High performance of chlorophyll-a prediction algorithms based on simulated OLCI Sentinel-3A bands in cyanobacteria-dominated inland waters

Fernanda Sayuri Yoshino Watanabe a,⇑, Enner Alcântara b, José Luiz Stech c

a Department of Cartography, Universidade Estadual Paulista (UNESP), Rua Roberto Simonsen, 305, Presidente Prudente, SP 19060-900, Brazil
b Department of Environmental Engineering, Universidade Estadual Paulista (UNESP), Rodovia Presidente Dutra, Km 137.8, Eugenio de Melo, 12247-004 São José dos Campos, SP, Brazil
c Remote Sensing Division, Instituto Nacional de Pesquisas Espaciais (INPE), Avenida dos Astronautas, 1758, São José dos Campos, SP 12227-010, Brazil

Received 15 December 2017; received in revised form 6 April 2018; accepted 19 April 2018
Available online 1 May 2018

Abstract

In this research, we have investigated whether the chlorophyll-a (chl a) retrieval algorithms based on OLCI Sentinel-3A bands are suitable for cyanobacteria-dominated waters. Phytoplankton assemblages model optical properties of the water, influencing the performance of bio-optical algorithms. Understanding these processes is important to improve the prediction of photoactive pigments in order to use them as a proxy for trophic state and harmful algal bloom. So that, both empirical and semi-analytical approaches designed for different inland waters were tested. In addition, empirical models were tuned based on dataset collected in situ. The study was conducted in the Funil hydroelectric reservoir, where chl a ranged from 2.33 to 208.68 mg m⁻³ in May 2012 (austral fall) and 4.37 to 306.03 mg m⁻³ in October 2012 (austral spring). OLCI Sentinel-3A bands were tested in existing algorithms developed for other sensors and new band combinations were compared to analyze the errors produced. Normalized Difference Chlorophyll Index (NDCI) exhibited the best performance, with a Normalized Root Mean Square Error (NRMSE) of 9.30%. Result showed that wavelength at 665 nm is adequate to estimate chl a, although the maximum pigment absorption band is shifted due to phycocyanin fluorescence at approximately 650 nm.

Keywords: Harmful algal bloom; Case-2 waters; Remote sensing; Water quality

1. Introduction

The wastewater discharge in freshwater bodies causes unbalance in aquatic ecosystem such as eutrophication and fish kills (Anderson et al., 2002). Eutrophication leads to growing the primary productivity (Anderson et al., 2002), which may become the water inappropriate for multiple uses such as drinking water supply. In reservoirs, the eutrophication may be more intense due to water level control, increasing the water residence time and providing nutrients for longer time (Soares et al., 2008, 2012). In these condition it is common to detect harmful algal blooms and, therefore, their monitoring is important to keep the quality of the aquatic life and public health.

In natural waters, phytoplankton assemblage is composed for different species, but there can be dominance of specific species depending on the aquatic system conditions. Cyanobacteria dominance is commonly associated with high retention time (Carvalho et al., 2011; Romo et al., 2013), water column mixing in tropical waters (Calijuri and dos Santos, 1996) and high trophic state (Calijuri and dos Santos, 1996; Ogashawara et al., 2013; Mishra et al., 2014; Costa et al., 2014). Variations in phytoplankton communities

* Corresponding author.
E-mail address: fernanda@fct.unesp.br (F.S.Y. Watanabe).
consequently change the optical properties (Richardson, 1996; Ruiz-Verdú et al., 2008). Ruiz-Verdú et al. (2008) compared the absorption coefficients of cyanobacteria, chlorophyll-a and diatomaceous and showed that the maximum cyanobacteria absorption peak in the red region is shifted into longer wavelengths, probably caused by phycocyanin fluorescence at approximately 650 nm (Schalles and Yacobi, 2000).

Chlorophyll-a (chl a) content, pigment found in all phytoplankton species, has been used as proxy of their biomass and, due to its photoactive characteristic remotely sensed imagery becomes a powerful tool to monitor harmful algal blooms. Several bio-optical algorithms have been designed to retrieve the chl a concentration in inland waters adopting different band combinations (Dekker, 1993; Dall’Olmo et al., 2003; Mishra and Mishra, 2012). In inland waters, red and near-infrared (NIR) bands are widely used in order to reduce the interference of other optically significant constituents (OSC) (Gilerson et al., 2010).

Different remote sensors have been tested to achieve the best performance, such as Landsat (Watanabe et al., 2015; Bernardo et al., 2017), MODIS (Carder et al., 1999; Mishra and Mishra, 2010), MERIS (Moses et al., 2009; Gilerson et al., 2010; Gurlin et al., 2011; Mishra and Mishra, 2012) and hyperspectral systems (Olmanson et al., 2013; Tan et al., 2015). However, in general, the phytoplankton species are not always considered in parameterizing chl a content and therefore the hypothesis might be responded. The aim of this work was to investigate if phycocyanin fluorescence in the red light region caused by phycocyanin fluorescence.

With the launch of the Ocean and Land Colour Instrument (OLCI) onboard Sentinel-3A, there is a new alternative for mapping chl a in inland waters and, therefore, the hypothesis might be responded. The aim of this work was to investigate if phycocyanin fluorescence in the red spectral region impairs the performance of chl a prediction algorithms based on OLCI Sentinel-3A bands in the red and NIR spectral region. In turn, the specific objectives were (a) to evaluate existing empirical NIR-red algorithms and semi-analytical algorithms based on absorption at 665 nm; (b) to calibrate NIR-red algorithms considering the dataset acquired in a cyanobacteria-dominated aquatic system; (c) to design a chl a prediction algorithm based on phycocyanin feature at 620 nm; and (d) to compare the performance of all tested algorithms in estimating chl a. We expected that the findings of this study could be useful to understand the influence of the spectral features of phytoplankton pigments in estimating chl a content and therefore to monitoring the trophic status.

2. Data and methods

2.1. Study area

The study area was the Funil hydroelectric reservoir (FHR, 22°33′48.97″S, 44°36′13.24″W), located in the Paraíba do Sul River, in the southern Rio de Janeiro State, Brazil (Fig. 1). Operating since 1969, the FHR has 40 km² flooded area and 890 × 10⁶ km³ volume, with an average retention time of 32 days. Paraíba do Sul River is located in a populous and industrialized region, responsible for a high sewage discharge, causing eutrophication in the reservoir (Soares et al., 2012). Eutrophication and retention time have been pointed as responsible for intense cyanobacteria blooms in the Funil reservoir (Rangel et al., 2012; Soares et al., 2012; Ogashawara et al., 2013).

2.2. Field sampling

Two field campaigns were conducted on May 20–22, 2012 (austral autumn and period of high water level) and September 2–4, 2012 (austral spring and period of low water level). In May 19 samples were collected, while 10 samples were acquired in September. Location of the sampling points at each field campaign is shown in Fig. 1c. Water samples were collected to estimate the chl a concentration (Nusch, 1980) and suspended material (total, inorganic and organic) (APHA, 1998) in the laboratory.

Field spectroscopy data were measured using two RAMSES spectroradiometers (TriOS, Rastede, Germany), acquiring measurements of radiance (ARC-VIS sensor, with 7° field of view) and irradiance (ACC-VIS sensor, with cosine collector). The radiometers work between 320 and 950 nm, with a spectral sampling of 3.3 nm and spectral accuracy of 0.3 nm. These data were used to calculate the remote sensing reflectance (Rs, in sr⁻¹) according to Mobley (1999). OLCI Sentinel-3A bands were simulated from in situ Rs and spectral response function of the sensor as:

\[
R_s^{OLCI}(\lambda_k) = \frac{\int_{\lambda_k}^{\lambda_{k+1}} S(\lambda)R_s(\lambda) d\lambda}{\int_{\lambda_k}^{\lambda_{k+1}} S(\lambda) d\lambda}
\]  

(1)

where \(R_s^{OLCI}\) stands for the remote sensing reflectance convoluted from OLCI spectral bands; \(\lambda_i\) and \(\lambda_f\) are the lower and upper limit of the band \(\lambda_k\), respectively; and \(S(\lambda)\) is the spectral response function of the \(n\)th spectral band of OLCI.

2.3. Chlorophyll-a model calibration and validation

Semi-empirical NIR-red based algorithms and semi-analytical using data in the red spectral region were tested adopting the datasets collected in May and September 2012. The NIR-red algorithms were adopted to estimate chl a concentration: two-band simple ratio (2B; Dekker 1993), three-band (3B; Dall’Olmo et al., 2003), and Normalized Difference Chlorophyll Index (NDCI; Mishra and Mishra, 2012). Table 1 shows the structure for each band algorithm. Bands 4, 5 and 6 (central wavelength at \(\lambda_1 = 665\ \text{nm}, \lambda_2 = 708.75\ \text{nm}\) and \(\lambda_3 = 753.75\ \text{nm}\), respectively) from OLCI Sentinel 3A were adopted in the models.
These algorithms were chosen because they were parameterized for productive waters.

Two semi-analytical (SA) algorithms based on 665 nm (Gons et al., 2008; Gilerson et al., 2010) were tested to estimate chl \( a \) concentration. Such algorithms request values of the phytoplankton pigment absorption coefficient \( (a_u) \) and/or non-water constituent absorption coefficient \( (a_t/C_w) \), but which neither were collected during the fieldworks nor estimated in laboratory. Therefore, the Inherent optical properties Inversion Model of Inland Waters (IIMIW) proposed by Li et al. (2013) was used to retrieve the absorption coefficients required in semi-analytical modeling. IIMIW (Li et al. 2013) was selected because it was parameterized based on a wide dataset from four different freshwaters. In addition, chl \( a \) range used by them cover concentration of chl \( a \) found in our study area.

Table 1 shows the calibrations tested in this study and the selection were based on researches conducted in turbid productive waters. Existing calibrations were tested both band algorithms (Moses et al., 2009; Gilerson et al., 2010; Mishra and Mishra, 2012; Gurlin et al., 2011) and semi-analytical models (Ritchie, 2008; Gons et al., 2008). Besides calibrations proposed for other aquatic systems we recalibrated the algorithms considering dataset collected in May 2012. The adjustment was conducted using the least square method, attempting linear and polynomial fits.

In cyanobacteria-dominated waters it is expected that phycocyanin and chl \( a \) are highly correlated and consequently with the phytoplankton biomass (Shi et al., 2015). Therefore, although the absorption feature around 620 nm is associated with phycocyanin pigment and not

Table 1
NIR-red algorithms and calibrations applied to estimate chl \( a \) concentration.

<table>
<thead>
<tr>
<th>Model</th>
<th>Formula</th>
<th>Calibration</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2B</td>
<td>( R_{665}^{-1} \times R_{708.75} )</td>
<td>MO09</td>
<td>Moses et al. (2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GI10</td>
<td>Gilerson et al. (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GU11</td>
<td>Gurlin et al. (2011)</td>
</tr>
<tr>
<td>3B</td>
<td>( [R_{665}^{-1} - (708.75)] \times R_{753.75} )</td>
<td>MO09</td>
<td>Moses et al. (2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GU11</td>
<td>Gurlin et al. (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MI12</td>
<td>Mishra and Mishra (2012)</td>
</tr>
<tr>
<td>NDCI</td>
<td>( R_{665}^{-1}/(R_{708.75}-R_{665}) )</td>
<td>GO08</td>
<td>Gons et al. (2008)</td>
</tr>
<tr>
<td>SA</td>
<td>( a_u(665) )</td>
<td>GH10</td>
<td>Gilerson et al. (2010)</td>
</tr>
<tr>
<td></td>
<td>( a_u(665) )</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1. Location of the study area, showing (a) its position in Brazil and (b) Rio de Janeiro state and the distribution of sampling stations in the (c) Funil hydroelectric reservoir, located in Paraíba do Sul River.
chl a, new algorithms were also parameterized replacing $\lambda_1 = 665$ nm for 620 nm in 2B and NDCI algorithms, and calibrated using the dataset collected in both field surveys in the FHR. The new algorithms were labeled as M2B and MNDCI, where M means modified. The adjustment of the algorithms was based on the least square method (linear and polynomial fits).

The assessment of all the algorithms was conducted using the following statistical metrics: Root Mean Square Error (RMSE) Normalized Root Mean Square Error (NRMSE) and determination coefficient ($R^2$). The September dataset was used in assessment of all algorithms tested. Additionally, the validation of the algorithms calibrated using the FHR dataset was based on Leave-One-Out Cross Validation (LOOCV). Such validation statistical technique is rather useful and widely used in cases where the sampling size is limited.

3. Results and discussion

3.1. Application of existing bio-optical models

The Funil reservoir exhibited a large variability of chl a concentrations ranging from 2.33 to 208.68 mg m$^{-3}$ in May and 4.37 to 306.03 mg m$^{-3}$ in September. Meanwhile, variation of the total suspended material (TSM) was from 3.38 to 64 mg L$^{-1}$, with predominance of organic fraction (average of 83%), and, therefore, TSM was mainly composed of phytoplankton biomass. Fig. 2 shows the chl a content variation observed from May to September 2012.

In situ $R_{rs}$ spectra and simulated bands are shown in Fig. 3. Analyzing the in situ $R_{rs}$ spectra the phycocyanin features associated with absorption at 620 nm and fluorescence at 650 nm were quite remarkable (Fig. 3a and b), whilst the first feature is realized at OLCI Sentinel-3A simulated spectra and the second one is lost (Fig. 3c and d). In addition, features of chl a absorption and particles reflectance around 680 nm and 715 nm, respectively, are more highlighted in September dataset.

Fig. 4 shows the validation results obtained by NIR-red algorithms and IIWIM-based semi-analytical models. Among the existing 2B algorithms, MO09 calibration showed the best performance (NRMSE of 13.65% and $R^2 = 0.92$), followed by GI10 (NRMSE = 19.77% and...
Chl-α range reported at 18 sampling stations and used by Moses et al. (2009) was from 0.63 to 65.61 mg m$^{-3}$, while Gilerson et al. (2010) adopted two thousand $R_s$ spectra simulated using Hydrolight, varying the chl-α concentration between 1 and 100 mg m$^{-3}$. Although Gurlin et al. (2011) have used a chl-α range (2.3–200.8 mg m$^{-3}$) very similar to that reported in FHR, GU11 showed the poorest performance (NRMSE of 118.81% and $R^2$ = 0.74).

Considering the existing algorithms, overall, the insertion of a third band at 753.75 nm did not improve the chl-α estimation. Despite that, MO09 calibration for 3B algorithm exhibited the best performance in estimating chl-α, with a NRMSE of 14.41% and $R^2$ of 0.81. Taking into account just existing algorithms, NDCI-MI12 using second-degree polynomial exhibited performance similar to 3B (NRMSE of 18.04% and $R^2$ of 0.98). Mishra and Mishra (2012) have used a synthetic dataset considering a chl-α range of 1–60 mg m$^{-3}$.

Semi-analytical models also exhibited good results in predicting chl-α concentration. Actually, comparing all the tested algorithms, IIMIW (Li et al., 2013) based on GO10 algorithm (Gons et al., 2008) exhibited the lower error (NRMSE of 10.16% and $R^2$ of 0.92). IIMIW was developed based on different turbid and productive inland waters. The researchers have used a dataset whose chl-α ranged from 1.85 to 285.80 mg m$^{-3}$, i.e., very close to conditions found in the FHR. Although we do not have absorption data to compare with the absorption retrieved by IIMIW, it is assumed that the performance of semi-analytical algorithms is directly related to quality of inherent optical properties used in parameterization. Therefore,

Fig. 4. Validation of existing algorithms using calibration parameters on the basis of RMSE (mg m$^{-3}$), NRMSE (%), bias (mg m$^{-3}$) and $R^2$ for (a) 2B-MO09, (b) 2B-GI10, (c) 2B-GU11, (d) 3B-MO09, (e) 3B-GU11, (f) NDCI-MI12, (g) SA-GO08 and (h) SA-GI10.
the results obtained by semi-analytical algorithms indicate IIMIW worked suitability in retrieving absorption coefficient in FHR. Gons et al. (2008) designed their algorithms for oligotrophic and eutrophic inland waters, with chl \( a \) ranging from 0.4 to 131 mg m\(^{-3}\) and even though the chl \( a \) range is lower than our dataset, one of the lakes studied by the researchers has presented abundance of cyanobacteria as Funil reservoir.

### 3.2. Parameterized and calibrated algorithms

After testing some algorithms designed for productive inland waters, such algorithms were tuned using dataset collected in the FHR. Additionally, the wavelength at 620 nm was tested to retrieve chl \( a \), due to phycocyanin absorption features observed in most \( R_{rs} \) spectra. The performance of these new algorithms was compared with their 650 nm counterparts. However, before a correlation analysis between OLCI Sentinel-3A plus the wavelength at 620 nm and the chl \( a \) was conducted in order to show statistical relevance. The \( R_{rs}(708.75)/R_{rs}(6 2 0) \) and \( R_{rs}(708.75)/R_{rs}(6 6 5) \) band ratios exhibited high correlation with such pigment \( (r = 0.892 \text{ and } r = 0.891, \text{ respectively}) \).

**Fig. 5** shows the relationship among the different indexes and chl \( a \) concentration.

---

### Table 2

<table>
<thead>
<tr>
<th>Model</th>
<th>( \beta_2 )</th>
<th>( \beta_1 )</th>
<th>( \beta_0 )</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2B</td>
<td>-3.9621</td>
<td>74.202</td>
<td>-39.476</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(-4.6227; 7.6944)</td>
<td>(28.846; 80.695)</td>
<td>(-43.743; -13.524)</td>
<td></td>
</tr>
<tr>
<td>3B</td>
<td>-114.23</td>
<td>363.44</td>
<td>41.601</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(-126.59; 401.48)</td>
<td>(253.48; 566.85)</td>
<td>(28.148; 43.616)</td>
<td></td>
</tr>
<tr>
<td>NDCI</td>
<td>294.49</td>
<td>119.51</td>
<td>19.688</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(196.79; 306.54)</td>
<td>(111.52; 129.6)</td>
<td>(17.579; 21.908)</td>
<td></td>
</tr>
<tr>
<td>M2B</td>
<td></td>
<td>40.836</td>
<td>-6.7387</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(38.426; -60.686)</td>
<td>(21.467; -5.9881)</td>
<td></td>
</tr>
<tr>
<td>MNDCI</td>
<td>234.59</td>
<td>152.53</td>
<td>32.019</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>(218.94; 244.55)</td>
<td>(146.97; 160.54)</td>
<td>(29.408; 35.061)</td>
<td></td>
</tr>
</tbody>
</table>
Table 2 shows the calibration coefficients obtained from the algorithms tuned for FHR using bands centered at 665, 708.75 and 753.75 nm (2B, 3B and NDCI) and 620 and 708.75 nm (M2B and MNDCI). In addition, it is presented the confidence interval of each coefficient determined from LOOCV. The validation has shown that all the algorithms tuned for FHR were statistically valid, since $b_0$, $b_1$ and $b_2$ were inserted within the 0.95 confidence interval.

Fig. 6 shows plots of measured versus modeled chl $a$ for new and tuned algorithms. All the algorithms fitted for FHR have shown good performance, with low NRMSE and bias values. Among the tuned algorithms, NDCI index (polynomial fit) exhibited the best performance (NRMSE = 3.06% and $R^2 = 0.983$), followed by MNDCI calibrated for FHR (NRMSE = 3.29% and $R^2 = 0.983$). The paired $t$-test showed that NDCI and MNDCI were statistically equal. Therefore, there would be no impairment in replacing $\lambda_2$ from 620 to 665 nm. The other algorithms exhibited similar performance, all of them not exceeding a NRMSE higher than 8%.

Fig. 6. Plots of measured versus modeled chl $a$ for tuned algorithms (a) 2B, (b) 3B, (c) NDCI as well as (d) M2B and (e) MNDCI fitted using 620 nm. Assessment of the algorithms was done based on RMSE (mg m$^{-3}$), NMRSE (%), bias (mg m$^{-3}$) and $R^2$.

Fig. 7. Relationships between (a) 2B and M2B indexes and (b) NDCI and MNDCI.
4. Discussion

The results obtained in this work showed that shifting the maximum phytoplankton absorption wavelength at red spectral region to longer wavelengths do not affect the chl \( a \) estimation using 665 nm. Comparing the algorithms with best performance, MNDCI with NDCI both calibrated for FHR using polynomial fit, they were considered statistically equal. Such results indicate that using either 665 nm or 620 nm present the same performance in estimating chl \( a \) in cyanobacteria-dominated waters. Fig. 7 shows the strong relationship between 2B and M2B indexes as well as NDCI and MNDCI indexes, with \( R^2 \) close to 1. In other words, analogous indexes were proportional, indicating there is no statistical difference in adopting either 620 nm or 665 nm. Furthermore, the results showed that OLCI Sentinel-3A bands are capable of retrieving chl \( a \) concentration accurately in cyanobacteria-dominated inland waters.

5. Conclusion

Considering the results obtained in this present research we concluded that NIR-red and SA algorithms are not impaired by shifting the maximum chl \( a \) absorption peak in the red spectral region. Both NIR-red and SA algorithms exhibited accurate performance in estimating chl \( a \) in FHR. Additionally, the OLCI Sentinel-3A data were rather suitable to estimate chl \( a \) content, encouraging its use in water quality monitoring. Bands centered at 665 nm and 709 nm were suitable to retrieve chl \( a \) in cyanobacteria-dominated waters, in spite of the maximum absorption by chl \( a \) is shifted at longer wavelengths around 680 nm. We have also verified that the use of 665 nm or 620 nm did not change remarkably the performance of M2B and MNDCI algorithms. The strong relationship between M2B and 2B ratios showed that both ratios are directly proportional and, therefore, they would produce similar results. Among the tested and recalibrated algorithms, NDCI index exhibited the best performance. Its structure was capable of highlighting the difference between chl \( a \) absorption in the red region and particles reflectance around 708.75 nm. Surprisingly, the SA algorithm presented better performance than the some existing band algorithms, even fitted to other aquatic systems. Although it has not been possible to evaluate the performance of IIMIW in estimating the inherent optical properties, the results indicated that the prediction of chl \( a \) was accurate, with the best performance among the tested existing algorithms using GO10 (NRMSE of 10.16%). Such algorithm used the wavelength at 665 nm, showing the efficiency of that spectral region in retrieving chl \( a \) content. Taking into account the findings in this research, more studies should be carried out in other cyanobacteria-dominated aquatic systems, especially, in relation to SA.

Acknowledgments

The authors thank to Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP No 2011/19523-8 and 2015/18525-8) and National Council of Technological and Scientific Development (CNPq No 471223/2011-5) for financial support under grants.

References


