the in-situ data was better compared to the other two classes and in particular in the blue and green region of the spectrum for which the biases of the GW94-based algorithms were closer to 0%. The STD algorithm was the most accurate algorithm for this class (RE ranging between 9 and 32% with a maximum at 667 nm and biases ranging between 0 and –28%). Except in the red region of the spectrum, the STD algorithm gave the lowest RE at all wavelengths (Fig. 5). A possible explanation for the good performance of the STD algorithm for Class 2 is the training dataset used to develop the bio-optical model included in the algorithm, notably, the NOMAD dataset. The latter covers indeed mainly mesotrophic to oligotrophic waters (Mélin et al., 2011) which corresponds to the definition of Class 2. Nevertheless, in

the red region of the spectrum, both the STD and NIR-SWIR algorithms highly underestimate $L_{wn}(\lambda)$ with RE of 32 and 31% and biases of –25 and –28%, respectively. The statistics of Class 2 are in agreement with the errors observed by Mélin et al. (2011) who validated the MODIS estimated $L_{wn}(\lambda)$ processed with the STD algorithm at the AAOT station. They observed RE varying from 11 to 40% with maxima at 412 and 667 nm. At the same AERONET-OC station, Zibordi et al. (2009b) also observed an underestimation of the MODIS estimated $L_{wn}(667)$ with the STD algorithm. In this region of the spectrum and for the three classes, the NN algorithm performed better than the GW94-based algorithms (Fig. 5).

While the NN algorithm showed a bias close to 0% in the blue region of the spectrum for Classes 1 and 2, it largely overestimated $L_{wn}(412)$ for Class 4 (bias of 34%). The STD and NIR-SWIR algorithms showed a positive bias in the blue too. This is in agreement with previous observations from Zibordi et al. (2009a) who observed biases of 25 and 18% with the STD algorithm at the HLT and GDLT stations, respectively. In contrast to the other algorithms, the SIMIL algorithm underestimated $L_{wn}(\lambda)$ at all wavelengths for Class 4 (biases ranging from –5 to –26%). A possible explanation is the lack of clear water pixels in the MODIS images. Indeed, for more than 90% of the cases in Class 4, the SIMIL algorithm used a climatological $c(748,869)$ to solve the atmospheric correction.

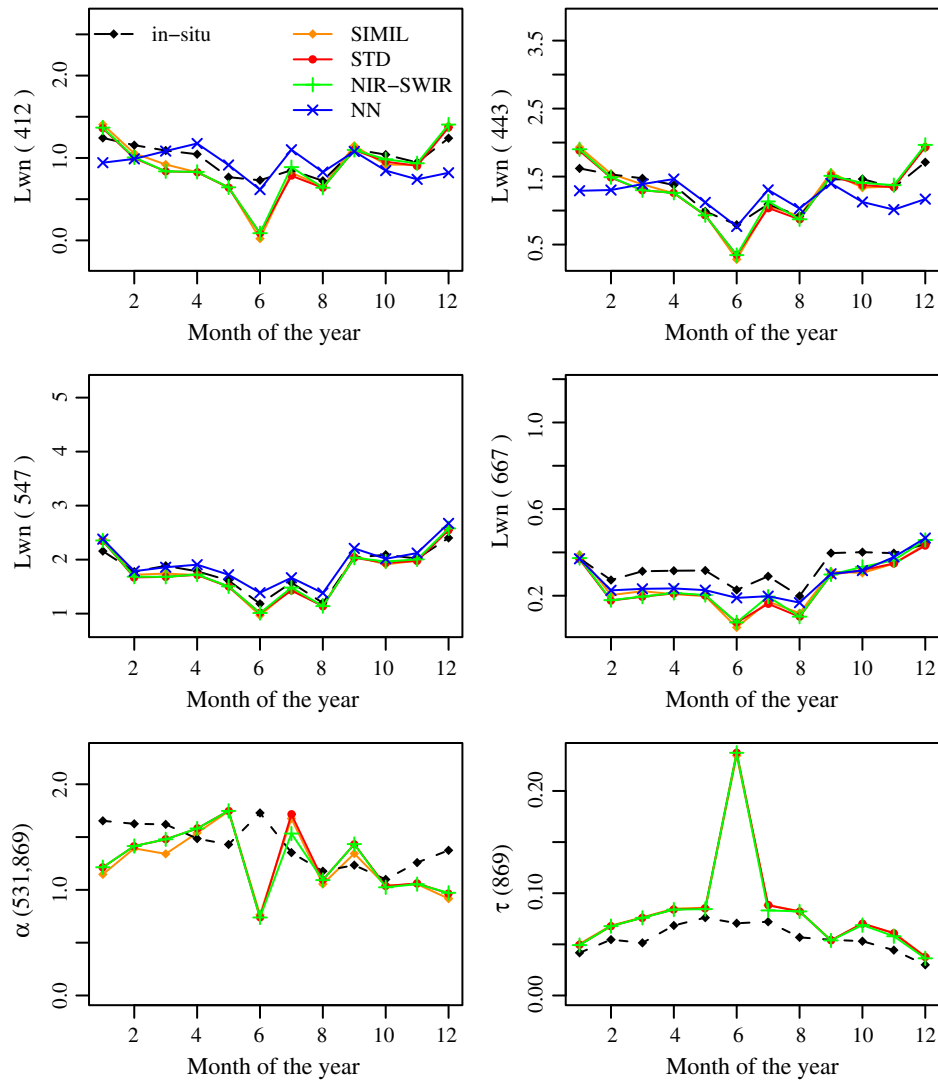


Fig. 6. Annual monthly averages of L_{wn} ($\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$) at four different wavelengths (412, 443, 547 and 667 nm) and two atmospheric parameters ($\alpha(531,869)$ and $\tau(869)$) for Class 2.

For Classes 1 and 2 a climatological $\epsilon(748,869)$ was used for 77 and 35% of the cases, respectively. The use of a climatological value for $\epsilon(748,869)$ might be less accurate and even inappropriate for some situations. As noticed by Jamet et al. (2011), an error in $\epsilon(748,869)$ has a significant impact on the retrievals of oceanic products and mainly on L_{wn} estimations at 412, 443 and 667 nm. These bands correspond to the one for which the SIMIL algorithm induced the largest RE, RMSE and negative biases (Table 8).

From Fig. 5 it seems that the errors in $L_{wn}(\lambda)$ at 531 and 547 nm are independent of the water type. Compared to the other wavelengths, at 531 and 547 nm the RE remained quite stable whatever water type was considered (Fig. 5). Accordingly, instead of forcing the algorithms with water type-function relationships to improve atmospheric correction, the algorithms might be constrained by spectral relationships valid for any water type. Therefore, spectral relationships including wavelengths at which the errors in $L_{wn}(\lambda)$ retrievals are independent of the water type (e.g. 531 and 547 nm) should be investigated.

4.5. Algorithm validation over time

Figs. 6 and 7 show the annual monthly averages of $L_{wn}(\lambda)$ at four different wavelengths and the two aerosol parameters for Classes 2 and 4. Time series for Class 1 are not discussed as the number of

observations per month is too low and three months only were available. For both classes, the $L_{wn}(\lambda)$ signal is higher and lower in the boreal winter and summer, respectively. This is the consequence of a larger input of sediments in the winter from nearby rivers due to more frequent rainfall events (Berthon & Zibordi, 2004; Mélin et al., 2011). During this period of the year the GW94-based algorithms performed better in the blue region of the spectrum, while in the summer, when the signal is lower, the NN algorithm performed better.

As remarked previously (in Section 4.3.), $\tau(869)$ is also mostly overestimated when $\alpha(531,869)$ is underestimated and an overestimation in $\tau(869)$ is often associated with an underestimation in $L_{wn}(\lambda)$ in the blue region of the spectrum. For both classes, a disagreement in $\tau(869)$ often led to erroneous $L_{wn}(\lambda)$ retrievals (e.g. in June on Fig. 6 and May on Fig. 7). For Class 4, the disparity between the algorithms in terms of $L_{wn}(\lambda)$ retrievals was mostly apparent in the blue and mainly in May and December when $\tau(869)$ was largely overestimated and $\alpha(531,869)$ underestimated. During these two months, the STD and NIR-SWIR algorithms performed better in the blue, while in the green and the red the SIMIL algorithm performed better.

5. Conclusion

Several algorithms have been developed in order to bypass the inappropriate application of the black pixel assumption in turbid

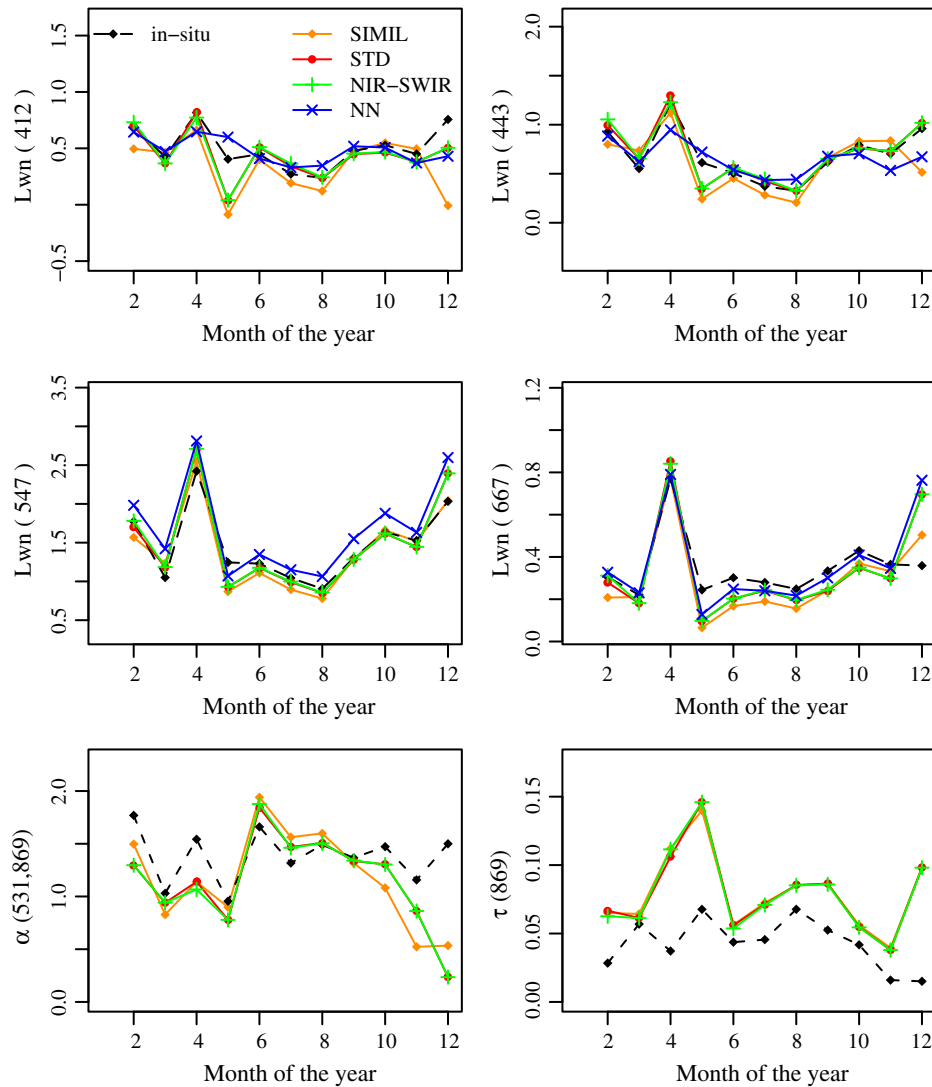


Fig. 7. Same as Fig. 6 for Class 4.

coastal waters. This research reported the accuracy of four MODIS-Aqua atmospheric correction algorithms in contrasted coastal waters using global coastal AERONET-OC data and cruise campaign measurements. The four algorithms were (1) the STD algorithm from NASA, (2) the SIMIL algorithm assuming spatial homogeneity within the area of interest of the reflectance of the atmosphere and the ocean in the NIR region of the spectrum, (3) the NIR-SWIR algorithm which, in highly turbid waters, switches towards the SWIR bands for the aerosol model selection, and, (4) the NN algorithm which applies a direct inversion of $L_t(\lambda)$. The former three algorithms are based on the GW94 atmospheric correction approach. Overall, the atmospheric correction algorithms performed well with RE ranging from 11 to 45% and biases from -25 to $+16\%$. For the GW94-based algorithms, the largest errors were encountered in the blue and red part of the spectrum (at 412 and 667 nm), confirming the challenge of using these bands for retrieving bio-optical parameters. In contrast, the NN algorithm performed slightly better in these bands. The SIMIL algorithm tended to underestimate $L_{wn}(\lambda)$, while the NN algorithm tended to overestimate it.

A closer look to the dataset conducted to an evaluation of the algorithms as a function of the water type. In-situ $L_{wn}(\lambda)$ spectra were classified according to a classification scheme developed by Vantrepotte et al. (2012) which focussed on coastal waters and distinguished the

classes based on the water type rather than on the water clarity. For water masses mainly influenced by detrital and mineral material (Class 1), the NN algorithm performed the best at all wavelengths followed by the SIMIL algorithm. For the two other classes, the STD algorithm seemed to perform better followed by the NIR-SWIR algorithm. The NN algorithm could be improved in order to increase the number of match-up pairs and to correct for the large overestimations of $L_{wn}(\lambda)$ in the green part of the spectrum. When considering only the GW94-based algorithms, the SIMIL algorithm performed slightly better for water masses mainly influenced by detrital and mineral material indicating that this algorithm is valuable for a specific range of turbidity only. The largest errors were encountered for water masses dominated by significant concentrations of phytoplankton and CDOM. Overall, the NN algorithm performed better where the GW94-based algorithms failed (mainly in the blue and red region of the spectrum and at all wavelengths for Class 1) and when considering all wavelengths and all water types the STD algorithm performed the best.

Atmospheric correction remains a challenging task and in particular in complex optical coastal waters. The present research provides an overview and validation of the already existing algorithms which is highly valuable for further improvement essential to meet the requirements of the ocean color end-users.

Acknowledgments

The authors would like to thank Dr. Vincent Vantrepotte for providing his classification method and Quinten Vanhellemont for his help with the match-up extraction. Also thanks go to the Ocean Biology Processing Group of NASA for the distribution of the MODIS-Aqua data and the development and support of the SeaDAS Software, as well as to the PIs and maintenance staffs of the AERONET-OC sites used in the present research. The anonymous reviewers are also acknowledged for their comments and suggestions. This work has been supported by the French Spatial Agency (CNES) through the TOSCA program and the “Ministère de l'Enseignement et de la Recherche Française” which provided a PhD scholarship. It has been performed in the frame of the GlobCoast project (www.foresea.fr/globcoast) which is funded by the Agence Nationale de la Recherche (ANR 2011 BS56 018 01). The GlobCoast project is affiliated to the LOICZ and AQUIMER project.

References

- Ahmad, Z., Franz, B. A., McClain, C. R., Kwiatkowska, E. J., Werdell, P. J., Shettle, E. P., et al. (2010). New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from the SeaWiFS and MODIS sensors over coastal regions and open oceans. *Applied Optics*, 49, 5545–5560.
- Bailey, S. W., Franz, B. A., & Werdell, P. J. (2010). Estimations of near-infrared water-leaving reflectance for satellite ocean color data processing. *Optics Express*, 18, 7521–7527.
- Bailey, S. W., & Werdell, P. J. (2006). A multi-sensor approach for the on-orbit validation of ocean color satellite data products. *Remote Sensing of Environment*, 102, 12–23.
- Berthon, J.-F., & Zibordi, G. (2004). Bio-optical relationships for the northern Adriatic Sea. *International Journal of Remote Sensing*, 25, 1527–1532.
- Cox, C., & Munk, W. (1954). Measurements of the roughness of the sea surface from photographs of the sun's glitter. *Journal of the Optical Society of America*, 44, 838–850.
- D'Alimonte, D., Mélin, F., Zibordi, G., & Berthon, J.-F. (2003). Use of the novelty detection technique to identify the range of applicability of the empirical ocean color algorithms. *IEEE Transactions on Geoscience and Remote Sensing*, 41, 2833–2843.
- Darecki, M., & Stramski, D. (2004). An evaluation of MODIS and SeaWiFS bio-optical algorithms in the Baltic Sea. *Remote Sensing of Environment*, 89, 326–350.
- Dogliotti, A. I., Ruddick, K., Nechad, B., & Lasta, C. (2011). Improving water reflectance retrieval from MODIS imagery in the highly turbid waters of La Plata River. *Proceedings of VI International Conference "Current Problems in Optics of Natural Waters"*, St. Petersburg, Russia: Publishing House "Nauka" of RAS.
- Doron, M., Bélanger, S., Duxoran, D., & Babin, M. (2011). Spectral variations in the near-infrared ocean reflectance. *Remote Sensing of Environment*, 115, 1617–1631.
- Feng, H., Vandemark, D., Campbell, J. W., & Holben, B. N. (2008). Evaluation of MODIS ocean color products at a northeast United States coast site near the Martha's Vineyard Coastal Observatory. *International of Remote Sensing*, 29, 4479–4497.
- Frouin, R., Schwindling, M., & Deschamps, P.-Y. (1996). Spectral reflectance of sea foam in the visible and near-infrared: In situ measurements and remote sensing implications. *Journal of Geophysical Research*, 101, 14361–14371.
- Gordon, H. R. (1987). Calibration requirements and methodology for remote sensors viewing the ocean in the visible. *Remote Sensing of Environment*, 22, 103–126.
- Gordon, H. R. (1997). Atmospheric correction of ocean color imagery in the Earth Observing system era. *Journal of Geophysical Research*, 102, 17081–17106.
- Gordon, H. R., & Wang, M. (1994). Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: A preliminary algorithm. *Applied Optics*, 33, 443–452.
- Jamet, C., Loisel, H., Kuchinke, C. P., Ruddick, K., Zibordi, G., & Feng, H. (2011). Comparison of three SeaWiFS atmospheric correction algorithms for turbid waters using AERONET-OC measurements. *Remote Sensing of Environment*, 115, 1955–1965.
- Kuchinke, C. P., Gordon, H. R., & Franz, B. A. (2009). *Spectral optimization for constituent retrieval in Case 2 waters I: Implementation and performance*, Vol. 113. (pp. 571–587).
- Loisel, H., Lubac, B., Dessailly, D., Duforet-Gaurier, L., & Vantrepotte, V. (2010). Effect of inherent optical properties variability on the chlorophyll retrieval from ocean color remote sensing: An in situ approach. *Optics Express*, 18, 20949–20952.
- Loisel, H., Mériaux, X., Poteau, A., Artigas, L., Lubac, B., Gardel, A., et al. (2009). Analyze of inherent optical properties of French Guiana coastal waters for remote sensing applications. *Journal of Coastal Research*, SI 56, 1532–1536.
- Lubac, B., & Loisel, H. (2007). Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea. *Remote Sensing of Environment*, 110, 45–58.
- Lubac, B., Loisel, H., Guiselin, N., Astoreca, R., Artigas, L., & Mériaux, X. (2008). Hyperspectral and multispectral ocean color inversions to detect *Phaeocystis globosa* blooms in coastal waters. *Journal of Geophysical Research*, 113, C06026.
- Mélin, F., Vantrepotte, V., Clerici, M., D'Alimonte, D., Zibordi, G., Berthon, J.-F., et al. (2011). Multi-sensor satellite time series of optical properties and chlorophyll-a concentration in the Adriatic Sea. *Progress in Oceanography*, 91, 229–244.
- Mélin, F., Zibordi, G., & Djavidnia, S. (2007). Development and validation of a technique for merging satellite derived aerosol optical depth from SeaWiFS and MODIS. *Remote Sensing of Environment*, 108, 436–450.
- Nechad, B., Ruddick, K., & Park, Y. (2010). Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sensing of Environment*, 114, 854–866.
- Neukermans, G., Ruddick, K., Bernard, E., Ramon, D., Nechad, B., & Deschamps, P.-Y. (2009). Mapping total suspended matter from geostationary satellites: A feasibility study with SEVIRI in the Southern North Sea. *Optics Express*, 17, 14029–14052.
- O'Reilly, J., Maritorena, S., O'Brien, M., Siegel, D., Toole, D., Menzies, D., et al. (2000). Ocean color chlorophyll a algorithms for seawifs, oc2, and oc4: Version 4. In J. E. O'Reilly (Ed.), *SeaWiFS Postlaunch Calibration and Validation Analyses, Part 3*. NASA/Tech. Memo. 2000-206892. Greenbelt, Maryland, NASA Goddard Space Flight Center volume 11. (pp. 9–23).
- Robinson, W. D., Franz, B. A., Patt, F. S., Bailey, S. W., & Werdell, P. J. (2003). Masks and flags updates. In S. B. Hooker, & E. R. Firestone (Eds.), *SeaWiFS Postlaunch Technical Report Series, Chap.6*, NASA/TM-2003-206892. Greenbelt, Maryland, NASA Goddard Space Flight Center volume 22. (pp. 34–40).
- Ruddick, K., Cauwer, V. D., Park, Y., & Moore, G. (2006). Seaborne measurements of near infrared water-leaving reflectance: The similarity spectrum for turbid waters. *Limnology and Oceanography*, 51, 1167–1179.
- Ruddick, K., Ovidio, F., & Rijkeboer, M. (2000). Atmospheric correction of SeaWiFS imagery for turbid coastal and inland waters. *Applied Optics*, 39, 897–912.
- Schroeder, T., Behnert, I., Schaale, M., Fischer, J., & Doerffer, R. (2007). Atmospheric correction algorithm for MERIS above case-2 waters. *International Journal of Remote Sensing*, 28, 1469–1486.
- Shi, W., & Wang, M. (2007). Detection of turbid waters and absorbing aerosols for the MODIS ocean color data processing. *Remote Sensing of Environment*, 110, 149–161.
- Shi, W., & Wang, M. (2009). An assessment of the black ocean pixel assumption for MODIS SWIR bands. *Remote Sensing of Environment*, 113, 1587–1597.
- Siegel, D. A., Wang, M., Maritorena, S., & Robinson, W. (2000). Atmospheric correction of satellite ocean color imagery: The black pixel assumption. *Applied Optics*, 39, 3582–3591.
- Stumpf, R. P., Arone, R. A., Gould, R. W., & Ransibrahmanakul, V. (2003). A partially coupled ocean-atmosphere model for retrieval of water-leaving radiance from SeaWiFS in coastal waters. In S. B. Hooker, & E. R. Firestone (Eds.), *SeaWiFS Postlaunch Technical Report Series, Chap.9*, NASA/TM-2003-206892. Greenbelt, Maryland, NASA Goddard Space Flight Center volume 22. (pp. 51–59).
- Vantrepotte, V., Loisel, H., Dessailly, D., & Mériaux, X. (2012). Optical classification of contrasted coastal waters. *Remote Sensing of Environment*, 123.
- Vantrepotte, V., Loisel, H., Mériaux, X., Jamet, C., Dessailly, D., Neukermans, G., et al. (2011). Seasonal and inter-annual (1998/2010) variability of the suspended particulate matter as retrieved from satellite ocean color sensors over the French Guiana coastal waters. *Journal of Coastal Research*, SI64, 1750–1754.
- Wang, M., & Shi, W. (2007). The NIR-SWIR combined atmospheric correction approach for MODIS ocean color data processing. *Optics Express*, 15, 15722–15733.
- Wang, M., Son, S., & Shi, W. (2009). Evaluation of MODIS SWIR and NIR-SWIR atmospheric correction algorithms using SeaBASS data. *Remote Sensing of Environment*, 113, 635–644.
- Werdell, P. J., & Bailey, S. W. (2005). An improved bio-optical data set for ocean color algorithm development and satellite data product validation. *Remote Sensing of Environment*, 98, 122–140.
- Werdell, P. J., Franz, B. A., & Bailey, S. W. (2010). Evaluation of shortwave infrared atmospheric correction for ocean color remote sensing of Chesapeake bay. *Remote Sensing of Environment*, 114, 2238–2247.
- Werdell, P. J., Franz, B. A., Bailey, S. W., Harding, L. W., Jr., & Feldman, G. C. (2007). Approach for the long-term spatial and temporal evaluation of ocean color satellite data products in a coastal environment. *Proceeding of SPIE*, Vol. 6680. (pp. 12).
- Wolfe, R. E., Nishihama, M., Fleig, A., Kuyper, J. A., Roy, D. P., Storey, J. C., et al. (2002). Achieving sub-pixel geolocation accuracy in support of MODIS land pixel. *Remote Sensing of Environment*, 83, 31–49.
- Zibordi, G., Berthon, J.-F., Mélin, F., D'Alimonte, D., & Kaitala, S. (2009). Validation of satellite ocean color primary products at optically complex coastal sites: Northern Adriatic Sea, Northern Baltic Proper and Gulf of Finland. *Remote Sensing of Environment*, 113, 2574–2591.
- Zibordi, G., Holben, B., Hooker, S. B., Mélin, F., Berthon, J.-F., Slutsker, I., et al. (2006). A network for standardized ocean color validation measurements. *EOS Transactions*, 87, 293–297.
- Zibordi, G., Holben, B., Slutsker, I., Giles, D., D'Alimonte, D., Mélin, F., et al. (2009). Aeronet-oc: A network for the validation of ocean color primary products. *Journal of the Atmospheric and Oceanic Technology*, 26, 1634–1651.
- Zibordi, G., Stroembeck, N., Mélin, F., & Berthon, J.-F. (2006). Tower-based radiometric observations at a coastal site in the Baltic Proper. *Estuarine, Coastal and Shelf Science*, 69, 649–654.