MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2022/06.15.17.57-TDI

# CYANOBACTERIA MONITORING ON URBAN RESERVOIRS USING HYPERSPECTRAL ORBITAL REMOTE SENSING DATA AND MACHINE LEARNING

Felipe Nincao Begliomini

Master's Dissertation of the Graduate Course in Remote Sensing, guided by Drs. Cláudio Clemente Faria Barbosa, and Vitor Souza Martins, approved in May 30, 2022.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/474PTSB>

INPE
São José dos Campos
2022

**PUBLISHED BY:**

Instituto Nacional de Pesquisas Espaciais - INPE
Coordenação de Ensino, Pesquisa e Extensão (COEPE)
Divisão de Biblioteca (DIBIB)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2022/06.15.17.57-TDI

# CYANOBACTERIA MONITORING ON URBAN RESERVOIRS USING HYPERSPECTRAL ORBITAL REMOTE SENSING DATA AND MACHINE LEARNING

Felipe Nincao Begliomini

Master's Dissertation of the Graduate Course in Remote Sensing, guided by Drs. Cláudio Clemente Faria Barbosa, and Vitor Souza Martins, approved in May 30, 2022.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/474PTSB>

INPE
São José dos Campos
2022

Cataloging in Publication Data

Begliomini, Felipe Nincao.

B394c Cyanobacteria monitoring on urban reservoirs using hyperspectral orbital remote sensing data and machine learning / Felipe Nincao Begliomini. – São José dos Campos : INPE, 2022.

xxvii + 90 p. ; (sid.inpe.br/mtc-m21d/2022/06.15.17.57-TDI)

Dissertation (Master in Remote Sensing) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2022.

Guiding : Drs. Cláudio Clemente Faria Barbosa, and Vitor Souza Martins.

1. Cyanobacteria. 2. C-Phycocyanin. 3. Remote Sensing. 4. PRISMA. 5. Machine Leaning. I.Title.

CDU 528.8:561.232

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÕES

**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**
Serviço de Pós-Graduação - SEPGR

## DEFESA FINAL DE DISSERTAÇÃO DE FELIPE NINCAO BEGLIOMINI
## BANCA Nº 170/2022 , REG. 852825/2020

No dia 30 de maio de 2022, as 14h, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi **APROVADO(A)** pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Mestre em Sensoriamento Remoto. O trabalho precisa da incorporação das correções sugeridas pela Banca e revisão final pelo(s) orientador(es).

**Título: "CYANOBACTERIA MONITORING ON URBAN RESERVOIRS USING HYPERSPECTRAL ORBITAL REMOTE SENSING DATA AND MACHINE LEARNING"**

**Membros da Banca:**

Dra. Evlyn Márcia Leão de Moraes Novo - Presidente - INPE
Dr. Cláudio Clemente Faria Barbosa - Orientador - INPE
Dr. Vitor Souza Martins - Orientador - Mississippi State University
Dra. Áurea Maria Ciotti - Membro Externo - USP
Dra. Cristina Souza Freire Nordi - Membro Externo - UNIFESP
Dra. Marta Condé Lamparelli - Membro Externo - CETESB

Documento assinado eletronicamente por **Evlyn Marcia Leão de Moraes Novo**, **Pesquisador**, em 31/05/2022, às 14:59 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Cláudio Clemente Faria Barbosa**, **Tecnologista**, em 02/06/2022, às 15:19 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Aurea Maria Ciotti (E)**, **Usuário Externo**, em 22/06/2022, às 16:18 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Vitor Souza martins (E)**, **Usuário Externo**, em 22/06/2022, às 16:59 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Marta condé lamparelli (E)**, **Usuário Externo**, em 23/06/2022, às 08:40 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Cristina Souza Freire Nordi (E)**, **Usuário Externo**, em 29/06/2022, às 13:40 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site http://sei.mctic.gov.br/verifica.html, informando o código verificador **9938580** e o código CRC **10C0504E**.

**Referência:** Processo nº 01340.004034/2022-07 SEI nº 9938580

“A vida é a arte do encontro,

embora haja tanto desencontro pela vida”

*Vinícius de Moraes*

A minha mãe Sônia,

Meu irmão Lucas e minha querida Gabriela.

## ACKNOWLEDGEMENTS

Primeiramente, gostaria de agradecer à minha mãe, Sônia Nincao, por todo apoio e amor que me oferece desde que cheguei a este mundo. Agradeço também à Gabriela Netto, com certeza sem seu apoio o fardo da pós-graduação teria sido muito pesado para carregar sozinho. Em seguida, devo minha total gratidão ao meu orientador, Dr. Cláudio Barbosa, por dois anos de apoio incondicional, confiança, recursos e conhecimentos oferecidos. Assim como ao meu coorientador, Dr. Vitor Martins, que não deixou 'minha peteca cair' nos momentos mais difíceis. Obrigado por sempre me mostrar o lado positivo e soluções criativas para os problemas encontrados. A Dra. Evlyn Novo agradeço pelas palavras estimulantes e conhecimentos que fortalecem a alma de quem os ouve, mesmo quando as vezes ela nem sabe que está mudando a vida de quem ela toca. Aos meus queridos colegas do LabISA Rejane, Daniel, Thainara e Rogério com certeza vocês comprovam que a união no INPE faz a força. Sem a contribuição de cada um de vocês, chegar ao fim desta caminhada não teria sido possível.

Também agradeço à minha querida ex-orientadora, que hoje tenho o prazer de chamar de amiga, Dra. Pilar Villar, que possibilitou firmar parcerias importantes para a pesquisa. Meu carinho e agradecimentos ao meu eterno companheiro de casa, Elton Escobar, que propiciou conversas que transformavam manhãs desagradáveis em tardes mais leves. Não poderia deixar agradecer a Lucas Nincao, Cássia Begliomini, Bianca Doné, Luisa Pessanha, Giovana Ciantelli, Gustavo Perterlevitz, Luiza Simões e Mariana Forgaça. Agradeço também a todos os outros amigos e familiares que sempre me apoiaram durante toda minha vida. Obrigado ao Instituto Nacional de Pesquisas Espaciais e todos os professores do Programa de Mestrado em Sensoriamento Remoto por abrir às portas para um Felipe que nem eu acreditava que poderia existir. Também agradeço todos os outros colegas do PGSER que contribuíram muito com essa caminhada.

Sem a colaboração da equipe da Companhia Ambiental do Estado de São Paulo (CETESB), com certeza este trabalho não teria saído do papel. Muito obrigado à toda equipe de campo e técnica, que não mediu esforços para me ajudar em todas as demandas que levei. Também agradeço ao barqueiro Rodrigo, por nos guiar pelas águas da represa Billings. Obrigado ao Dr. Isaias de Oliveira e a todos integrantes do Laboratório de Química por oferecerem equipamentos para a realização deste trabalho. Agradeço à Dra. Laura Borma e à Daniel Meneghetti por disponibilizarem a infraestrutura do LAQUATEC para a realização das análises necessárias. Agradeço também a Dra. Cristina Nordi por toda a contribuição durante o trabalho. Aproveito também para agradecer a Dra. Áurea Ciotti pelo empréstimo de equipamentos e contribuições oferecidas. Assim, como a Dra. Marta Lamparelli que fez parte da banca deste trabalho e no apoio dado pela CETESB. Agradeço à Igor Ogashawara pelas contribuições na determinação analítica da Ficocianina, que possibilitou a realização deste trabalho. Ofereço um agradecimento especial a todos os outros que não foram citados, mas contribuíram com este trabalho. A realização desta pesquisa não seria possível sem a mobilização enorme de todas estas pessoas.

Por fim, gostaria de agradecer às agencias que financiaram esta pesquisa. A Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) que ofereceu recursos por meio do projeto 2020/14613-8. O presente trabalho foi realizado com apoio da Coordenação

de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

## ABSTRACT

Urban Reservoirs provide relevant ecosystem services to the population worldwide. Although its recognized importance, there is an increasing degradation trend of metropolitan water systems' due to anthropical impacts. Cultural eutrophication is highlighted as a negative effect of human activities, with severe consequences such as the intensification of algae blooms. Cyanobacteria are the most concerning bloom-forming species for inland waters due to the environmental impacts and potential to produce toxic compounds. Therefore, this research presents a state-of-art methodology for monitoring Cyanobacteria based on orbital hyperspectral images and Machine Learning Algorithms (MLA) in tropical urban reservoirs. The photosynthetic pigment C-Phycocyanin (PC) was used as a proxy for the Cyanobacteria biomass once this billiprotein is specific from this algae group. Billings reservoir was chosen as the study area due to the constant presence of Cyanobacteria and its importance to the regional urban water supply. Eight field campaigns were made for collecting radiometric, photosynthetic pigments, and taxonomical samples. A hyperspectral image from the PRISMA was acquired in match-up condition, and tree atmospheric correction algorithms were assessed (ASI, ACOLITE, and 6SV). Synthetic multispectral Landsat-8/OLI and Worldview-3 images were generated from PRISMA's best surface reflectance product. Random Forest (RF), Extreme Gradient Boost (XgBOOST), and Support Vector Machine (SVM) were chosen to retrieve PC from Remote Sensing data. Previously published PC algorithms, Normalized Index, and Line Heights were generated from resampled *in-situ* radiometry for each sensor. A data-driven feature selection followed by a decorrelation procedure was used to identify the most informative layers. The Grid Search algorithm tuned the hyperparameters. PC was modeled from *in-situ* data through Monte Carlo simulations for all assessed sensors and MLA. Then, the best combinations were used for mapping PC in the hyperspectral and synthetic multispectral images. The results for *in-situ* and orbital modeling were compared with the state-of-art PC algorithm Mixture Density Network (MDN) (O'SHEA et al., 2021). PC from 0 to 301.81 μg/L were found, with mean and median values of 20.28 and 2.9 μg/L. Cyanobacteria species were at least abundant in 96% of the taxonomical samples. ASI was the best surface reflectance product (MAE < 20% for the visible spectrum). ACOLITE and 6SV underperformed ASI's product by two to ten folds. MDN has sharply overestimated PC in both orbital and *in-situ* assessments. RF had the best estimates for all assessed sensors using *in-situ* data, with MAE ranging from 59-86%. The best result from orbital data was achieved by PRISMA/RF (MAE = 45%). XgBOOST produced the best results for Worldview-3 (MAE = 49%) and Landsat-8/OLI (MAE = 74%) synthetic images. Those are the best-reported results for low PC concentrations and reduced PC:Chl*a* ratios. The low PC:Chl*a* ratios are also the most likely explanation for MDN's errors once the model was trained with samples with 6 times higher the mean PC:Chl*a* found in this study. Specked noise was identified in hyperspectral mapping and is probably due to the reduced Signal-to-Noise ratio. More studies assessing PC in tropical waters are recommended to understand the effects of different latitudes on PC production. Finally, Landsat-8/OLI was identified as the most feasible sensor for monitoring PC due to the reasonable accuracy, the increased temporal resolution (8 days with Landsat-9), and the free access data policy.

Keywords: Cyanobacteria, C-Phycocyanin, Remote Sensing, PRISMA, Machine Leaning, Inland Water, Eutrophication, Urban Reservoir.

**MONITORAMENTO DE CIANOBACTÉRIAS EM RESERVATÓRIOS URBANOS UTILIZANDO DADOS ORBITAIS DE SENSORIAMENTO REMOTO E ALGORITMOS DE APRENDIZADO DE MÁQUINA**

## RESUMO

Os reservatórios urbanos oferecem importantes serviços ecossistêmicos. Contudo, esses sistemas aquáticos têm a qualidade de suas águas impactada pela antropização. A eutrofização cultural é destacada como um efeito negativo das ações humanas e intensifica a ocorrência de florações de algas. As Cianobactérias são as espécies formadoras de florações mais preocupantes em águas continentais devido aos impactos ambientais causados e o potencial para produzir compostos tóxicos. Portanto, esse estudo apresenta uma metodologia para monitorar Cianobactérias por meio de imagens orbitais hiperespectrais e Algoritmos de Aprendizado de Máquina (AAM) em reservatórios tropicais urbanos. O pigmento fotossintético C-Ficocianina (PC) foi usado como proxy para a biomassa de Cianobactérias. O reservatório Billings serviu como área de estudo devido à presença constante de Cianobactérias e o uso para o abastecimento público. Oito campanhas foram realizadas para coletar dados radiométricos, pigmentos fotossintentizantes, e taxonomia. Uma imagem hiperespectral do sensor PRISMA foi adquirida concomitantemente com uma das amostragens, e três algoritmos de correção atmosférica foram avaliados (ASI, ACOLITE e 6SV). Imagens sintéticas dos sensores Landsat-8/OLI e Worldview-3 foram geradas pelo melhor produto de reflectância de superfície do sensor PRISMA. *Random Forest* (RF), *Extreme Gradient Boost* (XgBOOST), e *Support Vector Machine* (SVM) foram escolhidos para modelar a PC. Algoritmos de PC, Índices Normalizados, e *Line Heights* foram gerados por meio de dados radiométricos reamostrados para cada sensor. Uma metodologia de seleção de atributos baseada em dados foi utilizada para selecionar as feições mais informativas. O algoritmo *Grid Search* foi aplicado para ajustar os hiperparâmetros. A PC foi modelada com dados de campo por meio de Simulações Monte Carlo para todos os sensores e AAM avaliados. As melhores combinações foram usadas para mapear a PC nas imagens multiespectrais sintéticas e na hiperespectral. Os resultados foram comparados com o algoritmo Mixture Density Network (MDN) (O'SHEA et al., 2021). Foram encontrados valores de PC entre 0 to 301,81 μg/L, com uma média e mediana de 20,28 e 2,9 μg/L. As Cianobactérias foram pelo menos abundantes em 96% das amostras taxonômicas. A ASI teve o melhor produto de reflectância de superfície (MAE < 20% para o espectro do visível). ACOLITE e 6SV tiveram resultados de duas a dez vezes piores que o da ASI. O MDN superestimou os valores de PC tanto nas análises *in-situ* como orbitais. O RF obteve as melhores estimativas para todos os sensores com dados *in-situ*, com MAE entre 59-86%. O melhor resultado para dados orbitais foi obtido pelo PRISMA/RF (MAE = 45%).

O XgBOOST teve os melhores resultados para as imagens sintéticas do Worldview-3 e (MAE = 49%) e Landsat-8/OLI (MAE = 74%). Esses são os melhores resultados reportados para baixas concentrações de PC e baixas razões PC:Chla. A razão PC:Chla também é a melhor explicação para os erros do MDN, uma vez que o modelo foi treinado com amostras 6 vezes maiores do que a PC:Chla deste estudo. Mais estudos avaliando a PC em águas tropicais devem ser realizados para entender o impacto de diferentes latitudes na produção de PC. Finalmente, o sensor Landsat-8/OLI foi identificado com o sensor mais adequado para o monitoramento de PC devido suas métricas de predição razoáveis, alta resolução temporal e acesso de dados gratuito.

Palavras-chave: Cianobactérias, C-ficocianina, Sensoriamento Remoto, PRISMA, Aprendizado de Máquina, Águas Continentais, Eutrofização, Reservatórios Urbanos.

## LIST OF FIGURES

**Page**

## LIST OF TABLES

**Page**

## LIST OF ABREVIATIONS

| | |
|---|---|
| 6SV | Second Simulation of the Satellite Signal in the Solar Spectrum |
| AC | Atmospheric Correction |
| ACS | Absorption and Attenuation meters |
| AHSI | Advanced Hyperspectral Imager |
| AIS | Airborne Imaging Spectrometer |
| AOPs | Apparent Optical Properties |
| AOT | Aerosol Optical Thickness |
| ARE | Absolute Relative Error |
| ASI | *Agenzia Spaziale Italiana* |
| AVIRIS | Airborne Visible/Infrared Imaging Spectrometer |
| BIL | Billings reservoir |
| CA | Chromatic Adaptation |
| CASI | Compact Airborne Spectrographic Imager |
| CCA | Complementary Chromatic Adaptation |
| CDOM | Colored Dissolved Organic Matter |
| CE | Cultural Eutrophication |
| CETESB | *Companhia Ambiental do Estado de São Paulo* |
| CHIME | Copernicus Hyperspectral Imaging Mission for the Environment |
| CHIRS | Compact High Resolution Imaging Spectrometer |
| Chl*a* | Chlorophyll-*a* |
| CyHABs | Cyanobacteria Harmful Algae Blooms |
| DESIS | DLR (German Aerospace Center) Earth Sensing Imaging Spectrometer |
| DNN | Deep Neural Network |
| EnMAP | Environmental Mapping and Analysis Program |
| EO-1 | Earth Observation-1 |
| FOV | Field-of-View |
| GMT | Greenwich Mean Time |
| HABs | Harmful Algae Blooms |
| HCl | Hydrochloric acid |
| HICO | Hyperspectral Imager for the Coastal Ocean |
| HUN10 | PC estimation algorithm developed by Hunter et al. (2010) |

| | |
|---|---|
| HyspIRI | Hyperspectral Infrared Imager |
| IOPs | Inherent Optical Properties |
| ISS | International Space Station |
| L1 | PRISMA top-of-atmosphere radiance product |
| L2C | PRISMA bottom-of-atmosphere reflectance product, without geocoding |
| L2D | PRISMA bottom-of-atmosphere reflectance product, with geocoding |
| LH | Line Height |
| LIU17 | PC estimation algorithm developed by Liu et al. (2017) |
| LUT | Look-up-Tables |
| MAE | Mean Absolute Error |
| MAPE | Mean Absolute Percent Error |
| MDN | Mixture Density Network |
| MERIS | Medium Resolution Imaging Spectrometer |
| MLA | Machine Learning Algorithms |
| MM14 | PC estimation algorithm developed by Mishra and Mishra (2014) |
| MODIS | Moderate-Resolution Imaging Spectroradiometer |
| MSI | Multispectral Instrument |
| NAP | Non-Algal Particles |
| NASA | National Aeronautics and Space Administration |
| NI | Normalized Index |
| NIR | Near-Infrared region |
| OACs | Optically Actives Constituents |
| OCS | Ocean Color Satellites |
| OGA19 | PC estimation algorithm developed by Ogashawara and Li (2019) |
| OLCI | Ocean and Land Colour Imager |
| OLI | Operational Land Imager |
| OWTs | Optical Water Types |
| PACE | Plankton, Aerosol, Cloud, ocean Ecosystem |
| PC | C-Phycocyanin |
| PC:Chl*a* | Chlorophyll-*a* to C-Phycocyanin ratio |
| PE | Phycoerythrin |
| pH | Hydrogen Potential |
| PPS | Predictive Power Score |
| PRISMA | *PRecursore IperSpettrale della Missione Applicativa* |

| | |
|---|---|
| PROBA-1 | Project for On-Board Autonomy-1 |
| QAA | Quasi-Analytical Algorithm |
| $R^2$ | Coefficient of determination |
| RF | Random Forest |
| RMSE | Root Mean Squared Error |
| rpm | Rotation per Minute |
| $R_{rs}$ | Remote Sensing Reflectance |
| RS | Remote Sensing |
| SNR | Signal to Noise Ratio |
| SPMR | São Paulo Metropolitan Region |
| SRF | Spectral Response Function |
| SRTM | Shuttle Radar Topography Mission |
| STE | Standard Error |
| SVM | Support Vector Machine |
| TSS | Total Suspended Solids |
| USGS | United States Geological Survey |
| VNIR | Visible and Near-Infrared Regions |
| XgBOOST | Extreme Gradient Boost |

## LIST OF SYMBOLS

| | |
|---|---|
| $\hat{b}_i$ | *i*-th estimated band |
| $\hat{x}_i\,\phi$ | Azimuthal angle (Degrees) |
| $a^*_{pc}(\lambda)$ | C-Phycocyanin specific absorption coefficient at wavelength λ |
| $a_{CDOM}(\lambda)$ | Colored Dissolved Organic Matter absorption at wavelength λ ($m^{-1}$) |
| $a_{chl}(\lambda)$ | Chlorophyll-*a* absorption coefficient at 665 nm |
| $a_{pc}(\lambda)$ | C-Phycocyanin absorption coefficient at wavelength λ |
| $a_{PC}(\lambda)$ | C-Phycocyanin absorption at wavelength λ ($m^{-1}$) |
| $a_W(\lambda)$ | Pure Water absorption at wavelength λ ($m^{-1}$) |
| $b_b$ | Backscattering coefficient |
| $b_i$ | *i*-th simulated band from field radiometry |
| $d$ | Degree of similarity between the simulated and estimated spectrum |
| $DIF_{Rrs}$ | Absolute spectral difference between the sample and the median spectrum |
| $e$ | Median Symmetric Accuracy |
| $E_s$ | Downward Irradiance ($W/m^2$) |
| $i$ | *i*-th sample |
| $j$ | *j*-th sampling station |
| $LH(\lambda_{-0}, \lambda_0, \lambda_{+0})$ | Line height for bands centered at wavelenghts $\lambda_{-0}, \lambda_0$ and $\lambda_{+0}$ |
| $L_{sky}$ | Sky Radiance ($W/m^2.sr$) |
| $L_t$ | Total water-leaving radiance ($W/m^2.sr$) |
| $L_w$ | Water-leaving radiance ($W/m^2.sr$) |
| m | Meter |
| $M_i$ | Modeled Value |
| $n$ | Number of samples |
| $nb$ | Number of bands |
| $NI(\lambda_1, \lambda_2)$ | Normalized Index for bands centered at wavelenghts $\lambda_1$ and $\lambda_2$ |
| nm | Nanometer |
| $O_i$ | Observed Value |
| $R(0^{-1})_\lambda$ | Irradiance Reflectance just below the water surface at wavelength λ |
| $R(\lambda)$ | Reflectance Factor just below the water surface at wavelength |
| $R_{rs}(\lambda)$ | Remote Sensing Reflectance at wavelength λ ($sr^{-1}$) |

| | |
|---|---|
| $Rrs_{band}$ | Simulated target sensor band |
| $Rrs_{median}$ | Median Remote Sensing Reflectance measured spectrum |
| $\varepsilon$ | Conversion factor from Chlorophyll-*a* absorption coefficient at 665 nm to Chlorophyll-*a* absorption coefficient at 665 nm |
| $\theta$ | Zenithal angle (Degrees) |
| $\Theta$” | Nadiral angle (Degrees) |
| $\lambda$ | Wavelength (nm) |
| $\rho$ | Air-water interface reflectance |
| $\varphi_1$ | Correction coefficient for Chlorophyll-*a* absorption on 620 nm |
| $\varphi_2$ | Correction coefficient for C- Phycocyanin absorption on 665 nm |
| $\psi$ | Correction coefficient for Chlorophyll-*a* absorption on 620 nm |
| $\delta$ | Optimization coefficient |

**CONTENTS**

**Page**

# 1 INTRODUCTION

Urban reservoirs provide relevant ecosystem services worldwide (LUNDY; WADE, 2011). Water supply and hydroelectric power are the most listed, but recreation, fishery, and irrigation are other relevant benefits, typically depreciated due to the anthropic impacts (SCHOLES; FAULKNER; TAPSELL, 2008). Derived from human activities, cultural eutrophication leads to a primary production increase in the water column (SCHINDLER, 2006) and has been highlighted as one of the most threatening processes for metropolitan water supply systems (LEI et al., 2014). The rising eutrophication of inland waters has guided a global trend of Cyanobacteria Harmful Algae Blooms (CyHABs) intensification (HO; MICHALAK; PAHLEVAN, 2019). Furthermore, Cyanobacteria toxins have been described as a threat to environmental and human health (WOOD, 2016). They can cause neurotoxic, hepatotoxic, and even carcinogenic effects to organisms that ingest contaminated water (CARMICHAEL, 2001). Urban Reservoirs are among the most suitable environments for Cyanobacteria development due to the high concentration of nutrients, increased residence time, recurrent stratification, and low biodiversity (SCHINDLER, 2006; OLIVER; CORBURN; RIBEIRO, 2019). Thus, monitoring those aquatic ecosystems has become essential to assure the health of cities' populations.

Microscopy is recurrently used for Cyanobacteria identification and cell counting on traditional water monitoring (HILLEBRAND et al., 1999). Despite being accurate, this monitoring approach is time-consuming and might lead to poor spatial and temporal representation. Remote sensing-based methods are widespread for monitoring phytoplankton in ocean and inland water ecosystems, providing timely and synoptic Earth surface observations (BLONDEAU-PATISSIER et al., 2014; PALMER; KUTSER; HUNTER, 2015). However, Chlorophyll-*a* (Chl*a*) based approaches are imprecise for Cyanobacteria monitoring since Chl*a* is present in all phytoplankton species (SIMIS; PETERS; GONS, 2005). On the other hand, C-Phycocyanin (PC) is a photosynthetic pigment with a major presence in Cyanobacteria, making it a better proxy for monitoring CyHABs (KIRK, 2010). PC capacity for retrieving biomass information about Cyanobacteria through remote sensing data has been evaluated by several studies (RUIZ-VERDÚ et al., 2008; OGASHAWARA et al., 2013).

Monitoring PC from space in urban reservoirs is challenging due to the lack of sensors with adequate spectral and spatial resolutions. PC has an absorption feature around 620 nm (SIMIS; PETERS; GONS, 2005), which is unusual in the spectral range of most multispectral satellites. One of the most feasible sensors for monitoring phytoplankton in inland waters, Sentinel-2/MSI, has the applicability for estimating Cyanobacteria biomass hindered by the absence of that specific band. The ocean color sensors typically have a band at 620 nm (e.g., Sentinel-3/OLCI, Aqua/MODIS, Envisat/MERIS), but the spatial resolution is often medium or course for monitoring urban reservoirs. To overcome those limitations, Castagna et al. (2020) have proposed a methodology for deriving a virtual orange band (590-635 nm) for Landsat-8/OLI. This OLI channel has shown the potential to monitor PC at a 30 meters scale with a revisit period of 8 days considering Landsat-9 (KUMAR; MISHRA; ILANGO, 2020). Worldview-2 and Worldview-3 have adequate spectral and spatial resolutions for monitoring CyHABs in an urban reservoir. A Worldview-2 synthetic image simulated from airborne hyperspectral data was successfully used to derive PC and showed the platform's potential for estimating Cyanobacteria biomass (BECK et al., 2017).

With hundreds of narrow and contiguous spectral bands, hyperspectral sensors provide opportunities for assessing the unique features of the water constituents (GIARDINO et al., 2019). Many studies were developed using airborne hyperspectral sensors to derive water quality parameters (KUDELA et al., 2015; BECK et al., 2016; YIM et al., 2020), but assessments based on orbital spectrometry are still relatively scarce (GIARDINO et al., 2019). However, with the increasing number of instruments launched recently (e.g., DESIS, PRISMA, Geofen-5/AHSI) and soon to come (e.g., EnMAP, PACE, CHIME), the applications for hyperspectral remote sensing data are to increase. Phytoplankton community structure analysis will be one of the most benefited from radiometric data with high spectral resolution (GIARDINO et al., 2019). Besides more precise Chl*a* estimates, hyperspectral images have proven to be able to retrieve PC from freshwater ecosystems (DEV et al., 2022). The Italian PRISMA hyperspectral satellite is suitable for recording accurate data from water ecosystems (GIARDINO et al., 2020) and has been successfully used for deriving Chl*a*, Total Suspended Solids, Colored Dissolved Organic Matter, and PC (NIROUMAND-JADIDI; BOVOLO; BRUZZONE, 2020; BRESCIANI et al., 2022). Besides the increased spectral resolution, PRISMA's Landsat-like spatial

sampling (30 meters) and the off-nadir revisit of 7 days provide the appropriate specifications for PC monitoring in urban reservoirs.

Accurate PC quantification from remote sensing data can be difficult to achieve even when adequate sensor resolutions are available. When PC concentration is low (PC < 50 µg/L) and/or the PC:Chl*a* ratio is reduced (PC:Chl*a* < 1.5), a sharp accuracy decrease was reported on most previously published PC algorithms (RUIZ-VERDÚ et al., 2008). Precise predictions of the PC absorption coefficient in 620 nm may be hindered by the spectral properties of other compounds in optically complex waters, increasing the errors of semi-analytical and quasi-analytical algorithms (SIMIS et al., 2007). Additionally, the relatively low PC absorption capacity superposed by other spectral features may produce non-linear relationships between radiometric data and PC concentrations (SHI et al., 2019). The lack of linearity can explain the failure of traditional statistic models to predict the Cyanobacteria's photosynthetic pigment, especially in low concentrations. Able to predict linear and non-linear relationships, Machine Learning Algorithms (MLA) are recurrently overperforming traditional approaches for deriving water quality products (BELGIU; DRĂGU, 2016). Sun et al. (2012) reported the first use of MLA to retrieve PC concentration from hyperspectral data and observed the potential of Support Vector Machine to produce better estimates than previously published algorithms. Recently, O'Shea et al. (2021) provided an important advance on building a global PC algorithm capable of predicting precisely even in low concentrations. The authors used the largest reported bio-optical PC dataset ($N = 939$) and the MLA Mixture Density Network (MDN) to retrieve estimates from hyperspectral HICO and PRISMA data. This study introduces the potential of joining orbital hyperspectral data with the complex statistics of MLA for increasing the performance of PC predictions.

The Billings reservoir is the largest water body in São Paulo Metropolitan Region and contributes to the water security of millions of habitants (MOSCHINI-CARLOS et al., 2009). Although important for the regional population, the water quality has been degraded over the last decades due to the intense urbanization on its margins (MILZ et al., 2022). Furthermore, the frequent algae bloom of potentially toxic cyanobacteria species evidence the risk of water contamination (CARVALHO et al., 2007; GEMELGO et al., 2008; RIBEIRO et al., 2020). Thus, Monitoring PC at Billings from orbital images will improve the water management of the São Paulo Metropolitan Region, giving the

importance of the reservoir for the public supply. However, few efforts have been made to validate the remote sensing methodologies for estimating PC in optically complex tropical waters (OGASHAWARA et al., 2013). The dendritic shape with relatively narrow branches makes the use of ocean color satellites unfeasible, remaining the hyperspectral and few multispectral sensors for monitoring the reservoir. Therefore, this study has evaluated the potential of different platforms and MLA for estimating PC in the Billings reservoir. We collected a PRISMA image in match-up with a field campaign, which was used to generate synthetic Landsat-8/OLI and Worldview-3 images. The same boundary conditions enabled the comparison of multispectral and hyperspectral data for predicting PC. Three MLA algorithms were calibrated using *in-situ* radiometry and validated thought orbital data for each of the assessed sensors, and the results were compared with the state-of-art MDN algorithm produced by O'Shea et al. (2021). MDN was calibrated with samples from high latitudes aquatic ecosystem and its applicability to predict PC for complex tropical waters is unknown. This study also reported the results of one of the first atmospheric correction assessments of PRISMA data.

## 1.1 Hypothesis

C-Phycocyanin can be estimated from orbital hyperspectral and multispectral remote sensing data using machine learning models.

## 1.2 Objective

This study aimed to develop a remote sensing-based approach for mapping cyanobacteria and Cyanobacteria Harmful Algae Blooms, through C-Phycocyanin spectral features using orbital image.

## 1.3 Specific objectives

The following research questions were addressed in this dissertation:

- *What is the best surface reflectance product for PRISMA hyperspectral data?*
- *Are Machine Leaning Algorithms able to predict C-Phycocyanin accurately? Among the three tested, which is the best ?*
- *Is there an accuracy gain in using hyperspectral data over multispectral data for predicting C-Phycocyanin?*

Based on these questions, three specific objectives were proposed in this research:

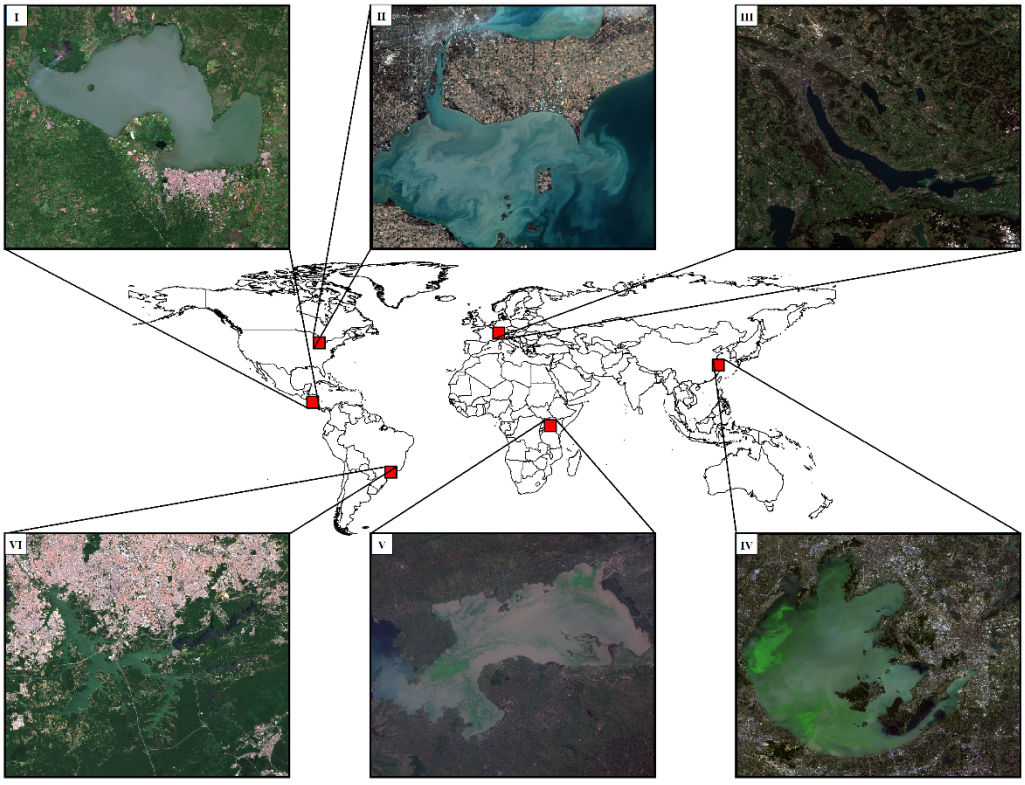
- Evaluate different atmospheric correction processors to estimate PRISMA surface reflectance data.
- Assess three Machine Learning Algorithms for predicting C-Phycocyanin concentrations.
- Compare the C-Phycocyanin estimates from PRISMA hyperspectral data and synthetic multispectral data.

# 2 THEORETICAL BACKGROUND

## 2.1 Eutrophication of urban reservoirs and implications on water quality

Urban Reservoirs have become one of the most important sites for water management and an opportunity gap for increasing the environmental and social quality of cities population (LUNDY; WADE, 2011). They can provide a wide array of ecosystem services, like water supply, flood control, fishing, hydropower, swimming sites, and scenic beauty (Figure 2.1). However, metropolitan water bodies are usual targets of anthropogenic impacts (SCHOLES; FAULKNER; TAPSELL, 2008). Contamination by heavy metals, biodiversity loss, fish mortality, anoxia, and unpleasant odors are frequently identified (MCGRANE, 2016). Among those negative processes, eutrophication is highlighted as the most concerning for urban reservoirs worldwide (OLIVER; CORBURN; RIBEIRO, 2019).

Figure 2.1 - Urban reservoirs and lakes.




Some of the urban reservoirs in the world. I – Lake Xalotlán/Nicaragua. II – Lake Erie/United States. III – Lake Zurich/Switzerland. IV – Billings Reservoir/Brazil. V – Lake Victoria/Kenya. IV – Lake Taihu/China.

Source: Author's elaboration.

Eutrophication causes an increase in primary production and is associated with the rising availability of photosynthesis limiting factors (CHISLOCK et al., 2013). This phenomenon is an inherent process of lake succession (CARPENTER, 1981). Natural eutrophication is recorded within a geological time scale, even though some events might cause acute trophic state changes for short periods (e.g., wildfires, landslides). However, anthropogenic activities can also trigger eutrophication, called Cultural Eutrophication (CE). CE differs from Eutrophication in both time scale and capacity of environmental recovery. The former can cause rapid changes in the trophic state, and those alterations will last until the cause ceases. Several drivers can induce CE, but one of the most common is increasing the nutrient input (SCHINDLER, 2006). CE's consequences are hypoxia, economic loss, impact on recreational activities, and algae blooms (CHISLOCK et al., 2013). Urban Reservoirs are commonly affected by Cultural Eutrophication, and hence, by frequent algae blooms (OLIVER; CORBURN; RIBEIRO, 2019).

## 2.2 Cyanobacteria Harmful Algae Blooms

Algae Blooms are potentially dangerous to aquatic environments and human activities (CHISLOCK et al., 2013). However, the intense growth of phytoplankton in the water is insufficient to be classified as a Harmful Algae Bloom (HABs). To be classified as HABs, they must offer potential harmful effects to aquatic biota, terrestrial animals, or even humans through toxins production (CARMICHAEL; BOYER, 2016). HABs are usually characterized by a monospecific toxic-phytoplankton bloom, mostly identified as Diatoms, Dinoflagellates, and Cyanobacteria (WYATT, 2014). Although responsible for the marine red tides, toxic Dinoflagellate species are not registered in freshwater ecosystems (VASCONCELOS, 2006). In contrast, Diatoms HABs' can occur in inland waters and change their organoleptic characteristics. Still, Diatoms metabolites were not identified as toxic for human beings (VASCONCELOS, 2006). The most common and concerning HABs for freshwaters are caused by Cyanobacteria (CARMICHAEL; BOYER, 2016), frequently called Cyanobacteria Harmful Algae Blooms (CyHABs).

Cyanobacteria, or blue-green algae, are ancient phytoplankton with first registers dating billions of years ago. They are responsible for forming an oxygen-rich atmosphere and the ozone layer (ALLEN; MARTIN, 2007). This group of algae is adapted for adverse environmental conditions and inhabits every aquatic environment on Earth

(CASTENHOLZ, 2015). Cyanobacteria have characteristics that provide ecological advantages over other taxa, such as: (i) Akinetes, the resist cells that can be dormant for long periods; (ii) Heterocyst, which are specialized cells that can fixate Nitrogen from the atmosphere; (iii) Storing high volumes of Phosphorus as intracellular Polyphosphate granules; (iv) Control their position in the water column through gas vacuoles; (v) Production of toxins and other strategies for avoiding grazers (DOKULIL; TEUBNER, 2000). Nevertheless, it is worth nothing that Cyanobacteria is a large phylum with various described species. Therefore, each taxon will gather defined structures, rather than all of them occurring simultaneously.

In some CyHABs species, the gas vacuoles provide a passive buoyancy capacity that responds to environmental conditions. When photosynthesis rates are high, the fixed carbon makes the cells heavy, so they sink. Otherwise, CyHABs develop gas vesicles and and float to increase the light-harvesting. The persistence of low photosynthetic rates continually activates the buoyancy system making the cells float to the water surface, creating scums (HUMPHRIES; LYNE, 1988). Scum formation is an efficient strategy for maintaining the CyHAB and Cyanobacteria domination. This condition allows Cyanobacteria to fixate carbon dioxide directly from the atmosphere and to get the advantage of capturing solar irradiation. Scums also shade the water column and prevent competition with other phytoplankton, prolonging the CyHAB (Figure 2.2) (DOKULIL; TEUBNER, 2000).

Phytoplankton growth is often limited by Phosphorus and Nitrogen availability (ELSER; MARZOLF; GOLDMAN, 1990). As Cyanobacteria can fix Nitrogen directly from the atmosphere, they succeed in growing in Phosphorus rich but Nitrogen poor environments (low Nitrogen:Phosphorus ratios). Besides the allochthonous sources (e.g., sewage), Phosphorus inputs might also come from inside the water body. Frequently caused by algae blooms, anoxia can induce the resuspension of Phosphorus from the hypolimnion, provoking positive feedback for other bloom events (SØNDERGAARD; JENSEN; JEPPESEN, 2003). Furthermore, Cyanobacteria dominance in inland water can be further aggravated by climate changes and water warming (EL-SHEHAWY et al., 2012; VISSER et al., 2016). First, Cyanobacteria usually have their temperature optimum above other phytoplankton species ( >25ºC) (PAERL; HUISMAN, 2009). Then, warmer waters are highly susceptible to thermal stratification. Stratified waters usually restrict

phytoplankton movement through the water column, limiting their capacity to harvest light and nutrients. However, some bloom-forming Cyanobacteria are least affected by stratification due to the high mobility offered by the gas vacuoles (WAGNER; ADRIAN, 2009).

Figure 2.2 - Cyanobacteria scum.

Source: Adapted from Grahan et al. (2008).

CyHABs cause adverse effects on aquatic ecosystems (HAVENS, 2008). Some of the most listed are anoxia, unpleasant odor, and taste of water (e.g., methylisoborneal and geosmin), reduction of phytoplankton biodiversity, and pH rise. However, the reason for those blooms to be called “Harmful” is the production of Cyanotoxins (VASCONCELOS, 2006). Since Francis (1878) related CyHABs and cattle intoxication, water managers have been worried about Cyanobacteria. Cyanotoxins are toxic compounds produced inside Cyanobacteria cells that are usually released into the environment after the cell death or zooplankton predation (CARMICHAEL, 2001). However, some species can also excrete those metabolites into the extracellular medium to prevent competition (DOKULIL; TEUBNER, 2000). There were two well established cases of human intoxication by cyanotoxins. The first one happened in Australia, intoxicating 148 persons per contaminated water supply (HAWKINS et al., 1985). The second one occurred in Brazil in 1996, causing injuries to 131 persons, of which 76 of

them died from dialysis water contamination (CARMICHAEL et al., 2001). The Brazilian case is the first confirmed case of human death involving cyanotoxins (WOOD, 2016), although the intoxication of fishes, birds, and other terrestrial animals is usual (CODD; BELL; BROOKS, 1989).

The most common cyanotoxins are Microcystins, Cylindrospermopsins, Saxitoxins, and Anatoxins (CARMICHAEL, 2001). The first two are hepatotoxic and caused both described human intoxications cases (HAWKINS et al., 1985; CARMICHAEL et al., 2001). The last ones are neurotoxic, with the Saxitoxins responsible for causing the Paralytic Shellfish Poisoning (NEGRI; JONES, 1995; CARMICHAEL; BOYER, 2016). Although confirmed cases of human deaths have been registered once, there are reasons to believe that they are underreported (CARMICHAEL; BOYER, 2016), since some cyanotoxins are prone to bioaccumulation on aquatic organisms (NEGRI; JONES, 1995). Biodegradation of the metabolites occurs within several days after the CyHAB dissipation, and some can persist in the water (HO et al., 2012). In some cases, cyanotoxins can be present in water after traditional water treatment but usually in low concentration (LAMBERT; HOLMES; HRUDEY, 1996). The consumption of cyanotoxins in chronic doses has proven to be carcinogenic (CARMICHAEL, 2001). Besides, most cases are likely to happen in healthcare-limited communities, hampering the correct diagnosis and adequate report.

Urban Reservoirs are one of the most suitable environments for Cyanobacteria development (OLIVER; CORBURN; RIBEIRO, 2019). This can be explained by the high concentration of nutrients, increased residence time, recurrent stratification, and low biodiversity (SCHINDLER, 2006). Therefore, the risk of water contamination in reservoirs commonly used for water supply is evident. Furthermore, the environmental factors that lead Cyanobacteria to produce cyanotoxins are not well elucidated. It means that the presence of toxic species and CyHABs are not sufficient for assessing cyanotoxins (CARMICHAEL et al., 2001; CARMICHAEL; BOYER, 2016). The unpredictability of a CyHAB being poisonous and the health risk of cyanotoxins are the main reason for constant water monitoring. Consequently, the precise and fast identification of CyHABs can support the decision-makers to respond with timely actions.

## 2.3 Water quality assessment

### 2.3.1 *In situ* approaches

Monitoring based on *in situ* samplings is the most widespread approach for assessing eutrophication and cyanobacteria presence in aquatic ecosystems (DIXON; CHISWELL, 1996). Some of the most important parameters can only be determined by standard laboratory procedures, such as cyanotoxins identification and quantification. Field campaigns also enable profiling since the algae biomass distribution might be heterogeneous in the water column (KUTSER; METSAMAA; DEKKER, 2008). Furthermore, cell counting still is the most reliable alternative for quantifying cyanobacteria (SRIVASTAVA et al., 2013). However, *in situ* approaches also have some disadvantages because the field campaigns are expensive and time-consuming, limiting water quality parameters' spatial and temporal representation (STACHELEK; MADDEN, 2015). Moreover, laboratory protocols are costly and usually cannot provide timely results (ARNDT et al., 2022). Therefore, monitoring based on a limited number of sampling stations and time-consuming analyses is not enough to address the CyHABs problem correctly.

### 2.3.2 Remote sensing approaches

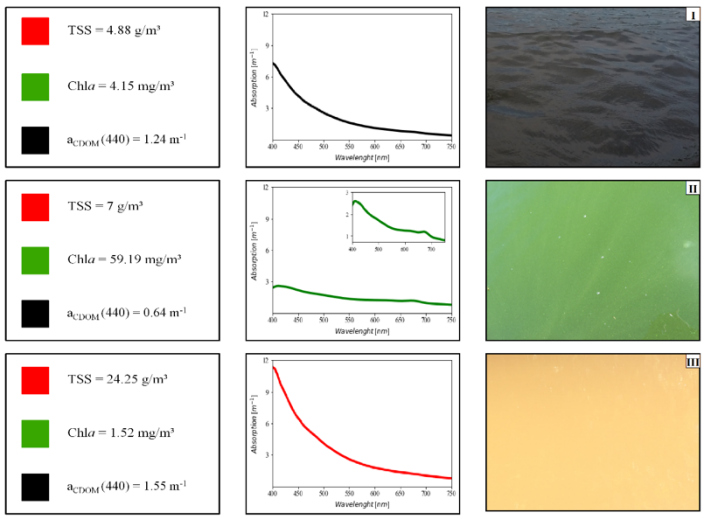
The synoptic view and time repeatability at a relatively low cost are advantages of incorporating Remote Sensing (RS) to monitor water bodies (BARBOSA; NOVO; MARTINS, 2019). This approach is based on the interactions between electromagnetic radiation and the water column (KIRK, 2010). Due to water's absorption bands, photons from shorter wavelengths than blue and longer than near-infrared cannot penetrate deeply (POPE; FRY, 1997; KIRK, 2010). Therefore, water quality assessments are restricted to those constituents that interact with visible and near-infrared (VNIR) radiation (KIRK, 2010; BARBOSA; NOVO; MARTINS, 2019). Satellites and field radiometers can continually record the spectral sign from water. Then, those radiometric measurements can estimate water quality parameters through calibrated bio-optical models, enabling near real-time results. Hence, integrating *in-situ* and RS assessments is recommended to achieve a representative diagnosis, subsidize the decision-making process, and ensure socio-ecological health (SRIVASTAVA et al., 2013).

## 2.4 Optical properties of the aquatic ecosystems

### 2.4.1 Inherent Optical Properties (IOPs)

The electromagnetic radiation field inside water ecosystems is subjected to two processes: Absorption and Scattering (KIRK, 2010). Absorption occurs when a molecule captures a photon, increasing its internal energy (Figure 2.3) (BARBOSA; NOVO; MARTINS, 2019). It is selective because each molecule will need a different amount of energy (corresponding to different wavelengths) to promote an electronic swift, creating a specific absorption spectrum (KIRK, 2010). Scattering may be defined as a deviation of a photon from its current propagation direction to another caused by a particle or molecule (Figure 2.4) (BARBOSA; NOVO; MARTINS, 2019). Larger particles are responsible for most scattering in aquatic ecosystems, even though they are of lower density than small particles (KIRK, 2010). Most photons diverted by particles are scattered by refraction in the forward direction, while few are backscattered (KIRK, 2010). However, backscattering is why an electromagnetic signal comes from the water column (MISHRA; OGASHAWARA; GITELSON, 2017).

Figure 2.3 - Absorption properties of natural inland waters.

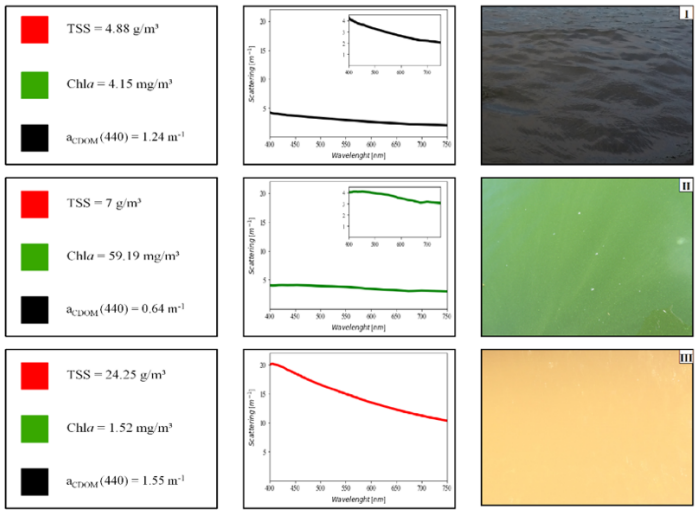



Spectral absorption properties from different Brazilian inland waters. I- Tapajós River/Pará State. II – Ibitinga Reservoir/São Paulo State. III – Lake Curuai/Pará State. Optically Active Constituents' values are shown on the left boxes. The line graphs show the mean absorption coefficient for the first 60 cm of the water column. The measures were taken *in situ* with an ACS. The photos are the visual aspects of waters with those optical properties. TSS = Total Suspended Solids; Chl*a* = Chlorophyll-*a;* $a_{CDOM}(440)$ = Colored Organic Dissolved Matter absorption coefficient at 440 nm.

Source: Author's elaboration.

The absorption and scattering measurements are called Inherent Optical Properties (IOPs) (MOBLEY, 1994). IOPs are those properties whose magnitudes depend only on the aquatic medium's substances and not on the geometric structure of the light fields (KIRK, 2010). The most important are the attenuation coefficient, the absorption coefficient, the scattering coefficient, and the volume scattering function. The first two are measurements of the absorbed or scattered fractions of an orthogonally incident amount of electromagnetic radiation within a thin water layer (KIRK, 2010). The unit is *meters*$^{-1}$, and the measurements are usually made spectrally. The attenuation coefficient is the sum of the absorption and scattering coefficients, while the volume scattering function is the angular distribution of the scattered photons (MOBLEY, 1994). The integration of the volume scattering function is the scattering coefficient (MOBLEY, 1994). Despite the importance of optical characterization, IOPs are difficult to determine directly. Most methods are based on water sampling or *in situ* measurements with expensive equipment (e.g., Absorption and Attenuation meters) (KIRK, 2010). Therefore, indirect methods were developed for estimating IOPs based on the Apparent Optical Properties, as the Quasi-Analytical Algorithm (LEE; CARDER; ARNONE, 2002).

Figure 2.4 - Scattering properties of natural inland waters.



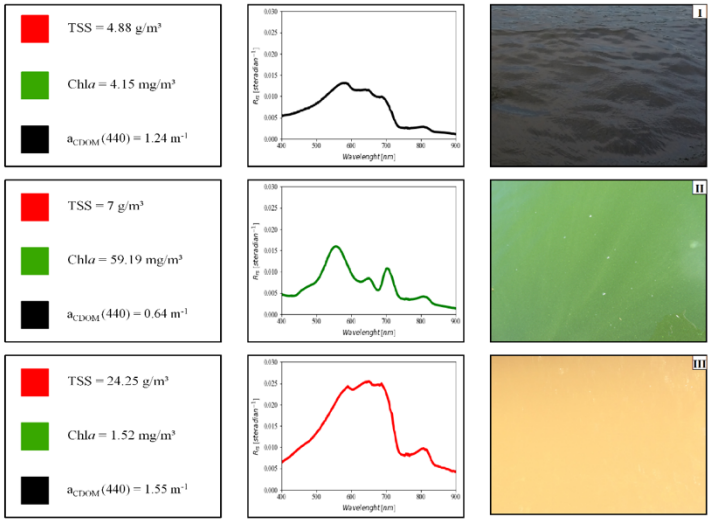

Spectral scattering properties from different Brazilian inland waters. I - Tapajós River/Pará State. II - Ibitinga Reservoir/São Paulo State. III - Lake Curuai/Pará State. Optically Active Constituents' values are shown on the left boxes. The line graphs show the mean scattering coefficient for the first 60 cm of the water column. The measures were taken *in situ* with an ACS. The photos are the visual aspects of waters with those optical properties. TSS = Total Suspended Solids; Chl*a* = Chlorophyll-*a*; $a_{CDOM}(440)$ = Colored Organic Dissolved Matter absorption coefficient at 440 nm.

Source: Author's elaboration.

### 2.4.2 Apparent Optical Properties (AOPs)

The other group of optical properties from aquatic ecosystems are the Apparent Optical Properties (AOPs). AOPs are combinations of radiometric quantities (e.g., quantities related to electromagnetic radiation) used for describing the alterations of an incident electromagnetic radiation field by the water (BARBOSA; NOVO; MARTINS, 2019). They are functions of IOPs and environmental variables, such as the electromagnetic field structure, wind, water surface, clouds, Sun's position, and others (MOBLEY, 1994). However, AOPs demonstrated to be most determined by IOPs and are quite stable to other variables changes (BARBOSA; NOVO; MARTINS, 2019). The most common AOPs are the reflectance values, especially the Remote Sensing Reflectance ($R_{rs}$, unit $steradian^{-1}$). It is a directional measure of the reflectance calculated by the ratio of water-leaving radiance and downward irradiance (MOBLEY, 1999). The variation in depth of the $R_{rs}$ is attributed to the vertical structure of IOPs, allowing IOPs modeling from AOP measures (Figure 2.5) (ZANEVELD, 1982).

Figure 2.5 - Apparent optical properties of natural inland waters.




Apparent Optical Properties from different Brazilian inland waters. I - Tapajós River/Pará State. II - Ibitinga Reservoir/São Paulo State. III - Lake Curuai/Pará State. Optically Active Constituents' values are shown on the left boxes. The line graph shows the Remote Sensing Reflectance (Rrs). The photos are the visual aspects of waters with those optical properties. TSS = Total Suspended Solids; Chla = Chlorophyll-*a*; $a_{CDOM}(440)$ = Colored Organic Dissolved Matter absorption coefficient at 440 nm.

Source: Author's elaboration.

### 2.4.3 Optically Active Constituents (OACs)

The Optically Active Constituents (OACs) are those particles or molecules in the water column that interact with the electromagnetic radiation (KIRK, 2010). OACs have been categorized based on spectral behavior similarities: pure water, Colored Dissolved Organic Matter (CDOM), tripton, and phytoplankton. Each OAC class has its specific IOPs values. Therefore, summing partial OACs contributions results in total IOPs' coefficients once this is a conservative feature. Pure water refers to molecular water spectral properties. It has minor importance for freshwater ecosystems once the visible region's spectral shape is most dominated by other OACs (BARBOSA; NOVO; MARTINS, 2019). CDOM is the organic compounds dissolved in the water that originated from allochthonous (e.g., plants decomposition) or autochthonous (e.g., decomposition of organic matter excreted by algae) processes (KIRK, 2010). Tripton or Non-Algal Particles (NAP) are the inanimate suspended solids that are not photosynthetically active, including inorganic and organic suspended compounds (KIRK, 2010). Phytoplankton have photosynthetic structures and other optically active compounds (BIDIGARE et al., 1990). There are different phytoplankton taxa, with several photosynthetic and protection pigments absorbing photons selectively (KIRK, 2010).

Researchers have been categorizing water bodies based on their water optical properties to facilitate bio-optical modeling. Morel and Prieur (1977) first separated the water bodies into blue and various green waters, subdivided into Case 1 and Case 2. Blue waters are those with spectral features like pure water. Case 1 waters were defined as water bodies whose radiometric spectral behavior was determined by Chlorophyll-a and other AOPs covary with the phytoplankton pigment. In Case 2 waters, the assumption of Chlorophyll-a spectral dominance is violated, and the optical properties depend on all OACs. This classification was important for developing IOPs modeling since bio-optical algorithms were created based on the assumptions made for each class (MOBLEY et al., 2004). However, Mobley et al. (2004) suggested that water must be seen as a mixed bulk of dissolved and suspended particles influenced by other environmental factors and should be analyzed individually. Recently, some efforts have been made to categorize the freshwater ecosystems in Optical Water Types (OWTs) to facilitate RS assessments (SPYRAKOS et al., 2017).

## 2.5 Bio-optical modeling

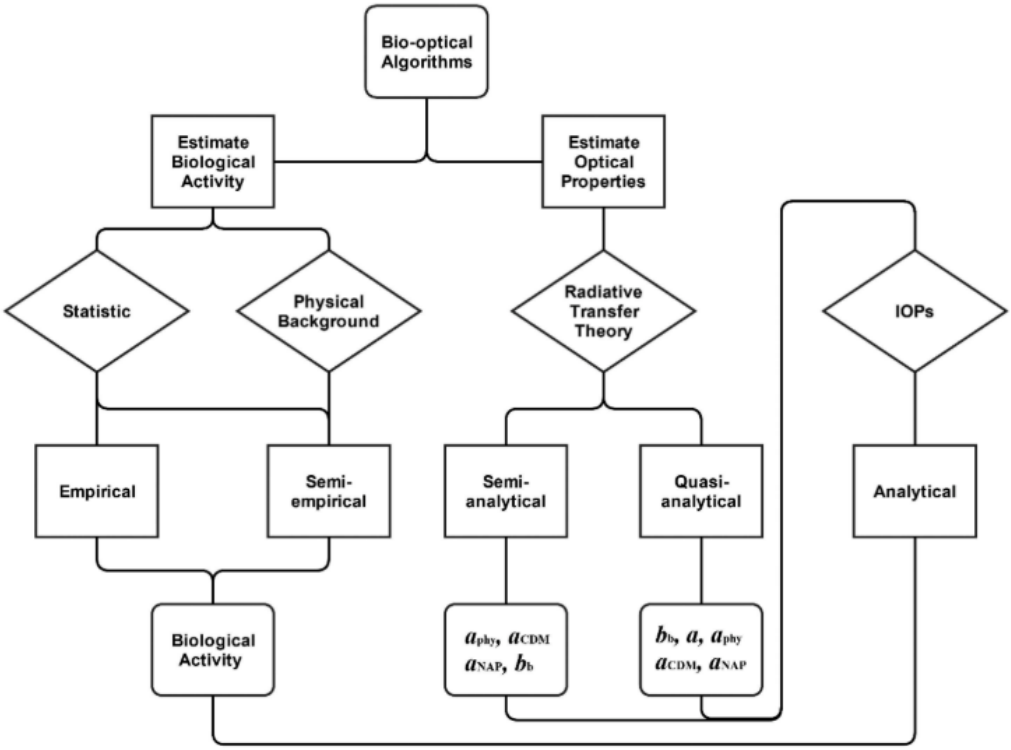
Bio-Optical models can be subdivided into five classes: empirical, semi-empirical, semi-analytical, quasi-analytical, and analytical (MOREL, 2001). The major difference between them is the use of physics or statistics to derive the OACs from radiometry. Empirical and semi-empirical algorithms are based on statistical relationships (YAN; BAO; SHAO, 2018). Empirical algorithm band selection is built only at the best correlations between radiometric data and OACs measurements (SHI et al., 2019). It means that the model can be constructed based on a ocasional relationship, limiting their application for different time and space observations. Semi-empirical models take advantage of a physical background to find the best correlations. The expected behavior of an OAC is used for limiting the spectral search, but statistical models still make the conversion of radiometric quantities to concentration.

The semi-analytical and quasi-analytical approaches rely on the inversion of the optical properties based on the radiative transfer theory (LEE; CARDER; ARNONE, 2002). They turn AOPs into IOPs and then analytically estimate the OACs. Quasi-analytical models derive total IOPs from AOPs and then partition them into the specific contributions of each OACs group. Otherwise, semi-analytical models directly derive IOPs for each OAC class. The analytical models are based only on physical properties, such as Gons (1999), which derived the concentration of Chlorophyll-*a* from total phytoplankton absorption and phytoplankton specific absorption ratio. A complete analytical approach may not be achieved once the radiative transfer equation isn't completely solved yet (MOBLEY, 1994). However, simplifications are made for analytically estimating OACs after deriving IOPs from semi-analytical or quasi-analytical approaches (Figure 2.6).

Bio-optical modeling has been used to estimate Phytoplankton, Non-Algal Particles (NAP), and Colored Dissolved Organic Matter (CDOM) in aquatic ecosystems (PALMER; KUTSER; HUNTER, 2015). Although this is the most used approach to derive water quality parameters from RS data, Machine Learn Algorithms (MLA) have recently gained some space (PYO et al., 2019; CAO et al., 2020). MLA are empirical algorithms once the estimates are also based on statistical relationships (SAGAN et al., 2020). The data blending capacity and non-linear relationship sensitivity are advantages

of using MLA. The greatest disadvantage of this modeling approach is that the equations for retrieving the output parameters from the input data are often unknown (LARY et al., 2016). However, MLA have shown solid results and outperformed bio-optical models (SAGAN et al., 2020).

Figure 2.6 - Flowchart of the bio-optical approaches classification.




Source: Ogashawara (2015).

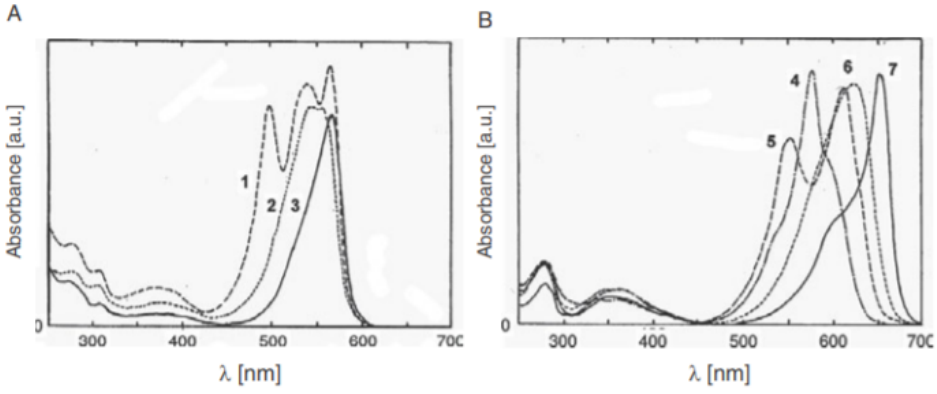
## 2.6 Advances in remote sensing-based methodologies for C-Phycocyanin estimation

Cyanobacteria and CyHABs monitoring based on RS were first based on by Chlorophyll-*a* (Chl*a*) estimation (REINART; KUTSER, 2006). This photosynthetic pigment is present in all phytoplankton species and is a widespread proxy for primary production (GURLIN; GITELSON; MOSES, 2011). Several studies focused on Chl*a* retrieval have succeeded in Cyanobacteria biomass estimation over inland water ecosystems (SHI et al., 2019). The use of Chl*a* as a proxy for cyanobacteria can be considered a good approximation for waters dominated by Cyanobacteria. However, this assumption does not work for sites with diverse phytoplankton communities (SIMIS et al., 2007). Therefore, Chl*a* cannot be

used as a Cyanobacteria biomass proxy as a rule. Phytoplankton has proteins other than Chl*a* that interacts with electromagnetic radiation, which could be used in RS assessments (KIRK, 2010). A variety of photosynthetic pigments have been described and were organized into three classes: Chlorophylls, Carotenoids, and Biliproteins. The first two are present in all phytoplankton, but biliproteins are only found on Rhodophyta (red algae), Cyanophyta (blue-green algae), and Cryptophyta (KIRK, 2010).

There are four kinds of biliproteins: phycoerythrins, phycoerythrocyanins, phycocyanins, and allophycocyanin (Figure 2.7). In Rhodophyta and Cyanophyta, biliproteins are organized in structures called phycobilisomes. They are responsible for harvesting the electromagnetic energy and transferring it to Chl*a,* where photosynthesis occurs (KIRK, 2010). For this reason, biliproteins are called accessory pigments. Despite being present in Rhodophyta and Cryptophytes, C-Phycocyanin (PC) is a major component only in Cyanophyta. Because it is correlated only to cyanobacteria biomass, PC is potentially better than Chl*a* for Cyanobacteria monitoring in inland waters (SIMIS; PETERS; GONS, 2005; KIRK, 2010). The main spectral feature of PC is the absorption maxima near 620 nm, which allows the detection of cyanobacteria employing RS methodologies (DEKKER, 1993; SIMIS; PETERS; GONS, 2005; MISHRA; MISHRA; LEE, 2013).

Figure 2.7 - Biliproteins absorption spectra.




Absorption spectra, in absorbance units, of Cryptophyta (3, 4, 6, 7) and Rhodophyta (1, 2, 4, 6, 7) biliproteins. Phycoerythrins (A) 1- R-Phycoerythrin; 2- B-Phycoerythrin; 3- C-Phycoerythrin. Phycocyanins (B) 4- Phycoerythrocyanin; 5- R-Phycocyanin; 6- C-Phycocyanin; 7- Allophycocyanin. C-Phycocyanin (6) spectral features are widely used as a Cyanobacteria biomass proxy in remote sensing methodologies (HUNTER et al., 2009).

Source: Adapted from Roy et al. (2011).

From the spectral features of PC, bio-optical algorithms were developed for monitoring CyHABs. Based on the published reviews (RUIZ-VERDÚ et al., 2008; OGASHAWARA et al., 2013; YAN; BAO; SHAO, 2018; SHI et al., 2019), the most cited are presented in Table 2.1. Dekker (1993) has made the first attempt to derive PC from RS data. Also, the author noticed that using only radiometric measurements from PC absorption maxima and the pigment concentration were not enough for accurate estimates. Therefore, he created a spectral baseline from two wavelengths with the minor influence of PC and then related it to the pigment absorption feature, reducing the effect of other OACs. Simis et al. (2005) adapted the Chl*a* semi-analytical algorithm of Gons (1999) and published one of the most successful bio-optical algorithms for PC retrieval. It was the first time that the spectral influence of Chl*a* in 620 nm was addressed in the modeling process. Liu et al. (2017) advanced on trying to isolate the spectral sign from PC and considered the influence of CDOM and other phytoplankton pigments in a semi-empirical four-bands algorithm. However, Sun et al. (2012) proved that MLA could be more accurate than bio-optical algorithms. Band ratios, correlated with PC concentration, were used as input features in a Support Vector Regression and the results overperformed previously published algorithms. More recently, O'Shea et al. (2021) calibrated a Mixture Density Network for predicting PC based on radiometric measurements simulated for HICO and PRISMA sensors. This study has achieved accurate estimates even for low PC concentrations, where other publications were usually inaccurate. Considering the accuracy consistency for different concentration ranges and the broad number of reservoirs used for calibrating and validating the algorithm, MDN is the closest to a global PC model ever published.

Table 2.1 - Review from remote sensing algorithms for C-Phycocyanin retrieval.

| Publication | Model Type | Model | PC Range | Accuracy |
|---|---|---|---|---|
| Dekker (1993) | Semi-Empirical | $PC \propto 0.5[R(0^{-1})_{600} + R(0^{-1})_{624}] - R(0^{-1})_{624}$ | 7 – 130 µg/L | STE = 2.34 µg/L |
| Schalles and Yacobi (2000) | Semi-Empirical | $PC \propto R(650)/R(625)$ | 0 – 530 µg/L | $R^2 = 0.612$ |
| Simis et al. (2005) | Semi-Analytical | $a_{pc}(620) = (\{[R(709)/R(620)] * [a_w(709) + b_b]\} - b_b - a_w(620)) * d^{-1}d^{-1}\delta^{-1}d^{-1} - [e * a_{chl}(665)][e * a_{chl}(665)][\varepsilon * a_{chl}(665)][e * a_{chl}(665)]$ $PC = a_{pc}(620)/a^*_{pc}(620)$ | 0 – 80 µg/L | RMSE = 6.5 µg/L |
| Hunter et al. (2010) | Semi-Empirical | $PC \propto [R_{rs}^{-1}(615) - R_{rs}^{-1}(600)] * R_{rs}(725)$ | 0 – 93.7 µg/L | RMSE = 2.65 µg/L |
| Sun et al. (2012) | Machine Learning | Support Vector Regression | 1.6 – 754.9 µg/L | RMSE = 38.4 µg/L |
| Mishra et al. (2013) | Quasi-Analytical | QAA | 68.13 – 3032.47 µg/L | ARE = 34.9 µg/L |
| Mishra and Mishra (2014) | Semi-Empirical | $PC \propto [R_{RS}^{-1}(629) - R_{RS}^{-1}(659)] * R_{RS}(724)$ | 68.13 – 3032.47 µg/L | STE = 150.38 µg/L |
| Liu et al. (2017) | Semi-Empirical | $PC \propto \left[\frac{1}{R_{rs}(620)} - \frac{0.4}{R_{rs}(560)} - \frac{0.6}{R_{rs}(709)}\right] R_{rs}(754)$ | 0 – 329.4 µg/L | RMSE = 27.691 µg/L |
| Ogashawara and Li (2019) | Semi-Empirical | $PC \propto \frac{\frac{R_{rs}(709)}{R_{rs}(620)} - \left(\frac{R_{rs}(709)}{R_{rs}(665)}\varphi_1\right)}{1 - (\varphi_1\varphi_2)}$ | 0.73 – 370.95 µg/L | RMSE ≤ 23.004 µg/L |
| O'Shea et al. (2021) | Machine Learning | Mixture Density Network | 0 – 1000 µg/L | $e$ = 44.3% |

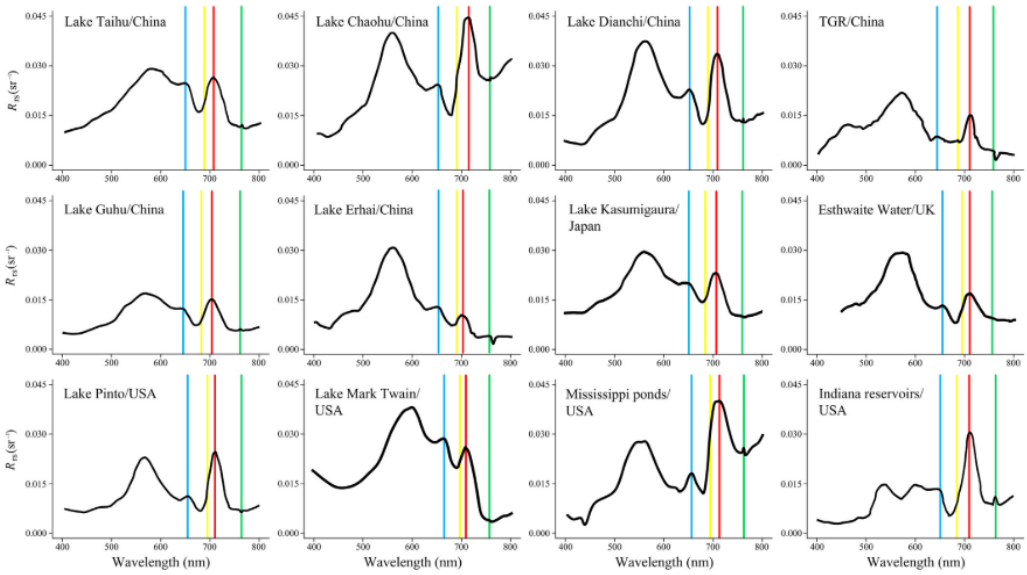
The accuracy metrics were chosen based on the most informative metrics presented by the authors. PC = C-Phycocyanin, RMSE = Root Mean Squared Error, STE = Standard Error, ARE = Absolute Relative Error, $R_{rs}(\lambda)$ = Remote Sensing Reflectance at wavelength λ, $R(0^{-1})_{\lambda}$ = Irradiance Reflectance just below the water surface at wavelength λ, R(λ) = Reflectance Factor just below the water surface at wavelength λ, QAA= Quasi Analytical Algorithm (LEE; CARDER; ARNONE, 2002), $a_{pc}(620)$ = PC absorption coefficient at 620 nm, $b_b$ = Backscattering coefficient, $a_w(\lambda)$ = Water absorption coefficient at wavelength λ, δ = Optimization coefficient, $e$ = Median Symmetric Accuracy (MORLEY; BRITO; WELLING, 2018), $a^*_{pc}(620)$ = PC specific absorption coefficient at 620 nm, $R^2$ = Coefficient of determination, $a_{chl}(\lambda)$ = Chl-*a* absorption coefficient at wavelength λ, $\varphi_1$ = Coefficient to correct the influence of Chl-*a* on PC absorption at 620 nm, $\varphi_2$= Coefficient to correct the influence of PC on Chl-*a* absorption at 665 nm, ε = Conversion factor from $a_{chl}(665)$ to $a_{chl}(620)$.

Source: Author's elaboration.

Despite several publications, there is no universal model for predicting PC (SHI et al., 2019). The major challenge is predicting low concentrations since algorithms are most accurate at moderate PC:Chl*a* ratios. For example, Simis et al. (2005) found optimum estimates at PC:Chla > 0.4. Increasing the proportion of PC over Chl*a* indicates

Cyanobacteria dominance and CyHABs establishment. This condition enhances PC absorption feature and enables a more precise estimate from RS data. However, the influence of other OAC's absorption signs around 620 nm at high and low PC:Chl*a* ratios are difficult to correct. When Chla is abundant in the water, its absorption spectrum is expanded toward longer and shorter wavelengths (ZHANG et al., 2012). The contribution of Chl*a* absorption at 620 nm is responsible for observed overestimations at high PC:Chl*a* (SIMIS; PETERS; GONS, 2005). Some of the reported algorithms provide coefficients for removing the effects of the Chl*a* absorption, but the results showed that a residual contribution might still be present after the correction. It was observed that coefficients calibrated based on *in vitro* or empirical approaches were prone to packaging effect and site dependence, respectively (SIMIS; PETERS; GONS, 2005; OGASHAWARA; LI, 2019). The packaging effect was also an underestimation factor in PC modeling at high algae concentrations (ALCÂNTARA et al., 2016).

Figure 2.8 - Spectral shape of Cyanobacteria dominated inland waters.




*In situ* measured Remote Sensing Reflectances ($R_{rs}$) from different Cyanobacteria-dominated inland waters. The scattering peak (red), the fluorescence peak (yellow), the C-Phycocyanin reflectance peak (blue), and Chlorophyll-a reflectance peak (green) are also shown.

Source: Shi et al. (2019).

Disregarding MLA, the most reasonable error metrics for low concentration ranges are provided by Simis et al. (2007) and are limited to samples with PC > 50 µg/L. Low

PC:Chl*a* ratios might indicate a more diversified phytoplankton community and increase the chance of the spectral influence of other photosynthetic pigments than Chl*a* (SIMIS et al., 2007). Chlorophyll-*b*, Chlorophyll-*c*, Phaeophytin, Fucoxanthin, and Allophycocyanin-*b* absorption spectra were described as overlapping PC features (Figure 2.8). Therefore, the increased absorption from other pigments than PC around 620 nm is a source of overestimations at low PC:Chl*a* ratios. Furthermore, the PC spectral sign is difficult to detect at low concentrations due to its reduced absorption rates. Dekker (1993) has reported PC-specific absorption coefficient threefold lower than Chl*a* for the same concentrations. In addition to the impact of other pigments, the optical complexity related to them and the remaining OACs is still to be addressed. Most PC algorithms make assumptions to simplify bio-optical modeling, such as invariant backscattering from red to NIR region or negligible CDOM and NAP absorption at PC absorption range (LIU et al., 2017). Those simplifications are also sources of error, which become more expressive at low PC concentrations (LIU et al., 2017). At last, the lack of a standard laboratory procedure for extracting PC from cyanobacteria cells reduces the reliability of the analytical determinations (RUIZ-VERDÚ et al., 2008). Those associated errors tend to be more relevant at low PC concentrations.

## 2.7 Orbital data for cyanobacteria monitoring

### 2.7.1 Multispectral orbital data

Satellite remote sensing data have grown in number over the last years. With the increasing number of open access policy missions, monitoring through RS has been democratized even for nations without a spatial program (BELWARD; SKØIEN, 2015). However, assessing information about inland water ecosystems from space is still challenging. RS applications for freshwater usually demand high spatial, spectral, high temporal resolutions and a high Signal to Noise Ratio (SRN) (BARBOSA; NOVO; MARTINS, 2019). However, no available sensors present all those characteristics, and the resolution's trade-off limits the application of the available orbital sensors. Still, the scientific community has found alternatives to suit the available data. Ocean Color Sensors (OCS) were designed for RS ocean water applications, with adequate spectral bands for assessing the most important OACs features and high SNR (YAN; BAO; SHAO, 2018; OGASHAWARA; LI, 2019). Considering that most free-data OCS have

been decommissioned or near the mission's end, the European Space Agency's Sentinel-3 is the most promising satellite family for water applications. This sensor has been extensively used for freshwater and ocean water quality assessments (BLIX et al., 2018; KRAVITZ et al., 2020; VANHELLEMONT; RUDDICK, 2021), and some authors have pointed out it as the most suitable available instrument for monitoring cyanobacteria and CyHABs (YAN; BAO; SHAO, 2018; SHI et al., 2019). The spectral band centered at 620 nm and almost daily images are the greatest advantages of using the Ocean and Land Color Instrument (OLCI) to monitor CyHABs. However, its application to urban reservoir monitoring is questionable due to its spatial resolution of 300 meters. In contrast, the Multispectral Instrument (MSI), abord Sentinel-2, has an adequate spatial resolution for monitoring Urban Reservoirs, and the inclusion of Red-Edge bands enhanced the retrieval of Chla in inland waters (BRAMICH; BOLCH; FISCHER, 2021). However, the lack of a spectral band centered at 620 nm limits the application of Sentinel-2/MSI to monitor PC.

The Landsat family is the oldest series of satellites designed for Earth Observation (LOVELAND; IRONS, 2016). The Landsat-8, equipped with the Operational Land Imager (OLI), has been successfully used for deriving freshwater ecosystems information worldwide, including in tropical reservoirs (MISHRA; OGASHAWARA; GITELSON, 2017; MACIEL et al., 2019). Using Landsat data also guarantees an almost half-century image archive for assessing water quality parameters, giving a unique opportunity to carry out long-term analysis and evaluate temporal environmental changes (SHI et al., 2019). Unfortunately, Landsat-8/OLI does not have a spectral channel at 620 nm. Still, Landsat platforms were already used to monitor PC through the original-designed spectral bands (VINCENT et al., 2004). However, Castagna et al. (2020) have developed a method of deriving a virtual contra-band from the panchromatic channel capable of detecting the spectral signature of PC. Kumar et al. (2020) have assessed the use of the virtual contra-band derived from Landsat-8/OLI for monitoring cyanobacteria in Lake Erie and concluded that the band has a similar radiometric response to OLCI orange band, with potential for use in CyHABs monitoring. This technique has never been used to evaluate tropical urban reservoirs, which can be one the most beneficed water bodies once they usually are not great enough to be assessed by OCS.

The Worldview is a series of high-resolution satellites which launched its first platform in 2007. The greatest advantage of using Worldview data is the multispectral bands at a resolution of fewer than 2 meters in VNIR (ZHAO et al., 2022). This spatial sampling capacity enables the application of Worldview data to anthropic environments (VERLIČ et al., 2014). Worldview radiometric data have proven suitable for water ecosystem applications (WILSON; WONG; DEVRED, 2022). Additionally, Worldview-2 images were used for deriving Chl*a* in optically complex inland water ecosystems with accuracy metrics compared with OCS (WANG; GONG; PU, 2018). Therefore, the increased spatial resolution, the short-revisit period (about 1 day), and the presence of a spectral band centered at 605 nm enables Worldview-2 and Worldview-3 to monitor cyanobacteria in urban reservoirs. Beck et al. (2017) estimated PC fluorescence values in a temperate reservoir using aircraft hyperspectral data simulated to Worldview-2/3 spectral response function. The results from the high-resolution satellite were comparable to or better than other Earth Observation programs. However, the on-demand imaging and the commercial data access police are limitations for applying Worldview data in continuous water monitoring programs. Additionally, the reduced swath width of the Worldview series (13-18 km) might limit the monitoring of large reservoirs.

### 2.7.2 Hyperspectral orbital data

Multispectral sensors have been used to monitor natural landscapes with a high degree of success since the launch of the first Landsat platform back in 1972 (WULDER et al., 2019). Concerning those sensors, specialists are responsible for designing the spectral resolution to maximize the applications and avoid atmospheric noises (e.g., bands placed on atmospheric windows). However, the growing interest in Remote Sensing and the investigation of new complex subjects might demand a better spectral resolution than the Multispectral satellites offer (DIERSSEN et al., 2021). Therefore, sensors with an ultra-high spectral resolution, called Hyperspectral (or Spectrometer), were meant to cover the electromagnetic spectrum with hundreds of narrow and contiguous channels (GIARDINO et al., 2019). The development of hyperspectral sensors occurred in the 1970s, conducted by NASA's Jet Propulsion Laboratory. At the time, the need for validating the data of the recently launched Landsat-1 pushed the creation of the firsts portable and airborne spectrometers (GOETZ, 2009). The Airborne Imaging Spectrometer (AIS) was the earliest hyperspectral imager, which covered 307 nm from

2030-2320 nm with a sampling interval of 9.6 nm and resolution of 32x32 pixels. AIS data was successfully used for identifying surface mineralogy, which wasn't possible using Multispectral data, and increased the investments and interest in hyperspectral imaging (GOETZ, 2009). Then, in 1987, the first flight of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) produced a hyperspectral image from 400 to 2400 nm with 224 bands (9.6 nm resolution). AVIRIS was a great advance for spectrometry due to its large spectral range (covering the VNIR) and improved radiometric quality (GREEN et al., 1998). Other airborne instruments came after AVIRIS; however, it was only in 2000 that the first hyperspectral orbital sensor was launched: NASA's Hyperion, abord the Earth Observation-1 (EO-1).

Brando and Dekker (2003) produced maps of Chl*a*, CDOM, and NAP from Hyperion data for Deception Bay (Australia) and marked the beginning of the use of orbital hyperspectral data for water applications. The authors used the semi-analytical approach to retrieve water quality parameters and concluded that Hyperion had enough radiometric quality to produce accurate estimates for water products. EO-1/Hyperion data was also used for retrieving Chl*a* and NAP in inland waters, using the Italian Lake Garda as a study case (GIARDINO et al., 2007). However, despite some successful attempts to use orbital images, most studies produced in the last 20 years using hyperspectral data were developed based on airborne platforms (GIARDINO et al., 2019). There are some reasonable explanations for the lack of assessments using satellite data. First, hyperspectral data have reduced SRN due to the improved spectral resolution (MOSES et al., 2012). This characteristic is especially problematic when analyzing water ecosystems that are considered 'dark targets', and even low radiometric noise introduction might greatly influence the final outputs (BARBOSA; NOVO; MARTINS, 2019). Secondly, a few years ago, the huge amount of data contained in a hyperspectral image was very hard to handle (GOETZ, 2009). Besides the challenge of processing and storing so much data, few techniques and software were available for manipulating the images obtained from hyperspectral sensors. Finally, the atmospheric correction had to be improved because of the increased presence of atmospheric effects on the spectral bands outside the atmospheric window (GAO et al., 2009).

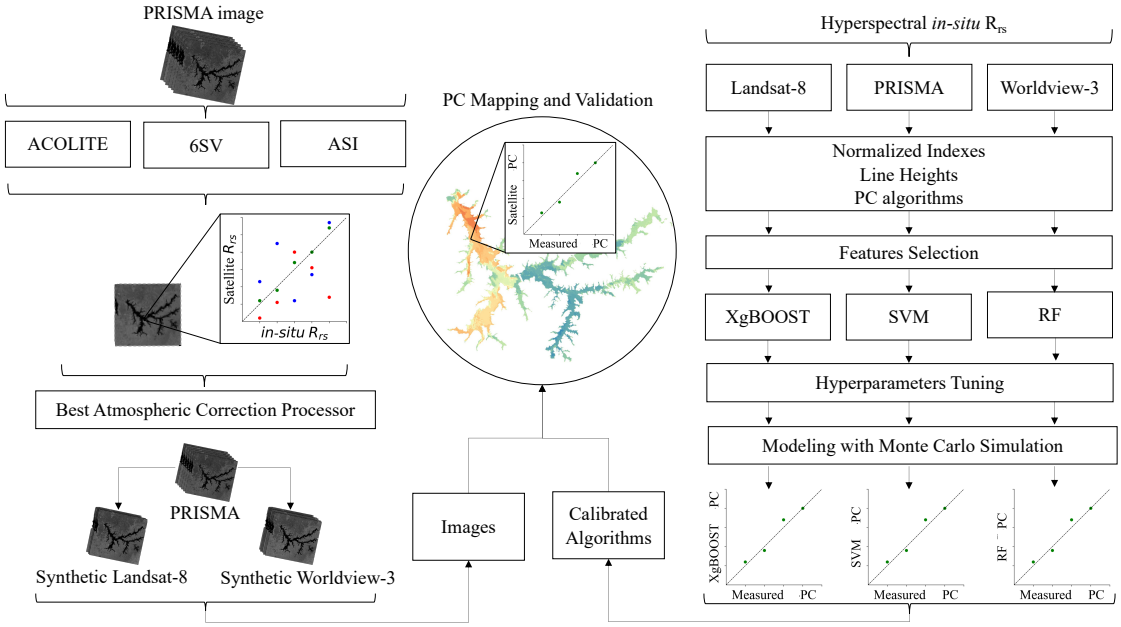
Even with the challenges presented above, some hyperspectral platforms were launched besides Hyperion in the millennium's first decade. PROBA-1/CHRIS and ISS/HICO were

also used for studying inland water ecosystems and contributed to the early development of hyperspectral instruments and applications (MOSES et al., 2014; TORBICK; CORBIERE, 2015). However, the last years of the 2010s decade have brought a new generation of hyperspectral sensors. The improvements in the engineering of detectors and the development of computational resources (e.g., cloud computing, petabyte databases, increase in hardware processing power) has enabled the construction of powerful instruments and reduced the problems reported in the past (DIERSSEN et al., 2021). Some of those missions are expected to revolutionize the Earth Observation applications, such as: PRISMA (*PRecursore IperSpettrale della Missione Applicativa*), EnMAP (Environmental Mapping and Analysis Program), PACE (Plankton, Aerosol, Cloud, ocean Ecosystem), DESIS (German Aerospace Center's Earth Sensing Imaging Spectrometer), HyspIRI (Hyperspectral Infrared Imager), CHIME (Copernicus Hyperspectral Imaging Mission for the Environment) and Geofen-5/AHSI (Advanced Hyperspectral Imager). Those satellites are also expected to enhance the monitoring of C-Phycocyanin from space. The applicability of hyperspectral data for retrieving PC in inland waters was already confirmed by studies using airborne imagery (KUDELA et al., 2015; BECK et al., 2017; YIM et al., 2020). However, few attempts were made to use orbital spectroscopy (TORBICK; CORBIERE, 2015; O'SHEA et al., 2021; BRESCIANI et al., 2022), evidencing the gap in the development of studies based on satellite images.

# 3 MATERIAL AND METHODS

The flowchart in Figure 3.1 resumes the applied methodology, and detailed information is given in the next items. On the left side of Figure 3.1, the image processing is explained, whereas on the right-side information about PC modeling is presented. The center of the flowchart regards PC mapping and validation.

Figure 3.1 - Methodology flowchart.



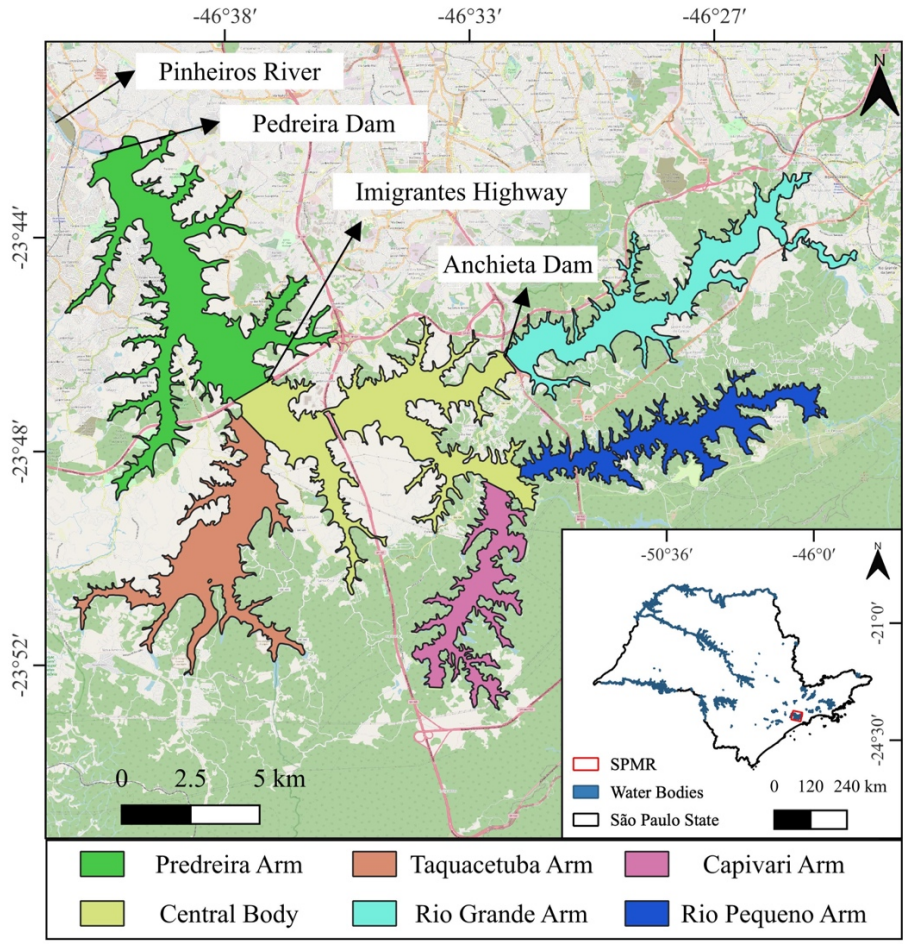

Source: Author's elaboration.

## 3.1 Study area

Billings Reservoir (BIL) is the largest reservoir in São Paulo Metropolitan Region (SPMR) (Figure 3.2). It is an artificial water body with an approximate surface area of 120 km² and a mean depth of 10 meters. BIL has a central role in the water supply of the SPMR, providing water directly for 1.2 million habitants (CÔRTES et al., 2015). Additionally, Billings indirectly supply other SPRM regions due to the transposition of its waters to other reservoirs. Despite its importance to water security, BIL suffers an intense eutrophication and water contamination, as evidenced by the long-term monitoring system (CETESB, 2019). Algae blooms are frequent in the reservoir, with regular CyHABs events reported (CARVALHO et al., 2007; GEMELGO et al., 2008). *Raphidiopsis raciborskii* and *Microcystis aeruginosa*, both potentially toxic

cyanobacteria species, have been constantly dominant or abundant in the phytoplankton community. The low N:P ratio, high residence time, low phytoplankton/fish biodiversity, intense load of contaminants, and water thermal stratification events are the most important factors that lead to continuous Cyanophyta presence at BIL (MOSCHINI-CARLOS et al., 2009; WENGRAT; BICUDO, 2011).

Figure 3.2 - Study area.




Source: Author's elaboration.

## 3.2 *In situ* data

Eight field campaigns between November/2020 and December/2021 collected the radiometric and limnologic data for calibrating the PC models (Table 3.1). The data acquisition was performed from 10:00 to 14:00h local time to avoid low Sun elevations. In addition, the campaigns were carried out during different periods of the year to capture seasonal variability in the water body composition.

Table 3.1 - Field campaigns description.

| Year | Month | *n* radiometric samples | *n* pigments samples | *n* taxonomical samples |
|---|---|---|---|---|
| 2020 | November | 17 | 17 | 11 |
| 2021 | July | 24 | 24 | 0 |
| 2021 | August | 30 | 30 | 9 |
| 2021 | August | 7 | 7 | 0 |
| 2021 | September | 8 | 8 | 0 |
| 2021 | October | 8 | 8 | 2 |
| 2021 | November | 24 | 13 | 4 |
| 2021 | December | 8 | 8 | 0 |

Source: Author's elaboration.

### 3.2.1 Chlorophyll-*a* and C-Phycocyanin determination

Water samples were collected from the subsurface (~0.2m), stored in dark bottles, and put into ice to prevent photo and thermal degradation. After no longer than six hours, subsamples (100-500 mL) were filtered in GF/F filters (Whatman, 47 mm diameter, 0.7 μm pore), under low light conditions and low vacuum, packaged on dark recipients, and kept in liquid nitrogen until the end of the campaign. Then, the filtered samples were stored at -70ºC until they were analyzed. For Chl*a* determination, extraction and determination followed the method described by APHA (1998). First, the filters were homogenized, and the pigments were extracted using 90% acetone. After a period of extraction of least than 12 hours at ~4ºC, the samples were centrifugated for 20 minutes at 3000 rpm. Next, the supernatant was placed on quartz cuvettes and read on a spectrophotometer. Then, the samples were acidified with hydrochloric acid (HCl 0.1M), and after a reaction time of 90 seconds, another spectrophotometric measurement was made. The Chl*a* concentrations were calculated by Lorenzen's equation (APHA, 1998), correcting for the presence of Pheophytin-*a*. PC extraction followed the protocol proposed by Sarada et al. (1999) and adapted by Horváth et al. (2013). The filters were suspended in a phosphate buffer (100 mM and pH 7.2) and submitted to 3 freeze-thaw cycles (-70ºC and 35ºC). Then, the samples were sonicated for 90 seconds at a frequency of 20 kHz and centrifugated for 15 minutes at 3000 rpm. The PC concentration was determined spectrophotometrically through the formula proposed by Bennet and Bogorad

(1973). Duplicates were made for all sampling stations, where a simple mean was used for further analysis.

### 3.2.2 Taxonomical analysis

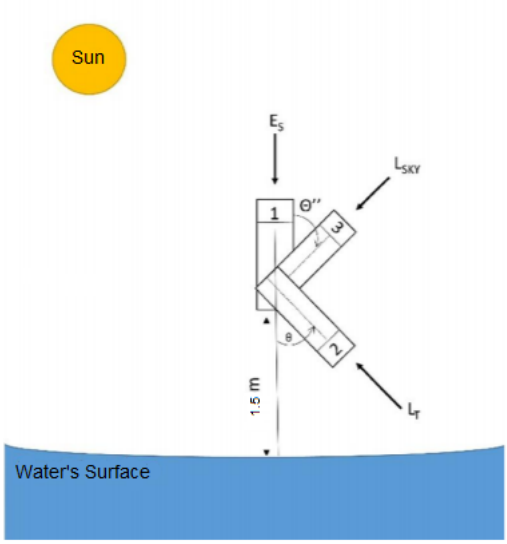
Some sample stations were selected for taxonomical identification and quantification of the phytoplankton community. The water samples were taken concomitantly for all the described limnological analyses. For the taxonomical identification, a volume was concentrated using a membrane of 0.55 μm pore and fixated with formalin (4%). As for the phytoplankton quantification, 100 mL of water from the subsurface was preserved with Lugol solution (1%). All samples were stored in the dark and at room temperature. The analysis was made using an inverted Carl Zeiss optic microscope with augmentation of 400X and sedimentation chambers of 5 mL and 10 mL. The phytoplankton cell density was expressed in individuals per milliliters. The densities were converted to biovolume ($mm^3/L$) as described by Hillebrand et al. (1999) and Fonseca et al. (2014). A specie was considered dominant when its biovolume was more than half of the total phytoplankton biovolume (LOBO; LEIGHTON, 1986). The abundant species are those that exceed the average biovolume of the sample.

### 3.2.3 Radiometric data

*In situ* radiometric measurements were carried out using TRIOS-RAMSES spectroradiometers, ranging from 400-900 nm with a sampling interval of approximately 3.3 nm. The three intercalibrated radiometric equipment were mounted in the boat bow, following the configuration proposed by Mobley (1999) (Figure 3.3):

- One radiance radiometer, with 7º Field-of-View (FOV), was used to measure total water-leaving radiance ($L_t$). This sensor was positioned at a sensor-viewing geometry of 45º zenithal angle and approximately 135º azimuth angle (taking the Sun position as reference);
- One irradiance radiometer equipped with a cosine collector pointed up, perpendicular to the water surface, to measure downward irradiance ($E_s$);
- With the same FOV as the first, another radiance radiometer measured the diffuse sky radiance ($L_{sky}$). It was positioned at 45º from the nadir direction and approximately 135º azimuth angle, forming a right angle with the $L_t$ radiometer.

Figure 3.3 - Sensor-viewing geometry.




The number one represents the downward irradiance sensor ($E_s$); the number 2 represents the total water-leaving radiance sensor ($L_T$); the number 3 refers to the sky radiance sensor ($L_{sky}$); θ is the zenithal angle from the total water leaving radiance sensor (45º) and Θ" the nadiral angle from the sky radiance sensor (45º).

Source: Adapted from Maciel (2019).

The 135º azimuthal angle was defined to reduce the Sun glint influence over the radiometric measurements (MOBLEY, 1999). The radiometers were placed at 1.5 meters from the water surface. All the radiometric quantities measurements were made concomitantly. The radiometric spectral measurements were processed following Mobley (1999) procedure. First, according to Equation 3.1, the water leaving radiance ($L_w$) was calculated as:

$$L_w(\theta, \phi, \lambda) = L_t(\theta, \phi, \lambda) - \rho(\theta, \phi)\, L_{sky}(\theta, \phi, \lambda) \quad (3.1)$$

Where $L_w$, $L_{t,}$ and $L_{sky}$ are a function of the zenithal angle ($\theta$), azimuthal angle ($\phi$), and wavelength ($\lambda$). The factor $\rho\ (\theta, \phi)$ was obtained in Mobley (2015), using Sun's zenithal angle and sensor-viewing geometry, and the wind speed, which was measured with an anemometer concomitantly with the radiometric sampling. Then, Remote Sensing Reflectance ($R_{rs}$) was estimated following Equation 3.2 (MOBLEY 1999):

$$R_{rs}(\theta, \phi, \lambda) = \frac{L_w(\theta, \phi, \lambda)}{E_s(\lambda)} \quad (3.2)$$

After processing all the radiometric data from each point, the outliers were removed through visual inspection, and one spectrum was chosen to represent the sample station based on the result of Equation 3.3 (MACIEL et al., 2019):

$$Dif_{Rrs(i,j)} = \sum_{\lambda=400}^{900} |Rrs(i,j,\lambda) - Rrs_{median}(j,\lambda)| \quad (3.3)$$

Where $Dif_{Rrs(i,j)}$ is the absolute spectral difference between an *i*-th sample and the median spectrum for a *j*-th sampling station. For some specific wavelength ($\lambda$), $Rrs(i,j,\lambda)$ is the sample $R_{rs}$ and $Rrs_{median}(j,\lambda)$ is the median $R_{rs}$ value for the *j*-th station. The chosen spectrum was the one that reached the lowest value of $Dif_{Rrs(i,j)}$ in each sampling station. After selection, the chosen $R_{rs}$ spectrum was interpolated for a resolution of 1 nm. Finally, all sample station spectra were simulated for PRISMA, Landsat-8/OLI, and Worldview-3 Spectral Response Function (SRF).

$$Rrs_{band(i)} = \frac{\int_{\lambda 1}^{\lambda 2} SRF(\lambda) * Rrs_{m(\lambda)} d\lambda}{\int_{\lambda 1}^{\lambda 2} SRF(\lambda) d\lambda} \quad (3.4)$$

Where $Rrs_{band(i)}$ is the *i*-th simulated band for the target sensor, $SRF(\lambda)$ is the Spectral Response Function in a given wavelength, and the $Rrs_{m(\lambda)}$ is the *in situ* measured Rrs for the same wavelength.

## 3.3 Hyperspectral PRISMA data processing

### 3.3.1 Hyperspectral PRISMA image

The *PRecursore IperSpettrale della Missione Applicativa* (PRISMA) is a hyperspectral platform launched by the Italian Space Agency (ASI, from Italian) in 2019. The mission combines two medium-resolution VNIR-SWIR hyperspectral instruments (30 meters) and a high-resolution panchromatic camera (5 meters). The platform operates in a Sun-

Synchronous low Earth orbit (615 km), with a nadir revisit period of 29 days that is enhanced by a roll system that enables re-looks for the same target every seven days. PRISMA can image longitudes from -180º to +180º and latitudes between -70º and +70º in the equinox (higher latitudes can be imaged depending on the Sun's Zenith angle). The swath width is 30 km, and the standard scene size has 30 x 30 kilometers, but it can be extended to a maximum of 1800 x 30 km. The VNIR spectrometer has 66 bands covering from 400 to 1010 nm (spectral sampling < 11 nm and bandwidth < 15 nm), whereas the SWIR instrument operates between 920-2500 nm resulting in 174 spectral bands (spectral sampling < 11 nm and bandwidth < 15). Table 3.2 presents information about the SNR of the instrument estimated after launch for different spectral ranges.

Table 3.2 - Signal to Noise Ratio between 400-2500 nm for the PRISMA sensor after launch.

| Spectral Range | SNR |
|---|---|
| 400-450 nm (VNIR) | 161 - 209 |
| 450-1000 nm (VNIR) | 200 - 450 |
| 1000-1300 nm (SWIR) | 300 - 800 |
| 1500-1750 nm (SWIR) | 200 – 400 |
| 1950-2500 nm (SWIR) | 100 - 200 |

Source: Cogliati et al. (2021).

PRISMA imaging must be required before the passage of the sensor on the official ASI's website (https://prisma.asi.it), where the user can inform the location, the period range, and the processing level. Then, the request is uploaded to the platform, and the acquisition is made inside the time window informed by the user if the weather (e.g., clouds) and geometrical requirements are achieved. In addition, there is an online platform to check the orbits of the sensor in a region of interest (http://prisma-prefeasibility.asi.it). After being acquired, the data can be downloaded in ASI's catalog in different processing levels: L1 (radiance at the top-of-atmosphere) and L2 (geolocated and geocoded

atmospherically corrected images). More information about the PRISMA mission and radiometric validation is described by Cogliati et al. (2021).

A PRISMA image was acquired on 23/11/2021, in match-up condition, during a cloudless day with a low turbidity atmosphere. The sensor's roll angle was -12.7 degrees from nadir. The scene has a size of 30 x 30 kilometers and covers the entire reservoir. The acquisition time was at 10:20h (GMT +3), creating a match-up window of ± 4 hours from the collected field data. The image has 234 continuous bands from 400 to 2500 nm. The VNIR cube has 63 spectral channels ranging from 400 to 970 nm. The data was downloaded from ASI's portal (https://prisma.asi.it) in two different processing levels: L1 (radiance at the top-of-atmosphere) and L2C (surface reflectance without geocoding).

### 3.3.2 Atmospheric correction

Three Atmospheric Correction (AC) processors were applied to the PRISMA image to evaluate the best surface reflectance product for mapping PC:

- ASI's Surface Reflectance Product (L2C): This product was acquired directly from the ASI's portal and contains reflectance values corrected for the atmospheric effects without geocoding. It was chosen over the final ASI's product (L2D) once the geolocation for the study area was displaced compared to a reference raster (Sentinel-2/MSI or Landsat-8/OLI images). This method uses hyperspectral bands for deriving atmospheric parameters, such as the water vapor and the Aerosol Optical Thickness (AOT). First, the water vapor is retrieved pixel-by-pixel using the water's absorption features at Near-Infrared bands (e.g., 940 nm). Then, a mean AOT is derived for the whole scene using the Dark Dense Vegetation method (GUARANI et al., 2018). Then, the radiative transfer model is inverted using MODTRAN, Look-Up Tables (LUT), and ancillary data.
- ACOLITE: This open-source, multi-sensor AC processor was developed for retrieving radiometric information from aquatic ecosystems (VANHELLEMONT; RUDDICK, 2018). For PRISMA data, ACOLITE takes as input both L1 and L2C products. L1 provides radiometric data, and the processor extracts the view geometry from L2C. The radiance measured by the sensor was converted to reflectance and then was corrected using the Dark Spectrum Fitting (VANHELLEMONT, 2019) with default options. The glint correction was

disabled to permit comparing with other AC processors. First, the AOT at 550 nm is derived from different spectral bands for different aerosols models (e.g., Continental and Maritime). Next, the band with the lowest AOT different from zero was selected for each aerosol model. Then, the best band and aerosol model are chosen by optimization. Finally, the other necessary parameters to perform the AC are retrieved from LUT, considering the calculated AOT at 550 nm for the selected band. ACOLITE processing is detailed by Vanhellemont (2019).

- 6SV: The Second Simulation of a Satellite Signal in the Solar Spectrum (6SV) has been successfully applied for different orbital sensors for inland waters applications (MARTINS et al., 2017; MACIEL et al., 2019; KRAVITZ et al., 2020). The 6SV radiative transfer model takes as input the top-of-atmosphere reflectance and some atmospheric and environmental parameters for simulating the bottom-of-atmosphere reflectance. The L1 PRISMA image was converted from radiance to reflectance and used as the radiometric input. The geometries of illumination and viewing were acquired from the PRISMA image metadata. The aerosol profile was set as 'Continental', and the AOT at 550 nm was extracted from the daily MODIS product MOD08. MOD08 also provided the water column information, and the ozone was measured by the Ozone Monitoring Instrument aboard the Aura satellite. Billings' elevation was provided by SRTM (30m). Given the input parameters, Surface Reflectance was calculated as described by Paulino et al. (2022).

After atmospherically corrected, all images were georeferenced. Then, the spectral values from the pixels coincident with the field match-up stations were extracted and compared with *in-situ* radiometric samples simulated for PRISMA bands. The spectral similarity will also be assessed by calculating the difference angles from the *in-situ* dataset and the AC products. Finally, the best AC was chosen by evaluating the accuracy metrics described in Section 3.5.

### 3.3.3 Synthetic multispectral data generation

Comparing estimative of water quality products from different orbital sensors is challenging due to the influence of other variables in the final output (e.g., differences in atmospheric conditions, geometries of illumination and acquisition, and the water body

trophic state). As PRISMA data have high-resolution spectral information, it allows for simulating images from other multispectral sensors and comparing the applicability of PC models in orbital data with the same boundary conditions (BECK et al., 2017). Therefore, Landsat-8/OLI and Worldview-3 synthetic images were generated using PRISMA hyperspectral image to identify the most suitable sensor for PC monitoring at BIL. Those platforms were chosen based on the potential for monitoring PC in urban reservoirs. First, the best atmospherically corrected product provided fifty-four PRISMA spectral channels, from 419 to 898 nm. Then, a linear interpolation was applied to achieve a resolution of 1 nm for each pixel. Finally, the interpolated data were resampled with SRF to target multispectral sensor bands. Information about the simulated bands for Landsat-8/OLI and Worldview-3 are described in Table 3.2. The synthetic data was assessed for the match-up points through comparisons of the field radiometric data resampled for the target sensor and the generated image.

Table 3.3 - Multispectral synthetic data description.

| Multispectral Sensor | Central Wavelength (nm) | Spectral Range (nm) | Spatial Resolution (m) |
|---|---|---|---|
| Landsat-8/OLI | 482 | 436-528 | 30 |
| Landsat-8/OLI | 561 | 512-610 | 30 |
| Landsat-8/OLI | 613 | 590-635 | 30 |
| Landsat-8/OLI | 655 | 625-691 | 30 |
| Landsat-8/OLI | 865 | 829-900 | 30 |
| WorldView-3 | 480 | 450-510 | 30 |
| WorldView-3 | 545 | 510-580 | 30 |
| WorldView-3 | 605 | 585-625 | 30 |
| WorldView-3 | 725 | 705-745 | 30 |
| WorldView-3 | 832 | 770-895 | 30 |

Source: Author's elaboration.

## 3.4 Machine learning for C-PC modeling

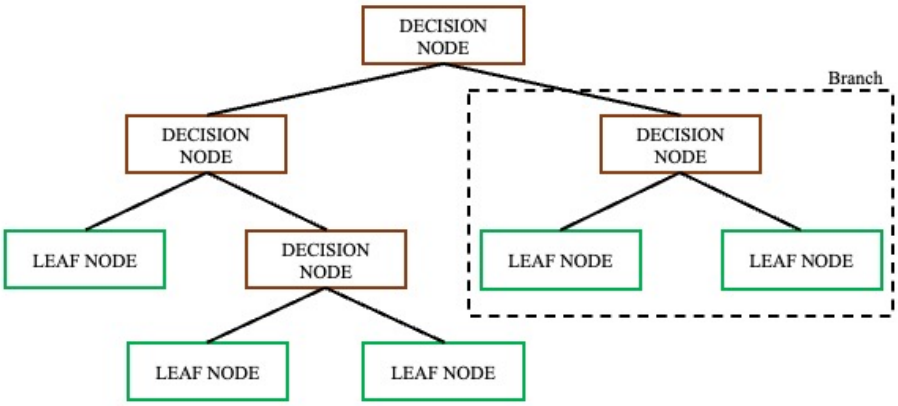
This study selected three Machine Learning Algorithms (MLA) for modeling PC: Random Forest, Extreme Gradient Boost, and Support Vector Machine. Firstly, the algorithms were trained and validated using *in-situ* data simulated for the different sensors. Secondly, the previously trained MLA were applied to the hyperspectral and synthetic multispectral images, and the PC estimates were validated with *in-situ* coincident measurements of the photosynthetic pigment. Finally, the PC results obtained

in this study were compared with the predictions made by the Mixture Density Network (MDN) for the Billings dataset using simulated and orbital data.

### 3.4.1 Models description

- Random Forest (RF) is a supervised MLA widely used for Remote Sensing applications (BELGIU; DRĂGU, 2016). It is an ensemble learning method that can be used to solve classification and regression problems. Ensemble algorithms are those which, instead of predicting based on only one classifier, generate several "weak" learners that are used together to produce a "strong" learner. First, RF creates different uncorrelated decision trees, each making an independent prediction of the target variable. Then, the response of all classifiers is used to define the final prediction. In the case of classification tasks, the predicted label is the one that receives the majority of votes from the learners. For regressions, the final output is the mean of all predictions. Each tree is constructed from a subset of the training dataset randomly sorted sample by sample (allowing repositions). This method is called "bagging" or "bootstrapping" and aim to produce uncorrelated trees to enhance the prediction. Each tree is composed of leaf nodes and decision nodes. Leaf nodes are the last component of each branch and represent a class label (e.g., "contaminated", or "not-contaminated"). The decision nodes are responsible for splitting the data until the final prediction on a terminal leaf node. They are generated based on the input features, and the splitting rule is defined based on minimalizing a cost function (e.g., what are the thresholds for a specific input feature that produces the minimum possible value for the cost function Mean Absolute Error?). The diagram in Figure 3.4 illustrates the structures of a decision tree. Overall, RF has a powerful generalization capacity and can provide good estimates even with small training datasets.

Figure 3.4 - Structures of a decision tree.

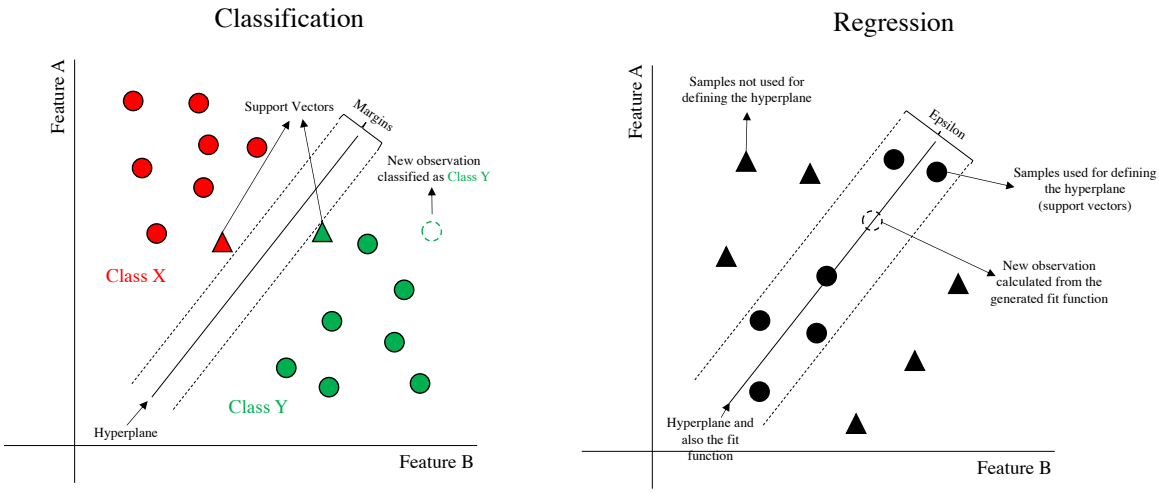



Source: Author's elaboration.

- The Extreme Gradient Boost (XgBOOST) is another supervised ensemble algorithm based on decision trees. XgBOOST shares most of the characteristics described for RF; however, the strategy for generating each tree is the greatest difference between those classifiers. While RF applies bootstrapping, XgBOOST uses the boosting approach. Instead of creating uncorrelated trees from random subsamples (bootstrapping), XgBOOST chooses the samples based on the predictions made from the previous trees. At each iteration, the algorithm evaluates the residuals of the generated tree and increases the chance of the misclassified samples being included in the subset that will create the next tree. This way, the algorithm can iteratively learn from the previous trees' errors to improve the next's accuracy. XgBOOST has been used to derive water quality parameters from radiometric data (CAO et al., 2020) due to its high performance in regression problems.
- Support Vector Machine (SVM) is also a supervised MLA that can be used for classification and regression tasks. First, SVM creates an n-dimensional space depending on the number of input features. Each informed feature will represent an axis coordinate in this newly created space. For classification problems, the algorithm calculates a decision boundary (or hyperplane) that can split the training samples into unique classes/labels. Then, when a new observation is presented to the algorithm, it uses the boundaries to classify the latest data. The hyperplane is constructed based on a pre-defined kernel function (a mathematical function that

might be linear or non-linear) to maximize the margin (distance) between the support vectors, which are the closest points of each label/class to the hyperplane (Figure 3.5). In regressions tasks, the algorithm uses the informed kernel to create a hyperplane that includes the greatest number of sample points (or support vectors). The epsilon defines the distance from the hyperplane (tolerated error) that a sample might have to be included as a support vector. The mathematical fit is then used to predict new observations (Figure 4.5).

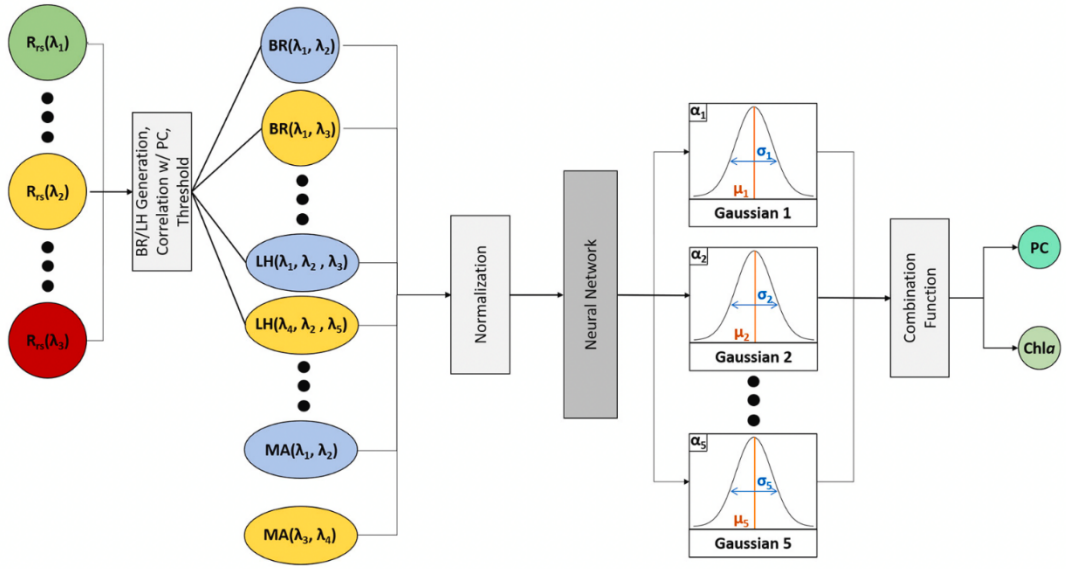
Figure 3.5 - Support vector machine for regression and classification tasks.




Source: Author's elaboration.

- The Mixture Density Network (MDN) combines a Deep Neural Network (DNN) and a mixture of contributions. The algorithm presupposes that any data distribution can be fitted using multiple Gaussians and the DNN is able to retrieve the adequate parameters for generating those curves. First, the user informs input layers that are provided to the DNN. Then, the DNN generates parameters (e.g., mean and standard deviation) to create *n* Gaussian curves to describe the input data. Mixing coefficients are also provided by the Neural Network and are used for combining the gaussian into a combinate function. Finally, this final function is used for predicting new observations. O'Shea et al. (2021) used multispectral algorithms, band ratios, and line heights as input layers and a combinate function using five Gaussians for calibrating their MDN PC algorithm (Figure 3.6).

Figure 3.6 - MDN calibration flowchart for estimating PC and Chl*a*.




Source: O'Shea et al. (2021).

### 3.4.2 Features generation

Each sample from the training dataset was composed of a $R_{rs}$ spectrum aligned with a PC measurement. PC concentrations from 0 to 0.1 µg/L were set to 0.1 µg/L, and then the dataset was log-transformed to reduce the asymmetry of the data distribution and prevent negative estimates. Although any of the applied MLA presupposes normal data distribution, predictions using high asymmetric datasets tend to be biased for some concentration range (QIANG; XINDONG, 2006). Then, the *in-situ* resampled spectral bands were selected for modeling PC. Wavelengths shorter than 500 nm are highly susceptible to atmospheric noise and are absent of PC features (O'SHEA et al., 2021), whereas wavelengths in the longer Near-Infrared and Mid-Infrared few contribute to OACs modeling due to the water's absorption (POPE; FRY, 1997). Therefore, forty bands from 500 to 877 nm were selected from the PRISMA dataset, as Landsat-8/OLI and Worldview-3 had all the bands beyond 500 nm. The individual spectral bands were not used as input for the MLA to avoid creating models sensitive to atmospheric-derived uncertainties. Instead, Normalized Indexes (NI), Line Heights (LH), and PC algorithms were used.

$$NI\ (\lambda_1, \lambda_2) = \frac{R_{rs}(\lambda_1) - R_{rs}(\lambda_2)}{R_{rs}(\lambda_1) + R_{rs}(\lambda_2)} \tag{3.5}$$

$$LH\ (\lambda_{-0}, \lambda_0, \lambda_{+0}) = R_{rs}(\lambda_0) - \left( R_{rs}(\lambda_{+0}) - \left( (R_{rs}(\lambda_{-0}) - R_{rs}(\lambda_{+0})) * \frac{\lambda_{+0} - \lambda_0}{\lambda_{+0} - \lambda_{-0}} \right) \right) \quad (3.6)$$

All possible combinations from NI were calculated for the evaluated sensors considering the selected bands for modeling PC (Equation 3.5). Since the same spectral information is contained in $NI\ (\lambda_1, \lambda_2)$ and $NI\ (\lambda_2, \lambda_1)$, only the NI where $\lambda_1 > \lambda_2$ were selected as input features. The LH is calculated based on three different wavelengths (or bands) in sequential ascending order. The first ($\lambda_{-0}$) and the last ($\lambda_{+0}$) are used to create a baseline for extracting valuable information from the central wavelength ($\lambda_0$) (QI et al., 2014). Different intervals from $\lambda_0$ might be used to calculate the LH, which might increase or decrease the relationship with some target variable. Therefore, LH was calculated using different intervals from the center wavelength to identify the best arrangements for predicting PC. For PRISMA, LH with 1 band, 3 bands, and 5 bands interval from $\lambda_0$ were derived. Worldview-3 had 1 band and 2 bands LH, while Landsat-8/OLI only had one-interval band LH. The PC algorithms were calculated considering the available bands for each sensor. The input features are described in Table 3.3. After generated, the features and PC values were normalized by subtracting the subset mean and dividing by the subset standard deviation.

Table 3.4 - Input features description.

| Orbital Sensor | *n* bands | Normalized Indexes | Line Heights | PC algorithms |
|---|---|---|---|---|
| PRISMA | 40 | 780 | 64 | MM14[1], LIU17[2], OGA19[3], HUN10[4] |
| Landsat-8/OLI | 4 | 6 | 2 | - |
| WordView-3 | 5 | 10 | 4 | OGA19[3] |

The formula for each PC algorithm is described in Table 2.1. References for the C-Phycocyanin algorithms: [1](MISHRA; MISHRA, 2014); [2](LIU et al., 2017); [3](OGASHAWARA; LI, 2019); [4](HUNTER et al., 2010).

Source: Author's elaboration.

### 3.4.3 Features selection

Features generated from hyperspectral data are prone to multicollinearity (KUMAR; GHOSH; CRAWFORD, 2001). MLA are sensitive to correlated features, with a sharp performance decrease observed in this situation (CHAN et al., 2022). Therefore, a feature

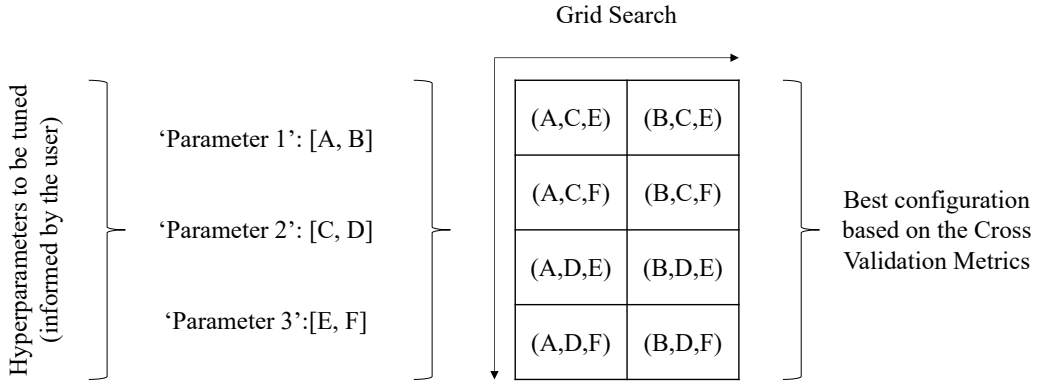
selection was made to remove the most correlated features. First, the calculated spectral layers (LH and NI) were ordered by their importance in predicting PC through the Predictive Power Score (PPS). The PPS is a univariate metric which scores each feature based on their relevance to predict a label (KIM et al., 2022). The PPS is calculated based on the loss function of a single decision tree, which was created from the evaluated feature and the target label. This metric can translate linear and non-linear relationships, overperforming traditional scores. PPS varies from 0 to 1, where features with no predictive power are penalized with 0, and a perfectly predictive feature achieves a unit score. Features that had a zero score were excluded, and the remain ones were organized in descending order relative to their PPS value. Then, a correlation matrix was calculated for every pair of input features. A Pearson Coefficient threshold was used to exclude the most correlated features, starting from the most important feature to predict PC in direction to the least important. To find the best Pearson value, a sensibility test was performed, varying the coefficient from 0.5 to 0.95 with an increment of 0.05 at each round.

### 3.4.4 Model training

The models were trained in a loop, varying the Pearson Coefficient value for selecting the input features. First, Grid Search was used to find the best hyperparameters for RF, XgBOOST, and SVM. Hyperparameters are used to control the learning rates of a machine learning algorithm (JIN, 2022). Each MLA is programmed with a default configuration, but fine adjustments of hyperparameters can increase the accuracy of the final predictions (LIN et al., 2021). Grid Search is one of the existent algorithms used to potentialize the search for the best combination (BELETE; HUCHAIAH, 2021). This methodology generates all configurations of a set of hyperparameters informed by the user and assesses each one at the target MLA (Figure 3.4). To find the best configuration, the dataset was divided into 5 equal parts (folds), and at each iteration, 4 were used for training, and 1 was used for testing (cross-validation). The best configuration is achieved with lowest mean accuracy metric (e.g., Mean Absolute Error) for all rounds. RF was tuned for the number of trees which informs the number of decision trees generated and maximum depth, which limits the number of splits allowed for each decision tree. For XgBOOST, different values for maximum depth, learning rate, and the number of trees were tested. The learning rate controls how much a generated decision tree will learn from

the previous misclassification. While for SVM, the kernel function, the epsilon, and the C values were optimized. The kernel is the mathematic function used to create the hyperplane (e.g., linear, radial basis function, polynomial) for splitting the samples. The C value is a parameter used to control the misclassifications from the samples used for training (might result in overfitting when C is too high, or underfitting for low C values). Epsilon control the margin of errors that are tolerated in the model.

Figure 3.7 - Grid search flowchart.




Source: Author's elaboration.

After tunning, a Monte Carlo Simulation with 1000 rounds was performed. For each iteration, the dataset was split into 80% for training and 20% for testing. The accuracy metrics were calculated for each round, and the median value for all rounds was used to assess the results. The best MLA for each orbital sensor was the one that resulted in the lowest error metrics aligned with the least number of selected features for a specific Pearson Coefficient value. Finally, all samples simulated to PRISMA were used to predict PC using the Mixture Density Network (MDN) model developed by O'Shea et al. (2021). This algorithm does not allow any extra training or parameters setting. Therefore, it was applied in default mode as provided by the authors (https://github.com/STREAM-RS/MDN-Phycocyanin). The model took 26 bands from 500 to 719 nm as input and calculated PC and Chl*a* concentrations. The MDN results were compared with the PC estimates from this study.

### 3.4.5 PC estimated from orbital data

After validated using *in-situ* radiometry, all MLA were applied to the available orbital data (Synthetic multispectral and hyperspectral images). The algorithms were re-trained using the field dataset, except the samples acquired in match-up condition with PRISMA. The same tunned hyperparameters from the previous modeling step were used to set the algorithms. Then, the selected features were generated from the spectral bands, and the models estimated PC concentrations. The PC maps were validated by comparing the match-up samples and the coincident pixel value. Then, the MDN was also used to estimate PC from the PRISMA image. The same validation strategy was used, and the outputs were compared with the modeled PC values from the MLA model.

## 3.5 Validation metrics

Two major validation metrics were used to assess the obtained results. The Mean Absolute Error (MAE) and Bias, both on a log scale, were calculated to compare atmospheric corrections, synthetic data generation, and PC modeling. MAE and Bias were calculated according to Equations 3.6 and 3.7, respectively.

$$MAE = 10 \wedge \left( \frac{\sum_{i=1}^{n} |log_{10}(M_i) - log_{10}(O_i)|}{n} \right) \quad (3.7)$$

$$Bias = 10 \wedge \left( \frac{\sum_{i=1}^{n} log_{10}(M_i) - log_{10}(O_i)}{n} \right) \quad (3.8)$$

Where $M_i$ is the modeled value, $O_i$ is the observed value, and $n$ is the sample size. MAE values will always be greater than 1.0, and any deviations are interpreted as the error magnitude (e.g., 1.4 means 40% of error). A Bias smaller than a unit indicates underestimation, while a greater one represents overestimation (e.g., a Bias of 0.7 means 30% of general underestimation, and 1.8 indicates an overestimation by 80%). Unit Bias is achieved when there is no trend of overestimation or underestimation. Those accuracy metrics were chosen due to the lack of data distribution assumptions and low sensibility to outliers (SEEGERS et al., 2018). Besides, they are adequate for comparing the results of data with the asymmetric distribution

The atmospherically corrected products had an additional metric for assessing the spectral similarity from *in situ* spectra and orbital data. This analysis does not account for the measurement intensity but rather the spectral shape similarity (KRUSE et al., 1993). The greater the calculated angle, the greater the difference between the measured and estimated $R_{rs}$.

$$d = \cos^{-1}\left(\frac{\sum_{i=1}^{nb} \hat{b}_i b_i}{\left(\sum_{i=1}^{nb} \hat{b}_i^2\right)^{1/2}\left(\sum_{i=1}^{nb} b_i^2\right)^{1/2}}\right) \quad (3.9)$$

Where $d$ is the degree (unit radians) of similarity from the measured and estimated spectrum, $\hat{b}_i$ is the i-th PRISMA's band from the evaluated AC product, $b_i$ is the simulated PRISMA's band from field data, and $nb$ is the number of bands. Finally, a summary of the number of samples used for validating and calibrating procedures is presented Table 3.4.

Table 3.5 - Summary of the number of samples used for validating and calibrating procedures.

| | N samples for Calibrating | N samples for Validating | Observations |
|---|---|---|---|
| Atmospheric Correction | - | 24 | Match-up samples with PRISMA passage (with and without PC measurements) |
| PC modeling (*in-situ* data) | 92 | 23 | 80% for calibrating and 20% for validating (For each round of the Monte Carlo Simulation) |
| PC modeling (orbital data) | 102 | 13 | Calibrated with all dataset but the match-up stations with PRISMA passage (with PC measurements) |
| MDN (*in-situ* data) | - | 115 | All in-situ PRISMA simulated dataset |
| MDN (orbital data) | - | 13 | Match-up samples with PRISMA passage (with PC measurements) |

Source: Author's elaboration.

# 4 RESULTS

## 4.1 Biological and radiometric *in-situ* data

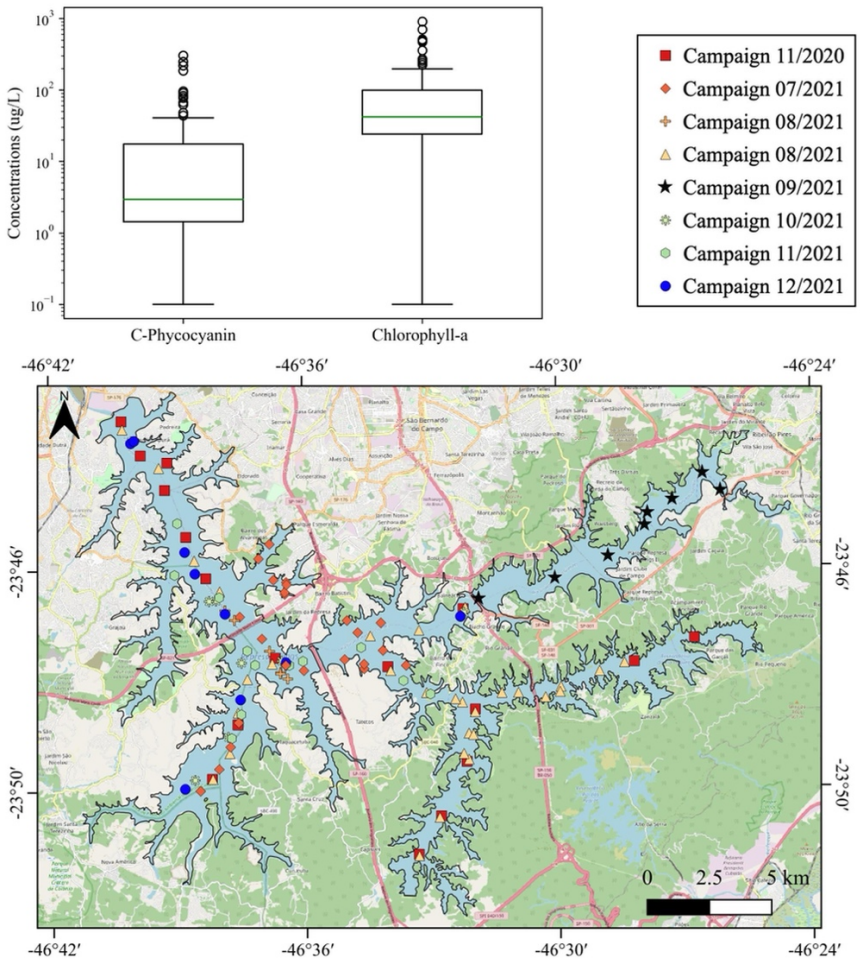
The biological analyses included photosynthetic pigments determination and Phytoplankton identification and quantification. Figure 4.1 shows the sampling stations' spatial distribution. A total of 115 samples of photosynthetic pigments concentrations were determined over eight different periods during 2020 and 2021. Table 4.1 provides descriptive statistics of PC and Chl*a* for each field campaign. The PC maximum was 301.81 μg/L, while the minimum was 0 μg/L. The mean PC value was 20.28 μg/L, the median was 2.9 μg/L, and the standard deviation was 46.81 μg/L. Regarding Chl*a*, the dataset ranged from 4.54 to 906.15 μg/L, with a mean value of 89.20 μg/L, a median value of 42.67 μg/L, and a standard deviation of 133.16 μg/L. The median PC:Chl*a* from this study was 0.09, and the mean was 0.14. Median Chl*a* concentrations were almost 14 times greater than median PC. The boxplot (Figure 4.1) shows that PC values above 40 μg/L are considered outliers, while for Chl*a,* the upper limit is 200 μg/L. Both pigment density distributions are positively skewed (Chl*a* = 3.80 and PC = 4.16). The pigment concentrations were compared with the estimated phytoplankton biovolume for some samples ($n$ = 26). Chl*a* had a strong linear positive relationship with the total phytoplankton biovolume, while the same was observed regarding PC and the Cyanophyta phylum (Figure 4.2).

Table 4.1 - Descriptive statistics from Phycocyanin (ug/L) and Chlorophyll-*a* (ug/L).

| Statistic | 11/2020 | 07/2021 | 08/2021 | 08/2021 | 09/2021 | 10/2021 | 11/2021 | 12/2021 |
|---|---|---|---|---|---|---|---|---|
| Chl*a* min | 29.34 | 16.04 | 16.04 | 40.10 | 4.54 | 7.68 | 15.37 | 40.43 |
| Chl*a* max | 486.16 | 114.94 | 149.69 | 225.87 | 256.61 | 103.80 | 196.91 | 906.15 |
| Chl*a* mean | 156.45 | 41.65 | 40.37 | 88.54 | 86.47 | 43.49 | 75.28 | 337.59 |
| Chl*a* median | 98.31 | 24.06 | 26.73 | 57.47 | 61.27 | 36.37 | 65.49 | 163.22 |
| Chl*a* STD | 153.41 | 8.83 | 31.77 | 64.96 | 90.70 | 30.33 | 51.56 | 325.72 |
| PC min | 1.92 | 0.05 | 0.00 | 2.86 | 0.00 | 0.25 | 1.68 | 3.26 |
| PC max | 186.49 | 33.16 | 47.76 | 95.49 | 2.59 | 17.51 | 77.68 | 301.81 |
| PC mean | 30.94 | 5.65 | 7.46 | 24.00 | 1.26 | 7.50 | 22.23 | 119.24 |
| PC median | 7.14 | 1.86 | 1.43 | 12.76 | 1.30 | 5.2 | 18.18 | 53.77 |
| PC STD | 49.50 | 30.84 | 14.06 | 32.59 | 0.83 | 6.31 | 20.72 | 117.50 |
| PC:Chl*a* median | 0.087 | 0.080 | 0.062 | 0.165 | 0.02 | 0.164 | 0.26 | 0.33 |

Source: Author's elaboration.

Figure 4.1- Photosynthetic pigments boxplot and sampling stations' spatial locations.

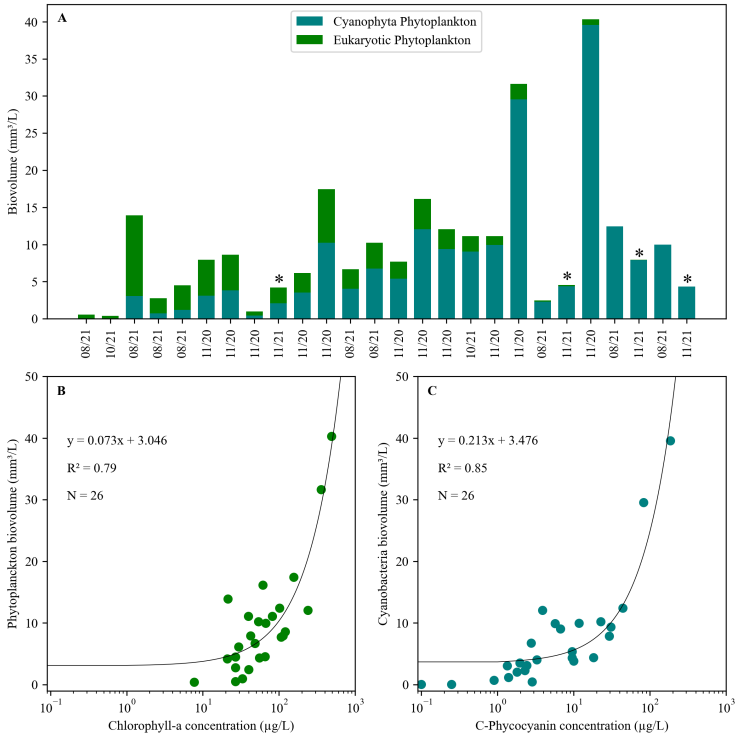



Source: Author's elaboration.

During the studied period, the phytoplankton community from Billings reservoir was marked by the presence of Cyanobacteria. Cyanophyta phylum represented more than half of the total biovolume in 65% of the samples ($n$ = 26). In 6 of them, Cyanobacteria represented the totality of phytoplankton in the water. The taxonomic analysis revealed that a single taxon was dominant on 14 samples, where 12 were Cyanobacteria species. Furthermore, in ten of these twelve, *Microcystis aeruginosa* dominated, while *Raphidiopsis raciborskii* and *Planktothrix isothrix* were the majority in each of the remaining two. The other two samples had the Dinoflagellate *Ceratium furcoides* as the dominating taxon. Furthermore, at least one Cyanobacteria species was abundant or dominant in 25 of the 26 analyzed samples. Likewise, Eukaryote species were observed 17 times. *Microcystis aeruginosa* was abundant in at least 80% of the samples (20), while *Raphidiopsis raciborskii* achieved the same condition in ten samples. Excluding

Cyanobacteria, the most represented phytoplanktonic classes were Zygnematophyceae, Coscinodiscophyceae, and Dinophyceae.

Figure 4.2 - Phytoplankton biovolume and comparison with photosynthetic pigments.



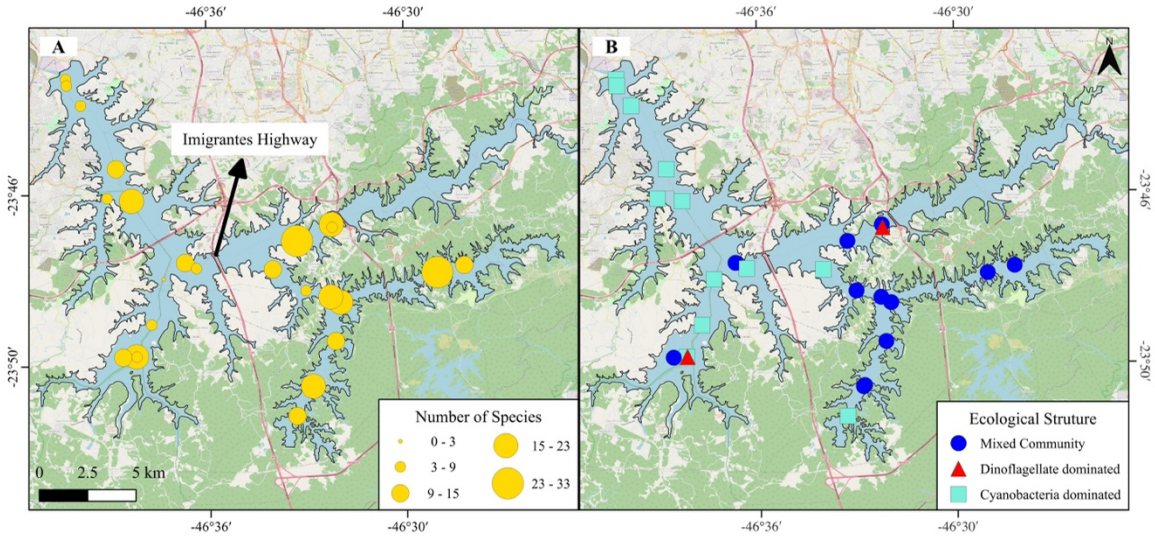

On the top graphic, PRISMA match-up stations have an asterisk (*). A) Total biovolume of each sample station split into Cyanophyta and Eukaryotic Phytoplankton. B) Measured Chl*a* concentration compared with measured total biovolume. C) Measured PC concentration compared with measured Cyanobacteria biovolume.

Source: Author's elaboration.

The species richness varied from 3 to 33, with a mean of 13 taxa per sample. This value decreased by 23% (10 species) when considering just the Cyanobacteria-dominated sample points. When analyzing just samples where no domination was identified (mixed community), there was an increase of 31% (17 species). A spatial trend was observed regarding the phytoplankton community structure (Figure 4.3). Taking the Imigrantes Highway as a reference, the western side of the reservoir (Pedreira Arm, Taquacetuba Arm and the western part of the Central Body) had 10 of the 13 samples dominated by Cyanobacteria with a mean species of 8. All sample stations in Pedreira Arm were dominated by *Microcystis aeruginosa*. Meanwhile, the eastern side of Imigrantes Highway (Rio Grande arm, Rio Pequeno Arm, Capivari Arm, and the eastern part of the

Central Body) had only 3 of its 13 sample stations presenting species domination and a mean richness of 16 taxa.

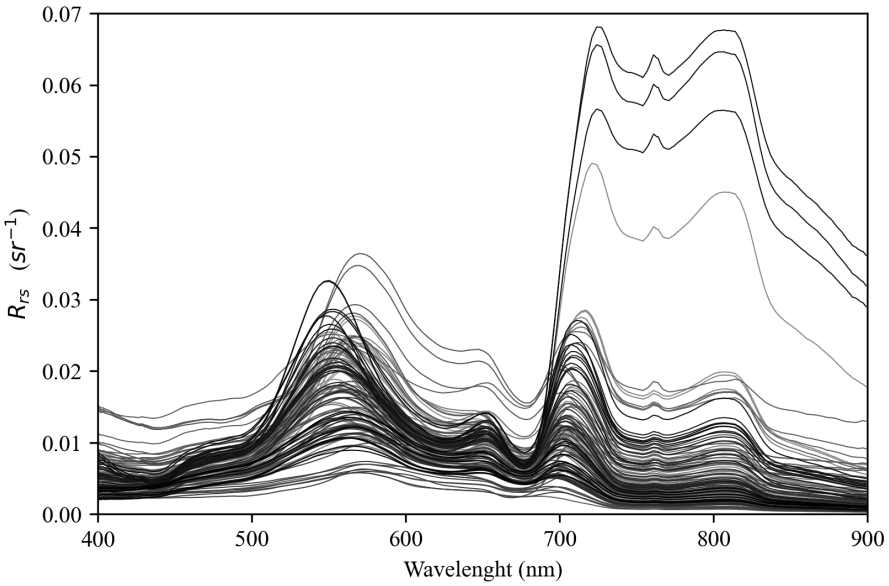
Figure 4.3 - Phytoplankton community structure and species richness.




Source: Author's elaboration.

*In-situ* radiometric samples were used to calibrate and validate the PC bio-optical algorithms. The radiometric measurements were carried out in a total of 126 sample stations, during the eight field campaigns. Figure 4.4 shows the estimated $R_{rs}$ for those sites. Of the total dataset, 115 stations had photosynthetic pigments concentrations concurrent with $R_{rs}$, and 11 had just radiometric *in-situ* data. The $R_{rs}$ without pigments reference were used for increasing the match-up rate with PRISMA and increase the atmospheric correction validation dataset.

Figure 4.4 - Remote sensing reflectance spectra.

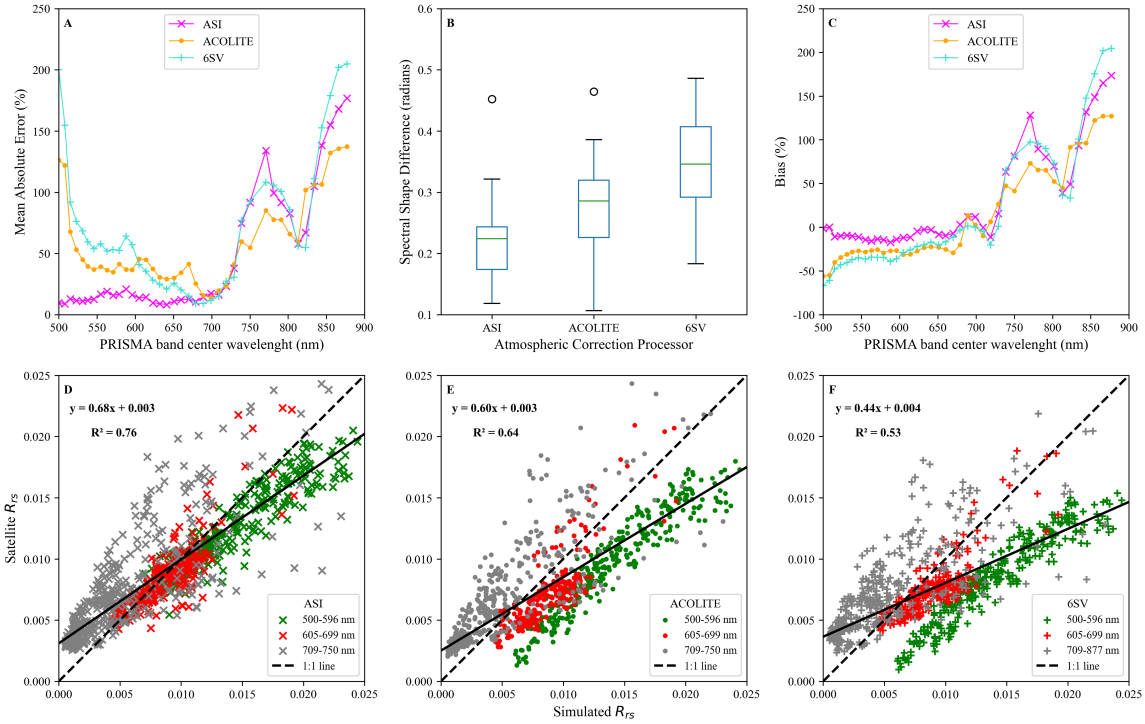



Source: Author's elaboration.

## 4.2 Atmospheric correction and synthetic data generation

Figure 4.5 shows the results for the validation of PRISMA's atmospheric corrected products. Only the bands used for modeling PC were assessed (500 to 877 nm). Twenty-four $R_{rs}$ samples acquired in match-up condition with PRISMA (± 4 hours) were used as a reference for evaluating the retrieved surface reflectance. ASI's product had the most consistent estimates for all the calculated accuracy metrics. ACOLITE followed PRISMA's official product, and 6SV had the worst performance from the tested methods. When considering all bands, ASI had a spectral MAE of 1.39, while ACOLITE and 6SV had 1.55 and 1.65, respectively. In addition, ASI had a 19% overestimation tendency, while the other AC processors had no general trend (Bias ~1).

Figure 4.5 - Atmospheric correction accuracy metrics.

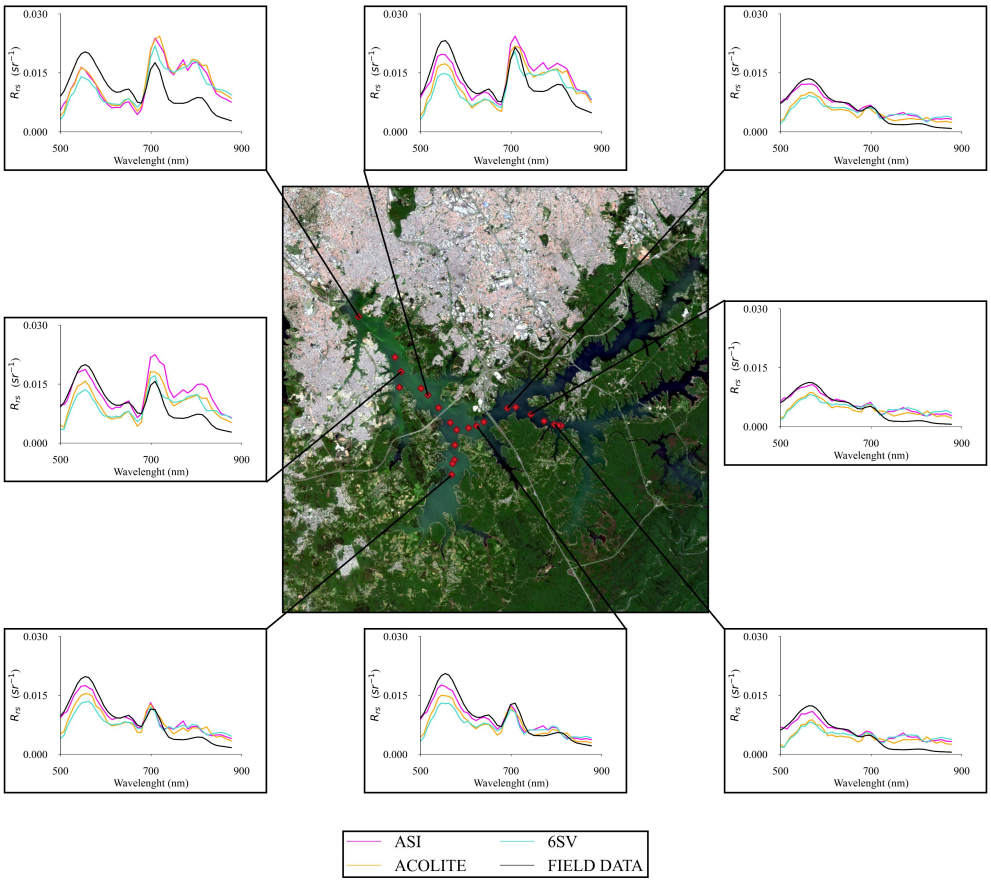



Metrics for the atmospheric Correction Processors (ASI, ACOLITE, and 6SV). A) Mean Absolute Error (MAE). B) Spectral difference between satellite and *in-situ* data. C) Bias. D) Scatter plot comparing ASI estimated $R_{rs}$ versus field data. E) Scatter plot comparing ACOLITE estimated $R_{rs}$ versus field data. F) Scatter plot comparing 6SV estimated $R_{rs}$ versus field data.

Source: Author's elaboration.

ASI had MAE lower than 20% for all bands from the visible region until the red edge, slightly underestimating the $R_{rs}$. ACOLITE and 6SV had MAE from two to more than ten times greater than ASI processor for the same spectral region, with a pronounced underestimation tendency. The accuracy metrics beyond 700 nm are similar for all AC processors, with a small advantage of ACOLITE. From 700 to approximately 780 nm, MAE increases, and Bias shows an overestimation trend in the Near-Infrared. Then, from 780 to 820 nm MAE sharply decreased, followed by another increase from 820 to 877 nm. ASI also had the most similar spectra compared with reference $R_{rs}$, with a median difference of 0.22 radians. ACOLITE achieved 0.29 radians, and 6SV exhibited the greatest difference with a 0.35 median value. ASI and ACOLITE had a common outlier sample, evidenced by the spectral difference boxplot. Considering the presented results, ASI was selected to proceed with PC modeling. A comparison from all the AC processors and reference *in-situ* data is provided for some sample stations (Figure 4.6).

Figure 4.6 - Comparison between orbital and *in-situ* $R_{rs}$.

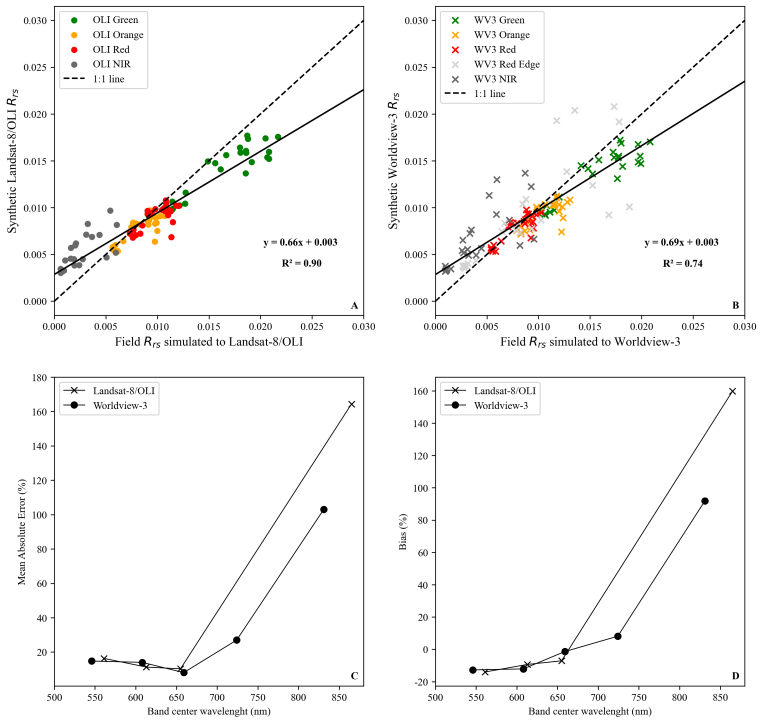



Source: Author's elaboration.

To understand how the radiometric noise has propagated to the Worldview-3 and Landsat-8 synthetic images, their data were compared with the *in-situ* $R_{rs}$ simulated to the respective sensor (Figure 4.7). The same 24 match-up $R_{rs}$ samples were used as reference data. The accuracy metrics revealed that the errors on both synthetic images were directly proportional to the errors from the integrated spectral region of the original PRISMA atmospheric corrected image. Therefore, there is no evidence that resampling the hyperspectral image has reduced the data's radiometric quality. Overall, MAE was lower than 16% in the visible spectrum for all the synthetic bands. As for Bias in the same spectral region, all channels were underestimated concerning the *in-situ* data with a maximum Bias value of -15%. The NIR bands showed more expressive errors than the visible range. The Landsat-8/OLI NIR band had a MAE of 2.64 and a Bias of 2.6, while Worldview bands had a MAE of 1.7 and 2.0, with a Bias of 1.08 and 1.92 for the Red Edge and NIR channels, respectively. The MAE increase and the overestimation tendency

in the NIR region followed the same pattern as observed in Figure 4.5 for the PRISMA image.

Figure 4.7 - Synthetic images accuracy metrics.




Metrics for the synthetic images generated from PRISMA orbital data. A) Scatter plot comparing Landsat-8/OLI synthetic $R_{rs}$ and *in-situ* data. B) Scatter plot comparing Worldview-3 synthetic $R_{rs}$ and *in-situ* data data. C) Mean Absolute Error (MAE). D) Bias.

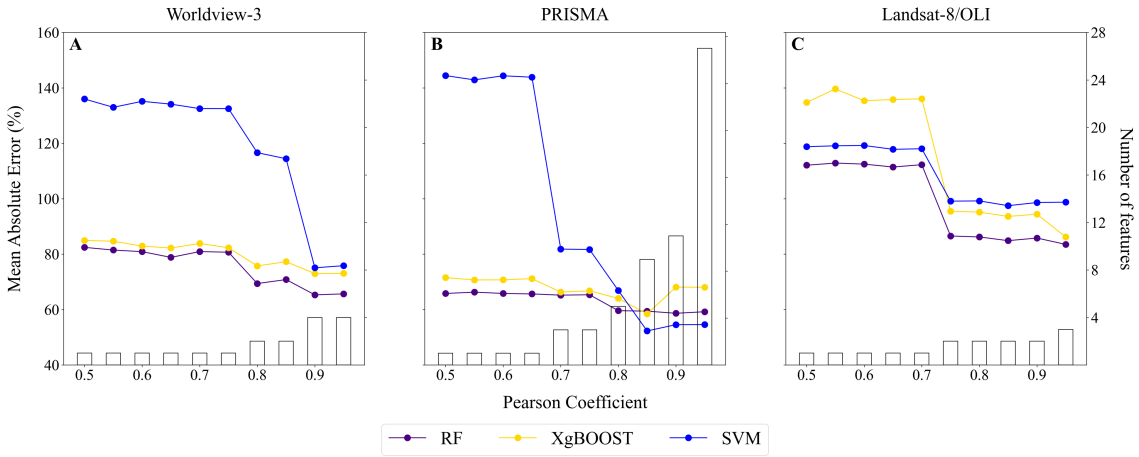
Source: Author's elaboration.

## 4.3 C-Phycocyanin estimated from radiometric data

The Machine Learning Algorithms for predicting C-Phycocyanin were first calibrated and validated using PRISMA, Worldview-3, and Landsat-8/OLI bands simulated from in-situ radiometric data. This approach was used to reduce the uncertainties derived from modeling using orbital data, especially due to the atmospheric noise. Figure 4.8 presents the results of the Pearson Coefficient sensibility test. This approach had the objective of defining the threshold for selecting the best input bands to estimate PC for each assessed sensor. For Worldview-3, the best combination was achieved with a Pearson of 0.9 and four input layers. Landsat-8/OLI had the best MAE for a threshold of 0.95, which resulted in three layers. Although PRISMA had the most accurate predictions for SVM with Pearson of 0.85 (9 input layers) (Figure 4.8-B), the difference between the error metrics

achieved in RF and XgBOOST on 0.8 (5 input layers) wasn't enough to justify adding almost double input features. Therefore, the five layers selected from the 0.8 coefficient were used for further analysis.

Figure 4.8 - Results of pearson sensibility test.




Each plot represents the results for PC modeling using a Monte Carlo simulation varying the Pearson Coefficient value used for selecting the input features. This assessment was made using *in-situ* hyperspectral data resampled for different orbital sensors. The left scale refers to MAE magnitude (line plot) while the right indicates the number of features in the bar plots; results for: A) Worldview-3. B) PRISMA. C) Landsat-8/OLI.

Source: Author's elaboration.

Table 4.2 presents the selected input features considering the different orbital sensors. ordered by their importance for predicting PC, considering the calculated Predictive Power Score. To be selected, the input feature had to offer some predictive power for estimating PC (PPS > 0) and present a degree of collinearity with the other features smaller than the Pearson threshold. This strategy maximized the information variance used for estimating the target label. Therefore, elements that did not contribute to PC modeling or contained the same spectral information of a previously selected feature were excluded. Figure 4.9 shows the importance of the features calculated by the Machine Learning Algorithms. The elements that most contributed for retrieving PC coincided with its relative PPS score for most MLA. As expected, features using bands centered around the PC absorption feature (620 nm) were select for predicting the target variable (e.g., LH (546,608,659), NI (608,546), MM14, LH (561,613,865)) However, the data-driven feature selection revealed that features based on the Green and Near-infrared regions were also sensitive to variations in the photosynthetic pigment concentrations.

PRISMA's most important feature for all MLA was the $NI(563,555)$. The LH considering three Near-Infrared wavelengths, also had foremost importance for the hyperspectral sensor ($LH(739,802,855)$). Predictions from Worldview-3 also used Near-Infrared information for retrieving PC ($LH(659,724,831)$). Due to its limited spectral resolution, Lansat-8 estimates were based only on the PC absorption band. After the input layers were selected, the hyperparameters were tuned for each MLA algorithm. The chosen values from Grid Search are in Table 4.3.

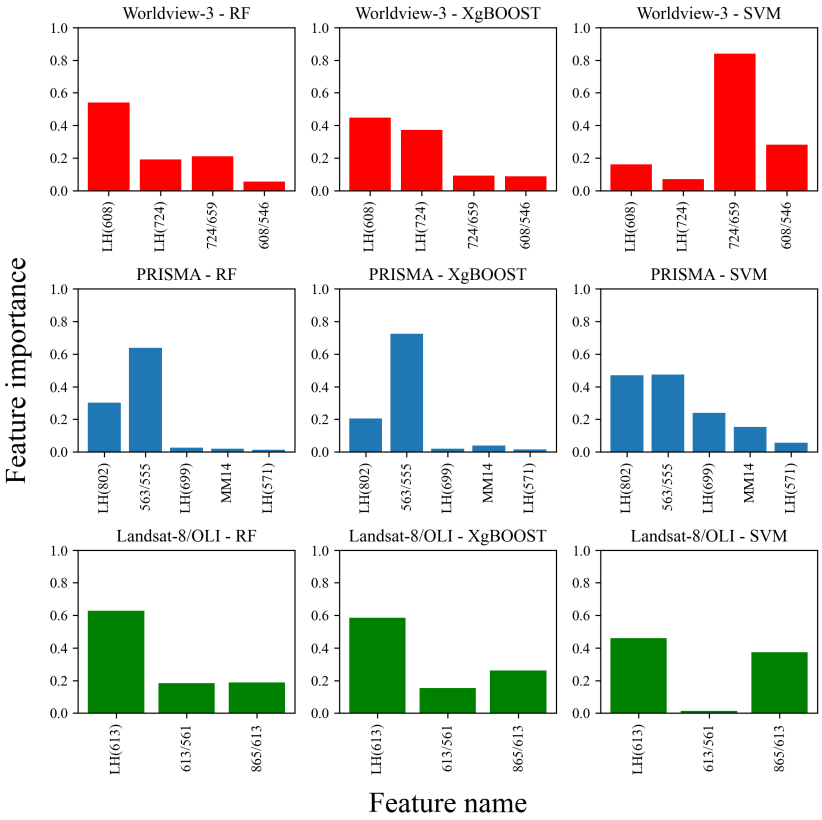
Table 4.2 - Selected input features for each orbital sensor.

| Sensor | Selected Features |
|---|---|
| Worldview-3 | $LH(546,608,659), LH(659,724,831), NI(724,659), NI(608,546)$ |
| PRISMA | $LH(739,802,855), NI(563,555), LH(651,699,750), MM14, LH(531,571,614)$ |
| Landsat-8/OLI | $LH(561,613,865), NI(613,561), NI(865,613)$ |

Features are ordered considering its Predictive Power Score for estimating PC. Features names are abbreviated considering their input spectral bands. $LH\ (\lambda_{-0}, \lambda_0, \lambda_{+0})$ refers to line heights. $NI\ (\lambda_1, \lambda_2)$ is an abbreviation for the Normalized index.

Source: Author's elaboration.

Figure 4.9 - Features importance calculated by the assessed machine learning algorithms considering the different orbital sensors.




Features names are abbreviated considering their input spectral bands. $LH(\lambda)$ refers to line heights centered at wavelength λ. $\lambda_1/\lambda_2$ is an abbreviation for the Normalized index $NI\ (\lambda_1, \lambda_2)$.

Source: Author's elaboration.

Table 4.3 - Tunned hyperparameters.

| Sensor / MLA | Hyperparameter tunned | | | | | |
|---|---|---|---|---|---|---|
| | *n* trees | Max depth | Learning rate | Kernel | C | Epsilon |
| Worldview-3 / RF | 120 | 4 | - | - | - | - |
| Worldview-3 / XgBOOST | 80 | 2 | 0.05 | - | - | - |
| Worldview-3 / SVM | - | - | - | Linear | 0.5 | 0.2 |
| | | | | | | |
| PRISMA / RF | 120 | 4 | - | - | - | - |
| PRISMA / XgBOOST | 120 | 2 | 0.15 | - | - | - |
| PRISMA / SVM | - | - | - | Linear | 0.5 | 0.2 |
| | | | | | | |
| OLI / RF | 80 | 4 | - | - | - | - |
| OLI / XgBOOST | 80 | 2 | 0.05 | - | - | - |
| OLI / SVM | - | - | - | Linear | 0.5 | 0.4 |

Source: Author's elaboration.

Considering the best Pearson Coefficient value for each sensor, Table 4.5 resumes the results for modeling PC using *in-situ* radiometric data. The best predictions were with PRISMA, followed by Worldview-3 and Landsat-8/OLI. Random Forest outperformed the other MLA for all the sensors. In contrast, SVM had the worst results. PRISMA's best performance was 25% more accurate than OLI's predictions with RF. Compared with the Worldview-3 best score, the hyperspectral sensor achieved accuracies 10% better than the ultra-resolution sensor. The results also revealed that the accuracy metrics were proportional to the number of input features available for predicting PC.

Table 4.4 - Results for PC modeling using *in-situ* simulated data for different sensors.

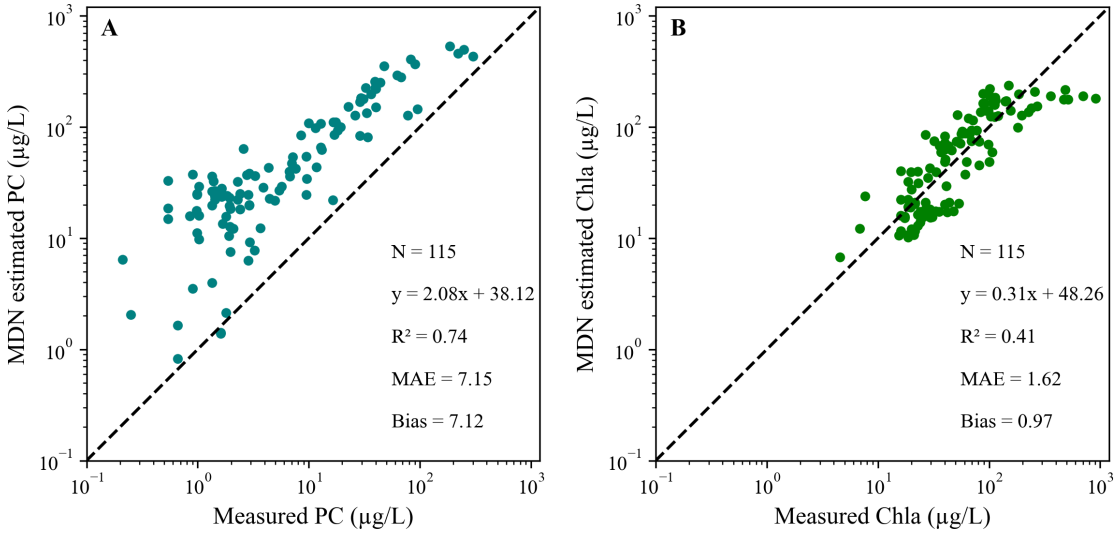
| Sensor / MLA | Pearson | Selected Bands | Median MAE | Median Bias |
|---|---|---|---|---|
| **Worldview-3 / RF** | **0.9** | **4** | **65%** | **0%** |
| Worldview-3 / XgBOOST | 0.9 | 4 | 72% | 0% |
| Worldview-3 / SVM | 0.9 | 4 | 75% | 9% |
| | | | | |
| **PRISMA / RF** | **0.8** | **5** | **59%** | **0%** |
| PRISMA / XgBOOST | 0.8 | 5 | 64% | 0% |
| PRISMA / SVM | 0.8 | 5 | 69% | 10% |
| | | | | |
| **OLI / RF** | **0.95** | **3** | **84%** | **-4%** |
| OLI / XgBOOST | 0.95 | 3 | 86% | 0% |
| OLI / SVM | 0.95 | 3 | 99% | 0% |

Best results for each sensor are in bold.

Source: Author's elaboration.

The Chl*a* and PC estimates from MDN are presented in Figure 4.10. Although PC estimate is linearly correlated to the measured values ($R^2 = 0.74$), the algorithm showed a strong overestimation trend (Bias = 612%). Considering the best achieved MAE for the calibrated MLA in this study, RF using PRISMA data as input features is 10.5 times more accurate than MDN. Regarding Chl*a* estimates, the model had more success in predicting this photosynthetic pigment (MAE = 64%). However, the scatter plot shows an accuracy decrease for high concentrations compared to low and medium values. Bias reveals that there is no clear trend of overestimation or underestimation.

Figure 4.10 - Chl*a* and PC estimated through the MDN algorithm.

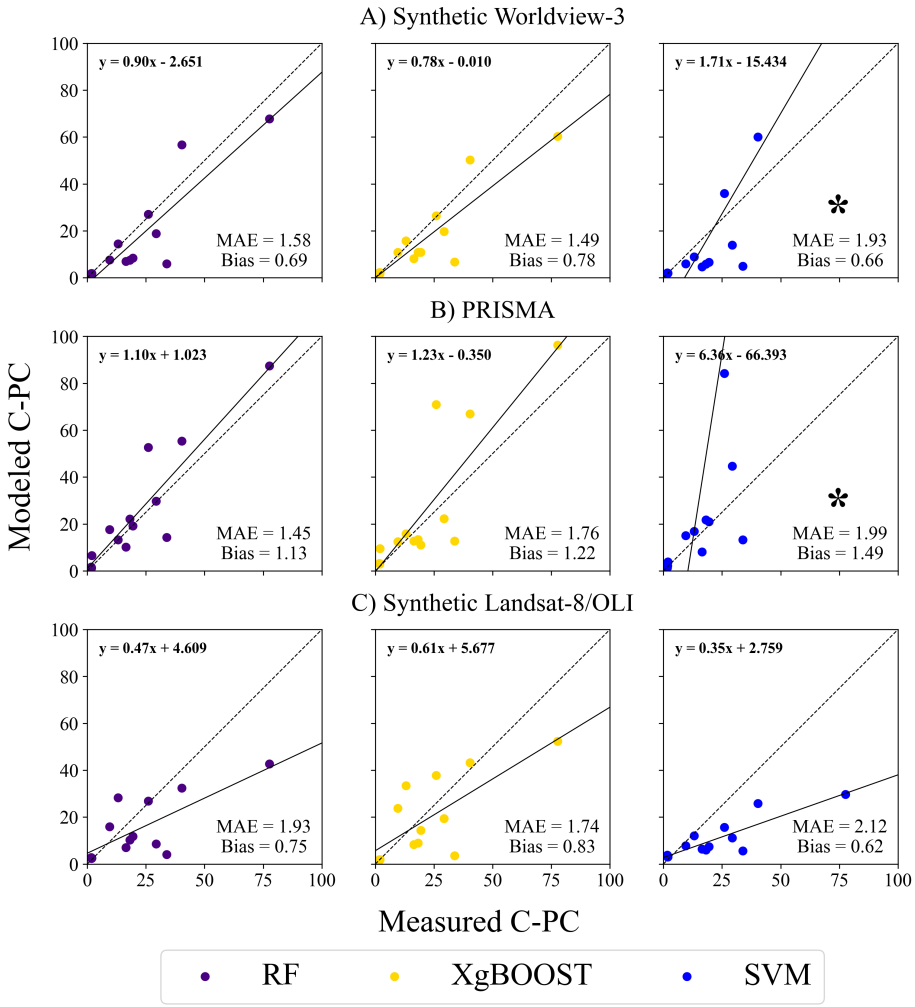



This model was implemented by O'Shea et al. (2021) and is the most global PC algorithm. This algorithm was used on default settings, as upload by the authors (https://github.com/STREAM-RS/MDN-Phycocyanin).

Source: Author's elaboration.

After being calibrated and validated using *in-situ* radiometric data, all evaluated Machine Learning Algorithms were used to predict PC from PRISMA, Synthetic Worldview-3, and Synthetic Landsat-8/OLI images. The generated PC maps were validated by comparing the measured and the estimated PC values for the match-up sampling stations. The validation metrics of the produced maps are presented in Figure 4.11. The combination of PRISMA data and Random Forest also had the most accurate results in image-based PC retrieval. This method has achieved a MAE of 45% and a Bias of 13%. MAE from orbital PC estimates was 15% lower than the obtained from *in-situ* radiometry for PRISMA/RF method. SVM had an expressive accuracy decrease compared with the first modeling step and was also the most inaccurate of the assessed MLA. For Worldview-3 and Landsat-8/OLI synthetic images, XgBOOST achieved the best results. When comparing with field modeling, there was an accuracy turn-over between RF and XgBOOST. Both sensors also had an MAE decrease when comparing the results obtained from simulated data and orbital data. The MAE difference between PRISMA and Worldview-3 predictions was only 4%. Landsat-8/OLI and Worldview-3 estimates were biased for underestimation.

Figure 4.11 - Results for PC retrieval using orbital data.




Results for PC modeling using orbital data and different Machine Learning Algorithms. Plots marked with an asterisk (*) have points out of the scale's range. A) Validation using the Synthetic Worldview-3 image. B) Validation using the PRISMA image. C) Validation using the Synthetic Landsat-8/OLI image.

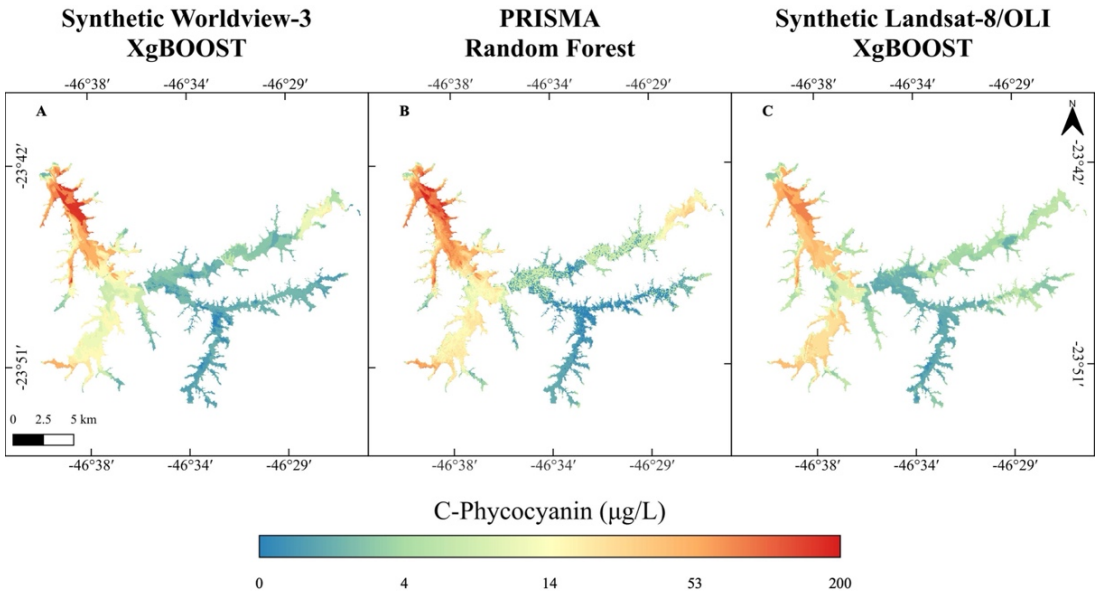
Source: Author's elaboration.

Figure 4.12 presents the PC maps with the best accuracy metrics for each assessed orbital sensor. A median filter with a 3x3 window was applied to smooth the results. Overall, the same spatial pattern is observed for the three methods throughout the Billings reservoir. Transects were made to illustrate differences in PC estimates (Figure 4.13). The same spatial pattern from phytoplankton community structure and species richness was observed in PC maps. From the Imigrantes Highway, the western side of the reservoir was marked by relatively high and medium concentrations, as on the eastern side it was observed a sharp concentration decrease and relatively low PC estimates. This trend is

observed when comparing Transects #1 and #2 with #3, #4, and #5. PRISMA/RF results exhibited high spatial frequency, even after smoothed by a median filter. The high variability on PC estimates in short distances are observed in all transects but are more expressive for low concentrations (Transects #3 and #5). Worldview-3 and Landsat-8/OLI results showed spatial frequency lower than that of PRISMA, and the changes in PC concentration were observed at greater distances.

Overall, the same PC concentration distribution pattern was observed for the three methods throughout the Billings reservoir. The Pedreira arm (Transept #1) had the greatest estimated PC concentrations, also observed with *in-situ* measured values. PRISMA/RF achieved values as high as 230 µg/L, while the maximum peak of Worldview-3/XgBOOST and Landsat-8/XgBOOST were 165 µg/L and 80 µg/L, respectively. Taquacetuba Arm and Central Body, represented by Transept #2, had the second-highest PC concentrations, even though considerably lower than the Pedreira arm. However, all methods have identified a high concentration spot near point D. Also, an increasing trend in cyanobacteria cells was observed on Transept #4, in the direction to point H. Transept #3 and #5 had smaller PC ranges and exhibited lower concentrations when compared with Transepts #1 and #2.

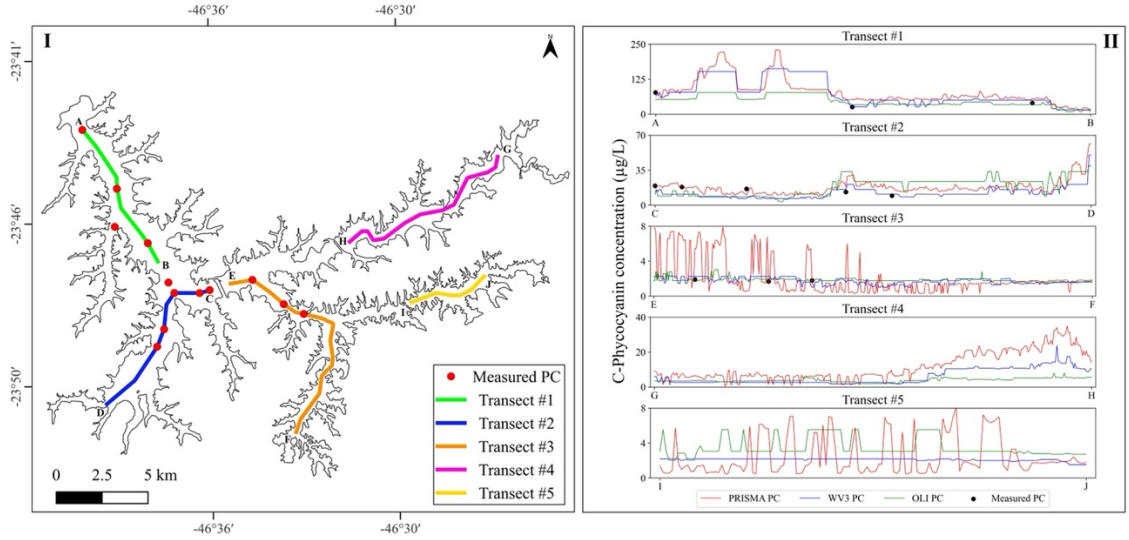
Figure 4.12 - PC mapping for different orbital sensors and MLA.




PC maps for best results for each of the assessed orbital sensors. A) Results for sensor Worldview-3 and the MLA XgBOOST. A) Results for sensor PRISMA and the MLA Random Forest. A) Results for sensor Landsat-8/OLI and the MLA XgBOOST.

Source: Author's elaboration.

Figure 4.13 - PC concentration transects.

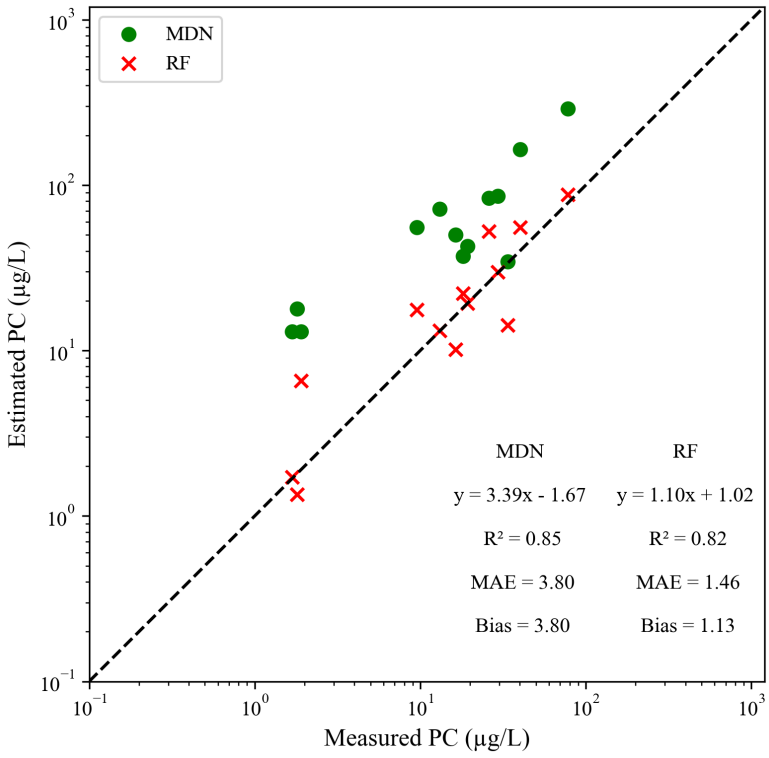



PC concentration transects comparing the different methodologies for the photosynthetic pigment retrieval (PRISMA/RF, Worldview-3/XgBOOST, and Landsat-8/XgBOOST).

Source: Author's elaboration.

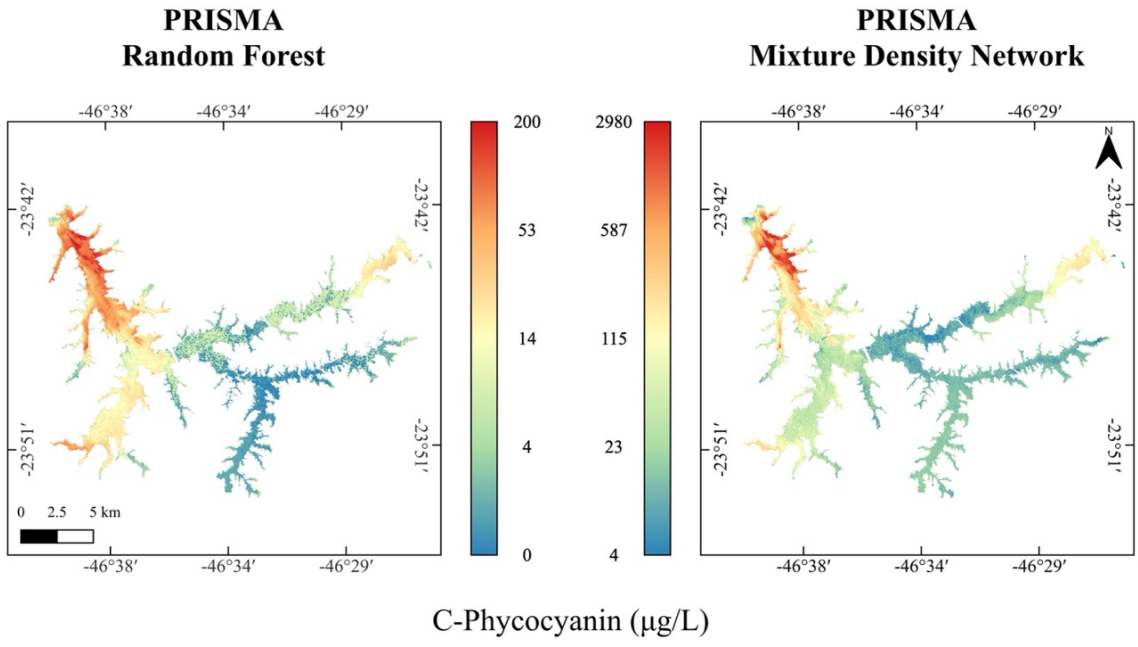
Finally, the results from PRISMA/RF were compared with MDN's prediction for the same orbital sensor. Figure 4.14 shows the same validation strategy applied before, where the match-up sampling stations were used to derive accuracy metrics from PC mapping. Regarding MDN's results, the overestimation trend observed in *in-situ* radiometry was repeated for the orbital sensed data (Bias = 280%). Even though MAE has fallen by half compared with MDN applied to *in-situ* data, it was still six times greater than RF prediction errors. However, both methods have shown a strong linear trend between the modeled and observed PC values ($R^2 > 0.80$). Billings reservoir PC maps derived from MDN and RF using PRISMA data are shown in Figure 4.15. The same median filter was applied in MDN's results to reduce high variations in PC estimates in short distances. Although the same spatial pattern is observed for both methodologies, the disparities in the concentration range are evidenced by the different scales used to represent PC estimates. However, MDN has achieved smoother results than Random Forest, reducing the spatial frequency in PC concentrations observed in the last method.

Figure 4.14 - Comparisons between C-Phycocyanin estimated obtained by MDN and RF using a PRISMA hyperspectral image.




Source: Author's elaboration.

Figure 4.15 - PC maps for PRISMA hyperspectral image using MDN and RF.




Source: Author's elaboration.

# 5 DISCUSSION

## 5.1 Billings history of anthropization

Billings (BIL) is the iconic example of an urban reservoir degraded by anthropogenic activities. It is part of the Serra Project that aimed to revert waters from the Upper Tietê Basin to São Paulo's coastal zone and produce hydroelectric energy taking advantage of the 720 meters height difference (SMA, 2010). BIL volume results from damming of ten rivers for storing and controlling the water flux to Cubatão City, and was finalized in the 1940's. To enable the increase of energy production, part of the Pinheiros River flux was pumped to BIL (CAPOBIANCO; WHATELY, 2002), which would also reduce São Paulo's floodplain (BRAGA, 2000). By that time, São Paulo had one of the greatest urban growth rates of the country, but the sanitation network did not follow the city expansion. Therefore, the water quality from most São Paulo's water bodies rapidly degraded (JACOBI, 1997), including BIL that was receiving water transposed from the Pinheiros River (CAPOBIANCO; WHATELY, 2002). However, BIL started to supply cities from SPMR to handle with the increasing population growing rates. The habitants supplied by the reservoir were worried about BIL's water quality and pressed to stop the Pinheiros' transposition (BRAGA; PORTO; SILVA, 2006). After some years of intense discussions, in 1992, the pumping was interrupted unless for flood control emergencies (SMA, 2010). Still, this action couldn't cease the pollution of the reservoir, since the intense human occupation of its margins continued to be a source of punctual and diffuse pollution (MILZ et al., 2022). Besides, the occasional pumping of Pinheiros' contaminated waters worsens the scenario (WENGRAT; BICUDO, 2011).

## 5.2 Billings cyanobacteria content assessed from space

This study's findings reflect the historical anthropic impacts on the reservoir. Cyanobacteria species' constant dominance is a known symptom of Cultural Eutrophication (DOKULIL; TEUBNER, 2000). Figure 4.3 shows the prevalence of Cyanobacteria species and low phytoplankton biodiversity near the most densely occupied spots and where the water from Pinheiros River is pumped into the reservoir (Pedreira Dam). The same pattern is observed in the PC maps produced by all the methods, with the highest concentrations estimated for this region. Cyanobacteria scums were recorded in Pedreira at all field campaigns, evidencing the perennial presence of

CyHABs. This information can be confirmed by the taxonomical analysis, which revealed that the potentially toxic *Microcystis aeruginosa* dominated all the samples collected in that region. Maciel et al. (2021) observed that Pedreira and Taquacetuba arms had the lowest Secchi Disk Depth for all periods in the monthly average. The authors concluded that the phytoplankton growth dominated Billings' water clarity regime, and the blooms intensities followed the precipitation pattern. The same was observed for the present photosynthetic pigments' dataset. The highest mean values of PC and Chl*a* were obtained in November and December (begging of the rainy season), and the lowest were found in July (peak of drought season). One of the explanations for this phenomenon is the increasing volume of the contaminated city runoff and the rising probability of Pinheiros' reversion (WENGRAT; BICUDO, 2011). Therefore, the great amount of nutrients, increased residence time (e.g.,720 days), and higher temperatures are the most common causes of the phytoplankton growth rates in the Pedreira Arm (PIRES et al., 2015; POMPÊO; MOSCHINI-CARLOS, 2020).

The maps also revealed an environmental quality gradient, with PC concentration reducing in the direction of most conserved areas. The Capivari and Rio Pequeno arms are near the Serra do Mar conservation Unit, one of the most preserved fragments of the Atlantic Rainforest (CAPOBIANCO; WHATELY, 2002). Those branches had the lowest PC concentrations for the assessed image. Besides, taxonomical analysis showed that 86% of these branches' samples had a mixed community structure and high species richness. Therefore, the reduced anthropical presence in those areas is the most reasonable explanation for the decrease in Cyanobacteria concentration (SILVA et al., 2014). The Rio Grande arm also had lower concentrations of PC when compared with most contaminated branches. This reservoir water mass is isolated from the others through Anchieta Dam (Figure 3.2). The barrier was constructed to separate Rio Grande from the remaining reservoir to protect the public water supply from contamination (SMA, 2010). Transept #4 (Figure 4.13) reveals low PC concentrations near Anchieta Dam (where the water is collected), increasing towards Rio Grande's upstream. The use of algicides (e.g., Copper Sulfate and hydrogen peroxide) to control the Cyanobacteria populations might explain the low PC values (CARVALHO et al., 1997; MOSCHINI-CARLOS; FREITAS; POMPÊO, 2010).

## 5.3 Impact of biological effects in RS assessments of Cyanobacteria

The PC:Chl*a* ratio is one of the most important parameters for accurate predictions of PC using Remote Sensing data (SIMIS; PETERS; GONS, 2005). Elevated ratios indicate the spectral dominance of PC features, reducing the influence of Chl*a* and other OAC. The optimum range for PC predictions from RS data was described as greater than 1.5 (RUIZ-VERDÚ et al., 2008), whereas the values below this threshold had increased errors. This study has one of the smallest PC:Chl*a* ratios observed in Cyanobacteria RS assessments. O'Shea et al. (2021), which gathered the greatest reported PC dataset, had median and mean PC:Chl*a* of 0.46 and 0.82, respectively. These are 5 and 5.8 folds greater than this study's median and mean values (0.09 and 0.14). The disparity is even greater when considering the dataset used by Simis et al. (2007), where PC:Chl*a* up to 8 were registered. At a first look, those low PC:Chl*a* values might induce to consider the absence or reduced presence of Cyanobacteria species. However, the taxonomical analysis refutes that hypothesis once 96% of the samples ($n$ = 26) had at least an abundant Cyanobacteria species. Another explanation is a possible low extraction efficiency obtained from PC samples. Biliproteins extraction standards are not as developed as Chl*a* methods, and the specialized literature has reported underestimation trends in most techniques (SARADA; PILLAI; RAVISHANKAR, 1999; HORVÁTH et al., 2013). Zimba (2012) observed that PC extractions using asolectin-CHAPS as solvent had achieved concentrations 28% higher than the Phosphate Buffer used in this study. Therefore, underestimations of PC concentrations might have incorrectly reduced the PC:Chl*a* ratios.

Even though PC standard determination procedures have been appointed as a limitation for Cyanobacteria assessments (RUIZ-VERDÚ et al., 2008), this study has used the same method as previously published articles (SIMIS; PETERS; GONS, 2005; MISHRA et al., 2013; LIU et al., 2017). In addition, when comparing the PC concentrations with Cyanobacteria biovolume found in Billings Reservoir, a strong linear relationship was observed ($R^2$ = 0.85). Therefore, at least a regular extraction efficiency was achieved in PC determination. Even if it was added 28% to the estimated PC values as reported by Zimba (2012), the PC:Chl*a* won't be as great as reported in other studies. Another possible explanation for the reduced values of PC:Chl*a* is the Chromatic Adaptation (CA). Cyanobacteria can control biliproteins production based on environmental factors (GROSSMAN et al., 1993). Nutrient deprivation and mechanical confinement of

Cyanobacteria cells can lead to alterations in PC:Chl*a* ratios (MOORE et al., 2020). In the case of CA, those alterations are induced by differences in the incident electromagnetic radiation field (TANDEAU DE MARSAC, 1977). Once biliproteins main function is to harvest light beyond the Chl*a* absorption spectrum, when the amount of energy captured by Chl*a* is enough for maintaining the cell's basal metabolism, it triggers a response for reducing the production of phycobilisomes (LÖNNEBORG et al., 1985). Thus, when Cyanobacteria cells are exposed to high illumination intensities, they might have a reduced PC concentration than others grown in shaded environments (HOTOS, 2021).

The Complementary Chromatic Adaptation (CCA) is related to the spectral properties of the illumination field (HATTORI; FUJITA, 1959). Some studies have reported that Cyanobacteria growing in green light were prone to increase Phycoerythrin (PE) concentrations and reduce PC levels (DE MARSAC; HOUMARD, 1988). The opposite was observed for cells illuminated with monochromatic red light. Cyanobacteria alter the rate of those produced pigments to better explore the available light outside the Chl*a* absorption range (LÖNNEBORG et al., 1985). Also, great light intensities increased Carotenoids production in Cyanobacteria cells (ŚLIWIŃSKA-WILCZEWSKA et al., 2019). Those proteins are recognized as protection pigments against the deleterious consequences of the cell's exposition to high light intensities (SIEFERMANN-HARMS, 1987). This study is one of the first PC assessments by RS in tropical reservoirs. Most of the study areas are in mid-latitude regions, where the Sun's illumination conditions are different than those observed in Billings. Maybe, the irradiance increase caused by lower Sun's zeniths throughout the year can unleash CA in Billings' Cyanobacteria cells. In addition, the reflectance peak located at the green region in all collected samples possibly indicates a predominant green underwater radiation field (Figure 4.4). The green light's great availability might provoke CCA and increase PE concentrations over PC. Although PE concentrations were not determined in this study, the spectrophotometric analysis (not shown) revealed an absorbance increase at 560 nm, where the Phycoerythrin absorption peak is reported (BRYANT, 1982). To confirm those hypotheses, more studies evaluating CA and CCA must be done in environmental conditions once all available studies are designed in controlled laboratories and with monospecific cultures.

The low PC:Chl*a* ratio is also the most reasonable explanation for MDN's failure in predicting PC for Billings Reservoir. Once MDN uses the Chl*a* as an input layer for predicting PC, the algorithm might be biased for the PC:Chl*a* ratios of the samples used for training. Billings PC:Chl*a* is, at mean, almost six times lower than the O'Shea et al. (2021) dataset, nearly the same magnitude of overestimation reported by Bias (7.12). It also can explain why the estimated Chl*a* was more accurate than PC values. Therefore, MDN should receive a representative size of training samples with a low PC:Chl*a* ratio to achieve better results for Billings. Also, more studies evaluating tropical and equatorial reservoirs need to be done to assess the impact of CA and CCA on PC production. This is the only path to achieving a truly global algorithm for predicting PC, where a representative dataset can be used for extensive validation work. Furthermore, if the hypothesis of CA and CCA were confirmed, low PC concentrations must be better interpreted in RS assessments (e.g., Low PC might not necessarily represent absence/low concentration of Cyanobacteria). Additionally, studies related to the potential of Phycoerythrin to estimate Cyanobacteria biomass might also elucidate the potential of this pigment to perform RS evaluations.

The results achieved for PC modeling in this study differ from most previously published. Besides the elevated PC:Chl*a* ratio, the specialized literature considers that values beyond 50 μg/L are more susceptible to produce accurate estimates using RS data (SIMIS et al., 2007). However, only ten samples would remain in the present dataset if this threshold was applied. Even when compared with the dataset used by O'Shea et al. (2021), which has found the best-reported estimates for low PC concentrations, the values obtained by the authors are, at the mean, 2.4 and, at the median, 4.9 folds greater than the present study. Ruiz-Verdú et al. (2008) noticed a sharp accuracy decrease when predicting PC below 50 μg/L compared with the reported optimum range of 50-200 μg/L. The authors registered a considerable percentage of negative predictions within 2 of the 3 assessed algorithms for PC < 50 μg/L. The capacity of predicting high concentrations more accurately might explain why the error metrics were reduced in the orbital data validation when compared with the *in-situ* data modeling. The dataset used in the Monte Carlo simulation had a median PC of 2.9 μg/L against 18.18 μg/L for PRISMA match-up stations.

## 5.4 Machine Learning Algorithms for predicting PC

Comparing the results obtained in this study with others is difficult due to the differences in the metrics used for assessing the estimates. However, some previous achievements are presented below. It is important to note that studies conceived using hyperspectral *in-situ* data (e.g., 1 nm resolution) or based on Ocean Color Satellites data weren't considered due to the incompatibility with this study's objectives (e.g., spatial resolution unable to monitor urban reservoirs). Using hyperspectral field data resampled for Landsat-8/OLI, Ogashawara et al. (2022) assessed the linear relationship between the band ratios available for the sensor (including the virtual-orange band) and measured PC values. The authors achieved a maximum $R^2$ of 0.08 considering all obtained samples (Mean PC = 40.61 μg/L, Median PC = 21.23 μg/L). Kumar et al. (2020) have found a stronger linear relationship ($R^2$ = 0.55) between PC concentrations and the band ratio virtual-orange/red from a Landsat-8/OLI image acquired in match-up condition (PC from 0.23 to 170.39 μg/L). Beck et al. (2017) created synthetic images from different sensors using a hyperspectral image from Compact Airborne Spectrographic Imager (CASI) and used it to estimate match-up PC fluorescence measurements. The algorithm developed using the synthetic Worldview-2 image had the best linear relationship with the measured fluorescence ($R^2$ = 0.79), followed by CASI ($R^2$ = 0.78) and Landsat/8/OLI ($R^2$ = 0.301). Just for comparison, $R^2$ values were calculated for *in-situ* PC modeling. Those values are the median $R^2$, from all Monte Carlo iterations, obtained from the 20% of samples used for validating the MLA predictions. PRISMA/RF achieved a median $R^2$ of 0.85, while Worldview-3/RF and Landsat-8/RF scored 0.83 and 0.79, respectively. Therefore, the worst linear relationship achieved in this study was equal to the best value from the reviewed articles.

The Machine Learning Algorithms capacity of translating non-linear relationships between the input features and output labels might explain the success of this approach (LARY et al., 2016). They are able to explore the prediction potential of multiple input features, which makes them often more precise when compared with a univariate linear regressions (SAGAN et al., 2020). Sun et al. (2012) also highlighted the advantages of using SVM over linear models to predict PC in productive Chinese waters (MAPE = 29.5%, PC mean = 106.3 μg/L, and mean PC:Chl*a* = 2.45). The results obtained in this study also supported that there is an accuracy gain when using MLA over the previously

published algorithms based on linear estimates. The applied algorithms were able to predict low PC concentrations with reduced PC:Chl*a* ratios at a reasonable accuracy, an achievement that was never reached before. Ensemble algorithms based on Decision trees (RF and XgBOOST) presented great applicability in retrieving PC concentrations for the Billings Reservoir. They also provided accurate estimates using orbital data, showing that they are robust to radiometric noise inputs. Cao et al. (2020) elucidated the advantages of using XgBOOST for predicting Chl*a* from Landsat-8/OLI data, while Maciel et al. (2021) proved that RF can retrieve Secchi Disk Depth from Sentinel-2/MSI imagery. The prediction capacity of RF and XgBOOST may be due to the high generalization and non-linear predictive power of both algorithms (BELGIU; DRĂGU, 2016). Instead, SVM might have underperformed due to the linear approach used for predicting PC. The chosen kernel also seems to be the most sensitive to radiometric noise of orbital images. Despite of the elucidated advantages of using MLA, some disadvantages were evident during the study. As shown by the predictions obtained by MDN, MLA are prone to a sharp accuracy decrease when predicting samples outside the range of the training dataset (CAO et al., 2020). The demand for large datasets might be the greatest challenge of using those algorithms (CHAN et al., 2020).

MLA can also be used to potentialize the use of hyperspectral data in RS assessments. Usually, PC assessments using hyperspectral uses only features related to the pigments' absorption feature (near 620 nm) (LI et al., 2010; BECK et al., 2017). To optimize PC prediction, O'Shea et al. (2021) used the MDN to explore the spectral range from 500-710 nm. The capacity of MDN to learn with multiple input layers has also permitted the algorithm to predict Chl*a* accurately from Sentinel-2/MSI and Sentinel-3/OLCI images (PAHLEVAN et al., 2020). The Predictive Power Score has shown the potential to identify input layers to predict PC that didn't use the band in 620 nm. PPS has highlighted LH and NI in the green and NIR regions as important for the prediction of PC. Dekker (1993) has noticed that the location of the maximum green reflectance peak was strongly related to PC concentrations. Also, green bands might have indirectly predicted PC based on PE absorption features. If CCA occurs in Billings, the increased PE concentrations are likely to produce more identifiable spectral characteristics, and MLA might have translated those relationships in PC concentration. Spectral indexes based only on NIR bands for predicting PC weren't reported in previously published studies. The predictive

power of this spectral region might be related to the increased algae scattering in the NIR region and Chl*a* fluorescence peaks (LEHMUSKERO; SKOGEN CHAUTON; BOSTRÖM, 2018; SANTABARBARA et al., 2020). Also, some Cyanobacteria can expand their absorbance spectrum to longer wavelengths by Near-Infrared photoacclimation (MIAO et al., 2016). However, more investigation needs to be done to understand the potential of isolated NIR bands to predict photosynthetic pigments.

### 5.5 Atmospheric correction of PRISMA hyperspectral data

The present study presents one of the first evaluations of PRISMA atmospherically corrected products. Atmospheric Correction (AC) is one of the most limiting factors for monitoring water quality parameters from inland water bodies (WARREN et al., 2019). Unfortunately, few efforts were made to estimate PC from real satellite images (OGASHAWARA, 2019; KUMAR; MISHRA; ILANGO, 2020; O'SHEA et al., 2021). Therefore, another valuable contribution is evaluating MLA under residual atmospheric noise. The ASI product has shown the potential to deliver accurate bottom-of-atmosphere reflectance products, especially in the visible region. Still, an error increase was observed at some green bands (550-600 nm). The same happened using the other methods, ACOLITE and 6SV. Paulino et al. (2022) have found a similar trend for Sentinel-2/MSI images atmospherically corrected with 6SV over Billings Reservoir. The authors have attributed those errors to the great contribution of the water's signal to the top-of-atmosphere radiance, inducing underestimations even for slight corrections of the atmosphere additive noise. Additionally, the strong underestimation tendency observed in ACOLITE and 6SV in the visible spectrum might be derived from inaccurate estimates from AOT at 550 nm, overcorrecting the output images (PAULINO et al., 2022). NIR bands have shown great importance for retrieving accurate PC estimates in this study. Furthermore, hyperspectral images are an opportunity for exploring the applicability of this spectral region to produce water quality products. All assessed methods had a parabolical-shaped error in the NIR (700-813 nm). This wavelength range is characterized by variations in atmospheric transmittance (80-100%) (GAO et al., 2009). The errors might traduce the reduced capacity of all assessed methods to accurately predict the variations of the effects of atmospheric constituents in this range (FROUIN et al., 2019). In addition, increased errors in the NIR region were observed in various RS assessments in inland waters (SIEGEL et al., 2000; WARREN et al., 2019; TAVARES et al., 2021)

and needed to be better addressed to fully explore the potential of hyperspectral images (FROUIN et al., 2019). Other studies should evaluate PRISMA atmospheric corrected images under different atmospheric conditions.

### 5.6 Multispectral and Hyperspectral data for predicting PC

The results showed a performance increase with the rising availability of spectral bands, leading to more input features for MLA predicting PC. Landsat-8/OLI had four spectral bands and three input features. The virtual-orange band permitted to estimate the same PC spatial pattern in Billings reservoir, although it had the lowest accuracy disregarding the sensors. Kumar et al. (2020) reported that Landsat-8/OLI overestimated PC in regions with high inorganic solids presence. However, this tendency couldn't be evaluated in this study once the Billings reservoir has low concentrations of this OAC (MOSCHINI-CARLOS et al., 2009). Landsat-8/OLI synthetic data was unable to predict the highest PC values, which can be observed by comparing the sensors estimates in the Pedreira arm (Transept #1). This might be explained due to the near-infrared bandwidth and position (829-900 nm), which reduces the applicability of the sensor for estimating photosynthetic pigments (BECK et al., 2016). This limitation is overcome by Worldview-3, with the additional Red-Edge band. This extra band has considerably reduced the errors in both *in-situ* and orbital PC retrievals. Sentinel-2/MSI red-edge bands were also responsible for enhancing Chl*a* estimates (BRAMICH; BOLCH; FISCHER, 2021). Worldview-3 has achieved accuracy metrics comparable with PRISMA hyperspectral sensor. With 40 spectral channels and five input features, the ASI's sensor had the best-assessed metrics in this study. The spectral information offered by the hyperspectral bands is probably the motive for overperforming the other sensors. However, PRISMA mapping had high variations in PC prediction in low spatial distances. Those errors are probably explained by the low Signal-to-Noise Ratio (SRN) from hyperspectral sensor bands (MOSES et al., 2012). O'Shea et al. (2021) also noticed speckled noise in PRISMA maps and observed that LH and Band Ratios might have amplified the errors. The applied median filter reduced the effect of low SRN, even though it has reduced the effective spatial resolution of the mapping. The synthetic multispectral images presented smoother results once the larger bandwidths improved the PRISMA image's SNR and radiometric quality (JORGE et al., 2017).

All assessed orbital sensors showed to be suitable for monitoring PC in urban reservoirs, each one with its pros and cons. PRISMA had the best-assessed results, but the on-demand image acquisition and reduced SNR are limitations of using the hyperspectral sensor for continuous monitoring. Worldview-3 had accurate predictions and has the most improved resolution for monitoring almost all urban reservoirs. However, the commercial data policy increases the cost of a monitoring plan based on the Worldview images. Although presenting the greater MAE, Landsat has the most consolidated Earth Observation program. Considering the data acquired from Landsat-8 and Landsat-9, orbital images are recorded every eight days and are freely available in the USGS portal. Also, the scene size (185x185 km) is the most adequate for different water bodies' sizes and shapes. Therefore, Landsat can be considered the most feasible orbital sensor for monitoring PC and CyHABs in the Billings reservoir. More validation studies should be done using real Landsat-8/OLI and Landsat-9/OLI-2 images to assess the proposed methodologies in this study.

# 6 FINAL CONSIDERATIONS

This study has assessed the potential of Machine Learning Algorithms and hyperspectral imagery to estimate C-Phycocyanin concentrations. The obtained results confirmed the initial hypothesis that hyperspectral and multispectral orbital images produce accurate PC mappings. Furthermore, accuracy varied according to Machine Learning Algorithms, and the type of images, hyperspectral PRISMA, synthetic multispectral Landsat-8/OLI and Worldview-3 images. This way, the following questions could be answered:

*1. What is the best surface reflectance product for PRISMA hyperspectral data?*

Three Atmospheric Correction processors were assessed for estimating bottom-of-atmosphere reflectance for a PRISMA hyperspectral image(ASI's Analise Ready Data, ACOLITE, and 6SV products, from 500 to 877 nm). ASI product had the best-observed results, especially in the visible spectrum. For this spectral range, the PRISMA original product had errors lower than 20% for all assessed bands. ACOLITE and 6SV underperformed ASI's product with a strong underestimation trend in the visible region. In Near-Infrared, the methods had a decreased performance and presented similar outputs, with a slight advantage of ACOLITE. Besides, an overestimation trend was observed beyond 700 nm against the previously described underestimation tendency from 500-700 nm. Overall, ASI's product was the most reliable atmospherically corrected product to estimate water quality parameters and was used to estimate PC concentrations in a productive tropical urban reservoir. More studies in different atmospheric conditions must be pursued to validate the ASI's surface reflectance product.

*Are Machine Leaning Algorithms able to predict C-Phycocyanin accurately? Among the three tested, which is the best?*

Yes. Machine Learning Algorithms (MLA) achieved accurate PC estimates based on *in-situ* and orbital $R_{rs}$. Random Forest and Extreme Gradient Boost had the most reliable results, both ensemble algorithms based on Decision Trees. In contrast, the Support Vector Machine had the worst predictions in all tested scenarios and wasn't a feasible option to estimate PC in the present study. *In-situ* PC modeling had results ranging from 59-99% Mean Absolute Error, considering the different sensors and MLA, while orbital

estimates had values from 45%-112%. Those results are expressive considering the low PC:Chl*a* and PC concentrations obtained in this study.

*2. Is there an accuracy gain in using hyperspectral data over multispectral data for predicting C-Phycocyanin?*

Yes. Hyperspectral data has achieved the most accurate results for orbital and *in-situ* PC modeling. PRISMA retrieved PC concentrations with a Mean Absolute Error (MAE) of 59% and 45% for resampled *in-situ* data and an orbital image, respectively. While, Landsat-8/OLI had the most inaccurate results, which might be attributed to the most limited spectral resolution from the assessed sensors. Worldview-3 presented results comparable with PRISMA, with 65% of MAE in *in-situ* modeling and 49% with a synthetic image generated from PRISMA data. This accuracy increase might be due to the presence of a Red-Edge band in the ultra-resolution satellite, enhancing the applications for algae monitoring. However, due to the low Signal-to-Noise Ratio of PRISMA data, a specked noise was observed in the PC mapping. The same was not observed in the synthetic multispectral results.

Finally, considering the results obtained in this research, Landsat-8/OLI has the greatest potential for monitoring PC in urban reservoirs. Even though the USGS sensor had accuracy metrics lower than the other assessed platforms, the high temporal resolution (8 days considering Landsat-8 and Landsat-9) and the free access data policy are advantages of using Landsat data in a continuous monitoring program. However, more studies should be done to validate the proposed methodologies for monitoring PC, as well as the impact of different PC:Chl*a* ratios in Remote Sensing assessments.

## REFERENCES


ALCÂNTARA, E.; WATANABE, F.; RODRIGUES, T.; BERNARDO, N. An investigation into the phytoplankton package effect on the chlorophyll-a specific absorption coefficient in Barra Bonita reservoir, Brazil. **Remote Sensing Letters**, v. 7, n. 8, p. 761–770, 2016.

ALLEN, J. F.; MARTIN, W. Evolutionary biology: out of thin air. **Nature**, v. 445, p. 610–612, 2007.

AMERICAN PUBLIC HEALTH ASSOCIATION (APHA). **Standard methods for the examination of water and wastewater**. 20. ed. Washington: American Public Health Association, 1998. 874 p.

ARNDT, J.; KIRCHNER, J. S.; JEWELL, K. S.; SCHLUESENER, M. P.; WICK, A.; TERNES, T. A.; DUESTER, L. Making waves: time for chemical surface water quality monitoring to catch up with its technical potential. **Water Research**, v. 213, n. February, p. 118168, 2022.

BARBOSA, C. C. F.; NOVO, E. M. L. M.; MARTINS, V. S. **Introdução ao sensoriamento remoto de sistemas aquáticos:** princípios e aplicações. São José dos Campos: Instituto Nacional de Pesquisas Espaciais, 2019. 178 p. ISBN(9788517000959).

BECK, R.; XU, M.; ZHAN, S.; LIU, H.; JOHANSEN, R. A.; TONG, S.; YANG, B.; SHU, S.; WU, Q.; WANG, S.; BERLING, K.; MURRAY, A.; EMERY, E.; REIF, M.; HARWOOD, J.; YOUNG, J.; MARTIN, M.; STILLINGS, G.; STUMPF, R.; SU, H.; YE, Z.; HUANG, Y. Comparison of satellite reflectance algorithms for estimating phycocyanin values and cyanobacterial total biovolume in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations. **Remote Sensing**, v. 9, n. 6, 2017.

BECK, R.; ZHAN, S.; LIU, H.; TONG, S.; YANG, B.; XU, M.; YE, Z.; HUANG, Y.; SHU, S.; WU, Q.; WANG, S.; BERLING, K.; MURRAY, A.; EMERY, E.; REIF, M.; HARWOOD, J.; YOUNG, J.; NIETCH, C.; MACKE, D.; MARTIN, M.; STILLINGS, G.; STUMP, R.; SU, H. Comparison of satellite reflectance algorithms for estimating chlorophyll-a in a temperate reservoir using coincident hyperspectral aircraft imagery and dense coincident surface observations. **Remote Sensing of Environment**, v. 178, p. 15–30, 2016.

BELETE, D. M.; HUCHAIAH, M. D. Grid search in hyperparameter optimization of machine learning models for prediction of HIV/AIDS test results. **International Journal of Computers and Applications**, 2021.

BELGIU, M.; DRĂGU, L. Random forest in remote sensing: a review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016.

BELWARD, A. S.; SKØIEN, J. O. Who launched what , when and why ; trends in global land-cover observation capacity from civilian earth observation satellites. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 103, p. 115–128, 2015.

BENNETT, A.; BOGORAD, L. Complementary chromatic adaptation in a filamentous blue-green alga. **The Journal of Cell Biology**, v. 58, n. 38, p. 419–435, 1973.

BIDIGARE, R. R.; ONDRUSEK, M. E.; MORROW, J. H.; KIEFER, D. A. In-vivo absorption properties of algal pigments. **Ocean Optics X**, v. 1302, p. 290–302, 1990.

BLIX, K.; PÁLFFY, K.; TÓTH, V. R.; ELTOFT, T. Remote sensing of water quality parameters over Lake Balaton by using Sentinel-3 OLCI. **Water (Switzerland)**, v. 10, n. 10, p. 1–20, 2018.

BLONDEAU-PATISSIER, D.; GOWER, J. F. R.; DEKKER, A. G.; PHINN, S. R.; BRANDO, V. E. A review of ocean color remote sensing methods and statistical techniques for the detection, mapping and analysis of phytoplankton blooms in coastal and open oceans. **Progress in Oceanography**, v. 123, p. 123–144, 2014.

BRAGA, B. P. F. The management of urban water conflicts in the metropolitan region of São Paulo. **Water International**, v. 25, n. 2, p. 208–213, 2000.

BRAGA, B. P. F.; PORTO, M. F. A.; SILVA, R. T. Water management in metropolitan São Paulo. **International Journal of Water Resources Development**, v. 22, n. 2, p. 337–352, 2006.

BRAMICH, J.; BOLCH, C. J. S.; FISCHER, A. Improved red-edge chlorophyll-a detection for Sentinel 2. **Ecological Indicators**, v. 120, p. 106876, 2021.

BRANDO, V. E.; DEKKER, A. G. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. **IEEE Transactions on Geoscience and Remote Sensing**, v. 41, n. 6, p. 1378–1387, 2003.

BRESCIANI, M.; GIARDINO, C.; FABBRETTO, A.; PELLEGRINO, A.; MANGANO, S.; FREE, G.; PINARDI, M. Application of new hyperspectral sensors in the remote sensing of aquatic ecosystem health: exploiting PRISMA and DESIS for four italian lakes. **Resources**, v. 11, n. 2, 2022.

BRYANT, D. A. Phycoerythrocyanin and phycoerythrin: properties and occurrence in cyanobacteria. **Journal of General Microbiology**, v. 128, n. 4, p. 835–844, 1982.

CAO, Z.; MA, R.; DUAN, H.; PAHLEVAN, N.; MELACK, J.; SHEN, M.; XUE, K. A machine learning approach to estimate chlorophyll-a from Landsat-8 measurements in inland lakes. **Remote Sensing of Environment**, v. 248, 2020.

CAPOBIANCO, J. P. R.; WHATELY, M. **Billings 2000:** ameaças e perspectivas para o maior reservatório de água da região metropolitana de São Paulo. São Paulo: Instituto Socioambiental, 2002. 59 p. ISBN(8585994088).

CARMICHAEL, W. W. Health effects of toxin-producing cyanobacteria: “the cyanoHABs”. **Human and Ecological Risk Assessment**, v. 7, n. 5, p. 1393–1407, 2001.

CARMICHAEL, W. W.; AZEVEDO, S. M. F. O.; AN, J. S.; MOLICA, R. J. R.; JOCHIMSEN, E. M.; LAU, S.; RINEHART, K. L.; SHAW, G. R.; EAGLESHAM, G. K. Human fatalities from cyanobacteria: chemical and biological evidence for cyanotoxins. **Environmental Health Perspectives**, v. 109, n. 7, p. 663–668, 2001.

CARMICHAEL, W. W.; BOYER, G. L. Health impacts from cyanobacteria harmful algae blooms: implications for the north american great lakes. **Harmful Algae**, v. 54, p. 194–212, 2016.

CARPENTER, S. R. Submersed vegetation: an internal factor in lake ecosystem succession. **The American Naturalist**, v. 118, n. 3, p. 372–383, 1981.

CARVALHO, L. R.; SANT'ANNA, C. L.; GEMELGO, M. C. P.; AZEVEDO, M. T. D. P. Cyanobacterial occurrence and detection of microcystin by planar chromatography in surface water of Billings and Guarapiranga reservoirs, SP, Brazil. **Revista Brasileira de Botanica**, v. 30, n. 1, p. 141–148, 2007.

CARVALHO, M. C.; COELHO-BOTELHO, M. J.; LAMPARELLI, M. C.; ROQUETTI-HUMAITÁ, M. H.; SALVADOR, M. E. P.; SOUZA, R. C. R.; TRUZZI, A. Spatial and temporal variations of chlorophyll a , plankton and some physico-chemical factors at Billings Complex, São Paulo, Brazil. **SIL Proceedings**, v. 26, n. 2, p. 452–457, 1997.

CASTAGNA, A.; SIMIS, S.; DIERSSEN, H.; VANHELLEMONT, Q.; SABBE, K.; VYVERMAN, W. Extending landsat 8: retrieval of an orange contra-band for inland water quality applications. **Remote Sensing**, v. 12, n. 4, p. 1–27, 2020.

CASTENHOLZ, R. W. General characteristics of the cyanobacteria. **Bergey's Manual of Systematics of Archaea and Bacteria**, p. 1–23, 2015.

COMPANHIA AMBIENTA DO ESTADO DE SÃO PAULO (CETESB). **Qualidade das águas interiores no Estado de São Paulo 2018**. São Paulo: CETESB, 2019. 302 p. ISBN(9781626239777).

CHAN, J. Y.; MUN, S.; LEOW, H.; BEA, K. T.; CHENG, W. K.; PHOONG, S. W.; HONG, Z.; CHEN, Y. Mitigating the multicollinearity problem and its machine learning approach : a review. **Mathematics**, p. 17, 2022.

CHAN, S.; REDDY, V.; MYERS, B.; THIBODEAUX, Q.; BROWNSTONE, N.; LIAO, W. Machine learning in dermatology: current applications, opportunities, and limitations. **Dermatology and Therapy**, v. 10, n. 3, p. 365–386, 2020.

CHISLOCK, M. F.; DOSTER, E.; ZITOMER, R. A.; WILSON, A. E. Eutrophication : causes, consequences, and controls in aquatic ecosystems. **Nature Education**, v. 4, p. 8, 2013.

CODD, G. A.; BELL, S. G.; BROOKS, W. P. Cyanobacterial toxins in water. **Water Science and Technology**, v. 21, n. 3, p. 1–13, 1989.

COGLIATI, S.; SARTI, F.; CHIARANTINI, L.; COSI, M.; LORUSSO, R.; LOPINTO, E.; MIGLIETTA, F.; GENESIO, L.; GUANTER, L.; DAMM, A.; PÉREZ-LÓPEZ, S.;

SCHEFFLER, D.; TAGLIABUE, G.; PANIGADA, C.; RASCHER, U.; DOWLING, T. P. F.; GIARDINO, C.; COLOMBO, R. The PRISMA imaging spectroscopy mission: overview and first performance analysis. **Remote Sensing of Environment**, v. 262, 2021.

CÔRTES, P. L.; TORRENTE, M.; ALVES FILHO, A. P.; RUIZ, M. S.; DIAS, A. J. G.; RODRIGUES, R. Crise de abastecimento de água em São Paulo e falta de planejamento estratégico. **Estudos Avançados**, v. 29, n. 84, p. 2015, 2015.

DE MARSAC, N. T.; HOUMARD, J. Complementary chromatic adaptation: physiological conditions and action spectra. **Methods in Enzymology**, v. 167, n. C, p. 318–328, 1988.

DEKKER, A. G. **Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing**. 1993. 369–375 p. Thesis (PhD) - Vrije Universiteit, Amsterdam, 1993.

DEV, P. J.; SUKENIK, A.; MISHRA, D. R.; OSTROVSKY, I. Cyanobacterial pigment concentrations in inland waters: novel semi-analytical algorithms for multi- and hyperspectral remote sensing data. **Science of the Total Environment**, v. 805, p. 150423, 2022.

DIERSSEN, H. M.; ACKLESON, S. G.; JOYCE, K. E.; HESTIR, E. L.; CASTAGNA, A.; LAVENDER, S.; MCMANUS, M. A. Living up to the hype of hyperspectral aquatic remote sensing: science, resources and outlook. **Frontiers in Environmental Science**, v. 9, p. 1–26, 2021.

DIXON, W.; CHISWELL, B. Review of aquatic monitoring program design. **Water Research**, v. 30, n. 9, p. 1935–1948, 1996.

DOKULIL, M. T.; TEUBNER, K. Cyanobacterial dominance in lakes. **Hydrobiologia**, v. 438, p. 1–12, 2000.

EL-SHEHAWY, R.; GOROKHOVA, E.; FERNÁNDEZ-PIÑAS, F.; DEL CAMPO, F. F. Global warming and hepatotoxin production by cyanobacteria: what can we learn from experiments? **Water Research**, v. 46, n. 5, p. 1420–1429, 2012.

ELSER, J. J.; MARZOLF, E. R.; GOLDMAN, C. R. Phosphorus and nitrogen limitation of phytoplankton growth in the freshwaters of North America: a review and critique of experimental enrichments. **Canadian Journal of Fisheries and Aquatic Sciences**, v. 47, n. 7, p. 1468–1477, 1990.

FONSECA, B. M.; FERRAGUT, C.; TUCCI, A.; CROSSETTI, L. O.; FERRARI, F.; BICUDO, D. DE C.; SANT'ANNA, C. L.; BICUDO, C. E. DE M. Biovolume de cianobactérias e algas de reservatórios tropicais do Brasil com diferentes estados tróficos. **Hoehnea**, v. 41, n. 1, p. 9–30, 2014.

FRANCIS, G. Poisonous australian lake. **Nature**, 1878.

FROUIN, R. J.; FRANZ, B. A.; IBRAHIM, A.; KNOBELSPIESSE, K.; AHMAD, Z.; CAIRNS, B.; CHOWDHARY, J.; DIERSSEN, H. M.; TAN, J.; DUBOVIK, O.;

HUANG, X.; DAVIS, A. B.; KALASHNIKOVA, O.; THOMPSON, D. R.; REMER, L. A.; BOSS, E.; CODDINGTON, O.; DESCHAMPS, P. Y.; GAO, B. C.; GROSS, L.; HASEKAMP, O.; OMAR, A.; PELLETIER, B.; RAMON, D.; STEINMETZ, F.; ZHAI, P. W. Atmospheric correction of satellite ocean-color imagery during the PACE era. **Frontiers in Earth Science**, v. 7, 2019.

GAO, B. C.; MONTES, M. J.; DAVIS, C. O.; GOETZ, A. F. H. Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean. **Remote Sensing of Environment**, v. 113, 2009.

GEMELGO, M. C. P.; SANT'ANNA, C. L.; TUCCI, A.; BARBOSA, H. R. Population dynamics of Cylindrospermopsis raciborskii (Woloszynska) Seenayya & Subba Raju, a Cyanobacteria toxic species, in watersupply reservoirs in São Paulo, Brazil. **Hoehnea**, v. 35, n. 2, p. 297–307, 2008.

GIARDINO, C.; BRANDO, V. E.; DEKKER, A. G.; STRÖMBECK, N.; CANDIANI, G. Assessment of water quality in lake Garda (Italy) using hyperion. **Remote Sensing of Environment**, v. 109, n. 2, p. 183–195, 2007.

GIARDINO, C.; BRANDO, V. E.; GEGE, P.; PINNEL, N.; HOCHBERG, E.; KNAEPS, E.; REUSEN, I.; DOERFFER, R.; BRESCIANI, M.; BRAGA, F.; FOERSTER, S.; CHAMPOLLION, N.; DEKKER, A. Imaging spectrometry of inland and coastal waters: state of the art, achievements and perspectives. **Surveys in Geophysics**, v. 40, n. 3, p. 401–429, 2019.

GIARDINO, C.; BRESCIANI, M.; BRAGA, F.; FABRETTO, A.; GHIRARDI, N.; PEPE, M.; GIANNINETTO, M.; COLMBO, R. First evaluation of PRISMA level 1 data for water applications. **Sensors**, v. 20, n. 16, p. 4553, 2020.

GOETZ, A. F. H. Three decades of hyperspectral remote sensing of the Earth: a personal view. **Remote Sensing of Environment**, v. 113, p. S5–S16, 2009.

GONS, H. J. Optical teledetection of chlorophyll a in turbid inland waters. **Environmental Science and Technology**, v. 33, n. 7, p. 1127–1132, 1999.

GRAHAM, J. .; ZIEGLER, A. C.; MEYER, M. T. **Guidelines for design and sampling for cyanobacterial toxin and taste-and-odor studies in lakes and reservoirs**. Reston: United States Geological Service, 2008. 39 p.

GREEN, R. O.; EASTWOOD, M. L.; SARTURE, C. M.; CHRIEN, T. G.; ARONSSON, M.; CHIPPENDALE, B. J.; FAUST, J. A.; PAVRI, B. E.; CHOVIT, C. J.; SOLIS, M.; OLAH, M. R.; WILLIAMS, O. Imaging spectroscopy and the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). **Remote Sensing of Environment**, v. 65, n. 3, p. 227–248, 1998.

GROSSMAN, A. R.; SCHAEFER, M. R.; CHIANG, G. G.; COLLIER, J. L. The phycobilisome, a light-harvesting complex responsive to environmental conditions. **Microbiological Reviews**, v. 57, n. 3, p. 725–749, 1993.

GUARANI, R.; LOIZZO, R.; FACCHIENETTI, C.; LONGO, F.; PONTICELLI, B.; FARACI, M.; DAMI, M.; COSI, M.; AMORUSO, L.; DE PASQUALE, V.;

COLANDREA, P.; MIOTTI, E.; DI NICOLANTONIO, W. PRISMA Hyperspectral mission products. **IEEE Transactions on Geoscience and Remote Sensing**, v. 1, n. 18, p. 179–182, 2018.

GURLIN, D.; GITELSON, A. A.; MOSES, W. J. Remote estimation of chl-a concentration in turbid productive waters: return to a simple two-band NIR-red model? **Remote Sensing of Environment**, v. 115, n. 12, p. 3479–3490, 2011.

HATTORI, A.; FUJITA, Y. Formation of phycobilin pigments in a blue-green alga, Tolypothrix tenuis, as induced by illumination with colored lights. **The Journal of Biochemestrty**, v. 46, n. 4, p. 521–524, 1959.

HAVENS, K. E. Cyanobacteria blooms: effects on aquatic ecosystems. **Advances in Experimental Medicine and Biology**, v. 619, n. 2004, p. 733–747, 2008.

HAWKINS, P. R.; RUNNEGAR, M. T. C.; JACKSON, A. R. B.; FALCONER, I. R. Severe hepatotoxicity caused by the (blue-green alga) Cylindrospermopsis raciborskii (Woloszynska) Seenaya and Subba Raju isolated from a domestic water supply reservoir. **Applied and Environmental Microbiology**, v. 50, n. 5, p. 1292–1295, 1985.

HILLEBRAND, H.; DÜRSELEN, C. D.; KIRSCHTEL, D.; POLLINGHER, U.; ZOHARY, T. Biovolume calculation for pelagic and benthic microalgae. **Journal of Phycology**, v. 35, n. 2, p. 403–424, 1999.

HO, J. C.; MICHALAK, A. M.; PAHLEVAN, N. Widespread global increase in intense lake phytoplankton blooms since the 1980s. **Nature**, v. 574, n. 7780, p. 667–670, 2019.

HO, L.; TANG, T.; MONIS, P. T.; HOEFEL, D. Biodegradation of multiple cyanobacterial metabolites in drinking water supplies. **Chemosphere**, v. 87, n. 10, p. 1149–1154, 2012.

HORVÁTH, H.; KOVÁCS, A. W.; RIDDICK, C.; PRÉSING, M. Extraction methods for phycocyanin determination in freshwater filamentous cyanobacteria and their application in a shallow lake. **European Journal of Phycology**, v. 48, n. 3, p. 278–286, 2013.

HOTOS, G. N. Culture growth of the cyanobacterium phormidium sp. In various salinity and light regimes and their influence on its phycocyanin and other pigments content. **Journal of Marine Science and Engineering**, v. 9, n. 8, 2021.

HUMPHRIES, S. E.; LYNE, V. D. Cyanophyte blooms: the role of cell buoyancy. **Limnology and Oceanography**, v. 33, n. 1, p. 79–91, 1988.

HUNTER, P. D.; TYLER, A. N.; CARVALHO, L.; CODD, G. A.; MABERLY, S. C. Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. **Remote Sensing of Environment**, v. 114, n. 11, p. 2705–2718, 2010.

HUNTER, P. D.; TYLER, A. N.; GILVEAR, D. J.; WILLBY, N. J. Using remote sensing to aid the assessment of human health risks from blooms of potentially toxic cyanobacteria. **Environmental Science and Technology**, v. 43, n. 7, p. 2627–2633,

2009.

JACOBI, P. Environmental problems in Sao Paulo: the challenge for co-responsibility and innovative crisis management. **Journal of Contingencies and Crisis Management**, v. 5, n. 3, p. 131–139, 1997.

JIN, H. **Hyperparameter importance for machine learning algorithms**. 2022.

JORGE, D. S. F.; BARBOSA, C. C. F.; CARVALHO, L. A. S.; AFFONSO, A. G.; LOBO, F. DE L.; NOVO, E. M. L. DE M. M. L. D. M. SNR (signal-to-noise ratio) impact on water constituent retrieval from simulated images of optically complex Amazon lakes. **Remote Sensing**, v. 9, n. 7, p. 1–18, 2017.

KIM, D.; PHAM, K.; OH, J. Y.; LEE, S. J.; CHOI, H. Classification of surface settlement levels induced by TBM driving in urban areas using random forest with data-driven feature selection. **Automation in Construction**, v. 135, p. 104109, 2022.

KIRK, J. T. O. **Light and photosynthesis in aquatic ecosystems**. Cambridge: Cambridge University Press, 2010. 665 p. ISBN(9780521151757).

KRAVITZ, J.; MATTHEWS, M.; BERNARD, S.; GRIFFITH, D. Remote sensing of environment application of Sentinel 3 OLCI for chl-a retrieval over small inland water targets: successes and challenges. **Remote Sensing of Environment**, v. 237, p. 111562, 2020.

KRUSE, F. A.; LEFKOFF, A. B.; BOARDMAN, J. W.; HEIDEBRECHT, K. B.; SHAPIRO, A. T.; BARLOON, P. J.; GOETZ, A. F. H. The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. **Remote Sensing of Environment**, v. 44, n. 2–3, p. 145–163, 1993.

KUDELA, R. M.; PALACIOS, S. L.; AUSTERBERRY, D. C.; ACCORSI, E. K.; GUILD, L. S.; TORRES-PEREZ, J. Application of hyperspectral remote sensing to cyanobacterial blooms in inland waters. **Remote Sensing of Environment**, v. 167, p. 196–205, 2015.

KUMAR, A.; MISHRA, D. R.; ILANGO, N. Landsat 8 virtual orange band for mapping cyanobacterial blooms. **Remote Sensing**, v. 12, n. 5, 2020.

KUMAR, S.; GHOSH, J.; CRAWFORD, M. M. Best-bases feature extraction algorithms for classification of hyperspectral data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 39, n. 7, p. 1368–1379, 2001.

KUTSER, T.; METSAMAA, L.; DEKKER, A. G. Influence of the vertical distribution of cyanobacteria in the water column on the remote sensing signal. **Estuarine, Coastal and Shelf Science**, v. 78, n. 4, p. 649–654, 2008.

LAMBERT, T. W.; HOLMES, C. F. B.; HRUDEY, S. E. Adsorption of microcystin-LR by activated carbon and removal in full scale water treatment. **Water Research**, v. 30, n. 6, p. 1411–1422, 1996.

LARY, D. J.; ALAVI, A. H.; GANDOMI, A. H.; WALKER, A. L. Machine learning in geosciences and remote sensing. **Geoscience Frontiers**, v. 7, n. 1, p. 3–10, 2016.

LEE, Z.; CARDER, K. L.; ARNONE, R. A. Deriving inherent optical properties from water color: a multiband quasi-analytical algorithm for optically deep waters. **Applied Optics**, v. 41, n. 27, p. 5755, 2002.

LEHMUSKERO, A.; SKOGEN CHAUTON, M.; BOSTRÖM, T. Light and photosynthetic microalgae: a review of cellular- and molecular-scale optical processes. **Progress in Oceanography**, v. 168, p. 43–56, 2018.

LEI, L.; PENG, L.; HUANG, X.; HAN, B. P. Occurrence and dominance of Cylindrospermopsis raciborskii and dissolved cylindrospermopsin in urban reservoirs used for drinking water supply, South China. **Environmental Monitoring and Assessment**, v. 186, n. 5, p. 3079–3090, 2014.

LI, L.; SENGPIEL, R. E.; PASCUAL, D. L.; TEDESCO, L. P.; WILSON, J. S.; SOYEUX, A. Using hyperspectral remote sensing to estimate chlorophyll-a and phycocyanin in a mesotrophic reservoir. **International Journal of Remote Sensing**, v. 31, n. 15, p. 4147–4162, 2010.

LIN, N.; CHEN, Y.; LIU, H.; LIU, H. A comparative study of machine learning models with hyperparameter optimization algorithm for mapping mineral prospectivity. **Minerals**, v. 11, n. 2, p. 1–31, 2021.

LIU, G.; SIMIS, S. G. H.; LI, L.; WANG, Q.; LI, Y.; SONG, K.; LYU, H.; ZHENG, Z.; SHI, K. A four-band semi-analytical model for estimating phycocyanin in inland waters from simulated MERIS and OLCI data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 56, n. 3, p. 1374–1385, 2017.

LOBO, E.; LEIGHTON, G. Estructuras comunitarias de la fitocenosis planctónicas de los sistemas de desembocaduras de ríos y esteros de la Zona Central de Chile. **Revista de Biología Marina**, v. 22, n. 1, p. 1–29, 1986.

LÖNNEBORG, A.; LIND, L. K.; KALLA, S. R.; GUSTAFSSON, P.; ÖQUIST, G. Acclimation processes in the light-harvesting system of the Cyanobacterium Anacystis nidulans following a light shift from white to red light. **Plant Physiology**, v. 78, n. 1, p. 110–114, 1985.

LOVELAND, T. R.; IRONS, J. R. Landsat 8: the plans, the reality, and the legacy. **Remote Sensing of Environment**, v. 185, p. 1–6, 2016.

LUNDY, L.; WADE, R. Integrating sciences to sustain urban ecosystem services. **Progress in Physical Geography**, v. 35, n. 5, p. 653–669, 2011.

MACIEL, D. A. **Quantificação remota da concentração de sólidos totais e inorgânicos em suspensão em lagos da planície de inundação do baixo Amazonas - Uma abordagem Multi-Sensor**. 2019. 196 p. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2019.

MACIEL, D. A.; BARBOSA, C. C. F.; NOVO, E. M. L. DE M.; FLORES JÚNIOR, R. Water clarity in brazilian water assessed using Sentinel-2 and machine leaning methods. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 182, p. 134–152, 2021.

MACIEL, D.; NOVO, E. M. L. DE M.; CARVALHO, L. S.; BARBOSA, C. C. F.; FLORES JÚNIOR, R.; LOBO, F. L. Retrieving total and inorganic suspended sediments in Amazon floodplain lakes: a multisensor approach. **Remote Sensing**, v. 11, n. 15, p. 1–33, 2019.

MARTINS, V. S.; BARBOSA, C. C. F.; DE CARVALHO, L. A. S.; JORGE, D. S. F.; LOBO, F. L.; DE MORAES NOVO, E. M. L. Assessment of atmospheric correction methods for sentinel-2 MSI images applied to Amazon floodplain lakes. **Remote Sensing**, v. 9, n. 4, 2017.

MCGRANE, S. J. Impacts of urbanisation on hydrological and water quality dynamics, and urban water management: a review. **Hydrological Sciences Journal**, v. 61, n. 13, p. 2295–2311, 2016.

MIAO, D.; DING, W. L.; ZHAO, B. Q.; LU, L.; XU, Q. Z.; SCHEER, H.; ZHAO, K. H. Adapting photosynthesis to the near-infrared: Non-covalent binding of phycocyanobilin provides an extreme spectral red-shift to phycobilisome core-membrane linker from Synechococcus sp. PCC7335. **Biochimica et Biophysica Acta - Bioenergetics**, v. 1857, n. 6, p. 688–694, 2016.

MILZ, B.; AQUINO, P. O.; ORTEGA, J. C. G.; BITENCOURT, A. L. V.; NORDI, C. S. F. Spatio-temporal variability of water quality in Billings reservoir central body - São Paulo,Brazil. **Revista Ambiente e Agua**, v. 17, n. 3, p. 16, 2022.

MISHRA, D. R.; OGASHAWARA, I.; GITELSON, A. A. **Bio-optical modeling and remote sensing of inland waters**. Amsterdam: Elsevier, 2017. 1–316 p. ISBN(9780128046548).

MISHRA, S.; MISHRA, D. R. A novel remote sensing algorithm to quantify phycocyanin in cyanobacterial algal blooms. **Environmental Research Letters**, v. 9, n. 11, 2014.

MISHRA, S.; MISHRA, D. R.; LEE, Z. Bio-optical inversion in highly turbid and cyanobacteria-dominated waters. **IEEE Transactions on Geoscience and Remote Sensing**, v. 52, n. 1, p. 375–388, 2013.

MISHRA, S.; MISHRA, D. R.; LEE, Z.; TUCKER, C. S. Quantifying cyanobacterial phycocyanin concentration in turbid productive waters: A quasi-analytical approach. **Remote Sensing of Environment**, v. 133, p. 141–151, 2013.

MOBLEY, C. D. **Light and water: radiative trasnfer in natural waters**. Cambridge: Academic Press, 1994. 592 p.

MOBLEY, C. D. Estimation of the remote-sensing reflectance from above-surface measurements. **Applied Optics**, v. 38, n. 36, p. 7742–7455, 1999.

MOBLEY, C. D. Polarized reflectance and transmittance properties of windblown sea surfaces. **Applied Optics**, v. 54, n. 15, p. 4828, 2015.

MOBLEY, C. D.; STRAMSKI, D.; PAUL BISSETT, W.; BOSS, E. Optical modeling of ocean waters: is the case 1 - case 2 classification still useful? **Oceanography**, v. 17,

p. 60–67, 2004.

MOORE, K. A.; ALTUS, S.; TAY, J. W.; MEEHL, J. B.; JOHNSON, E. B.; BORTZ, D. M.; CAMERON, J. C. Mechanical regulation of photosynthesis in cyanobacteria. **Nature Microbiology**, v. 5, n. 5, p. 757–767, 2020.

MOREL, A. Bio-optical models. **Encyclopedia of Ocean Sciences**, p. 317–326, 2001.

MOREL, A.; PRIEUR, L. Analysis of variations in ocean color. **Limnology and Oceanography**, v. 22, n. 4, p. 709–722, 1977.

MORLEY, S. K.; BRITO, T. V.; WELLING, D. T. Measures of model performance based on the log accuracy ratio. **Space Weather**, v. 16, n. 1, p. 69–88, 2018.

MOSCHINI-CARLOS, V.; BORTOLI, S.; PINTO, E.; NISHIMURA, P. Y.; DE FREITAS, L. G.; POMPÊO, M. L. M.; DÖRR, F. Cyanobacteria and cyanotoxin in the billings reservoir (São Paulo, SP, Brazil). **Limnetica**, v. 28, n. 2, p. 273–282, 2009.

MOSCHINI-CARLOS, V.; FREITAS, L. G.; POMPÊO, M. Limnological evaluation of water in the Rio Grande and Taquacetuba branches of the Billings Complex (São Paulo, Brazil) and management implications. **Revista Ambiente e Agua**, v. 5, n. 3, p. 47–59, 2010.

MOSES, W. J.; BOWLES, J. H.; LUCKE, R. L.; CORSON, M. R. Impact of signal-to-noise ratio in a hyperspectral sensor on the accuracy of biophysical parameter estimation in case II waters. **Optics Express**, v. 20, n. 4, p. 4309, 2012.

MOSES, W. J.; GITELSON, A. A.; BERDNIKOV, S.; BOWLES, J. H.; POVAZHNYI, V.; SAPRYGIN, V.; WAGNER, E. J.; PATTERSON, K. W. HICO-based NIR-red models for estimating chlorophyll-a concentration in productive coastal waters. **IEEE Geoscience and Remote Sensing Letters**, v. 11, n. 6, p. 1111–1115, 2014.

NEGRI, A. P.; JONES, G. J. Bioaccumulation of paralytic shellfish poisoning (PSP) toxins from the cyanobacterium Anabaena circinalis by the freshwater mussel Alathyria condola. **Toxicon**, v. 33, n. 5, p. 667–678, 1995.

NIROUMAND-JADIDI, M.; BOVOLO, F.; BRUZZONE, L. Water Quality Retrieval from PRISMA Hyperspectral Images : First Experience in a Turbid Lake and Comparison with Sentinel-2. **Remote Sensing**, v. 12, 2020.

O'SHEA, R. E.; PAHLEVAN, N.; SMITH, B.; BRESCIANI, M.; EGERTON, T.; GIARDINO, C.; LI, L.; MOORE, T.; RUIZ-VERDU, A.; RUBERG, S.; STEFAN, G.; SIMIS, H.; STUMPF, R.; VAI, D. Advancing cyanobacteria biomass estimation from hyperspectral observations : Demonstrations with HICO and PRISMA imagery. **Remote Sensing of Environment**, v. 266, 2021.

OGASHAWARA, I. Terminology and classification of bio-optical algorithms. **Remote Sensing Letters**, v. 6, n. 8, p. 613–617, 2015.

OGASHAWARA, I. The use of sentinel-3 imagery to monitor cyanobacterial blooms. **Environments - MDPI**, v. 6, n. 6, 2019.

OGASHAWARA, I.; LI, L. Removal of chlorophyll-a spectral interference for improved phycocyanin estimation from remote sensing reflectance. **Remote Sensing**, v. 11, n. 15, p. 19, 2019.

OGASHAWARA, I.; LI, L.; HOWARD, C.; DRUSCHEL, G. K. Monitoring phycocyanin with Landsat 8/operational land imager orange contra-band. **Environments**, v. 9, n. 3, p. 40, 2022.

OGASHAWARA, I.; MISHRA, D. R.; MISHRA, S.; CURTARELLI, M. P.; STECH, J. L. A performance review of reflectance based algorithms for predicting phycocyanin concentrations in inland waters. **Remote Sensing**, v. 5, p. 4774–4798, 2013.

OLIVER, S.; CORBURN, J.; RIBEIRO, H. Challenges regarding water quality of eutrophic reservoirs in urban landscapes: a mapping literature review. **International Journal of Environmental Research and Public Health**, v. 16, n. 1, 2019.

PAERL, H. W.; HUISMAN, J. Climate change: a catalyst for global expansion of harmful cyanobacterial blooms. **Environmental Microbiology Reports**, v. 1, n. 1, p. 27–37, 2009.

PAHLEVAN, N.; SMITH, B.; SCHALLES, J.; BINDING, C.; CAO, Z.; MA, R.; ALIKAS, K.; KANGRO, K.; GURLIN, D.; HÀ, N.; MATSUSHITA, B.; MOSES, W.; GREB, S.; LEHMANN, M. K.; ONDRUSEK, M.; OPPELT, N.; STUMPF, R. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: a machine-learning approach. **Remote Sensing of Environment**, p. 111604, fev. 2020.

PALMER, S. C. J.; KUTSER, T.; HUNTER, P. D. Remote sensing of inland waters: challenges, progress and future directions. **Remote Sensing of Environment**, v. 157, p. 1–8, 2015.

PAULINO, R. S.; MARTINS, V. S.; NOVO, E. M. L. M.; BARBOSA, C. C. F.; CARVALHO, L. A. S.; BEGLIOMINI, F. N. Assessment of adjacency correction over inland waters using Sentinel-2 MSI images. **Remote Sensing**, v. 14, n. 8, p. 30, 2022.

PIRES, D. A.; TUCCI, A.; DO CARMO CARVALHO, M.; LAMPARELLI, M. C. Water quality in four reservoirs of the metropolitan region of São Paulo, Brazil. **Acta Limnologica Brasiliensia**, v. 27, n. 4, p. 370–380, 2015.

POMPÊO, M.; MOSCHINI-CARLOS, V. **Reservatórios que abastecem São Paulo:** problemas e perspectivas. São Paulo: USP, 2020. v. 1 136 p. ISBN(9788585658823).

POPE, R. M.; FRY, E. S. Absorption spectrum (380–700 nm) of pure water. **Applied Optics**, v. 36, p. 8710–8723, 1997.

PYO, J. C.; DUAN, H.; BAEK, S.; KIM, M. S.; JEON, T.; KWON, Y. S.; LEE, H.; CHO, K. H. A convolutional neural network regression for quantifying cyanobacteria using hyperspectral imagery. **Remote Sensing of Environment**, v. 233, p. 111350, 2019.

QI, L.; HU, C.; DUAN, H.; CANNIZZARO, J.; MA, R. A novel MERIS algorithm to

derive cyanobacterial phycocyanin pigment concentrations in a eutrophic lake: theoretical basis and practical considerations. **Remote Sensing of Environment**, v. 154, p. 298–317, 2014.

QIANG, Y.; XINDONG, W. 10 Challenging problems in data mining research. **International Journal of Information Technology and Decision Making**, v. 5, n. 4, p. 597–604, 2006.

REINART, A.; KUTSER, T. Comparison of different satellite sensors in detecting cyanobacterial bloom events in the Baltic Sea. **Remote Sensing of Environment**, v. 102, n. 1–2, p. 74–85, 2006.

RIBEIRO, M. S. F.; TUCCI, A.; MATARAZZO, M. P.; VIANA-NIERO, C.; NORDI, C. S. F. Detection of cyanotoxin-producing genes in a eutrophic reservoir (Billings reservoir, São Paulo, Brazil). **Water**, v. 12, n. 903, p. 10–12, 2020.

ROY, S.; LLEWELLYN, C. A.; EGELAND, E. S.; JOHNSEN, G. **Phytoplankton pigments:** characterization, chemotaxonomy and applications in oceanography pigments. Cambridge: Cambridge University Press, 2011. 890 p.

RUIZ-VERDÚ, A.; SIMIS, S. G. H.; DE HOYOS, C.; GONS, H. J.; PEÑA-MARTÍNEZ, R. An evaluation of algorithms for the remote sensing of cyanobacterial biomass. **Remote Sensing of Environment**, v. 112, n. 11, p. 3996–4008, 2008.

SAGAN, V.; PETERSON, K. T.; MAIMAITIJIANG, M.; SIDIKE, P.; SLOAN, J.; GREELING, B. A.; MAALOUF, S.; ADAMS, C. Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. **Earth-Science Reviews**, v. 205, n. April, p. 103187, 2020.

SANTABARBARA, S.; REMELLI, W.; PETROVA, A. A.; CASAZZA, A. P. Influence of the wavelength of excitation and fluorescence emission detection on the estimation of fluorescence-based physiological parameters in different classes of photosynthetic organisms. In: GRIGORYEVA, N. (Ed.). **Fluorescence methods for investigation of living cells and microorganisms**. London: IntechOpen, 2020. p. 28.

SÃO PAULO. SECRETARIA DO MEIO AMBIENTE (SMA). **Billings**. São Paulo: SMA, 2010. 156 p. ISBN(9788562251047).

SARADA, R.; PILLAI, M. G.; RAVISHANKAR, G. A. Phycocyanin from Spirulina sp: influence of processing of biomass on phycocyanin yield , analysis of efficacy of extraction methods and stability studies on phycocyanin. **Process Biochemistry**, v. 34, p. 795–801, 1999.

SCHALLES, J. F.; YACOBI, Y. Z. Remote detection and seasonal patterns of phycocyanin, carotenoid and chlorophyll pigments in eutrophic waters. **Ergebnisse Der Limnologie**, v. 55, n. October, p. 153–168, 2000.

SCHINDLER, D. W. Recent advances in the understanding and management of eutrophication. **Limnology and Oceanography**, v. 51, p. 356–363, 2006.

SCHOLES, L.; FAULKNER, H.; TAPSELL, S. Urban rivers as pollutant sinks and sources: a public health concern for recreational river users? **Water, Air, & Soil Pollution: Focus**, v. 8, p. 543–553, 2008.

SEEGERS, B. N.; STUMPF, R. P.; SCHAEFFER, B. A.; LOFTIN, K. A.; WERDELL, P. J. Performance metrics for the assessment of satellite data products: an ocean color case study. **Optics Express**, v. 26, n. 6, p. 7404, 2018.

SHI, K.; ZHANG, Y.; QIN, B.; ZHOU, B. Remote sensing of cyanobacterial blooms in inland waters: present knowledge and future challenges. **Science Bulletin**, v. 64, n. 20, p. 1540–1556, 2019.

SIEFERMANN-HARMS, D. The light-harvesting and protective functions of carotenoids in photosynthetic membranes. **Physiologia Plantarum**, v. 69, n. 3, p. 561–568, 1987.

SIEGEL, D. A.; WANG, M.; MARITORENA, S.; ROBINSON, W. Atmospheric correction of satellite ocean color imagery: the black pixel assumption. **Applied Optics**, v. 39, n. 21, p. 3582, 2000.

SILVA, S. C.; NISHIMURA, P. Y.; PADIAL, P. R.; MARIANI, C. F.; POMPÊO, M. L. M. Compartimentalização e qualidade da água: o caso da Represa Billings. **Bioikos**, v. 28, n. 1, p. 31–43, 2014.

SIMIS, S. G. H.; PETERS, S. W. M.; GONS, H. J. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. **Limnology and Oceanography**, v. 50, n. 1, p. 237–245, 2005.

SIMIS, S. G. H.; RUIZ-VERDÚ, A.; DOMÍNGUEZ-GÓMEZ, J. A.; PEÑA-MARTINEZ, R.; PETERS, S. W. M.; GONS, H. J. Influence of phytoplankton pigment composition on remote sensing of cyanobacterial biomass. **Remote Sensing of Environment**, v. 106, n. 4, p. 414–427, 2007.

ŚLIWIŃSKA-WILCZEWSKA, S.; CIESZYŃSKA, A.; KONIK, M.; MACULEWICZ, J.; LATAŁA, A. Environmental drivers of bloom-forming cyanobacteria in the Baltic Sea: Effects of salinity, temperature, and irradiance. **Estuarine, Coastal and Shelf Science**, v. 219, n. June 2018, p. 139–150, 2019.

SØNDERGAARD, M.; JENSEN, J. P.; JEPPESEN, E. Role of sediment and internal loading of phosphorus in shallow lakes. **Hydrobiologia**, v. 506–509, p. 135–145, 2003.

SPYRAKOS, E.; O'DONNELL, R.; HUNTER, P. D.; MILLER, C.; SCOTT, M.; SIMIS, S. G. H.; NEIL, C.; BARBOSA, C. C. F.; BINDING, C. E.; BRADT, S.; BRESCIANI, M.; DALL'OLMO, G.; GIARDINO, C.; GITELSON, A. A.; KUTSER, T.; LI, L.; MATSUSHITA, B.; MARTINEZ-VICENTE, V.; MATTHEWS, M. W.; OGASHAWARA, I.; RUIZ-VERDÚ, A.; SCHALLES, J. F.; TEBBS, E.; ZHANG, Y.; TYLER, A. N. Optical types of inland and coastal waters. **Limnology and Oceanography**, v. 63, n. 2, p. 846–870, 2017.

SRIVASTAVA, A.; SINGH, S.; AHN, C.-Y.; OH, H.-M.; ASTHANA, R. K. Monitoring approaches for a toxic cyanobacterial bloom. **Environmental Science and**

**Technology**, v. 47, n. 16, p. 8999–9013, 2013.

STACHELEK, J.; MADDEN, C. J. Application of inverse path distance weighting for high-density spatial mapping of coastal water quality patterns. **International Journal of Geographical Information Science**, v. 29, n. 7, p. 1240–1250, 2015.

SUN, D.; LI, Y.; WANG, Q.; LE, C.; LV, H.; HUANG, C.; GONG, S. A novel support vector regression model to estimate the phycocyanin concentration in turbid inland waters from hyperspectral reflectance. **Hydrobiologia**, v. 680, n. 1, p. 199–217, 2012.

TANDEAU DE MARSAC, N. Occurrence and nature of chromatic adaptation in cyanobacteria. **Journal of Bacteriology**, v. 130, n. 1, p. 82–91, 1977.

TAVARES, M. H.; LINS, R. C.; HARMEL, T.; FRAGOSO, C. R.; MARTÍNEZ, J. M.; MOTTA-MARQUES, D. Atmospheric and sunglint correction for retrieving chlorophyll-a in a productive tropical estuarine-lagoon system using Sentinel-2 MSI imagery. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 174, p. 215–236, 2021.

TORBICK, N.; CORBIERE, M. A multiscale mapping assessment of lake champlain cyanobacterial harmful algal blooms. **International Journal of Environmental Research and Public Health**, v. 12, n. 9, p. 11560–11578, 2015.

VANHELLEMONT, Q. Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives. **Remote Sensing of Environment**, v. 225, p. 175–192, 2019.

VANHELLEMONT, Q.; RUDDICK, K. Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications. **Remote Sensing of Environment**, v. 216, p. 586–597, 2018.

VANHELLEMONT, Q.; RUDDICK, K. Remote sensing of environment atmospheric correction of Sentinel-3 / OLCI data for mapping of suspended particulate matter and chlorophyll-a concentration in Belgian turbid coastal waters. **Remote Sensing of Environment**, v. 256, p. 112284, 2021.

VASCONCELOS, V. Eutrophication, toxic cyanobacteria and cyanotoxins: when ecosystems cry for help. **Limnetica**, v. 25, n. 1–2, p. 425–432, 2006.

VERLIČ, A.; DURIĆ, N.; KOKALJ, Ž.; MARSETIČ, A.; SIMONČIČ, P.; OŠTIR, K. Tree species classification using worldview-2 satellite images and laser scanning data in a natural urban forest. **Sumarski List**, v. 138, n. 9–10, p. 477–488, 2014.

VINCENT, R. K.; QIN, X.; MCKAY, R. M. L.; MINER, J.; CZAJKOWSKI, K.; SAVINO, J.; BRIDGEMAN, T. Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie. **Remote Sensing of Environment**, v. 89, n. 3, p. 381–392, 2004.

VISSER, P. M.; VERSPAGEN, J. M. H.; SANDRINI, G.; STAL, L. J.; MATTHIJS, H. C. P.; DAVIS, T. W.; PAERL, H. W.; HUISMAN, J. How rising CO2 and global warming may stimulate harmful cyanobacterial blooms. **Harmful Algae**, v. 54, p. 145–

159, 2016.

WAGNER, C.; ADRIAN, R. Cyanobacteria dominance: quantifying the effects of climate change. **Limnology and Oceanography**, v. 54, p. 2460–2468, 2009.

WANG, X.; GONG, Z.; PU, R. Estimation of chlorophyll a content in inland turbidity waters using WorldView-2 imagery: a case study of the Guanting Reservoir, Beijing, China. **Environmental Monitoring and Assessment**, v. 190, n. 10, 2018.

WARREN, M. A.; SIMIS, S. G. H.; MARTINEZ-VICENTE, V.; POSER, K.; BRESCIANI, M.; ALIKAS, K.; SPYRAKOS, E.; GIARDINO, C.; ANSPER, A. Assessment of atmospheric correction algorithms for the Sentinel-2A MultiSpectral Imager over coastal and inland waters. **Remote Sensing of Environment**, v. 225, p. 267–289, 2019.

WENGRAT, S.; BICUDO, D. C. Spatial evaluation of water quality in an urban reservoir (Billings Complex, southeastern Brazil). **Acta Limnologica Brasiliensia**, v. 23, n. 2, p. 200–216, 2011.

WILSON, K. L.; WONG, M. C.; DEVRED, E. Comparing Sentinel-2 and WorldView-3 imagery for coastal bottom habitat mapping in Atlantic Canada. **Remote Sensing**, v. 14, n. 5, 2022.

WOOD, R. Acute animal and human poisonings from cyanotoxin exposure: a review of the literature. **Environment International**, v. 91, p. 276–282, 2016.

WULDER, M. A.; LOVELAND, T. R.; ROY, D. P.; CRAWFORD, C. J.; MASEK, J. G.; WOODCOCK, C. E.; ALLEN, R. G.; ANDERSON, M. C.; BELWARD, A. S.; COHEN, W. B.; DWYER, J.; ERB, A.; GAO, F.; GRIFFITHS, P.; HELDER, D.; HERMOSILLA, T.; HIPPLE, J. D.; HOSTERT, P.; HUGHES, M. J.; HUNTINGTON, J.; JOHNSON, D. M.; KENNEDY, R.; KILIC, A.; LI, Z.; LYMBURNER, L.; MCCORKEL, J.; PAHLEVAN, N.; SCAMBOS, T. A.; SCHAAF, C.; SCHOTT, J. R.; SHENG, Y.; STOREY, J.; VERMOTE, E.; VOGELMANN, J.; WHITE, J. C.; WYNNE, R. H.; ZHU, Z. Current status of Landsat program, science, and applications. **Remote Sensing of Environment**, v. 225, p. 127–147, 2019.

WYATT, T. Margalef's mandala and phytoplankton bloom strategies. **Deep-Sea Research II**, v. 101, p. 32–49, 2014.

YAN, Y.; BAO, Z.; SHAO, J. Phycocyanin concentration retrieval in inland waters: a comparative review of the remote sensing techniques and algorithms. **Journal of Great Lakes Research**, v. 44, n. 4, p. 748–755, 2018.

YIM, I.; SHIN, J.; LEE, H.; PARK, S.; NAM, G.; KANG, T.; CHO, K. H.; CHA, Y. K. Deep learning-based retrieval of cyanobacteria pigment in inland water for in-situ and airborne hyperspectral data. **Ecological Indicators**, v. 110, 2020.

ZANEVELD, J. R. V. Remotely sensed reflectance and its dependence on vertical structure: a theoretical derivation. **Applied Optics**, v. 21, n. 22, p. 4146–4150, 1982.

ZHANG, Y.; YIN, Y.; WANG, M.; LIU, X. Effect of phytoplankton community

composition and cell size on absorption properties in eutrophic shallow lakes: field and experimental evidence. **Optics Express**, v. 20, n. 11, 2012.

ZHAO, Q.; YU, L.; DU, Z.; PENG, D.; HAO, P.; ZHANG, Y. An overview of the applications of Earth observation satellite data: impacts and future trends. **Remote Sensing**, v. 14, n. 8, 2022.

ZIMBA, P. V. An improved phycobilin extraction method. **Harmful Algae**, v. 17, p. 35–39, 2012.