International Journal of Applied Earth Observation and Geoinformation 137 (2025) 104408

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

Recovery of pixels with extremely turbid waters and intensive floating algae from false cloud masking in satellite ocean color remote sensing

Menghua Wang [a,*], Lide Jiang [a,b]

[a] NOAA National Environmental Satellite, Data, and Information Service, Center for Satellite Applications and Research, E/RA3, 5830 University Research Ct., College Park, MD 20740, USA
[b] CIRA at Colorado State University, Fort Collins, CO, USA

ARTICLE INFO

Keywords:
Cloud masking
Floating algae
Turbid waters
Ocean color
Remote sensing

ABSTRACT

We describe our work to improve the cloud masking for satellite ocean color data processing over extremely turbid waters and intensive algae blooms (or floating algae), which are often identified as cloud mistakenly. An improved cloud masking approach is proposed using additional information of the Alternate Floating Algae Index (AFAI) and a new normalized AFAI (nAFAI), as well as ratios of the Rayleigh-corrected reflectance $\varepsilon^{(RC)}(\lambda_i, \lambda_j)$ from the blue and near-infrared bands. Specifically, the proposed algorithm adds a recovery procedure after the original cloud masking to retrieve falsely masked pixels and identifies these pixels as turbid waters, floating algae or absorbing aerosols, from which ocean color products can be further derived. The new cloud masking algorithm has been implemented in the NOAA Multi-Sensor Level-1 to Level-2 (MSL12) ocean color data processing system for routine global data processing for various satellite ocean color sensors, e.g., the Visible Infrared Imaging Radiometer Suite (VIIRS), the Ocean and Land Colour Instrument (OLCI), the Geostationary Ocean Color Imager (GOCI), etc. Results show that the new cloud masking has remarkably improved ocean color data coverage, particularly over highly turbid coastal and inland waters, as well as intensive floating algae, eliminating almost all false cloud pixels. For example, using the new cloud masking algorithm, the falsely masked pixels are recovered, reducing cloud masking pixels by ~30–40% and ~40–50% over highly turbid China East Coast and China's Lake Taihu, respectively.

1. Introduction

Satellite remote sensing provides the essential capability for routine measurements of global ocean/water properties over open oceans, coastal waters, and inland lakes (IOCCG, 2008). Current satellite ocean color sensors include the Visible Infrared Imaging Radiometer Suite (VIIRS) (Goldberg et al., 2013) on the Suomi National Polar-orbiting Partnership (SNPP), NOAA-20, and NOAA-21, the Moderate Resolution Imaging Spectroradiometer (MODIS) (Salomonson et al., 1989) on the Terra and Aqua, the Ocean and Land Colour Instrument (OLCI) (Donlon et al., 2012) on the Sentinel-3A (S3A) and Sentinel-3B (S3B), the Second-Generation Global Imager (SGLI) on the Global Change Observation Mission-Climate (GCOM-C), and the Geostationary Ocean Color Imager (GOCI) on the Communication, Ocean, and Meteorological Satellite (COMS) (Groom et al., 2019). Satellite ocean color data processing starts with the two major data masks, i.e., cloud masking (Ackerman et al., 1998; Robinson et al., 2003; Wang and Shi, 2006) and land masking (Carroll et al., 2017; Mikelsons et al., 2021). Pixels identified as cloud or land are masked without further data processing (no ocean color products derived). Therefore, it is crucial to have accurate cloud and land masks to obtain complete coverage of satellite ocean/water products. While accuracy of land mask mostly impacts products over coastal oceans and inland waters, cloud masking can have significant effects on satellite products over everywhere globally. In fact, clouds change all the time both spatially and temporally (King et al., 2013) and their variations are usually unpredictable. Thus, it is generally challenging to accurately identify clouds from various atmosphere and ocean/water properties for ocean color remote sensing. Specifically, there are still issues in cloud masking for some complicated cases, particularly over extremely turbid coastal and inland waters, as well as for intensive algae blooms or floating algae (Banks and Mélin, 2015).

In this study, we introduce a new cloud masking scheme which can recover most of the false cloud masking pixels due to highly turbid waters, intensive floating algae (or algae blooms) or absorbing aerosols,

* Corresponding author.
E-mail address: Menghua.Wang@noaa.gov (M. Wang).

https://doi.org/10.1016/j.jag.2025.104408
Received 14 October 2024; Received in revised form 18 January 2025; Accepted 2 February 2025
Available online 17 February 2025

and further identify their respective property types with setting up associated flags properly. Therefore, ocean color products can be further processed from these recovered pixels, e.g., deriving the normalized water-leaving radiance $nL_w(\lambda)$ spectra over all kinds of floating algae for understanding their properties (Wei and Wang, 2024). In this paper, we first provide a brief background for the topic, followed by the detailed description of rationales for development of the new recovery algorithm from false cloud masking in Section 3. In Section 4, the proposed cloud masking procedure is described and discussed. We then provide several case studies using the proposed new cloud masking over highly turbid coastal and inland waters, as well as cases from intensive floating algae, showing significantly improved results. We finalize the study with discussions in Section 6 and summary in Section 7.

2. Background

The cloud masking algorithm implemented in the NOAA Multi-Sensor Level-1 to Level-2 (MSL12) ocean color data processing system is simple and efficient (Robinson et al., 2003; Wang and Shi, 2006), i.e., an albedo threshold $\alpha(\lambda)$ for cloud masking is defined as:

$$\rho_{rc}(\lambda) = [\rho_t(\lambda) - \rho_r(\lambda)]/t(\lambda) \ > \alpha(\lambda), \tag{1}$$

where $\rho_{rc}(\lambda)$ is the Rayleigh-corrected reflectance, $\rho_t(\lambda)$ is the top-of-atmosphere (TOA) reflectance (after ozone absorption correction), $\rho_r(\lambda)$ is the Rayleigh (air molecules) scattering reflectance (Wang, 2016), $t(\lambda)$ is the atmospheric transmittance (no ozone absorption component), and $\alpha(\lambda)$ is the albedo threshold, above which the pixel is masked as cloud, and λ is the wavelength used for cloud masking, at the near-infrared (NIR) or shortwave infrared (SWIR) band. It should be noted that, in $\rho_{rc}(\lambda)$ computation using Eq. (1), we have to appropriately account for the ozone aborption in $\rho_t(\lambda)$ and $t(\lambda)$.

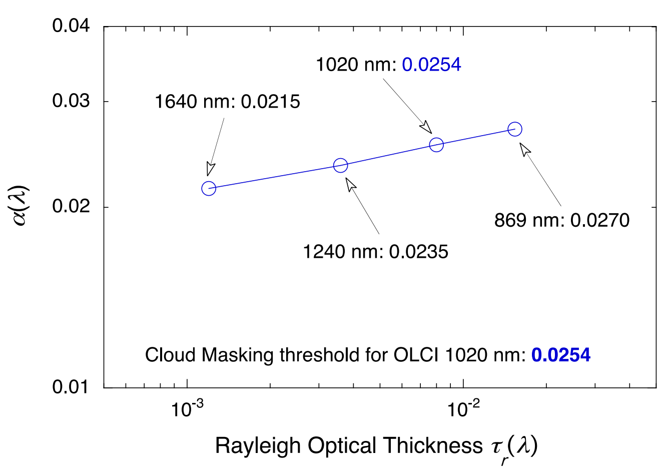

Fig. 1. $\alpha(\lambda)$ as a function of Rayleigh optical thickness $\tau_r(\lambda)$ in the log-log scale for the spectral bands from 869 nm (NIR) to SWIR 1640 nm, showing an excellent relationship between these two.

For sensors with only the NIR band, such as the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) (McClain et al., 2004) and GOCI (Choi et al., 2012), the NIR band at around 865 nm is used for cloud masking (Robinson et al., 2003; Wang and Shi, 2006), while for those with the SWIR bands (such as VIIRS and MODIS), more options are available, such as around 1240 nm, 1640 nm, or 2130 nm. Wang and Shi (2006) provide the following $\rho_{rc}(\lambda)$ threshold $\alpha(\lambda)$ values at the NIR (865 nm) and SWIR (1240, 1640, and 2130 nm) bands, i.e.,

$$\alpha(865) = 0.0270;\ \alpha(1240) = 0.0235;\ \alpha(1640) = 0.0215;$$
$$\text{and } \alpha(2130) = 0.0175, \tag{2}$$

which are used in the MSL12 cloud masking. For OLCI sensors, there is only one short SWIR band around 1020 nm, and we use $\alpha(1020) = 0.0254$, which was derived from the interpolation of $\alpha(\lambda)$ [or $\rho_{rc}(\lambda)$] as a function of Rayleigh optical thickness $\tau_r(\lambda)$ (in log-log scale) as showing in Fig. 1. In fact, Fig. 1 shows that $\alpha(\lambda)$ can be well related to $\tau_r(\lambda)$ from the NIR 869 nm to SWIR 1640 nm, i.e.,

$$\log[\alpha(\lambda)] = a_0 + a_1 \log[\tau_r(\lambda)], \tag{3}$$

where $\tau_r(\lambda)$ is Rayleigh optical thickness at the wavelength λ (from the NIR to SWIR bands), and a_0 and a_1 are the fitting coefficients of −1.4073 and 0.0896, respectively. Therefore, for the OLCI 1020 nm band, the cloud masking threshold $\alpha(1020)$ is 0.0254. Other SWIR $\alpha(\lambda)$ between 869 nm and 1640 nm can also be derived using Eq. (3). It should be noted that the same threshold of 0.0270 has been used for the NIR band around 865 nm for almost all satellite ocean color sensors, e.g., SeaWiFS, MODIS, VIIRS, OLCI, GOCI, etc.

For the MSL12 data processing, the SWIR bands near 1240 nm has been used as a default for cloud masking (Wang and Shi, 2006). However, in some extreme cases, water reflectance contributions at the SWIR 1240 nm are not negligible and cloud masking threshold $\alpha(\lambda)$ at 1640 nm (or even 2130 nm) should be used (Shi and Wang, 2009, 2014). In addition to turbid waters, there are also other situations that can cause false cloud masking, especially for sensors without the SWIR bands like GOCI, with high loading aerosols and intensive floating algae being the two other examples (Banks and Mélin, 2015).

In this study, the Level-1B data from VIIRS, OLCI, and GOCI were obtained from their respective official data sources and used as inputs to NOAA-MSL12 data processing system to generate Level-2 ocean color

Table 1
The nominal center wavelengths at the blue, red, and NIR bands for various satellite sensors, which are used for computations of AFAI and corresponding c value, as well as $\varepsilon^{(RC)}(\lambda_i, \lambda_j)$ [e.g., $\varepsilon^{(RC)}(443, 862)$].

Satellite Sensor	λ(Blue) (nm)	λ(Red) (nm)	$\lambda(NIR_1)$ (nm)	$\lambda(NIR_2)$ (nm)	c
GOCI	443	660	745	865	0.4146
VIIRS-SNPP	443	671	745	862	0.3874
VIIRS-NOAA-20	445	667	746	868	0.3930
VIIRS-NOAA-21	445	671	747	868	0.3827
OLCI-S3A/S3B	443	674	754	865	0.4188
SGLI-GCOM-C	443	672	763	867	0.4667

data (Wang et al., 2013a; Wang et al., 2013b). The Rayleigh-corrected reflectance $\rho_{rc}(\lambda)$ spectra are an intermediate product necessary to this work and therefore are output to the Level-2 data together with ocean color products. $\rho_{rc}(\lambda)$ data are also used to generate near-surface true-color images for intuitive visualization. In this work, we use measurements from various satellites, including VIIRS (on SNPP and NOAA-21), OLCI-S3A, and GOCI, representing cloud masking results from both polar and geostationary satellites. It should be noted that in this study the original cloud masking refers to the simple NIR or SWIR threshold cloud masking method, i.e., $\rho_{rc}(1240) > 0.0235$, $\rho_{rc}(1020) > 0.0254$, and $\rho_{rc}(865) > 0.027$ for VIIRS, OLCI, and GOCI, respectively.

3. New cloud masking algorithm development

3.1. AFAI for identification of floating algae

We first start by defining Blue, Red, NIR$_1$, and NIR$_2$ for spectral bands at around 443, 671, 745, and 865 nm (depending on sensors, see Table 1), respectively. The Alternative Floating Algae Index (AFAI) is modified from the Floating Algae Index (AFI) (Hu, 2009; Wang and Hu, 2016), which can be defined as,

$$AFAI = [\rho_{rc}(\mathrm{NIR}_1) - \rho_{rc}(\mathrm{Red})] - c\,[\rho_{rc}(\mathrm{NIR}_2) - \rho_{rc}(\mathrm{Red})], \tag{4}$$

where coefficient c is computed from actual wavelengths of the three bands used (Red, NIR$_1$, and NIR$_2$), i.e.,

$$c = \frac{\lambda(\mathrm{NIR}_1) - \lambda(\mathrm{Red})}{\lambda(\mathrm{NIR}_2) - \lambda(\mathrm{Red})}, \tag{5}$$

and $\lambda(\mathrm{NIR}_1)$, $\lambda(\mathrm{NIR}_2)$, and $\lambda(\mathrm{Red})$ are wavelengths at the two NIR bands and red band, respectively. Table 1 provides nominal center wavelengths for the spectral bands used at the red and NIR wavelengths with the corresponding c values [Eq. (5)], as well as an atmospheric correction parameter $\varepsilon^{(\mathrm{RC})}(\lambda_i, \lambda_j)$ (see Section 3.5), for various satellite sensors.

Essentially, AFAI can be regarded as reflectance difference between sensor-measured $\rho_{rc}(\mathrm{NIR}_1)$ and a baseline value computed, which is from a linear interpolation of $\rho_{rc}(\lambda)$ between at the red band $\rho_{rc}(\mathrm{Red})$ and at NIR$_2$ band $\rho_{rc}(\mathrm{NIR}_2)$ (Hu, 2009; Wang and Hu, 2016). In case with algae blooms, AFAI values are elevated due to higher (lower) $\rho_{rc}(\lambda)$ at the NIR (red) bands (by algae), thus making it possible to differentiate algae blooms from surrounding non-algae waters.

3.2. Normalization of AFAI for high $\rho_{rc}(NIR_2)$ situation

To recover floating algae pixels for cases with $\rho_{rc}(\mathrm{NIR}_2) > 0.15$ (see details in next section), we define a new index, i.e., normalized AFAI, nAFAI, which equals AFAI divided by $\rho_{rc}(\mathrm{NIR}_2)$, i.e.,

$$nAFAI = AFAI/\rho_{rc}(\mathrm{NIR}_2) = \left[\frac{\rho_{rc}(\mathrm{NIR}_1)}{\rho_{rc}(\mathrm{NIR}_2)} - \frac{\rho_{rc}(\mathrm{Red})}{\rho_{rc}(\mathrm{NIR}_2)}\right] - c\left[1 - \frac{\rho_{rc}(\mathrm{Red})}{\rho_{rc}(\mathrm{NIR}_2)}\right], \tag{6}$$

where $\rho_{rc}(\mathrm{Red})$, $\rho_{rc}(\mathrm{NIR}_1)$, and $\rho_{rc}(\mathrm{NIR}_2)$ are Rayleigh-corrected reflectance at red, short NIR, and longer NIR bands, respectively. We use $\rho_{rc}(\mathrm{NIR}_2)$ in the AFAI normalization due to its role as the cloud masking criterium for global ocean color data processing (Robinson et al., 2003; Wang and Shi, 2006). With its normalization, we can cancel the impact of AFAI increased intensity due to simultaneous increase in $\rho_{rc}(\lambda)$ at all bands in thick cloud situations, to enhance spectral differences between intensive algae and heavy clouds.

3.3. Using AFAI and nAFAI for identification of floating algae

To correctly recover algae pixels from falsely masked cloud pixels, the AFAI and nAFAI signatures from algae pixels need to be compared with those from cloud pixels, and to a lesser extent, with those from non-algal water pixels, from which a method to separate floating algae from cloud can be developed. Before moving forward, it is important to first separate the pixels masked as cloud into two categories due to their distinct characteristics that require different ways of handling in the MSL12 ocean color data processing.

Following the Wang and Shi (2006) approach, we separate pixels into three reflectance categories based on $\rho_{rc}(\mathrm{NIR}_2)$ value, i.e.,

$$\begin{cases} \text{Low reflectance (clear sky)}: \rho_{rc}(\mathrm{NIR}_2) \le 0.027, \\ \text{Medium reflectance}: 0.027 < \rho_{rc}(\mathrm{NIR}_2) \le 0.15, \\ \text{High reflectance}: \rho_{rc}(\mathrm{NIR}_2) > 0.15. \end{cases} \tag{7}$$

The first category in Eq. (7) is for cases with clear sky (low reflectance) that all ocean color products are derived. The second (medium reflectance) and third (high reflectance) categories are for thin and thick clouds, but also include cases with floating algae, turbid waters, heavy aerosols, etc. It should be noted that in this study we have specifically excluded sun glint cases that can be accurately identified (Cox and Munk, 1954; Wang and Bailey, 2001).

Although all pixels with $\rho_{rc}(\mathrm{NIR}_2) > 0.027$ are masked as cloud in the original NIR cloud masking scheme (Robinson et al., 2003), they may not necessarily be cloud, and it is our purpose to recover any falsely masked pixels in such cases. For cases with $0.027 < \rho_{rc}(\mathrm{NIR}_2) \le 0.15$, it is expect to include non-cloud cases (in addition to thin clouds), e.g., turbid waters, heavy aerosols (smokes, dusts, fogs, etc.), floating algae (or algae blooms), etc. For those with $\rho_{rc}(\mathrm{NIR}_2) > 0.15$, there is a high possibility that they are non-clear sky cases (e.g., heavy/thick clouds or aerosols). However, it is still possible that such high reflectance cases are from ocean/water surface features (e.g., intensive-algae or extremely turbid waters).

To understand the AFAI and nAFAI signatures for the mentioned three categories in Eq. (7), we investigate two OLCI-S3A scenes shown in Figs. 2 and 3. In Fig. 2, cloud masked pixels are all true clouds, as shown visually in the true color image (Fig. 2a and enlarged Box D). The pixels flagged in blue, yellow, and red in Fig. 2b (and enlarged Box D in Fig. 2b) correspond to the categories defined in Eq. (7), i.e., clear sky, thin cloud (medium reflectance), and thick cloud (high reflectance) pixels (see also Fig. 4a for histogram results), respectively.

Fig. 3 presents a sub-region from one of the cases from OLCI-S3A measurements on 25 June 2016 over the Yellow Sea (Fig. 3a for true color image), showing cloud masked pixels are mostly floating algae and turbid waters, and they are flagged according to Eq. (7) (Fig. 3b). This case demonstrates that intensive floating algae can have $\rho_{rc}(\mathrm{NIR}_2) > 0.15$ (shown in red), which are quite distinguishable, often near the center of each small patches, but relatively low in pixel count. After excluding real clouds (shown in white in Fig. 3c and d), the histograms of the pixels corresponding to the medium and high reflectance definitions in Eq. (7) are shown in Fig. 4b. It is noted that there are seven sites marked in Fig. 3a, including the two highlighted Jihongtan Reservoir (pink circle) and Chanzhi Reservoir (pink rectangle) in Qingdao of China, for further data analysis and demonstration. Enlarged true color images of these two reservoirs, which were acquired from Landsat-8 on 16 June 2016 at around 2:35 UTC, are also presented in Fig. 3 for clear visualization.

The green curve in Fig. 4b is the AFAI histogram for the medium reflectance category (yellow blotch in Fig. 3b). This includes most of the floating algae and turbid waters (shown in green and brown in Fig. 3d). Two modes in histogram (Fig. 4b) can be identified at AFAI values of −0.0038 and 0.0036, corresponding to the turbid waters and floating algae in histogram, respectively. For example, AFAI values for Sites 4, 5, and 7 noted in Fig. 3a are 0.0111, −0.0017, and −0.0127, respectively, corresponding to cases from floating algae and turbid waters in the Jihongtan Reservoir in Qingdao (Yang et al., 2024) and Jiaozhou Bay, respectively. On the other hand, the purple curve is the histogram for the high reflectance category in Eq. (7) mostly from intensive floating algae or highly turbid waters (red blotch excluding the cloud in Fig. 3b). For example, AFAI values for Sites 3 and 6 noted in Fig. 3a are 0.227 and

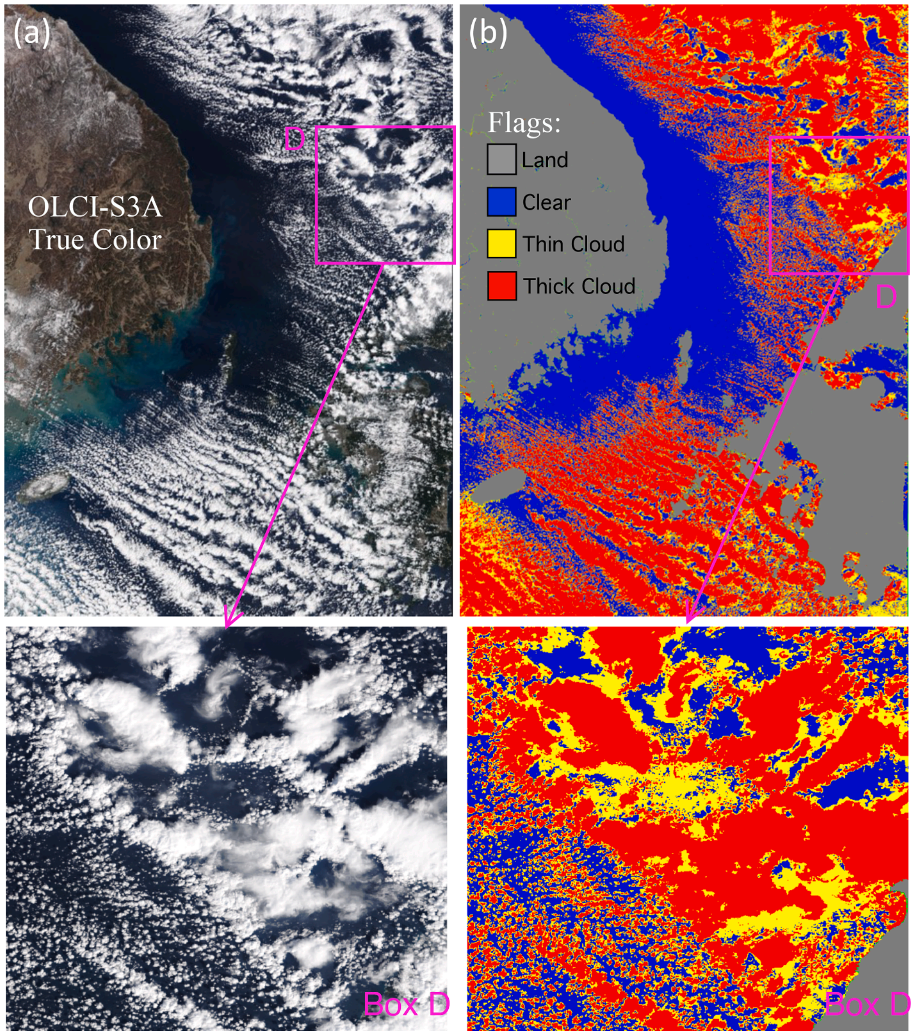

Fig. 2. A case study from OLCI-S3A on 12 February 2017 at 1:47 UTC near southern part of Korean Peninsula and Kyushu Island of Japan for (a) true color image and (b) pixels flagged in different colors (see flag color in panel b) showing clear sky (blue), thin cloud (yellow), and thick cloud (red) pixels based on Eq. (7). Note that the selected Box D in both panels a and b is enlarged to show more detailed features for the true color and pixel flagged images.

0.022, respectively, corresponding to cases with intensive floating algae and extremely turbid waters from the Chanzhi Reservoir in Qingdao of China (Xing et al., 2022), which are shown in green (algae) and brown (turbid waters) in Fig. 3d.

The three curves in Fig. 4a all resemble the Gaussian distribution, with AFAI modes being very close to zero, i.e., −0.0007, −0.0007, and 0.0001 for categories of the low reflectance (clear sky), medium reflectance (thin clouds), and high reflectance (thick clouds), respectively. The high reflectance category has a higher mode value and the most spreading out among the three. To determine a reasonable AFAI threshold for recovery of floating algae pixels from false cloud masks, we list in Table 2 the percentage of pixels meeting the AFAI threshold for various categories in Figs. 2 and 3. Basically, a good choice of AFAI threshold 'X' will give low percentage for real cloud and high percentage for floating algae (or turbid waters), though all percentages decrease with increase of 'X'. For recovery of floating algae (or turbid waters) with $0.027 < \rho_{rc}(\mathrm{NIR}_2) \le 0.15$, a threshold of 0.001 seems to be a good choice. However, since there is still more than a quarter of the thick clouds with AFAI > 0.001, the same threshold does not work well for the thick cloud, and should only be applied to cases with $0.027 < \rho_{rc}(\mathrm{NIR}_2) \le 0.15$.

Similarly, Figs. 3c, 4c, and 4d show nAFAI results. We notice one major difference between Fig. 4a and 4c, i.e., the relatively spreading in histograms among the clear sky, medium reflectance (thin cloud), and high reflectance (thick cloud) categories has changed. The clear-sky histogram has the widest spreading now (standard deviation (STD) = 0.1782), the thin cloud histogram spreading is next (STD = 0.0548), and the thick cloud histogram has the narrowest spreading (STD = 0.0083) with their mode values all close to 0, which is desirable. It is noted that in Fig. 4c the thick cloud histogram has a peak value at about 40% (much higher than other two). The modes for clear sky (low reflectance), thin cloud (medium reflectance), and thick cloud (high reflectance) histograms are −0.0425, −0.0075, and 0.0025, respectively. For thick cloud (high reflectance), histogram results of AFAI and nAFAI are significantly different. Similarly, the relative spreading in histograms (green curves in Fig. 4b and 4d for turbid waters/floating algae) between AFAI and nAFAI have also changed. For example, nAFAI values for Sites 4, 5, and 7 noted in Fig. 3a are 0.2519, −0.0361, and −0.3195, respectively, compared with AFAI values of 0.0111, −0.0017, and −0.0127, respectively. For the high reflectance with $\rho_{rc}(\mathrm{NIR}_2) > 0.15$ (purple curve in Fig. 4d), the nAFAI value ranges from ~0.45–0.65 for cases with intensive floating algae (e.g., the nAFAI value for Site 3 is 0.5667), and ~0–0.4 for pixels with extremely turbid waters (e.g., the nAFAI of 0.1424 for Site 6 from the Chanzhi Reservoir).

Therefore, we can set the nAFAI threshold to be 0.4, noting that the thin cloud histogram has a near-zero mean with STD of 0.0025. Although this criterium of nAFAI > 0.4 is developed to recover intensive floating algae from thick clouds, there is no need to limit its value only to

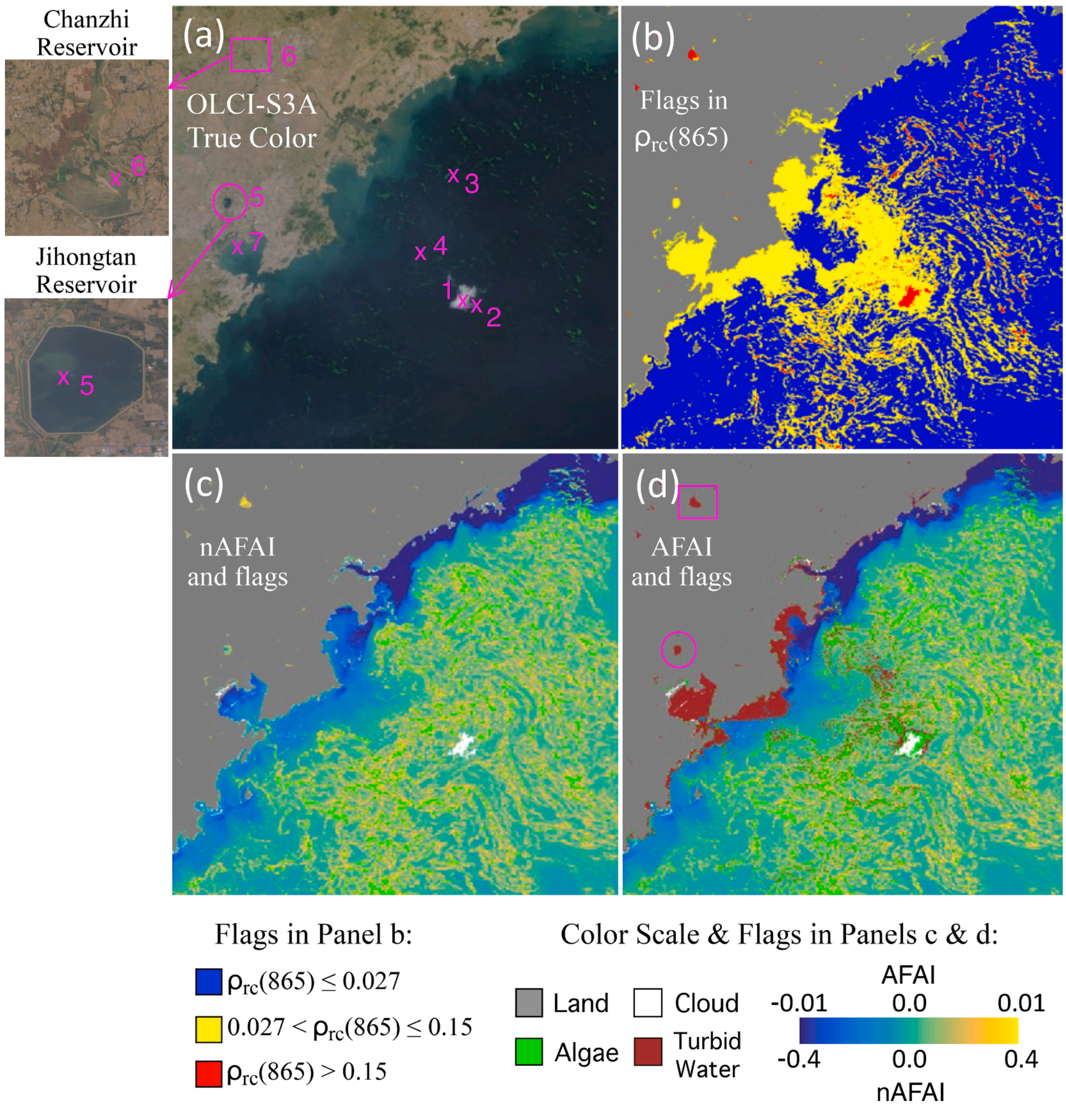

Fig. 3. A case study from OLCI-S3A on 25 June 2016 at 2:01 UTC in the Yellow Sea for (a) true color image, (b) pixels flagged in different colors according to Eq. (7) (noted in the bottom left), (c) nAFAI with recovered floating algae flagged in green, and (d) AFAI with recovered floating algae flagged in green and turbid water flagged in brown. Color scales (nAFAI and AFAI) and flags for panels c and d are noted in the bottom right. In panel a, there are seven sites marked for further data analysis and demonstration. Note that, in panels a and d, the Jihongtan Reservoir and Chanzhi Reservoir are marked with pink circle and rectangle, respectively, including their enlarged true color images acquired from Landsat-8 on 16 June 2016 at around 2:35 UTC.

the cases with $\rho_{rc}(\text{NIR}_2) > 0.15$. The reason is that for cases with $0.027 < \rho_{rc}(\text{NIR}_2) \leq 0.15$, pixels of nAFAI > 0.4 are floating algae (green curve in Fig. 4d). Thus, we do not need to distinguish between the medium and high reflectance categories for floating algae, but just apply this criterium of nAFAI >0.4 to all pixels with $\rho_{rc}(\text{NIR}_2) > 0.027$.

In summary, to determine whether a cloud masked pixel should be recovered as floating algae (or algae blooms), its nAFAI is first checked and will be recovered as a floating algae pixel if nAFAI > 0.4. If the test fails but its $\rho_{rc}(\text{NIR}_2) \leq 0.15$, we can go through a second test of AFAI and the pixel is recovered as floating algae if AFAI > 0.001.

3.4. Using AFAI for identification of turbid waters

Although AFAI is not originally intended to identify turbid waters, during the investigation of AFAI thresholds for identifying different water types, we found that turbid waters usually have strongly negative AFAI signature. While clear-sky water pixels and thin/thick clouds usually have AFAI > -0.005 (Fig. 4a), turbid waters can have more negative AFAI values (Fig. 4b, negative peak of the green curve). As a result, AFAI < -0.01 is a good candidate for recovering turbid waters from false cloud masks, and its value does not overlap with the floating algae criterium of AFAI >0.001. It should be noted that, although AFAI < -0.01 can identify most of turbid water pixels, it may still miss some turbid waters with positive AFAI, e.g., values from 0 to 0.04 portion of the purple curve in Fig. 4b, from the Chanzhi Reservoir mentioned previously (Fig. 3d). As a result, AFAI for identification of turbid waters is best used as a complementary criterium, i.e., using AFAI < -0.01 to identify some missed turbid water pixels, complementary to another well-established criterium for identification of turbid waters (see next).

3.5. Using $\varepsilon^{(RC)}(\lambda_1, \lambda_2)$ parameters for identification of turbid waters

Following Wang and Shi (2006), we define $\varepsilon^{(RC)}(\lambda_1, \lambda_2)$ as the ratio of $\rho_{rc}(\lambda)$ at two wavelengths λ_1 and λ_2, i.e.,

$$\varepsilon^{(RC)}(\lambda_1, \lambda_2) = \rho_{rc}(\lambda_2)/\rho_{rc}(\lambda_1), \tag{8}$$

for distinguishing highly turbid waters or absorbing aerosols from false cloud masking (Wang and Shi, 2006). This criterium is based on the fact that reflectance spectral shapes of cloud pixels are usually flat (i.e.,

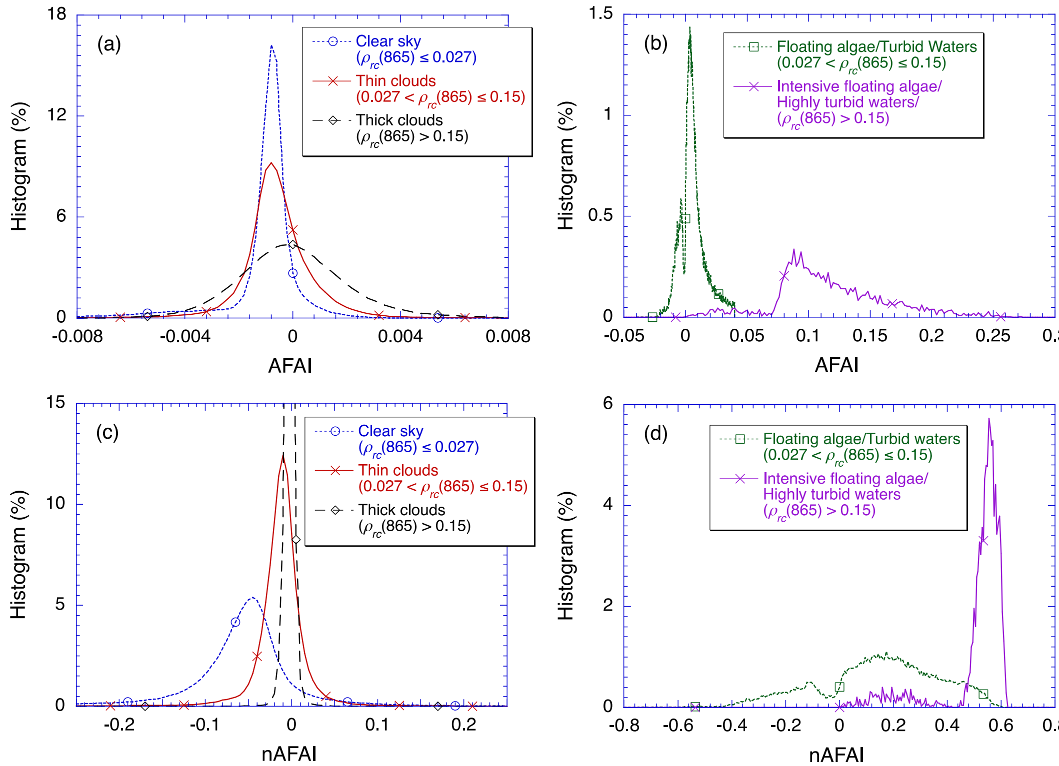

Fig. 4. AFAI and nAFAI histograms for various categories of pixels for (a) AFAI histograms for pixels flagged in blue, yellow, and red in Fig. 2b, (b) AFAI histograms for pixels flagged in yellow and red in Fig. 3b, (c) nAFAI histograms corresponding results in Fig. 2b, and (d) nAFAI histograms corresponding results in Fig. 3b. Note that the nAFAI histogram for thick clouds in plot c has a peak value of about 40% (cut it at 15% for clear demonstration).

Table 2
The relationship between AFAI threshold 'X' and pixel percentage for different categories flagged in Fig. 2b [i.e., low reflectance (clear sky), medium reflectance (thin clouds), and high reflectance (thick clouds)] and Fig. 3b [i.e., floating algae (FA) or turbid waters (TW) with $0.027 < \rho_{rc}(\mathrm{NIR}_2) \le 0.15$ and intensive floating algae (IFA) or highly turbid waters (HTW) with $\rho_{rc}(\mathrm{NIR}_2) > 0.15$]. Note that the second to fourth columns are for results from Fig. 2b and the last two (fifth and sixth) columns are for results from Fig. 3b.

AFAI > X	Clear sky	Thin cloud	Thick cloud	FA (or TW)	IFA (or HTW)
0.000	7.53%	23.54%	46.30%	76.96%	99.98%
0.001	2.04%	8.92%	27.15%	72.66%	99.40%
0.002	0.53%	3.66%	15.37%	67.14%	98.35%
0.003	0.14%	1.90%	7.38%	60.68%	96.52%

white), and their $\varepsilon^{(RC)}(\mathrm{NIR}_1, \mathrm{NIR}_2)$ is close to one. This scheme has also been implemented in the MSL12 package specifically for the processing of GOCI images (and other sensors without the SWIR bands) in the China East Coast to alleviate the false cloud-masking issue due to prevalent turbid waters or heavy aerosols in the region and gained much improved results.

However, there is still room for improvement. First of all, it is noticed that in extremely turbid water region, some pixels can have $\rho_{rc}(\mathrm{NIR}_2) > 0.10$, so the original limitation of $\rho_{rc}(\mathrm{NIR}_2) < 0.06$ in Wang and Shi (2006) reduces the capability to further identify and recover false cloud pixels. Second, some extremely turbid water pixels can have $\varepsilon^{(RC)}(\mathrm{NIR}_1, \mathrm{NIR}_2) < 1.15$, thus are not separable from clouds with this parameter alone. Since highly turbid waters over coastal and inland lakes usually have low reflectance at the shorter wavelengths, they can be separated from clouds using $\varepsilon^{(RC)}(\mathrm{Blue}, \mathrm{NIR}_2)$ (IOCCG, 2010).

Fig. 5 shows an example of such case, taken from OLCI-S3A along the coast of Jiangsu province in China around (33.5–34.5°N, 120.0–121.5°E) on February 15, 2017, at 2:09 UTC. All waters in this area have originally masked as cloud, shown as thin cloud in yellow and thick cloud in red (Fig. 5b). From the true color image (Fig. 5a), we notice that the red blotch also includes some turbid waters that are not really cloud. Fig. 5c shows a scatter plot of $\varepsilon^{(RC)}(443, 865)$ versus $\varepsilon^{(RC)}(745, 865)$ for pixels with $\rho_{rc}(865) > 0.15$ (red blotch in Fig. 5b). Indeed, we can see two distinct patterns in Fig. 5c that correspond to cloud and highly turbid water pixels, respectively, with both patterns are constrained within $\varepsilon^{(RC)}(745, 865) < 1.15$. From results in Fig. 5, it is observed that $\varepsilon^{(RC)}(443, 865)$ of 0.85 is a reasonable value to separate turbid waters from cloud pixels in such condition.

From some extensive tests and evaluations, we found that $\varepsilon^{(RC)}(\mathrm{NIR}_1, \mathrm{NIR}_2) \ge 1.15$ works quite well across all pixels with $\rho_{rc}(\mathrm{NIR}_2) > 0.027$, except for some pixels with $\rho_{rc}(\mathrm{NIR}_2) > 0.15$, in which $\varepsilon^{(RC)}(\mathrm{Blue}, \mathrm{NIR}_2) < 0.85$ can work well complementarily. Therefore, the criterium of $\varepsilon^{(RC)}(\mathrm{NIR}_1, \mathrm{NIR}_2) \ge 1.15$ or $\varepsilon^{(RC)}(\mathrm{Blue}, \mathrm{NIR}_2) < 0.85$ can be applied to all pixels with $\rho_{rc}(\mathrm{NIR}_2) > 0.027$ to recover turbid water pixels from those of false cloud masks.

4. Data recovery procedure

Based on our investigations described in the previous sections, we propose an improved cloud masking scheme in two steps, i.e., the original cloud flagging step (Wang and Shi, 2006) and the non-cloud pixel recovery step (change from cloud to turbid water/floating algae or absorbing aerosols). In other words, pixels identified as cloud in the original cloud masking scheme (including using the NIR or SWIR approach) are further tested to see if they meet the criteria to be recovered and assign different flags accordingly. Only the remaining cloud pixels (not recovered) are masked as cloud.

The criteria for the non-cloud pixel recovery are assessed one by one in the following order and priority for all pixels flagged as cloud with the original cloud masking scheme, i.e., using the NIR or SWIR cloud masking method in Eqs. (1) and (2):

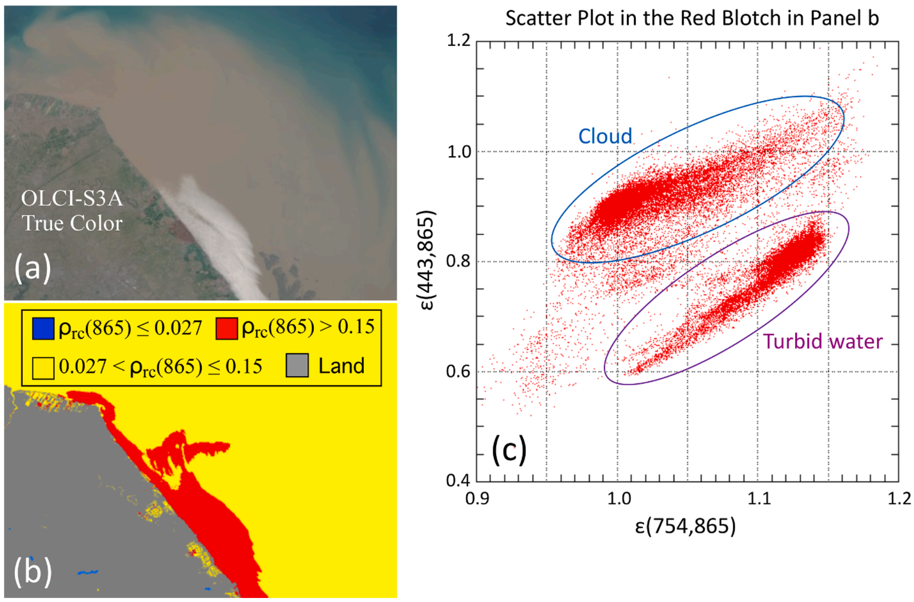

Fig. 5. Example of extremely turbid waters along the coast of Jiangsu province in China from OLCI-S3A observations on February 15, 2017, with all $\varepsilon^{(RC)}(745, 865) < 1.15$ for (a) true color image, (b) pixels flagged in different colors noted in the flag color (same as in Fig. 2b), and (c) scatter plot for $\varepsilon^{(RC)}(443, 865)$ versus $\varepsilon^{(RC)}(745, 865)$ in the red blotch in panel b.

(1) If nAFAI > 0.4, set both recovery and floating algae flags.
(2) If either $\varepsilon^{(RC)}(\mathrm{NIR}_1, \mathrm{NIR}_2) \geq 1.15$ or $\varepsilon^{(RC)}(\mathrm{Blue}, \mathrm{NIR}_2) < 0.85$, set both recovery and turbid water flags.
(3) If $\rho_{rc}(\mathrm{NIR}_2) \leq 0.15$ and AFAI > 0.001, set both recovery and floating algae flags.
(4) If $\rho_{rc}(\mathrm{NIR}_2) \leq 0.15$ and AFAI < −0.01, set both recovery and turbid water flags.

Note that we have specifically introduced a new recovery flag for those pixels recovered from the original cloud pixels using the NIR or SWIR cloud masking in the MSL12 data processing.

Fig. 3c and 3d visualize the procedure described above. Fig. 3c shows the floating algae pixels (green) recovered by the first criterium (1), with the background color showing nAFAI values. Before the recovery process, the total count of the original cloud masking is 48,630 pixels. After the first criterium (1) is applied, 16,917 (34.8%) pixels are recovered as floating algae. Fig. 3d shows AFAI values, with all the recovered floating algae flagged in green after the first and third criteria [(1) and (3)], which increased to 26,313 (54.1%), meaning applying the third criterium (3) recovered an additional 19.3% from the originally cloud masked pixels. In addition, a total of 42.4% of the false cloud masked pixels are recovered as turbid waters, which are almost all recovered from the second criterium (2), and only 3.5% of the original cloud masking remains unchanged. The fourth criterium (4) contributes to only additional 75 turbid water pixels, less than 0.4% of the total recovered turbid water pixels. Therefore, using the proposed recovery method, the total number of pixels recovered in this case (Fig. 3) is 46,909, which is about 96.5% of the total original cloud masking pixels.

All the recovered pixels from originally identified as cloud are then re-categorized their flag/mask status with appropriate new flags or masks. Therefore, the original cloud pixels are now split into recovered pixels and re-confirmed cloud pixels, with the former flagged (not masked) and latter masked. Unlike masked pixels, the recovered pixels have their ocean color products retrieved. It should be noted that all these recovered pixels with derived ocean color products have the recovery flag associated with. Therefore, these recovered pixels can be easily identified for further study and investigation.

For all pixels flagged as turbid waters through either the aforementioned non-cloud pixel recovery process or the original turbid water identification process (Shi and Wang, 2007), their absorbing aerosol

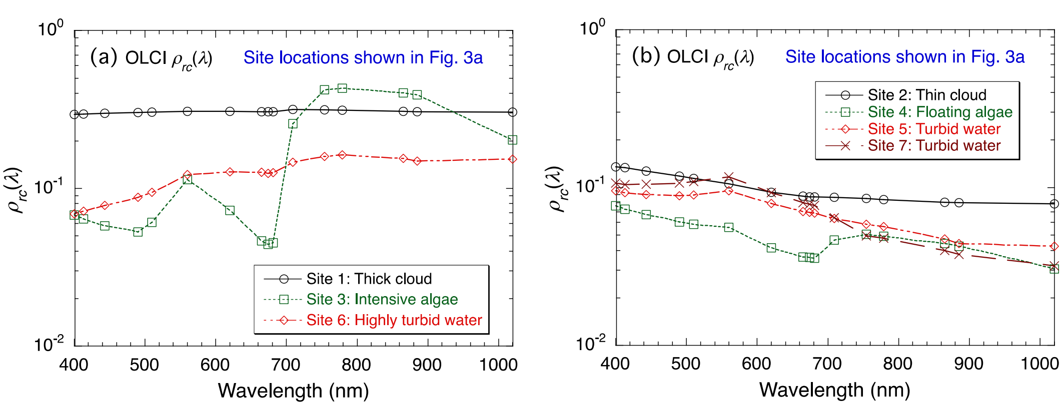

Fig. 6. OLCI-measured $\rho_{rc}(\lambda)$ as a function of OLCI spectral bands from 400–1020 nm for seven cases marked in Fig. 3a for (a) three cases with $\rho_{rc}(865) > 0.15$ for thick cloud at Site 1, intensive floating algae at Site 3, and extremely turbid water at Site 6 (Chanzhi Reservoir) and (b) four cases with $0.027 < \rho_{rc}(865) \leq 0.15$ for thin cloud at Site 2, floating algae at Site 4, and turbid waters at Sites 5 (Jihongtan Reservoir) and 7 (Jiaozhou Bay). Note that, for clear demonstration, we have removed OLCI data at 761, 764, and 768 nm covering oxygen A-band absorption.

Table 3
The criteria tested in the recovery procedure for seven cases (sites) marked in Fig. 3a. Note that the process of recovery procedure stops when test result is true.

Site # (Fig. 3a)	Criteria Tested in the Recovery Procedure: First	Second	Third	Fourth	Result
1	False	False	False	False	Cloud
2	False	False	False	False	Cloud
3	True	—	—	—	Floating Algae
4	False	False	True	—	Floating Algae
5	False	True	—	—	Turbid Water
6	False	True	—	—	Turbid Water
7	False	True	—	—	Turbid Water

indices are also calculated and compared with a threshold to see if they are absorbing aerosols rather than turbid waters. Consequently, based on the Shi and Wang (2007) method, flags for turbid water pixels may be changed accordingly, i.e., for the pixel identified as absorbing aerosols, their absorbing aerosol flag bit is set with also resetting the turbid water flag bit.

To provide further detailed information as how the recovery algorithm works, Fig. 6 shows OLCI-measured $\rho_{rc}(\lambda)$ as a function of OLCI spectral bands from 400–1020 nm for the seven cases marked in Fig. 3a. These include three cases in Fig. 6a with $\rho_{rc}(865) > 0.15$, i.e., Site 1 for thick cloud, Site 3 for intensive floating algae, and Site 6 for extremely turbid water in the Chanzhi Reservoir, and four cases in Fig. 6b with $0.027 < \rho_{rc}(865) \leq 0.15$, i.e., Site 2 for thin cloud, Site 4 for floating algae, and Sites 5 and 7 for turbid waters in the Jihongtan Reservoir and Jiaozhou Bay, respectively. As expected, both thick and thin clouds at Sites 1 and 2 have relatively flat $\rho_{rc}(\lambda)$, both intensive and non-intensive floating algae at Sites 3 and 4 show strong algae spectral features (i.e., reduced $\rho_{rc}(\lambda)$ at the red band and elevated $\rho_{rc}(\lambda)$ at the NIR band), and turbid waters at Sites 5–7 display high $\rho_{rc}(\lambda)$ features at green, red, and NIR bands. Using the proposed recovery algorithm, all seven cases have been correctly identified and pixels with floating algae and turbid waters are recovered from false cloud masks. Table 3 lists criteria test results of the recovery procedure for the seven cases marked in Fig. 3a. It is noted that the recovery data processing stops when a criterium test is truth. For example, for real cloud cases at Sites 1 and 2, all four criteria are tested, while for the case with floating algae at Site 3 only the first criterium is tested (Table 3). Similarly, Table 3 shows that, for cases with turbid waters at Sites 5–7 in Fig. 3a, only first and second criteria are tested.

5. Results from some case studies

5.1. GOCI and VIIRS-SNPP images on 15 February 2017

Fig. 7 shows a portion of images from GOCI obtained at 5:16 UTC on February 15, 2017, for true color (Fig. 7a) and the MSL12-derived normalized water-leaving radiance $nL_w(\lambda)$ at 555 nm [$nL_w(555)$] with pixels recovered from the original cloud masking and identified as turbid waters (brown), floating algae (green), and remaining cloud (white) (Fig. 7b). Due to its lack of the SWIR bands, GOCI can only rely on the NIR bands for cloud masking, and result in wide range of false cloud masks in highly turbid regions like the Bohai Sea, Yellow Sea, and East China Sea. The brown in Fig. 7b indicates the false cloud masks from the original cloud masking of $\rho_{rc}(\text{NIR}_2) > 0.027$, which is recovered to turbid water pixels by the proposed algorithm.

By comparing Fig. 7b with Fig. 7a, we confirm that the proposed algorithm successfully recovered most of the false cloud pixels, except for those along the slot boundary (between slot 7 and slot 10) in south of the Yellow Sea, west of the Korean Peninsula, due to a known GOCI issue of interslot radiometric discrepancy (ISRD), which is caused by the straylight generated from the misalignment of sensor elements (Kim et al., 2015).

On the other hand, although the real cloud masking is largely unchanged by the proposed method, we do notice a small amount of false recovery near the edges of real cloud masks, which can be identified to either floating algae (green) or turbid waters (brown). This is unavoidable, because as is shown in Table 2, there is still about 9% of the thin cloud which meets the criterium of AFAI > 0.001, and there are also some cloud pixels meeting $\varepsilon^{(RC)}(\text{NIR}_1, \text{NIR}_2) \geq 1.15$ or $\varepsilon^{(RC)}(\text{Blue}, \text{NIR}_2) < 0.85$, as shown in Fig. 5c. Specifically for GOCI, this issue is aggravated by the fact that each band image is taken at a slightly different time. For slot 7 in Fig. 7, e.g., the relative times (relative to the granule time, which was 5:16:42 UTC) of image capturing for bands at 660 nm, 745 nm, and 865 nm were 543, 557, and 587 s, respectively. This means that data at 745 and 865 nm bands were obtained 14 and 44 s after 660 nm data. The time difference is not insignificant when dealing with fast-moving cloud, which may distort the spectral shape at the cloud edges and make the AFAI value unreliable. Since the spectral distortion

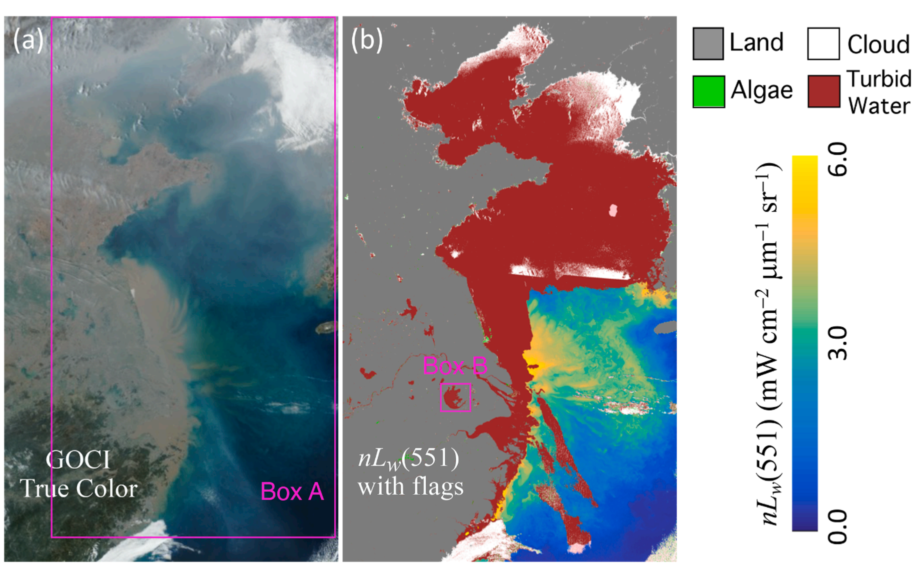

Fig. 7. A case study from GOCI on 15 February 2017 at 5:16 UTC over the western Pacific Ocean for (a) true color image and (b) MSL12-derived $nL_w(555)$ image with pixels recovered from the original cloud masking and identified as turbid waters (brown), floating algae (green), and remaining cloud (white). Note that there are marks in regions of Box A (panel a) and Box B (panel b, Lake Taihu) for further data analyses. It is also emphasized that the flag colors in panel b are only for the recovered pixels.

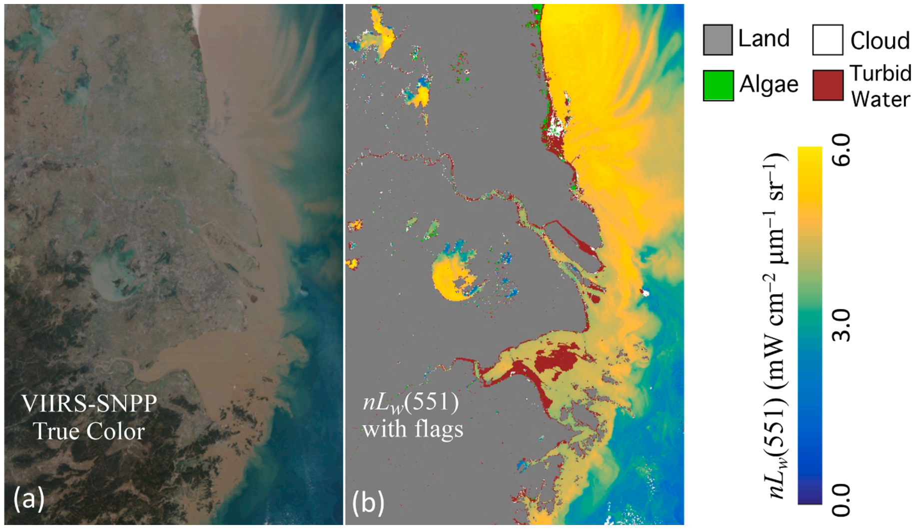

Fig. 8. A case study from VIIRS-SNPP on 15 February 2017 at 5:04 UTC over the China East Coast region for (a) true color image and (b) MSL12-derived $nL_w(555)$ image with pixels recovered from the original cloud masking and identified as (noted in flag color) turbid waters (brown), floating algae (green), and remaining cloud (white). Note that the flag colors in panel b are only for the recovered pixels.

happens near the cloud edge, the issue can be greatly alleviated by applying the straylight and cloud shadow flag (Jiang and Wang, 2013), which identifies those pixels near the cloud edge that are contaminated by straylight or cloud shadows. In fact, the straylight and cloud shadow detection algorithm has been implemented in the MSL12 for routine data processing and can be used to mask most of those problematic cloud edge pixels. It is noted that the straylight mentioned here is caused by the bright cloud (near cloud edge) (Jiang and Wang, 2013).

As a comparison for this GOCI case, Fig. 8 shows a portion of VIIRS-SNPP granule obtained on the same day around the same time at 5:04 UTC on February 15, 2017, which focused on the highly turbid coastal area around the Hangzhou Bay. Comparing with Fig. 7, the false cloud masking is remarkably reduced due to the use of the SWIR 1240 nm band for cloud masking ($\rho_{rc}(1238) > 0.0235$). In Fig. 8b, we notice that $nL_w(551)$ can be as high as ~6.0 mW cm^{-2} μm^{-1} sr^{-1}, indicating extremely turbid waters in the region (Shi and Wang, 2014). There are still some false cloud masks even though the SWIR-based cloud masking was used, and the proposed recovery method successfully recovered those pixels as turbid waters (brown). We notice some unrecovered false cloud pixels (shown in white) in the middle of a turbid water mass due to errors in land mask (Mikelsons et al., 2021), since these are not really water pixels if one closely examines the true color image. We also notice some pixels erroneously recovered as floating algae near the northern-most part of the image, which is due to the saturation issue of VIIRS-SNPP 745 nm band (Cao et al., 2013). It sometimes shows abnormally large TOA radiance values at 745 nm band when this band is saturated (e.g., over bright cloud pixels) (Cao et al., 2013), generating abnormally high nAFAI values. To deal with this issue, we have implemented a procedure to have these pixels masked as bad Level-1B data.

5.2. Cases of OLCI-S3A and GOCI on 25 June 2016

Fig. 9 shows a portion of OLCI-S3A image obtained at 2:03 UTC on June 25, 2016. There was a remarkable floating algal bloom in south of

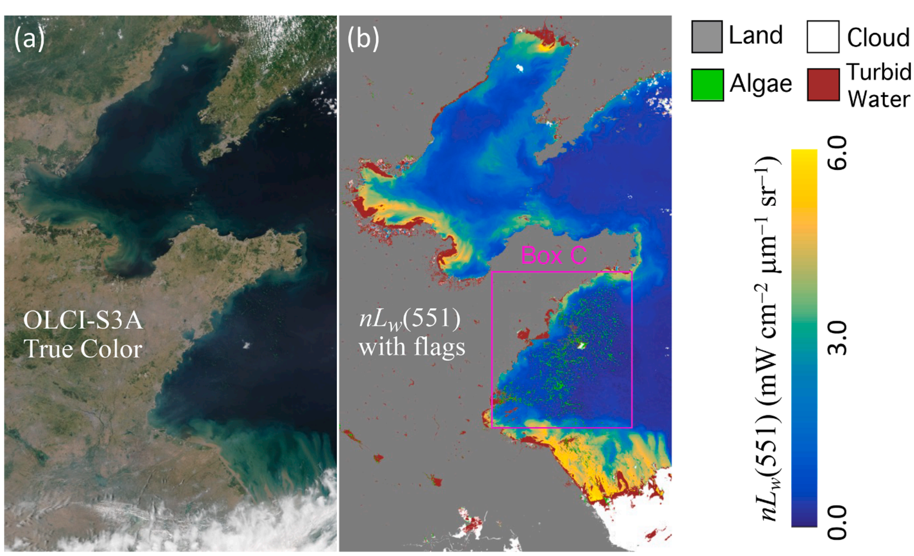

Fig. 9. A case study from OLCI-S3A on 25 June 2016 at 2:01 UTC over the Bohai Sea and Yellow Sea for (a) true color image and (b) MSL12-derived $nL_w(555)$ image with pixels recovered from the original cloud masking and identified as (noted in flag color) turbid waters (brown), floating algae (green), and remaining cloud (white). In panel b, it is highlighted floating algae pixels in Box C (flagged in green) that are recovered from the original cloud masking data processing. Note that the flag colors in panel b are only for the recovered pixels.

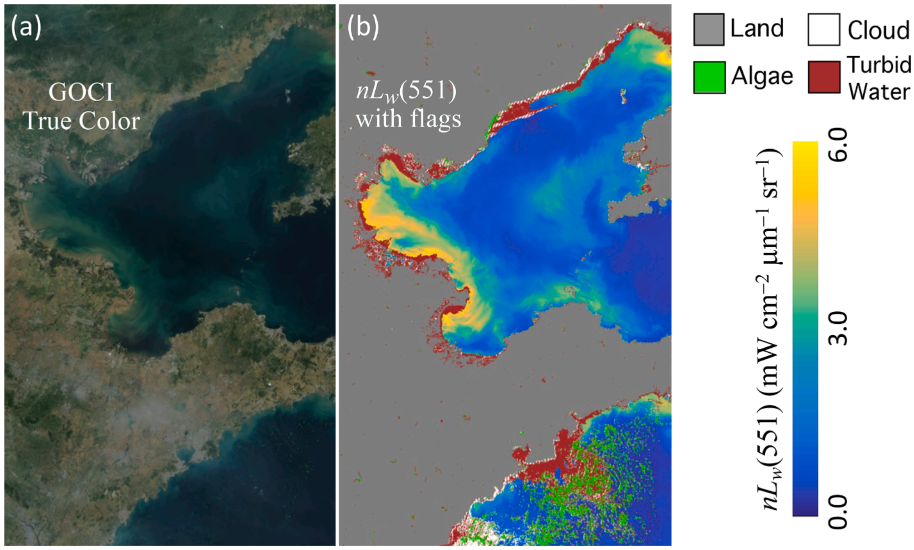

Fig. 10. A case study from GOCI on 25 June 2016 at 2:16 UTC over the Bohai Sea and near the Shandong Peninsula for (a) true color image and (b) MSL12-derived $nL_w(555)$ image with pixels recovered from the original cloud masking and identified as (noted in flag color) turbid waters (brown), floating algae (green), and remaining cloud (white). Note that the flag colors in panel b are only for the recovered pixels.

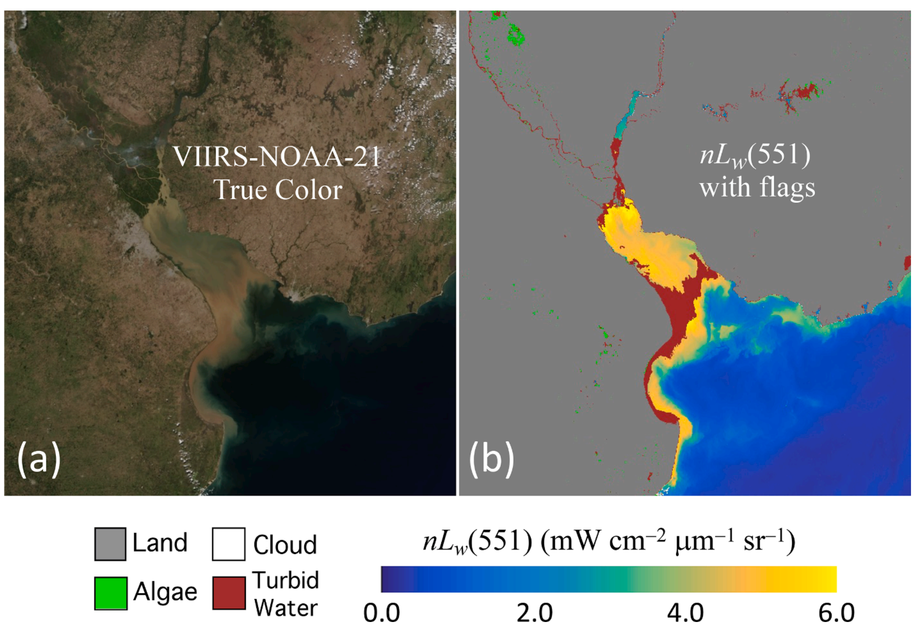

Fig. 11. A case study from VIIRS-NOAA-21 on 27 February 2023 at 17:45 UTC over the La Plata River Estuary in South America for (a) true color image and (b) MSL12-derived $nL_w(555)$ image with pixels recovered from the original cloud masking and identified as (noted in flag color) turbid waters (brown), floating algae (green), and remaining cloud (white). Note that the flag colors in panel b are only for the recovered pixels.

the Shandong Peninsula (shown in Box C), free from any interference from cloud, which makes this scene optimal to test the proposed cloud recovery algorithm. By comparing Fig. 9a and 9b, we confirm the success of the proposed algorithm in recovering false cloud masks from floating algae (green) and turbid water (brown) pixels, while keep the real cloud (white) pixels unchanged.

As a comparison, we show the same day GOCI image around the same area taken at 2:16 UTC on June 25, 2016 (Fig. 10). The proposed algorithm correctly recovered the floating algae in south of the Shandong Peninsula, as well as some turbid waters extended from coast. Compared with results from OLCI, which uses the longer wavelength (1020 nm) for cloud masking, the GOCI false cloud masking is more extensive. The proposed algorithm worked reasonably well in recovering most of turbid waters and floating algae, but left a small percentage of pixels unrecovered near the slot boundary (the southernmost part of the region) due to its inferior data quality resulted from the ISRD issue (Kim et al., 2015).

5.3. VIIRS-NOAA-21 measurements over the La Plata River Estuary

Fig. 11 shows an example of a more recent VIIRS-NOAA-21 image obtained at 17:45 UTC on February 27, 2023, over the La Plata River Estuary. As one of the most turbid water regions in the world, the La Plata River Estuary is a region with a highly changing environment on the ocean physical, optical, biological, and biogeochemical processes (Dogliotti et al., 2016; Moreira et al., 2013; Shi and Wang, 2020). Despite of using the SWIR 1241 nm band for cloud masking in VIIRS-NOAA-21, the extremely turbid waters in the La Plata River Estuary

Table 4
Pixel recovery statistics from the original cloud masking for the five study cases with image results presented in Figs. 7–11.

Cases	Satellite Sensor	Date/Time	# Total Cloud Mask	% Recovered Floating Algae	% Recovered Turbid Waters	% Remained Cloud Mask
Fig. 7	GOCI-COMS	2017.2.15/0516Z	1.87×10^6	0.65	84.12	15.23
Fig. 8	VIIRS-SNPP	2017.2.15/0504Z	1.64×10^4	7.74	77.31	14.95
Fig. 9	OLCI-S3A	2016.6.25/0203Z	4.45×10^5	11.13	47.13	41.74
Fig. 10	GOCI-COMS	2016.6.25/0216Z	9.85×10^4	21.19	62.38	16.43
Fig. 11	VIIRS-NOAA-21	2023.2.27/1745Z	1.27×10^4	10.13	88.59	1.28

have still triggered the false cloud masks as shown in Fig. 11b (brown), as well as for the turbid inland lakes and rivers in the region. Indeed, the recovered regions from false cloud masks in Fig. 11b generally have the highest turbidity in the La Plata River Estuary (Dogliotti et al., 2016; Shi and Wang, 2020). In this case, the sky in the study area was almost cloud free and the proposed algorithm worked quite well, recovered $\sim 1.25 \times 10^4$ false cloud pixels.

To get a more quantitative understanding of how the proposed algorithm performs, cloud pixel recovery statistics for the five cases described above are listed in Table 4. Results in Table 4 shows that, using the new proposed cloud masking algorithm, significant number of pixels are recovered from false cloud masks, with total recovery from original cloud pixels (the sum of columns 5 and 6 in Table 4) of 84.77%, 85.05%, 58.26%, 83.57%, and 98.72% for cases in Figs. 7–11, respectively. It should be noted that all these cases are over highly turbid coastal and inland waters, which most likely will have the cloud masking issue.

5.4. Pixel recovery statistics for the China East Coast

Although Table 4 shows cloud masking recovery rate of ~60–99% for these cases, the numbers may not be representative since these cases are subjectively selected, e.g., cases for easy explanation and demonstration. To get a more objective understanding, we proceed with a more systematic way of extracting the cloud recovery statistics (focusing on coastal and inland waters). We select a region of interest over the Bohai Sea, Yellow Sea, and East China Sea (Box A in Fig. 7a) covering region of (27°N–42°N, 117°E–127°E), and count the pixels recovered from the original cloud masking for all the pixels that fall within the region of interest for specify periods. For VIIRS-SNPP, all granules are used, while for GOCI only one observation at around 4:16 UTC (or 1:16 pm local time) out of eight per day is used, which is the closest to the VIIRS-SNPP local equator crossing time of 1:30 pm.

We select four months (January, April, July, and October) in 2017 representative of different seasons (winter, spring, summer, and autumn) throughout the year and listed in Table 5 the total pixel numbers from the original cloud masking and percentages of pixels recovered using the proposed algorithm for each month respectively, as well as for the sum of the four months. Both VIIRS-SNPP and GOCI show most cloud masks during January (winter), when waters are the most turbid due to strong winds (Shi and Wang, 2014), which increases chance of false cloud masking. The recoveries using the proposed algorithm are around 30% and 33% for VIIRS-SNPP and GOCI, respectively.

Table 5
Statistics in number of pixels from the original cloud masking (CM) (either in the NIR or SWIR method) and percentage (%) of recovered pixels using the proposed algorithm for VIIRS-SNPP and GOCI in the region of Box A in Fig. 7 for the four months of January, April, July, and October in 2017, as well as from 4-month total.

Time Period (2017)	VIIRS-SNPP Pixels in CM	% Recovered	GOCI Pixel in CM	% Recovered
January	3.42×10^7	27.70	8.65×10^7	36.72
April	2.94×10^7	29.02	7.20×10^7	32.71
July	2.93×10^7	30.49	7.31×10^7	27.94
October	3.12×10^7	31.97	7.76×10^7	33.28
4-Month	1.24×10^8	29.75	3.09×10^8	32.85

5.5. Statistics of recovered pixels in Lake Taihu

Here we focus on the highly turbid inland lake—Lake Taihu in China, to assess the performance of the proposed algorithm. Because of the important impacts of the lake water quality on this one of the world's most heavily populated regions with the highest economic development, Lake Taihu has been well studied in the last decade or so, particularly from satellite remote sensing (Duan et al., 2012; Hu et al., 2022; Shi et al., 2015; Wang et al., 2011). The region of interest is (30.9°N–31.6°N, 119.87°E–120.63°E) (Box B in Fig. 7b), and include OLCI-S3A in addition to VIIRS-SNPP and GOCI. All OLCI and VIIRS granules are used, while only 1:16 pm observation from GOCI is used in the calculation. The same four months of January, April, July, and October in 2017 are used, for which the statistics are calculated and shown in Table 6. Results in Table 6 show that the percentages of pixels recovered are remarkably high in this case, reflecting significantly higher chance of false cloud masking for Lake Taihu due to its extremely turbid waters. This also shows that the proposed cloud masking algorithm can significantly improve the coverage of ocean color data for such inland lakes globally from satellite measurements.

Table 6
Statistics in number of pixels from the original cloud masking (CM) (either in the NIR or SWIR method) and percentage (%) of recovered pixels using the proposed algorithm for VIIRS-SNPP, GOCI, and OLCI-S3A for Lake Taihu (Box B in Fig. 7b) in the four months of January, April, July, and October in 2017, as well as from 4-month total.

Time Period (2017)	VIIRS-SNPP Pixels in CM	% Recovered	GOCI Pixels in CM	% Recovered	OLCI-S3A Pixel in CM	% Recovered
January	9.78×10^4	40.22	2.77×10^5	49.20	4.23×10^5	40.30
April	1.04×10^5	54.95	2.68×10^5	51.29	4.32×10^5	45.07
July	8.42×10^4	50.48	2.49×10^5	49.19	3.76×10^5	46.55
October	1.02×10^5	46.92	2.76×10^5	41.16	3.59×10^5	38.22
4-Month	3.87×10^5	48.15	1.07×10^6	47.65	1.59×10^6	42.60

6. Discussion

The new cloud masking scheme has been shown effective, yet we are aware of some issues. Some thin cloud pixels may not be distinguishable from floating algae, particularly pixels nearby clouds (e.g., comparing results in Fig. 4a and 4b), and there can be misidentifications both ways. This issue can be greatly resolved using the straylight and cloud shadow flag (Jiang and Wang, 2013) to mask these data. Similarly, there may be an issue for overlapping in $\varepsilon^{(RC)}(\lambda_1, \lambda_2)$ values between cloud and turbid water pixels (Fig. 5c), although the rate of mis-identification is generally low. In addition, VIIRS has single-gain M6 band saturation issue. Such saturation usually occurs over bright clouds, generating erroneous $\rho_{rc}(\mathrm{NIR}_1)$ values used in calculation of AFAI, nAFAI, and $\varepsilon^{(RC)}(\lambda_i, \lambda_j)$. To deal with this issue, we have implemented a procedure to use the quality flag of VIIRS Level-1B data to identify saturated pixels flagged and exclude them from the cloud recovery process. It should also be noted that there are errors for recovery of real cloud pixels using the new cloud masking approach. For example, from the original cloud pixels of 2.17×10^6 in the Fig. 2 case, using the new cloud masking (after the straylight flag) resulted in mis-identifying (mis-recovery of) ~0.2% and ~0.5% cloud pixels as the floating algae and turbid waters, respectively.

Some multi-sensor case studies emphasizing on extremely turbid waters and intensive floating algae are presented, showing the effectiveness of the proposed cloud recovery algorithm in identifying falsely masked pixels. Further multi-sensor statistical results show that, in turbid water regions, such as the Bohai Sea, Yellow Sea, and East China Sea, the proposed algorithm can significantly reduce falsely masked pixels by ~30–40%. We also provide results from a case over the highly turbid La Plata River Estuary, showing that almost all falsely masked cloud pixels are recovered. For turbid inland lakes like Lake Taihu, the monthly statistical results of cloud mask reductions are significant, i.e., around 40–50%.

7. Conclusion

This study describes a new cloud masking scheme that has already been implemented into the MSL12 ocean color data processing for routine global ocean color data processing for various satellite ocean color sensors including VIIRS, OLCI, SGLI, GOCI, etc. Specifically, in the new proposed cloud masking algorithm, we have used algae indices AFAI and nAFAI, along with parameters $\rho_{rc}(\mathrm{NIR}_2)$, $\varepsilon^{(RC)}(\mathrm{NIR}_1, \mathrm{NIR}_2)$, and $\varepsilon^{(RC)}(\mathrm{Blue}, \mathrm{NIR}_2)$, to identify clear sky, clouds (thin/thick clouds), floating algae (including intensive ones), turbid waters (including extremely turbid waters), and absorbing aerosols. Particularly, the new parameter nAFAI has been developed to distinguish highly intensive floating algae from thick clouds. This is quite useful because there is effectively no maximum $\rho_{rc}(\lambda)$ limitation at the NIR or SWIR bands for recovery of falsely masked pixels. The new cloud masking scheme significantly improves the original cloud masking (using the NIR or SWIR method) by performing certain recovery procedures that remove false cloud masks based on specific criteria and identify either floating algae or turbid waters/absorbing aerosols with new appropriate flags. Therefore, global ocean/water properties and various surface algae features can be routinely derived using satellite ocean color remote sensing.

CRediT authorship contribution statement

Menghua Wang: Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Lide Jiang:** Writing – review & editing, Writing – original draft, Investigation, Formal analysis, Data curation.

Declaration of competing interest

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

Acknowledgments

This research was supported by the Joint Polar Satellite System (JPSS) funding. We thank four anonymous reviewers for their useful comments. The scientific results and conclusions, as well as any views or opinions expressed herein, are those of the author(s) and do not necessarily reflect those of NOAA or the Department of Commerce.

Data availability

Data will be made available on request.

References

Ackerman, S.A., Strabala, K.I., Menzel, W.P., Frey, R.A., Moeller, C.C., Gumley, L.E., 1998. Discriminating clear sky from clouds with MODIS. J. Geophys. Res. 103, 32141–32157.

Banks, A.C., Mélin, F., 2015. An assessment of cloud masking schemes for satellite ocean colour data of marine optical extremes. Int. J. Remote Sens. 36, 797–821.

Cao, C., Xiong, X., Blonski, S., Liu, Q., Uprety, S., Shao, X., Bai, Y., Weng, F., 2013. Suomi NPP VIIRS sensor data record verification, validation, and long-term performance monitoring. J. Geophys. Res. Atmos. 118, 11664–11678.

Carroll, M.L., DiMiceli, C.M., Townshend, J.R., Sohlberg, R.A., Elders, A.I., Devadiga, S., Sayer, A.M., Levy, R.C., 2017. Development of an operational land water mask for MODIS Collection 6, and influence on downstream data products. Int. J. Digit. Earth 10, 207–218.

Choi, J.K., Park, Y.J., Ahn, J.H., Lim, H.S., Eom, J., Ryu, J.H., 2012. GOCI, the world's first geostationary ocean color observation satellite, for the monitoring of temporal variability in coastal water turbidity. J. Geophys. Res. 117, C09004. https://doi.org/10.1029/2012JC008046.

Cox, C., Munk, W., 1954. Measurements of the roughness of the sea surface from photographs of the sun's glitter. Jour. Opt. Soc. of Am. 44, 838–850.

Dogliotti, A.I., Ruddick, K., Guerrero, R., 2016. Seasonal and inter-annual turbidity variability in the Rio de la Plata from 15 years of MODIS: El Nino dilution effect. Estuar. Coast. Shelf Sci. 182, 27–39.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.-H., Femenias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., Nieke, J., Rebhan, H., Seitz, B., Stroede, J., Sciarra, R., 2012. The global monitoring for environment and security (GMES) sentinel-3 mission. Remote Sens. Environ. 120, 37–57.

Duan, H., Ma, R., Hu, C., 2012. Evaluation of remote sensing algorithms for cyanobacterial pigment retrievals during spring bloom formation in several lakes of East China. Remote Sens. Environ. 126, 126–135.

Goldberg, M.D., Kilcoyne, H., Cikanek, H., Mehta, A., 2013. Joint polar satellite system: The United States next generation civilian polar-orbiting environmental satellite system. J. Geophys. Res. Atmos. 118, 13463–13475.

Groom, S., Sathyendranath, S., Ban, Y., Bernard, S., Brewin, R., Brotas, V., Brockmann, C., Chauhan, P., Choi, J., Chuprin, A., Ciavatta, S., Cipollini, P., Donlon, C., Franz, B.A., He, X., Hirata, T., Jackson, T., Kampel, M., Krasemann, H., Lavender, S.J., Pardo-Martinez, S., Melin, F., Platt, T., Santoleri, R., Skakala, J., Schaeffer, B., Smith, M., Steinmeta, F., Valente, A., Wang, M., 2019. Satellite ocean colour: Current status and future perspective. Front. Mar, Sci 6, 485. https://doi.org/10.3389/fmars.2019.00485.

Hu, C., 2009. A novel ocean color index to detect floating algae in the global oceans. Remote Sens. Environ. 113, 2118–2129.

Hu, M., Ma, R., Xiong, J., Wang, M., Cao, Z., Xue, K., 2022. Eutrophication state in the Eastern China based on Landsat 35-year observations. Remote Sens. Environ. 277, 113057. https://doi.org/10.1016/j.rse.2022.113057.

IOCCG, 2008. Why ocean colour? the societal benefits of ocean-colour technology. In: Platt, T., Hoepffner, N., Stuart, V., Brown, C. (Eds.), Reports of the International Ocean-Colour Coordinating Group. IOCCG, Dartmouth, Canada. https://doi.org/10.25607/OBP-97.

IOCCG, 2010. Atmospheric correction for remotely-sensed ocean-colour products. In: Wang, M. (Ed.), Reports of the International Ocean-Colour Coordinating Group. IOCCG, Dartmouth, Canada. https://doi.org/10.25607/OBP-101.

Jiang, L., Wang, M., 2013. Identification of pixels with stray light and cloud shadow contaminations in the satellite ocean color data processing. Appl. Opt. 52, 6757–6770.

Kim, W., Ahn, J.H., Park, Y.J., 2015. Correction of stray-light-driven interslot radiometric discrepancy (ISRD) present in radiometric products of geostationary ocean color imager (GOCI). Ieee T Geosci Remote 53, 5458–5472.

King, M.D., Platnick, S., Menzel, W.P., Ackerman, S.A., Hubanks, P.A., 2013. Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites. IEEE Trans. Geosci. Remote Sens. 51, 3826–3852.

McClain, C.R., Feldman, G.C., Hooker, S.B., 2004. An overview of the SeaWiFS project and strategies for producing a climate research quality global ocean bio-optical time series. Deep Sea Res. Part II 51, 5–42.

Mikelsons, K., Wang, M., Wang, X., Jiang, L., 2021. Global land mask for satellite ocean color remote sensing. Remote Sens, Environ 257, 112356. https://doi.org/10.1016/j.rse.2021.112356.

Moreira, D., Simionato, C.G., Gohin, F., Cayocca, F., Tejedor, L.C., M.,, 2013. Suspended matter mean distribution and seasonal cycle in the Río de La Plata estuary and the adjacent shelf from ocean color satellite (MODIS) and in-situ observations. Cont. Shelf Res. 68, 51–66.

Robinson, W.D., Franz, B.A., Patt, F.S., Bailey, S.W., Werdell, P.J., 2003. Masks and flags updates. NASA Goddard Space Flight Center, Greenbelt, Maryland, pp. 34–40.

Salomonson, V.V., Barnes, W.L., Maymon, P.W., Montgomery, H.E., Ostrow, H., 1989. MODIS: advanced facility instrument for studies of the Earth as a system. IEEE Trans. Geosci. Remote Sens. 27, 145–153.

Shi, K., Zhang, Y., Zhu, G., Liu, X., Zhou, Y., Xu, H., Qin, B., Liu, G., Li, Y., 2015. Long-term remote monitoring of total suspended matter concentration in Lake Taihu using 250m MODIS-Aqua data. Remote Sens. Environ. 164, 43–56.

Shi, W., Wang, M., 2007. Detection of turbid waters and absorbing aerosols for the MODIS ocean color data processing. Remote Sens. Environ. 110, 149–161.

Shi, W., Wang, M., 2009. An assessment of the black ocean pixel assumption for MODIS SWIR bands. Remote Sens. Environ. 113, 1587–1597.

Shi, W., Wang, M., 2014. Ocean reflectance spectra at the red, near-infrared, and shortwave infrared from highly turbid waters: A study in the Bohai Sea, Yellow Sea, and East China Sea. Limnol. Oceanogr. 59, 427–444.

Shi, W., Wang, M., 2020. Water properties in the La Plata River Estuary from VIIRS observations. Continent. Shelf Res. 198, 104100.

Wang, M., 2016. Rayleigh radiance computations for satellite remote sensing: Accounting for the effect of sensor spectral response function. Opt. Express 24, 12414–12429.

Wang, M., Ahn, J.H., Jiang, L., Shi, W., Son, S., Park, Y.J., Ryu, J.H., 2013a. Ocean color products from the Korean Geostationary Ocean Color Imager (GOCI). Opt. Express 21, 3835–3849.

Wang, M., Bailey, S., 2001. Correction of the sun glint contamination on the SeaWiFS ocean and atmosphere products. Appl. Opt. 40, 4790–4798.

Wang, M., Hu, C., 2016. Mapping and quantifying Sargassum distribution and coverage in the Central West Atlantic using MODIS observations. Remote Sens. Environ. 183, 350–367.

Wang, M., Liu, X., Tan, L., Jiang, L., Son, S., Shi, W., Rausch, K., Voss, K., 2013b. Impact of VIIRS SDR performance on ocean color products. J. Geophys. Res. Atmos. 118, 10347–10360.

Wang, M., Shi, W., 2006. Cloud masking for ocean color data processing in the coastal regions. IEEE Trans. Geosci. Remote Sens. 44, 3196–3205.

Wang, M., Shi, W., Tang, J., 2011. Water property monitoring and assessment for China's inland Lake Taihu from MODIS-Aqua measurements. Remote Sens. Environ. 115, 841–854.

Wei, J., Wang, M., 2024. Mapping ocean surface algal blooms with SWIR-derived satellite remote sensing reflectance. Int. J. Appl. Earth Obs. Geoinf. 131, 103921. https://doi.org/10.1016/j.jag.2024.103921.

Xing, W., Guo, B., Sheng, Y., Yang, X., Ji, M., Xu, Y., 2022. Tracing surface water change from 1990 to 2020 in China's Shandong Province using Landsat series images. Ecol. Ind. 140, 108993. https://doi.org/10.1016/j.ecolind.2022.108993.

Yang, Y., Ci, F., Xu, A., Zhang, X., Ding, N., Wan, N., Lv, Y., Song, Z., 2024. Seasonal dynamics of eukaryotic microbial communities in the water-receiving reservoir of the long-distance water diversion project. China. Microorganisms 12. https://doi.org/10.3390/microorganisms12091873.