Can chlorophyll-\(\text{a}\) in meso-oligotrophic shallow waters be estimated using statistical approaches and empirical models from MODIS imagery?

É possível estimar a Clorofila-\(\text{a}\) em águas rasas meso-oligotróficas usando abordagens estatísticas e modelos empíricos a partir de imagens MODIS?

Andres Mauricio Munar\(^1\), José Rafael Cavalcanti\(^1\), Juan Martin Bravo\(^1\), David Manuel Lelinho Da Motta Marques\(^1\) and Carlos Ruberto Fragoso Júnior\(^2\)

\(^1\)Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil

\(^2\)Centro de Tecnologia Maceió, Universidade Federal de Alagoas, AL, Brasil

*E-mails: mauriciomunar4@gmail.com (AMM); rafaelcavalcantii@gmail.com (JRC); jumarbra@iph.ufrgs.br (JMB); dmm@iph.ufrgs.br (DMLDMM); carlosruberto@gmail.com (CRFJ)

Received: August 30, 2017 - Revised: December 29, 2017 - Accepted: February 27, 2018

ABSTRACT

Accurate estimation of chlorophyll-\(\text{a}\) (Chl-\(\text{a}\)) concentration in inland waters through remote-sensing techniques is complicated by local differences in the optical properties of water. In this study, we applied multiple linear regression (MLR), artificial neural network (ANN), nonparametric multiplicative regression (NPMR) and four models (Appel, Kahru, FAI and O14a) to estimate the Chl-\(\text{a}\) concentration from combinations of spectral bands from the MODIS sensor. The MLR, NPMR and ANN models were calibrated and validated using in-situ Chl-\(\text{a}\) measurements. The results showed that a simple and efficient model, developed and validated through multiple linear regression analysis, offered advantages (i.e., better performance and fewer input variables) in comparison with ANN, NPMR and four models (Appel, Kahru, FAI and O14a). In addition, we observed that in a large shallow subtropical lake, where the wind and hydrodynamics are essential factors in the spatial heterogeneity (Chl-\(\text{a}\) distribution), the MLR model adjusted using the specific point dataset, performed better than using the total dataset, which suggest that would not be appropriate to generalize a single model to estimate Chl-\(\text{a}\) in these large shallow lakes from total datasets. Our approach is a useful tool to estimate Chl-\(\text{a}\) concentration in meso-oligotrophic shallow waters and corroborates the spatial heterogeneity in these ecosystems.

Keywords: Chl-\(\text{a}\); MLR; ANN; NPMR; Remote sensing; Lake Mangueira.

RESUMO

A estimativa da concentração de Clorofila–\(\text{a}\) (Chl-\(\text{a}\)) em ambientes aquáticos através de técnicas de sensoriamento remoto é complexa devido as diferenças entre as propriedades óticas da água. O objetivo deste trabalho foi estimar concentrações de Chl-\(\text{a}\) a partir das combinações de bandas espectrais do sensor MODIS, aplicando análise de regressão linear múltipla (MLR), redes neurais artificiais (ANN), regressão multiplicativa não paramétrica e quatro modelos (Appel, Kahru, FAI e O14a) para estimar a Chl-\(\text{a}\) concentração. Os modelos de regressão linear múltipla (MLR), NPMR e ANN foram calibrados e validados utilizando medições in-situ de Clorofila–\(\text{a}\). Os resultados demonstraram que um simples e eficiente modelo desenvolvido e validado através de regressão linear múltipla (MLR) produz vantagens (i.e., melhor performance e poucas variáveis de entrada) em comparação com os modelos ANN, NPMR e quatro modelos (Appel, Kahru, FAI e O14a). Além disso, foi observado que em um grande lago raso subtropical, no qual o vento (intensidade e direção) e a hidrodinâmica são fatores essenciais na heterogeneidade espacial (distribuição de Chl-\(\text{a}\)), o modelo MLR desenvolvido utilizando um ponto específico do lago teve melhor performance que utilizando o total de dados, sugerindo que não seria apropriado generalizar um único modelo para estimar Clorofila–\(\text{a}\) nesses grandes lagos rasos utilizando o total de dados. Esta abordagem é uma ferramenta útil para estimar concentrações de Chl-\(\text{a}\) em ambientes rasos meso–oligotróficos e corroborar a heterogeneidade espacial desses ecossistemas.

Palavras-chave: Chl-\(\text{a}\); MLR; ANN; NPMR; Sensoriamento remoto; Lagoa Mangueira.
INTRODUCTION

Chl-\(a\) is an important indicator of the trophic state in lakes and reservoirs, showing patterns associated with internal processes and natural stressors (HONEYWILL; PATERSO; HAGERTHEY, 2002; DUKA; CULLA, 2009; SCHALLES et al., 1998). Detection and quantification of Chl-\(a\) are essential for assessing water quality in these ecosystems, as the concentration of this compound provides critical knowledge of the phytoplankton community. However, most traditional methods used to retrieve Chl-\(a\) concentrations are based on \textit{in-situ} measurements, and are time-consuming and difficult to apply to large areas (GHOLIZADEH; MELESSE; REDDI, 2016; KASPRZAK et al., 2008), especially with frequent measurements. In recent years, Chl-\(a\) concentrations have often been estimated by remote sensing, which is an effective means of rapid, high-frequency data acquisition (ALLAN et al., 2011; OLMANSON; BREZONIK; BAUER, 2011; SHI et al., 2015). In addition, this technique allows one to obtain information on remote sites and larger areas.

Currently, bio-optical algorithms based on radiometric measurements (e.g., solar irradiance, sky radiance) have been used to obtain quantitative information of substances present on water bodies (OGASHAWARA; MISHRA; GITTELSON, 2017). These algorithms can be classified in empirical, semiempirical, semi-analytical, quasi-analytical or analytical algorithms (OGASHAWARA, 2015), being the empirical and empirical algorithms commonly used to retrieve Chl-\(a\) concentrations (DUAN et al., 2010; LE et al., 2013; RITCHIE; ZIMBA; EVERITT, 2003; ROSA NETO et al., 2015). Limitations of Chl-\(a\) detection by remote sensing include atmospheric correction methods and sensor limitations, as well as the influences of detritus, the presence of colored dissolved organic matter (CDOM), and scattering by Total Suspended Matter (TSM), which are difficult to detect because they affect the optical properties of water (DARECKI; STRAMSKI, 2004; HU, 2009; WU et al., 2009). Another limitation is the light reflected off the bottom, which may affect the accuracy of the empirical algorithm because the signal received by the sensor varies as a function of the wavelength and with the clarity of the water (CARDOSO et al., 2012; SHUBHA, 2000; LEF. et al., 2001).

Recently, data from MODIS (\textit{Moderate Resolution Imaging Spectroradiometer}) have been used successfully for assessment and continuous monitoring of water quality, including Chl-\(a\) concentrations, in lakes and reservoirs (DENG et al., 2017; OLMANSON; BREZONIK; BAUER, 2011; SONG et al., 2013; WU et al., 2014; ZHANG et al., 2011). Empirical algorithms based on MODIS data, such as O3M (O’REILLY et al., 2000), have been used in different environments (GILERSON et al., 2010; GURLIN; GITTELSON; MOSES, 2011; LESHT; BARBIERO; WARREN, 2013; MATTHEWS; BERNARD; ROBERTSON, 2012; OGASHAWARA et al., 2014). Nevertheless, this algorithm has a tendency to overestimate the Chl-\(a\) concentrations in oligotrophic environments, and to underestimate it in highly productive environments (KOMICK; COSTA; GOWER, 2009; MOORE; CAMPBELL; DOWELL, 2009).

Recent studies have applied empirical algorithms for retrieving Chl-\(a\) concentrations in lakes and reservoirs combining MODIS spectral bands (DUAN et al., 2017; CHANG et al., 2012; HUANG et al., 2014; SHI et al., 2017; WANG; SHI; TANG, 2011; WANG; LIU; TANG, 2017; WU et al., 2009; XIANG et al., 2015; ZHANG et al., 2016). However, these algorithms were developed for eutrophics waters (e.g., Lakes Taihu and Chaohu, China and Lake Okeechobee, USA) and for particular seasons, limiting their uses elsewhere and throughout the year (MATTHEWS, 2011; PALMER et al., 2015a). These approaches can be improved using the appropriate spectral remote-sensing reflectance values and combinations of spectral bands with either regression analysis (CHANG et al., 2012; OGASHAWARA et al., 2014) or an ANN (WU et al., 2009; XIANG et al., 2015). However, only the approach proposed by Ogashawara et al. (2014) was developed in a meso-oligotrophic environment (Itumbia Reservoir, Brazil).

Monitoring Chl-\(a\) concentrations from remote sensing in freshwater environments such as shallow lakes has limitations, which are associated with low signal noise ratio and the optical complexity of these waters being highly variable between and even within water bodies (GHOLIZADEH; MELESSE; REDDI, 2016; PALMER; KUTSER; HUNTER, 2015b). In addition, in large shallow lakes, which are rarely studied, especially in meso-oligotrophic environments submitted to intense wind regimes, limitations such as scale factors (e.g., representative monitoring at a single point), spatial gradients and hydrodynamics might significantly affect the interpretation of results from remote-sensing data and the performance of the empirical algorithms (CHAVULA et al., 2009; OGASHAWARA et al., 2014; RUIZ-VERDU et al., 2016).

In this study, we (1) proposed and compared multivariable empirical models to retrieve Chl-\(a\) concentrations, using MODIS data, (2) determined the accuracy of these models in a large shallow sub tropical, and (3) verified if generalized models can represent the spatial heterogeneity of the lake. We used multiple linear regression, artificial neural network, nonparametric multiplicative regression, and four models (Appel, Kahru, FAI and O14a) to retrieve the Chl-\(a\) concentration in the Lake Mangueira, a large shallow sub tropical lake located in Southern Brazil. This approach is the first attempt to assess the application of remote sensing techniques in such a system, which the oligotrophic conditions are common most of the time (CROSSETTI et al., 2013; LIMA et al., 2016), and the spatial heterogeneity has large differences between the littoral and pelagic zones (CARDOSO et al., 2012; CROSSETTI et al., 2013; THEY; MARQUES; SOUZA, 2012).

Differently from previous studies, our approach is a first step toward to obtain comprehensive and reliable empirical models for meso-oligotrophic shallow waters, which may be further tested with remote-sensed data in order to retrieve the Chl-\(a\) concentration and help to understanding the shallow lakes heterogeneity.

METHODS

Study area

Lake Mangueira is a large shallow sub tropical waterbody located in Rio Grande do Sul, Brazil. The lake is 820 km\(^2\) in area and has a mean depth of 2.6 m, a maximum depth of 6.9 m, and is 90 km long and 3–10 km wide (Figure 1). Its trophic state ranges from oligotrophic to mesotrophic (CROSSETTI et al., 2013; FRAGOSO JUNIOR et al., 2011). The regional climate is...
subtropical with a mean annual temperature of 16 °C and annual rainfall between 1,800 and 2,200 mm (KOTTEK et al., 2006).

Field data

The field data were collected over a period of 9 years (2001 to 2009) at three sampling points in the North, Central and South parts of Lake Mangueira, TAMAN, TAMAC and TAMAS respectively, measured two to three times a year. The dataset contained concentrations of Chl-\(a\), Secchi disk depth and other limnological variables (i.e., Total Suspended Solids (TSS), Dissolved Oxygen (DO), pH and Depth), collected from the three sampling points, and indicated a wide range of Chl-\(a\) concentrations at the Central point (Table 1). We used the data from the 2001 – 2005 period for calibration and from the 2006 – 2009 period for model validation.

For Chl-\(a\) determination, surface water samples were filtered through Whatman GF/F with 90% ethanol and measured by spectrophotometry (JESPERSEN; CHRISTOFFERSEN, 1987). Dissolved Oxygen, pH and Depth were measured with a multiparameter probe (YSI 6600). Total Suspended Solids were assessed gravimetrically by water evaporation in porcelain dishes (EATON; CLESCERI, 1999) and the Water Transparency (WT) was measured using a Secchi disk.

Satellite data and image processing

MODIS-Terra (MOD09GA) and MODIS-Aqua (MYD09GA) Level-2 Surface Reflectance products with a daily frequency and a 500-m spatial resolution were downloaded via the web interface The Land Processes Distributed Interface Active Archive Center (LP DAAC), from The US Geological Survey EROS Data Center (LPDAAC, 2015). These products have been successfully used in different inland waters, including oligo- to mesotrophic reservoirs (surface area: 778 km\(^2\)) (CURTARELLI et al., 2015; OGASHAWARA et al., 2014), Amazon floodplain lakes (surface area: 2000 km\(^2\)) (NOVO et al., 2006), Minnesota lakes (lake surface area > 160 ha) (KNIGHT; VOTH, 2012), Poyang Lake (surface area: 3000 km\(^2\)) (QI et al., 2016), and coastal lagoons (900 km\(^2\)) (SRICHANDAN et al., 2015). The cloud cover fraction was obtained using a cloud mask product from MODIS product level 2 (named M*D35L2) (ACKERMAN et al., 1998). In addition, MODIS quality product was also used to assess image quality (ROY et al., 2002). The sinusoidal projection images (ISIN) from the LP DAAC platform were transformed to geographic projection (Lat/Long) using the MODIS Reprojection Tool. The MOD09GA, MYD09GA and M*D35L2 products were processed and analysed using a MatLab\textsuperscript{®} routine. To retrieve Chl-\(a\) concentrations, we used the surface reflectance from the first six spectral bands of the MOD09GA and MYD09GA products, with the centre wavelength location respective (band 1: 648 nm, band 2: 858 nm; band 3: 470 nm; band 4: 555 nm; band 5: 1240 nm; and band 6: 1640 nm).

We downloaded a total of 50 MODIS images (32 MODIS-Terra and 18 MODIS-Aqua) matching the in-situ measurements of Chl-\(a\) monitoring days (Table 2). From this total, we selected 16 images for model development and calibration, and 10 images for model verification; the remaining 24 images were discarded due to the presence of clouds, shadows on the measurement

**Table 1.** Chl-\(a\) concentrations, Secchi disk depth, total suspended solids (TSS), dissolved oxygen (DO), pH and Depth measured in 2001-2009 (N=43) at three points (North, Central, South) in Lake Mangueira.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>North</th>
<th>Central</th>
<th>South</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ± SD (Min – Max)</td>
<td>Mean ± SD (Min – Max)</td>
<td>Mean ± SD (Min – Max)</td>
</tr>
<tr>
<td>Chl-(a) (µg.L(^{-1}))</td>
<td>10.88 ± 8.54 (1.04–36.9)</td>
<td>13.64 ± 13.52 (0.5–63.3)</td>
<td>5.54 ± 4.73 (0.2–19.30)</td>
</tr>
<tr>
<td>Secchi disk depth (m(^{-1}))</td>
<td>0.72 ± 0.33 (0.16–1.55)</td>
<td>0.79 ± 0.21 (0.55–1.36)</td>
<td>1.23 ± 0.40 (0.72–1.90)</td>
</tr>
<tr>
<td>TSS (µg.L(^{-1}))</td>
<td>33.04 ± 21.03 (3.0–80.0)</td>
<td>26.82 ± 14.74 (6.0–68.0)</td>
<td>19.85 ± 12.97 (3.0–60.0)</td>
</tr>
<tr>
<td>DO (µg.L(^{-1}))</td>
<td>9.44 ± 1.50 (6.89–13.45)</td>
<td>9.62 ± 1.26 (6.45–12.70)</td>
<td>9.94 ± 1.15 (8.12–12.70)</td>
</tr>
<tr>
<td>pH</td>
<td>8.06 ± 0.42 (7.07–8.60)</td>
<td>8.22 ± 0.47 (6.76–8.95)</td>
<td>8.25 ± 0.53 (6.84–8.92)</td>
</tr>
<tr>
<td>Depth (m)</td>
<td>2.88 ± 0.83 (1.60–4.30)</td>
<td>2.58 ± 0.74 (1.04–3.80)</td>
<td>2.43 ± 0.73 (1.26–3.80)</td>
</tr>
</tbody>
</table>

Figure 1. Bathymetry and locations of sampling stations (North - TAMAN, Central - TAMAC and South - TAMAS) in Lake Mangueira, Rio Grande do Sul, Brazil.
Can chlorophyll-α in meso-oligotrophic shallow waters be estimated using statistical approaches and empirical models from MODIS imagery?

Multivariable models to retrieve Chl-a concentration

Approaches based on MLR (using exclusion-inclusion sequential models), ANN and NPMR were used to retrieve Chl-α concentrations, using the MOD09GA and MYD09GA reflectance data products.

An empirical model based on MLR was developed, using two search criteria. The first used a step-wise forward regression from the reflectance for each of the spectral bands from MODIS. This approach was carried out taking into account that the F value was low and the p-value was high. This method generates a MLR of the bands with the highest correlation. The second development criterion was based on a sequential automatic search of exclusion-inclusion using the Akaike Information Criterion – AIC (BOZDOGAN, 1987). Using this criterion, an R code was developed to fit a sub-model for each of the three datasets (North, Central and South points) and for the total dataset in Lake Mangueira.

Table 2. Number of available MODIS images for calibration/validation periods for the North (N), Central (C) and South (S) points in Lake Mangueira.

<table>
<thead>
<tr>
<th>Id</th>
<th>Field Campaign Date</th>
<th>Station (N)</th>
<th>Station (C)</th>
<th>Station (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>21/03/2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>20/04/2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>23/05/2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>19/06/2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>17/07/2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>21/08/2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>27/09/2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>24/10/2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>20/11/2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>19/12/2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>29/01/2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>21/02/2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>20/03/2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>23/04/2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>27/05/2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>19/06/2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>23/07/2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>21/08/2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>30/10/2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>19/11/2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>17/12/2002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>04/05/2003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>04/08/2003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>01/10/2003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>01/12/2003</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Discarded
Calibration
Validation

An ANN was applied using a hidden layer (Figure 2). The reflectances for each of the spectral bands for three points (North, Central and South) and the total dataset were the input values and the Chl-α concentration was the output node. The training algorithm used was the back-propagation method with heuristic techniques acceleration (VOGL et al., 1988) performed in MatLab®. The Purelin function (DEMUTH; BEALE; HAGAN, 2008) was used as the activation function in the hidden and output layers.

NPMR using the software HyperNiche 1.0 was also used to select the best models describing the relationships between Chl-α and the reflectance values from the MODIS spectral bands (McCUNE, 2006; 2004). Two matrices were used as input in NPMR: (1) a matrix of in-situ Chl-α measurements, and (2) a matrix of reflectance values from the MODIS spectral bands. Here, the local mean NPMR, uniform weights (Species Occurrence model – SpOcc) was used (PETERSON, 2000; McCUNE et al., 2003). This model uses a simple kernel function to estimate the response variable (Chl-α in our case) and gives equal weight to all sampling points within the window, while all observations outside the window are given zero weight. In addition, we assessed the performance of four models (three empirical models (Kahru, Appel and O14a) and one semi-empirical model (FAI), Table 3), that are commonly used to estimate Chl-α concentrations in aquatic sites or to the bad quality of the images from MODIS quality product control.
environments (CURTARELLI et al., 2015; HUANG et al., 2014; HU et al., 2010; TARRANT; NEUER, 2009).

Model accuracy assessment

To evaluate the accuracy of the models, the coefficient of determination ($R^2$), Bias, Root-Mean-Square-Error (RMSE), and relative RMSE (RMSE%) were calculated between in-situ observed and model-estimated Chl-a concentrations.

RESULTS

Model development and calibration

The performance of the multiple linear regression, ANN, and NPMR models for the three points (North, Central and South) and the total dataset for Lake Mangueira showed good agreement for the Central and South datasets. In addition, the MLR model performed best for each dataset (Figure 3).

The MLR model assessment indicated that it performed best for the South point ($R^2 = 0.83$, Bias = 0.00, RMSE% = 10.6%) compared to the Central point ($R^2 = 0.77$, Bias = −0.31, RMSE% = 14.6%) and the North point ($R^2 = 0.61$, Bias = 0.00, RMSE% = 19.6%). Also, the MLR model indicated that the most important MODIS channels for the correlation with in-situ Chl-a values at North point were the spectral bands 2, 3 and 6, while that for the Central and South points, were the spectral bands 1, 4, 5, 6 and 1, 2, 5, 6 respectively. For the total dataset, spectral bands 1, 3, 5 and 6 were the most important MODIS channels.

The MLR models adjusted to the four datasets used are listed as follows:

$$MLR_{South} = 1.13 + [−584.84 * R(λ_{b6})] + (721.14 * R(λ_{b3})) + (33.33 * R(λ_{b6}))$$

$$MLR_{Central} = 30.26 + (950.36 * R(λ_{b6})) + (−1032.97 * R(λ_{b4})) + (510.97 * R(λ_{b3})) + (−406.31 * R(λ_{b6}))$$

$$MLR_{South} = −3.12 + (616.28 * R(λ_{b6})) + (−625.52 * R(λ_{b5})) + (127.10 * R(λ_{b3})) + (−33.21 * R(λ_{b6}))$$

$$MLR_{totaldataset} = 9.23 + (36.95 * R(λ_{b6})) + (−52.92 * R(λ_{b5})) + (14.47 * R(λ_{b5})) + (−35.73 * R(λ_{b6}))$$

where $R(λ_{b1}), R(λ_{b2}), R(λ_{b3}), R(λ_{b4}), R(λ_{b5}), R(λ_{b6})$ are the reflectance of bands 1 to 6 respectively with $λ_{b1} = 645 \text{ nm}$, $λ_{b2} = 859 \text{ nm}$, $λ_{b3} = 470 \text{ nm}$, $λ_{b4} = 555 \text{ nm}$, $λ_{b5} = 1250 \text{ nm}$ and $λ_{b6} = 1640 \text{ nm}$.

The ANN model performed best for the Central point ($R^2 = 0.62$, Bias = 0.01, RMSE% = 16.0%), compared to the South point ($R^2 = 0.41$, Bias = −5.19, RMSE% = 18.7%), the North point ($R^2 = 0.01$, Bias = 2.72, RMSE% = 19.6%), and the total dataset ($R^2 = 0.21$, Bias = 2.42, RMSE% = 18.4%). The ANN model architecture was composed of 3 layers (input, hidden and output layers): the input layer had the same MODIS channel inputs as the MLR models, the hidden layer had several neurons based on each dataset (North: 12, Central: 4; South: 10), and the output had one neuron (Chl-a). Thus, the ANN architecture was: North: 3–12–1; Central: 4–4–1 and South: 4–10–1.

The results from the NPMR models indicated the best performance for the South point ($R^2 = 0.54$, Bias = 0.43, RMSE% = 15.4%), compared to the Central point ($R^2 = 0.43$, Bias = 1.37, RMSE% = 11.3%), the North point ($R^2 = 0.06$, Bias = 1.57, RMSE% = 28.4%) and the total dataset ($R^2 = 0.17$, Bias = 0.70, RMSE% = 13.3%). For the three points in the lake, we used as the predictor matrix (input data), the reflectance values of MODIS spectral bands 1, 2, 3, 4, 5 and 6 and MLR is the estimated Chl-a in $\mu g.L^{-1}$.

Table 3. Kabru, APPEL, FAI, and O14a models to estimate Chl-a concentrations in Lake Mangueira.

<table>
<thead>
<tr>
<th>Model</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kabru</td>
<td>Hu and He (2008)</td>
</tr>
<tr>
<td>APPEl</td>
<td>El-Alem et al. (2012)</td>
</tr>
<tr>
<td>FAI</td>
<td>Kahrui et al. (2004)</td>
</tr>
<tr>
<td>O14a</td>
<td>Ogashawara et al. (2014)</td>
</tr>
</tbody>
</table>

Figure 2. Schematic diagram of an artificial neural-network model.
Can chlorophyll-a in meso-oligotrophic shallow waters be estimated using statistical approaches and empirical models from MODIS imagery?

Amongst the four models evaluated, the Appel model performed better than the Kahru, FAI and O14a models (Figure 4). The Appel model assessment indicated that it performed best for the Central point ($R^2 = 0.66$, Bias = 0.58, RMSE% = 35.2%), compared to the North point ($R^2 = 0.38$, Bias = 6.53, RMSE% = 43.9%), the South point ($R^2 = 0.57$, Bias = 5.64, RMSE% = 35.2%) and the total dataset ($R^2 = 0.29$, Bias = 11.6, RMSE% = 26%).

Despite the O14a model showing a good agreement for the North point ($R^2 = 0.52$, Bias = 6.93, RMSE% = 31.4%) and the South point ($R^2 = 0.52$, Bias = 4.04, RMSE% = 29.5%), this
model showed the worse results for the Central point ($R^2 \leq 0.10$, Bias $\leq 8.37$, RMSE% $\leq 28.1\%$).

Results from the Kahru and FAI models showed similar performance, being the adjustment for the Kahru model slightly better ($R^2 \leq 0.50$, Bias $\leq 6.69$, RMSE% $\leq 44.7\%$), that the FAI model ($R^2 \leq 0.40$, Bias $\leq 6.37$, RMSE% $\leq 35.3\%$).

**Model validation and comparative assessment**

The performance of the different models during the validation phase showed, in general, lower values of the accuracy metrics (Figure 5). However, the MLR and ANN models showed a smaller loss of performance ($R^2 = 0.50$, Bias = –6.69, RMSE% = 44.7\%) and maintained satisfactory accuracy compared to the NPMR model (e.g., Central point).

In all cases, the performance of our multivariable models using the total dataset was lower than with the specific point dataset. The same occurred in comparisons against four commonly used models, Appel, Kahru, FAI, and O14a (Figure 6).

We found that the MLR was reliable in terms of its estimation of Chl-a concentrations for the three points in the lake, although it was less accurate for the South point ($R^2 = 0.37$, Bias = –3.54, RMSE% = 34.40\%). The ANN model performed best for the Central point ($R^2 = 0.46$, Bias = –1.78, RMSE% = 24.60\%), but did not provide satisfactory results for the North and South points.

The NPMR model had limitations in retrieving the Chl-a concentration for the North (N), Central (C) and South (S) points ($R^2 \leq 0.19$, Bias $\leq 5.37$, RMSE% $\leq 42.50\%$). The assessments with the four
Can chlorophyll-a in meso-oligotrophic shallow waters be estimated using statistical approaches and empirical models from MODIS imagery?

Models (Appel, Kahru, FAI and O14a) showed lower accuracy than the MLR model (Figure 6).

The Kahru model gave the best Chl-a retrieval ($R^2 \leq 0.37$, Bias $\leq 9.09$, RMSE% $\leq 36.4\%$), followed by the O14a model ($R^2 \leq 0.35$, Bias $\leq 5.53$, RMSE% $\leq 39.97\%$); the Appel model gave the poorest results ($R^2 \leq 0.26$, Bias $\leq 9.07$, RMSE% $\leq 35.04\%$). Taking into account the total dataset, the accuracy yielded by the MLR, ANN and NPMR models and the four models (Appel, Kahru, FAI and O14a) also showed low performance compared with the in-situ Chl-a values.

**DISCUSSION**

**Performance of Chl-a algorithms**

For the models proposed here, the MLR and ANN models performed better than the other models ($R^2$ and RMSE%, Figure 6), in agreement with the performance found previously for oligo- to mesotrophic waters using MLR models (CURTARELLI et al., 2015; OGASHAWARA et al., 2014) and for eutrophic shallow waters using Back Propagation neural networks (BPs) and
Radial Basis Function neural networks (RBFs) (WU et al., 2009). Nevertheless, the performance of models proposed and tested during the calibration period was better than for the validation period. This loss of performance during the validation models is a behavior expected, which can be explained by the size of samples and the difference in the datasets used, as has been observed in others studies (GIANCRISTOFARO; SALMASO, 2007; ORTH et al., 2015).

Although the NPMR model is generally stronger than the MLR model (McCUNE et al., 2003; YOST, 2008), our results indicated a significant difference in the performance of the two models. Also, the NPMR model showed lower performance, which was apparent during both the calibration (Figure 3) and the validation (Figure 5) periods. This result indicates that the application of NPMR models to retrieve the Chl-a concentration from MODIS imagery in Lake Mangueira would not be recommended based on the dataset available. For the four models assessed (Appel, Kahru, FAI and O14a), our findings obtained from the total datasets were consistent with those of El-Alem et al. (2012), who obtained satisfactory results for Chl-a > 50 µg.L⁻¹ using the Kahru and FAI models.

In Lake Mangueira, the Chl-a concentration estimated using the total dataset indicated that the correlation was the lowest for every model proposed as shown in the calibration (Figure 3), validation (Figure 5) and the four models assessed (Figure 6). Therefore, it would not be appropriate to generalise a single model to retrieve the Chl-a concentration in Lake Mangueira from a total dataset, but rather would be necessary to subdivide the datasets into regions. This can be explained by the spatial heterogeneity, where the Chl-a concentration decreases from

Figure 6. Plots of the Appel, Kahru, FAI and O14a models values versus in-situ chlorophyll-a measurements at three points (North, Central and South) and the total dataset in Lake Mangueira during validation. The dotted lines represent the best fit.
Algorithm limitations for estimating Chl-a from MODIS imagery

Several studies have shown that algorithms for retrieving Chl-a from MODIS imagery may indeed be precluded due to the low signal-to-noise ratio and the noise-equivalent Chl-a concentration of MODIS (e.g., CARDER; CANNIZZARO; LEE, 2005; GOWER; BORSTAD, 2004). This is a limitation in meso-oligotrophic environments such as Lake Mangueira, which can clip the performance of algorithms based on MODIS imagery, due to low Chl-a contents, and the adjacency effects that can be caused by sun glints, as found in similar environments (CHAVULA et al., 2009; MATTHEWS; ODERMATT, 2015; WATANABE et al., 2015). Furthermore, at the South point of Lake Mangueira, located in the shallower area of the lake, water transparency is high (Table 1) and the light reflected off the bottom also affects the accuracy of the empirical algorithms assessed. Another limitation associated with the use of satellite remote-sensing techniques is the detection of Chl-a in surface waters only. However, this limitation is not applicable to polymeric shallow-water systems such as Lake Mangueira, where the Chl-a concentrations are relatively uniform throughout the water column except for short time periods (hours, see FRAGOSO JUNIOR et al., 2011).

CONCLUSION

This study evaluated the use of several models to retrieve Chl-a concentrations from MODIS Terra and Aqua reflectance data in a large shallow subtropical lake. Our findings suggested the MLR model as the better alternative compared to ANN and NPMR. In addition, the simple MLR model proposed in this study showed two main advantages: better performance and fewer input variables compared with others empirical and semi-empirical models that are commonly used (Appel, Kahr, FAI and O14a). Finally, although the purpose of this study was not to replace standard monitoring methods, the MLR model using MODIS reflectance data can provide a dynamic, broader-scale view of Chl-a concentrations. Further investigation could be focused on MLR model application to analyze in long-term scale, the temporal and spatial distribution of Chl-a concentration in Lake Mangueira, in order to understand the seasonal dynamic and the inter-annual variations. The overall methodology proposed here can be used as a tool to help reduce the effort and costs associated with conventional limnological monitoring of shallow waters, and as a data source to support distributed ecological mathematical modelling of these ecosystems.

ACKNOWLEDGEMENTS

We are grateful to the Brazilians funding agency Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, www.capes.gov.br) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ, www.cnpq.br) for doctoral scholarships awarded to A.M. Munar and J.R. Cavalcanti. The study is part of the research project ‘Impacto das mudanças climáticas em ambientes fluviais e lacustres do Rio Grande do Sul’, funded by CNPq. We are also grateful to the Global Lake Ecological Observatory Network (GLEON, www.gleon.org) for providing a venue and resources for lake-science discussions, and to Dr. Janet W. Reid (Trumansburg, NY, U.S.A.) for revising the English text.

REFERENCES


ALLAN, M. G.; HAMILTON, D. P.; HICKS, B. J.; BRABYN, L. Landsat remote sensing of chlorophyll a concentrations in central North Island lakes of New Zealand. International Journal
Can chlorophyll-\(a\) in meso-oligotrophic shallow waters be estimated using statistical approaches and empirical models from MODIS imagery?


MUNAR et al.


Can chlorophyll-\(a\) in meso-oligotrophic shallow waters be estimated using statistical approaches and empirical models from MODIS imagery?


YOST, A. C. Probabilistic modeling and mapping of plant indicator species in a Northeast Oregon industrial forest, USA. *Ecological...*


**Authors contributions**

Andrés Mauricio Munar, David da Motta Marques and Carlos Ruberto Fragoso Jr developed the idea for the study, as well as its design and paper review.

J. Rafael Cavalcanti and Juan Martin Bravo helped in the methodology for the data analysis, data analysis, construction and validation of the models, support in statistical evaluation and paper review.

All authors were responsible for reviewing the written text and approved the manuscript submission.