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

sid.inpe.br/mtc-m21d/2023/03.16.12.27-TDI

MAPPING OF CYANOBACTERIA IN THE PROMISSÃO RESERVOIR BY MULTISPECTRAL AND HYPERSPECTRAL IMAGES

Thainara Munhoz de Lima

Master's Dissertation of the Graduate Course in Remote Sensing, guided by Drs. Cláudio Clemente Faria Barbosa, and Cristina Souza Freire Nordi, approved in March 14, 2023.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/48NALQ2>

INPE
São José dos Campos
2023

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 INOVAÇÃO
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21d/2023/03.16.12.27-TDI

MAPPING OF CYANOBACTERIA IN THE PROMISSÃO RESERVOIR BY MULTISPECTRAL AND HYPERSPECTRAL IMAGES

Thainara Munhoz de Lima

Master's Dissertation of the Graduate Course in Remote Sensing, guided by Drs. Cláudio Clemente Faria Barbosa, and Cristina Souza Freire Nordi, approved in March 14, 2023.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/48NALQ2>

INPE
São José dos Campos
2023

Cataloging in Publication Data

Lima, Thainara Munhoz de.

L628m Mapping of cyanobacteria in the promissão reservoir by multispectral and hyperspectral images / Thainara Munhoz de Lima. – São José dos Campos : INPE, 2023.

xxviii + 96 p. ; (sid.inpe.br/mtc-m21d/2023/03.16.12.27-TDI)

Dissertation (Master in Remote Sensing) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2023.

Guiding : Drs. Cláudio Clemente Faria Barbosa e Dra. Cristina Souza Freire Nordi.

1. Harmful algal bloom. 2. Phycocyanin. 3. Inland water. 4. hybrid model. 5. OLCI. I.Title.

CDU 528.7:582.27

MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E INOVAÇÃO

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

DEFESA FINAL DE DISSERTAÇÃO DE THAINARA MUNHOZ DE LIMA
REG. 263871/2021, BANCA Nº 030/2023

No dia 14 de março de 2023, 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 Mestra em Sensoriamento Remoto, com a exigência de que o trabalho final a ser publicado deverá incorporar as correções sugeridas pela Banca Examinadora, com revisão pelo(s) orientador(es).

Título: "MAPPING OF CYANOBACTERIA IN THE PROMISSÃO RESERVOIR BY MULTISPECTRAL AND HYPERSPECTRAL IMAGES"

Membros da banca:

Dra. Evlyn Márcia Leão de Moraes Novo – Presidente – INPE
Dr. Cláudio Clemente Faria Barbosa – Orientador – INPE
Dra. Cristina Souza Freire Nordi – Orientadora – UNIFESP/ICAQF
Dra. Fernanda Sayuri Yoshino Watanabe – Membro Externo – FCT-UNESP/Departamento de Cartografia
Dra. Aurea Maria Ciotti – Membro Externo – USP/CEBIMAR
Dra. Claudia Giardino – Membro Externo – Nacional Research Council of Italy – CNR/IREA

Declaração de aprovação de Cláudia Giardino anexa ao processo.

Documento assinado eletronicamente por **Fernanda sayuri yoshino watanabe (E)**, **Usuário Externo**, em 22/03/2023, às 09:35 (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 22/03/2023, às 10:42 (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 **Evlyn Marcia Leão de Moraes Novo**, **Pesquisador**, em 22/03/2023, às 11:24 (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 22/03/2023, às 14:04 (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 27/03/2023, às 17:39 (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 https://sei.mcti.gov.br/verifica.html, informando o código verificador **10879917** e o código CRC **C0D40B06**.

Referência: Processo nº 01340.001702/2023-17 SEI nº 10879917

“O conhecimento nos faz responsáveis”.

Che Guevara

To my parents, Sonia e Dedé, and my dogs,
Vinícius, Theodoro and Cristal. Love, Thainara.

ACKNOWLEDGEMENT

Agradeço inicialmente a Deus pela vida, saúde, e por ter me colocado em uma posição de enorme privilégio. Agradeço também aos meus pais. Mãe, obrigada por sempre acreditar, cuidar, e ser tão presente. Eu dedico essa dissertação à senhora! Pai, obrigada por me apoiar, por se orgulhar, e por ser sempre tão preocupado. Agradeço imensamente ao meu namorado e melhor amigo, Vitor Henrique. Meu bem, só você sabe o quão difícil foi chegar até aqui, obrigada pela paciência. Em seguida gostaria de agradecer aos meus orientadores Dr. Claudio Clemente Faria Barbosa e Dra. Cristina Nordi por tantos conhecimentos, direcionamentos e oportunidades. Dr. Claudio, obrigada por confiar tanto em mim. Aos meus amigos queridos que o LaBISA me presenteou. Dra. Evlyn Novo obrigada pelas sempre sábias palavras, a senhora para sempre vai ser uma grande inspiração para mim, como mulher e como pesquisadora. Dr. Vitor Martins, muito obrigada por sempre se mostrar presente quando precisei, e por sempre saber exatamente a coisa certa a se dizer. Ao meu precioso amigo Felipe Nincao, você foi uma luz ao longo desses anos, eu jamais vou conseguir encontrar as palavras certas para expressar a gratidão que eu sinto por você. Muito além do mestrado, você se tornou meu amigo para a vida toda. Daniel Maciel, obrigada por ser sempre tão solicito, prestativo e parceiro. Você é uma joia, e daquelas bem raras. Também agradeço aos meus amigos de mestrado, Rômulo, Guilherme, Diego, Allan, e todos os outros. Eu jamais vou esquecer as boas festas e momentos que tivemos juntos. Obrigada!

I also dedicate special thanks to the CNR-IREA, Dra. Claudia Giardino, Dr. Mariano Bresciani, Alice Fabbretto, Andrea Pelegrino, and Erika Piaser. Dra. Claudia and Dr. Mariano, thank you for receiving me with so much affection, for all the knowledge shared, and for the good times we lived together. Alice, Andrea, and Erika, I will always take you with me, no matter where I am. Thank you for being a family to me. Thanks!

Ofereço um agradecimento especial a todos os outros que não foram citados, mas que contribuíram neste trabalho! Eu jamais teria consigo finalizar essa pesquisa sem a ajuda de todos.

Por fim, agraço a Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP) que ofereceu recursos por meio do projeto (Processo N: 2021/02714-7), e pelos recursos oferecidos durante o estágio-pesquisa no exterior (Processo N: 2022/01107-2). Agradeço a Pousada Baobá pelos recursos concedidos durante as realizações das campanhas de campo. E por último, mas não menos importante, ao INPE, que me proporcionou toda a infraestrutura necessária para o desenvolvimento deste trabalho.

ABSTRACT

Timely monitoring, detecting, and quantifying cyanobacterial blooms are important for controlling public health risks and understanding aquatic ecosystem dynamics. In this context, Remote sensing emerges as an important tool in monitoring these ecosystems, providing information about water quality parameters based on bio-optical algorithms. Since phycocyanin is a unique pigment of inland water cyanobacteria, the quantification of its concentrations from satelliltes obervations has been largely assessed for water quality monitoring. This work aims to test the state-of-the-art algorithms for retrieving phycocyanin concentrations in a Brazilian reservoir using multispectral and hyperspectral images and concurrent *in-situ* data. In this research, two main analyses were performed: *(i)* development of a hybrid algorithm to retrieve PC concentration in optically dynamic waters uing OLCI imagery data; and *(ii)* comparison between the PC estimate from semi-analytical and machine learning models applied in hyperspectral PRISMA observations. In the first analysis *(i)*, a cyanobacteria detection algorithm was firstly applied in three atmospherically corrected OLCI image (L2-WFR, 6SV, and ACOLITE) to mask the cyano-dominant waters. For each class, a suitable bio-optical algorithm was then applied, composing the final hybrid model. The PC retrieval uncertainties showed that the hybrid model (ζ = 25.35%) is superior to the single algorithms calibrated for the entire range of optical properties. The final application of the hybrid model proved to be highly sensitive to variations in the spectrum shape, and less sensitive to spectrally neural offset in the corrected R_{rs}. In the second analysis *(ii),* the semi-analytical model provided low median absolute error (MdAE) for PRISMA-derived (MdAE = 3.06 mg.m^{-3}) and OLCI-derived (MdAE = 3.93 mg.m^{-3}) PC concentrations, while it overestimated PRISMA-derived Chl-a (MdAE = 42.11 mg.m^{-3}). The random forest model for PC applied to PRISMA performed slightly worse than the semi-analytical (MdAE = 5.21 mg.m^{-3}). The Mixture Density Network (MDN) model showed a rather different performances with higher errors for PC (MdAE = 40.94 mg.m^{-3}) and lower errors for Chl-a (MdAE = 23.21 mg.m^{-3}). The results overall suggest that model calibrated with

site-specific measurement perform better and indicate that semi-analytical models could be applied to PRISMA and OLCI for remote sensing of PC in the Brazilian reservoirs. In conclusion, both evaluated approaches were capable of estimating the pigment concentration with the required accuracy, exploring the advantages and disadvantages of the multispectral and hyperspectral sensors. Future investigations on the synergy between the two sensors may be important for monitoring cyanobacteria in small water bodies with dynamic optical characteristics.

Keywords: harmful algal bloom, phycocyanin, inland water, hybrid model, OLCI, PRISMA.

MAPEAMENTO DE CIANOBACTÉRIAS NO RESERVATÓRIO DE PROMISSÃO POR IMAGENS MULTISPECTRAIS E HIPERESPECTRAIS

RESUMO

O monitoramento, detecção e quantificação de florações de cianobactérias são importantes para controlar os riscos à saúde pública e entender as dinâmicas dos ecossistemas aquáticos. Nesse contexto, o Sensoriamento Remoto surge como uma importante ferramenta no monitoramento desses ecossistemas, fornecendo informações sobre parâmetros de qualidade da água com base em algoritmos bio-ópticos. Uma vez que a ficocianina é um pigmento único de cianobactérias, a quantificação de suas concentrações a partir de observações de satélites tem sido amplamente avaliada para monitoramento da qualidade da água. Este trabalho tem como objetivo testar os algoritmos de última geração para recuperação das concentrações de ficocianina em um reservatório brasileiro usando imagens multiespectrais e hiperespectrais e dados *in-situ* concorrentes. Nesta pesquisa, duas análises principais foram realizadas: *(i)* desenvolvimento de um algoritmo híbrido para estimar a concentração de ficocianina em águas opticamente dinâmicas usando dados de imagens OLCI; e *(ii)* comparação entre a estimativa de ficocianina a partir de modelos semi-analíticos e de *machine learning* aplicados em observações hiperespectrais PRISMA. Na primeira análise *(i)*, um algoritmo de detecção de cianobactérias foi primeiramente aplicado em três imagens OLCI corrigidas atmosfericamente (L2-WFR, 6SV e ACOLITE) para identificar pixels com a presença de florações de cianobacterias. Para cada classe, um algoritmo bio-óptico adequado foi então aplicado, compondo o modelo híbrido final. As incertezas de estimativa do pigmento mostraram que o modelo híbrido (ζ = 25,35%) é superior aos algoritmos simples calibrados para toda a faixa de propriedades ópticas. A aplicação final do modelo híbrido provou ser altamente sensível a variações na forma do espectro e menos sensível ao deslocamento espectralmente neutro da R_{rs} corrigida. Na segunda análise *(ii)*, o modelo semi-analítico forneceu erro absoluto médio baixo (MdAE) para a concentração de ficocianina estimada a partir de dados PRISMA (MdAE = 3,06 mg.m^{-3}) e de

imagens do sensor OLCI (MdAE = 3,93 $mg.m^{-3}$), enquanto superestimou a concentração de Chl-a obtida a partir de dados do satélite PRISMA (MdAE = 42,11 $mg.m^{-3}$). O modelo *random forest* para estimativa da ficocianina aplicado as iamgens PRISMA teve desempenho ligeiramente inferior ao semi-analítico (MdAE = 5,21 $mg.m^{-3}$). O modelo *Mixture Density Network* (MDN) apresentou desempenhos bastante diferentes com erros maiores para PC (MdAE = 40,94 $mg.m^{-3}$) e erros menores para Chl-a (MdAE = 23,21 $mg.m^{-3}$). Os resultados gerais sugerem que o modelo calibrado com dados *in-situ* do local apresentou melhor desempenho, indicando que modelos semi-analíticos podem ser aplicados as iamgens PRISMA e OLCI para sensoriamento remoto de PC nos reservatórios brasileiros. Em conclusão, ambas as abordagens avaliadas foram capazes de estimar a concentração de pigmento com a precisão necessária, explorando as vantagens e desvantagens dos sensores multiespectrais e hiperespectrais. Futuras investigações sobre a sinergia entre os dois sensores podem ser importantes para o monitoramento de cianobactérias em pequenos corpos d'água com características ópticas dinâmicas.

Palavras-Chave: florações de algas tóxicas, ficocianina, águas continentais, modelo híbrido, OLCI, PRISMA.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABREVIATIONS

6SV	Second Simulation of the Satellite Signal in the Solar Vector
AC	Atmospheric Correction
AOD	Aerosol Optical Depth
AOPs	Apparent Optical Properties
AOT	Aerosol Optical Thickness
ASI	*Agenzia Spaziale Italiana*
BAC	OLCI Baseline Algorithm
BIL	Billings Reservoir
BR	Band Ratio
BOA	Bottom-of-Atmosphere
CDOM	Coloured Dissolved Organic Matter
CHIME	Copernicus Hyperspectral Imaging Mission for the Environment
CHRIS	Compact High-Resolution Imaging Spectrometer
Chl-a	Chlorophyll-a
CI	Cyanobacteria Index
CNR	National Research Council of Italy
CODA	Copernicus Online Data Access
CyHABs	Cyanobateria Harmful Algae Bloom
DDV	Dense Dark Vegetation
DESIS	Earth Sensing Imaging Spectrometer
EnMAP	Environmental Mapping and Analysis Program
ESA	European Space Agency
FOV	Field-of-View
HAB	*Harmful Algae Bloom*
HED05	Glint correction model developed by Hedley et al. (2005)
HICO	Hyperspectral Imager for the Coastal Ocean
ISS	International Space Station
IOPs	Inherent Optical Properties
KUT09	Glint correction model developed by Kutser et al. (2009)
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
L1-FR	*OLCI top-of-atmosphere digital number product*
L2-WFR	*OLCI water leaving radiance Full Resolution*
LH	Line Height
LIU17	PC estimation algorithm developed by Liu et al. (2017)
LUT	Look-up-Tables
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MC	Monte Carlo
MDN	Mixture Density Network
MERIS	Medium Resolution Imaging Spectrometer
ML	Machine Learning
MM14	PC estimation algorithm developed by Mishra and Mishra (2014)
NAP	Non-Algal Particles
NASA	National Aeronautics and Space Administration
NI	Normalized Index
NIR	Near-Infrared region
OACs	Optically Actives Constituents
OGA19	PC estimation algorithm developed by Ogashawara and Li (2019)
OLCI	Ocean and Land Colour Imager
OZA	Observation Zenith Angles
OWT	Optical Water Types
PACE	Plankton, Aerosol, Cloud, Ocean Ecosystem
PC	Phycocyanin
PC:Chl-a	Phycocyanin to Chlorophyll-a ratio
PPS	Predictive Power Score
PRISMA	*PRescursore IperSpettrale della Missione Applicativa*
PROM	Promissão Reservoir
R^2	Coefficient of determination
RF	Random Forest
RMSE	Root Mean Squared Error

RTE	Radiative Transfer Equation
SA	Spectral Angle
SAM	Spectral Angle Mapper
SIM05	PC estimation algorithm developed by Simis, Peters and Gons (2005)
SHALOM	Israeli/Italian Spaceborne Hyperspectral Applicative Land and Ocean Mission
SNR	Signal to Noise Ratio
SRF	Spectral Response Function
SS	Spectral Shape
SY00	PC estimation algorithm developed by Schalles and Yacobi (2000)
SWIR	Short Wave Infrared Regions
TCSR	Tietê Cascade System Reservoir
T-R	Transmitance-Reflectance
TOA	Top-of-Atmosphere
TSS	Total Suspended Solids
VNIR	Visible and Near-Infrared Regions
WS07	Glint correction model developed by Wang and Shi (2007)

LIST OF SYMBOLS

$a(\lambda)$	Absorption coefficient
$a_{CDOM}(\lambda)$	Colored Dissolved Organic Matter absorption at wavelength λ (m^{-1})
$a_{Chla}(\lambda)$	Chlorophyll-*a* absorption coefficient at wavelength λ (m^{-1})
$a_{phy}(\lambda)$	Phytoplankton absorption coefficient at wavelength λ (m^{-1})
$a_w(\lambda)$	Pure Water absorption at wavelength λ (m^{-1})
$a_{PC}(\lambda)$	Phycocyanin absorption coefficient at wavelength λ (m^{-1})
$\alpha^*_{pc}(\lambda)$	C-Phycocyanin specific absorption coefficient at wavelength λ
$b(\lambda)$	Backscattering coefficient (m^{-1})
b_b	Backscattering coefficient
b_i	i-th simulated band from field radiometry
$\beta(\psi,\lambda)$	Volume scattering function (m^{-2} sr^{-2})
$c(\lambda)$	Attenuation coeficient (m^{-1})
DIF_{Rrs}	Absolute spectral difference between the sample and the median spectrum
$E_d(\lambda)$	Descending spectral irradiance (Wm^{-2})
$E_s(\lambda)$	Downward Irradiance (Wm^{-2})
$E_u(\lambda)$	Upwelling spectral irradiance (Wm^{-2})
F_p	False positives
F_n	False negative
i	i-th sample
j	j-th sampling station
$K_d\ (z,\lambda)$	Diffuse attenuation coefficient (m^{-1})
$L_{sky}(\lambda)$	Sky Radiance (W m^{-2} sr^{-1} nm^{-1})
$L_t(\lambda)$	Total water-leaving radiance (W m^{-2} sr^{-1} nm^{-1})
$L_w(\theta,\phi,\lambda)$	Water-leaving radiance (W m^{-2} sr^{-1} nm^{-1})
m	Meter
N	Number of samples
$[PC]$	Phycocyanin concentration (µg L^{-1})
$R(z,\lambda)$	Reflectance Factor just below the water surface at wavelength

$R_{sr}(\theta, \phi, \lambda)$	Remote Sensing Reflectance at wavelength λ (sr^{-1})
$R_{rs_{m(\lambda)}}$	Median Remote Sensing Reflectance measured spectrum (sr^{-1})
$R_{rs_{sim}}(B_i)$	Simulated target sensor band (sr^{-1})
T_p	True positives
T_n	True negatives
x_i	in-situ R_{rs}
y_i	AC-glint corrected R_{rs}
β	Symmetric signed percentage bias
δ	Correction factor used in the PC estimation algorithm developed by Simis, Peters and Gons (2005)
γ	Correction factor used in the PC estimation algorithm developed by Simis, Peters and Gons (2005)
ε	Conversion factor from Chlorophyll-*a* absorption coefficient at 665 nm to Chlorophyll-*a* absorption coefficient at 665 nm
ζ	Median symmetric accuracy
ρ	Water-surface reflected factor used to correct the surface radiance reflection or glint effect
θ	Zenithal angle (Degrees)
ϕ	Azimuthal angle (degrees)
λ	Wavelength (nm)
ψ	Scattering angle

CONTENTS

1 INTRODUCTION

1.1 Overview

Continental aquatic systems provide critical and diverse habitats for a large number and diversity of species and ecosystem services, indispensable to support the maintenance of biodiversity (TONG et al., 2017). Access to clean water is one of the 17 Sustainable Development Goals of the 2030 Agenda elaborated and agreed by the member countries of the United Nations, to which Brazil is a signatory (MIOLA; SCHILTZ, 2019). However, despite the importance of continental aquatic systems to sustainable development, human activities have accelerated the eutrophication process, as a result of the excessive accumulation of nutrients in the water (e.g., phosphorus and nitrogen), particularly from agricultural lands through non-point source runoff (AZEVEDO; VASCONCELOS, 2006; XIA et al., 2016; OWODUNNI et al., 2003). The eutrophication of continental water bodies has become a global environmental challenge, due to its negative ecological and economic consequences related to the occurrence of Cyanobacterial Harmful Algae Blooms (CyHABs) (RANDOLF et al., 2008; VEERABADHRAN et al., 2023).

Cyanobacteria are primitive oxygen-evolving gram-negative photosynthetic prokaryotes inhabiting water, terrestrial, and epiphytic environments (RAI et al., 2018). The proliferation of cyanobacteria represents a major threat to the use of aquatic ecosystems for irrigation, fishing, and recreational purposes, in addition to a significant increase in the cost of obtaining drinking water (TUNDISI, 2008). The cyanobacteria blooms give color, odor, and different flavors to the water, causing a series of problems for public supply (BORTOLI; PINTO, 2015). Furthermore, cyanobacteria are capable of producing toxins (cyanotoxins) that can lead to a wide range of health risks when ingested by humans, fish, and birds (CARMICHAEL, 2001; MEREL, et al., 2013; SUN ET AL, 2010). A wide range of health risks has been associated with human exposure to cyanobacteria, including skin and eye irritation, adverse effects on liver and kidney function, and flu-like symptoms, including headache, nausea, diarrhea, and vomiting (PAERL; HUISMAN, 2009).

In association with in situ data, orbital remote sensing has been an effective tool in the studies of these algal communities (DEKKER, 1993; HUNTER et al., 2009; LI et al., 2012). Although satellite technology cannot currently be used to detect toxins (STUMPF et al., 2016), it can be used effectively to identify CyHABs, and to quantify cyanobacterial abundance. The use of remote sensing provides advantages to observe spatio-temporal dynamics, since it provides a synoptic and high temporal resolution view of a feature of interest, providing the basis for reliable and accountable scientific understanding and knowledge of the cyanobacterial dynamics (SHI et al., 2019; COFFER et al., 2020).

An indicator for the presence of CyHABs using orbital data is phycocyanin (PC) (AHN et al., 2007). PC consists of an auxiliary pigment that, unlike chlorophyll-a, is only present in considerable concentrations in cyanobacteria, being an assertive indicator for the presence of CyHABs (AHN et al., 2007; OGASHAWARA, 2020). Due to PC specific absorption peak at 620 nm, remote sensing of freshwater cyanobacteria has been largely focused on developing bio-optical algorithms to estimate PC concentration (DEKKER, 1993; SCHALLES; YACOBI, 2000; SIMIS; PETERS; GONS, 2005; SIMIS et al., 2007). Despite the advances, monitoring this pigment from space still faces some challenges. For instance, the specific absorption peak of the pigment (620 nm) is unusual in the spectral range of most multispectral sensors. Furthermore, the pigment concentration varies with environmental conditions (SIMIS et al., 2007). These factors require extensive validation of the pigment estimate to assess the quality of retrievals, particularly in optically complex waters.

In view of these topics, there is a great need for attention to the continuous monitoring of CyHABs, the early bloom alerts, and the study of cyanobacteria dynamics, especially in regions such as lakes, estuaries, and eutrophic reservoirs, where these species are commonly dominant (PAERL; HUISMAN, 2009). Regular and continuous CyHABs monitoring using orbital remote sensing is a key to assessing, understanding, and mitigating environmental changes, supporting informed decisions and evidence-based policies for efficient use of our planet’s resources (GIULIANI et al., 2017).

1.2 Objectives

The motivation of this study is to understand the limitations in the remote sensing estimated PC as an indicator for CyHABs. The PC spectral features in 620 nm are highly affected by other pigments, and may vary depending on specific environmental conditions. In view of these factors, changes in the dynamics of optically complex waters may impose challenges in applying multispectral sensors and developing bio-optical algorithms, highlighting the need for assessments to better understand how to predict and monitor CyHABs occurrence using orbital data.

In this context, two main objectives can be established in this work: *(i)* the development of a hybrid model to estimate PC concentration using Sentinel-3/OLCI (Ocean and Land Colour Instrument) images (Chapter 4); and *(ii)* assess different PC estimate approaches using hyperspectral PRISMA (PRecursore IperSpettrale della Missione Applicativa) satellite, comparing the hyperspectral and multispectral data performance (Chapter 5). Based on these objectives, the following research questions are addressed:

1. Is the hybrid approach feasible to better estimate the space-time variation of PC concentration in an optically complex reservoir? What are the challenges in monitoring CyHABs using remote sensing? (Chapter 4)
2. What are the gains of using hyperspectral data over multispectral data for predicting PC concentration? (Chapter 5)
3. Is the Machine Learning approach able to predict PC more accurately than the semi-analytical models? What are the difficulties in developing a global model for estimating PC concentration? (Chapter 5)

1.3 Outline

Chapter 1 introduces the negative environmental and health consequences of CyHABs, and how remote sensing has been used to assess cyanobacteria proliferation in inland waters.

Chapter 2 presents the theoretical background with topics including the cyanobacteria growing process, and the CyHABs estimates using remote

sensing. This chapter also describes the main fundamentals applied in estimating PC using orbital data, as well as the development of bio-optical models.

Chapter 3 describes the study area, and the dataset used. The dataset description is divided into *(i) in-situ* data, where the field campaigns, and the radiometric and limnological measurement protocols are described; and *(ii)* the satellite imagery data, which present the OLCI data collection used, and the pre-processing methodology applied, i.e., the atmospheric correction models.

Chapter 4 presents the development of a hybrid model to estimate the PC concentration in complex waters. The results also discuss about the performance of different atmospheric correction models in OLCI images.[1]

Chapter 5 describes the state-of-the-art algorithms to retrieve PC concentrations using machine learning and semi-analytical models from PRISMA hyperspectral images. An evaluation of different atmospheric correction and glint model approaches is also performed to estimate Remote sensing reflectance (R_{rs}) from PRISMA from Top-of-Atmosphere (TOA) radiances.[2]

Chapter 6 summarizes this study's finding and synthesizes them with respect to the general assessment of CyHABs using optical remote sensing.

[1] This chapter is an adapted version of the paper: LIMA et al. (2023). Cyanobacteria mapping using hybrid algorithms and Sentinel-3 (OLCI) data in a Brazilian Reservoir.

[2] This chapter is an adapted version of the paper: LIMA, T.M.A.; GIARDINO, C.; BRESCIANI, M.; BARBOSA, C.C.F.; BEGLIOMINI, F.N.; FABRETTO, A.; PELLEGRINO, A. (2023) Assessment of estimated phycocyanin concentration from PRISMA and Sentinel-3 (OLCI) in Brazilian inland waters: A comparison between semi-analytical and machine learning algorithms. **Remote Sensing**. (Under review).

2 BACKGROUND

2.1 Cyanobacteria Harmful Algal Bloom

Phytoplankton can be defined as the collective of photosynthetic microorganisms, adapted to live in the open sea, lakes (including reservoirs), estuaries, and rivers (DOKULIL; TEUBNER, 2000). The intense development of phytoplankton in aquatic environments is called Algae Bloom. It can be classified as harmful to an environment when they cause negative effects on aquatic biota, terrestrial animals, and humans (CARMICHAEL; BOYER, 2016). This growth, called Harmful Algae Blooms (HABs), is characterized by monospecific toxic-phytoplankton blooms that lead to poisonous seafood, mortality of fish and other animals, economic impacts, losses to aquaculture enterprises, and long-term ecosystem changes (CARMICHAEL; BOYER, 2016).

In addition to these detrimental effects, some HABs pose an additional threat due to the production of harmful toxins. The main toxic phytoplankton species are diatoms and cyanobacteria (VASCONCELOS, 2006; GELETU, 2023). Dinoflagellate species are not registered in freshwater ecosystems and Diatom's metabolites were not toxic for human being. On the other hand, cyanobacterial (or blue-green algae) blooms (Cyanobacteria Harmful Algae Bloom - CyHABs), occur most frequently in freshwater environments and have several impacts in human health (CARMICHAEL; BOYER, 2016). These algae present great floating capacity, a variety of pigments, and cyanotoxin production, which increase their ecological competition, ensuring successful development (DOKULIL; TEUBNER, 2000). Cyanotoxins are toxic compounds produced inside Cyanobacteria cells that are usually released into the environment after the cell death (CARMICHAEL, 2001). Some species are still capable of excreting those metabolites into the extracellular medium to prevent competition (DOKULIL; TEUBNER, 2000).

In general, the growth process of CyHABs is mainly a function of the increase in water temperature, light and nutrients availability, being directly associated with the eutrophication of the water body (PITOIS et al., 2000). The eutrophication of an environment consists of the process of nutrients enrichment in water bodies,

mainly phosphorus and nitrogen (HORN; GOLDMAN, 1994). The enrichment of nutrients in the water through the flow of urban and agricultural activities (pesticides and fertilizers), industrial waste and the inadequate management of watershed, causes an accelerated eutrophication that leads to the dominance of CyHABs (SCHINDLER, 2006).

Temperature increase is another factor that promotes CyHABs growth. Cyanobacteria typically thrive better at high temperatures than other phytoplankton species (PAERL; HUISMAN, 2009). Increasing the surface temperature of the water also strengthens the stratification process, reducing vertical mixing. Many cyanobacteria exploit this stratification condition by forming intracellular gas vesicles, which make the cells buoyant. Cyanobacteria float to the water surface can promote in dense surface blooms, creating scums (HUMPHRIES; LYNE, 1988). Scums formation shadowing non-floating phytoplankton, suppressing their opponents through light competition (PAERL; HUISMAN, 2008).

2.2 Estimative of CyHABs from orbital observations

Microalgae present a wide range of pigments responsible for photosynthesis. These photosynthetically active pigments are organized into discrete protein complexes that can be functionally divided into two groups: *(i)* photochemical reaction center containing Chlorophyll-a (Chl-a); *(ii)* light-gathering complexes that collect and transfer available light energy to the reaction center (e.g., carotenoids and phycobiliproteins) (LEE, 2008). Although all phytoplankton contain chlorophylls and carotenoids, only red, blue-green and cryptophytes algae contain phycobiliproteins. Based on the absorption properties, phycobiliproteins can be divided into: allophycocyanin – which absorbs in the range of 650-655 nm; phycocyanin – with maximum absorption at 610-620 nm; and phycoerythrin – which absorbs light at 495 nm and 540-570 nm (KIRK, 2011).

Among the phycobiliproteins, phycocyanin is produced only by Rhodophyta and Cryptophytes, being a major component only in Cyanophyta (C-Phycocyanin). Because it is correlated to cyanobacteria biomass, PC is potentially better than

Chl-a for cyanobacteria algae monitoring (SIMIS; PETERS; GONS, 2005; KIRK, 2011). The main spectral feature of PC (620 nm) forms the basis for detecting cyanobacteria by remote sensing (DEKKER, 1993; SIMIS et al., 2005).

2.2.1 Optical properties of inland waters

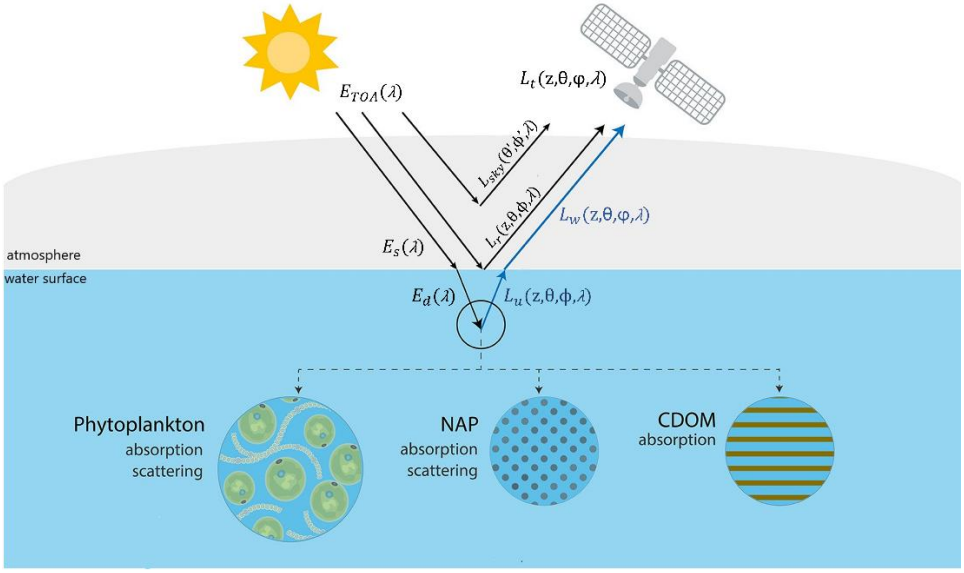
Remote sensing applied in aquatic environments is based on the processes of selective absorption and scattering of REM by water and its so-called optically active constituents (OAC). The OACs interact with the REM, modifying its magnitude and composition. The main OACs are: pure water; colored dissolved organic matter (CDOM); and the total particulate (TSS) that can be fractionated into microscopic chlorophyll organisms (phytoplankton) and non-algal particles (NAP) (inorganic and organic particles in suspension) (BARBOSA; NOVO; MARTINS, 2019). In addition to OACs, other important concepts in hydrological optics are the inherent optical properties (IOPs) and apparent optical properties (AOPs).

The IOPs (Figura 2.1) main characteristic is that their magnitudes depend only on the composition and concentration of OACs present in the water (MOBLEY, 1994). They are defined in terms of the absorption (a(λ), unit: m^{-1}) and scattering (b(λ), unit: m^{-1}) coefficients, and the volume scattering function (β(ψ,λ) , unit: $m^{-1}sr^{-1}$). The definition of these properties is made from an infinitesimal plane illuminated at right angles by a beam of monochromatic light, where part of the light, when crossing the plane, is absorbed, part scattered and the other part transmitted (MOBLEY, 1994). The fraction of incident flux that is absorbed and scattered, divided by the width of the plane, defines the absorption and scattering coefficients, respectively (KIRK, 2011).

The AOPs (Figure 2.1) are the properties in which the magnitude depends both on the characteristics of the medium and the geometric structure of the light field that crosses the water column (PREISENDORFER, 1961). Among the AOPs of interest can be highlighted: remote sensing spectral reflectance (R_{rs}(θ, ϕ, λ), unit: sr^{-1}), defined by the ratio of the radiance emerging from the water column (L_w(θ, ϕ, λ)) and the incident irradiance (E_s(λ)); the spectral irradiance reflectance (R(z, λ), unit: dimensionless), calculated by the ratio of the upwelling

spectral irradiance (E_u(λ)) and the descending spectral irradiance (E_d(λ)); and the diffuse attenuation coefficient (K_d(z, λ), unit: m-1), which determines the magnitude of E_d(λ) at a given depth z (BARBOSA; NOVO; MARTINS, 2019).

Figure 2.1 – Representation of the OACs and radiometric measurements.



Source: Adapted from Gupana et al. (2021).

Once the properties of the light field (AOPs) and the properties of the medium (IOPs) are defined, the relationship and integration between them are done through the application of the radiative transfer theory (RTE - Radiative Transfer Equation), which describes how radiance varies, in direction and magnitude, with depth (KIRK, 2011). RTE is used as the main basis for the development of bio-optical algorithms for estimating OACs.

2.2.2 Bio-optical modeling

The mathematical equations that describe the optical properties of a water body, relating them to the AOCs, are called bio-optical models. It is from the bio-optical modeling that the remote sensing data acquired over water bodies can be used to estimate the AOCs concentrations. The bio-optical models can be divided into three approaches: empirical, semi-empirical, and semi-analytical (MOBLEY, 1994). Empirical and semi-empirical algorithms are generally established from statistical regressions between *in-situ* OACs measurements

(e.g., PC concentrations) and radiometric data (SHI et al., 2019). One of the first semi-empirical algorithms to estimate PC concentrations was developed by Dekker (1993), and mainly exploits the unique absorption feature near 620 nm (DEKKER, 1993; SCHALLES; YACOBI, 2000). Despite the wide application, this algorithm does not account for the impact of Chl-a, which significantly reduces the accuracy of the estimate. Furthermore, considering the simplicity of the empirical and semi-empirical models, their applications are usually restricted to the specific locations and dates of the dataset collection, leading to a poor spatial transferability (GIARDINO et al., 2019).

More complex algorithms were later developed based on the derivation of direct and inverse relationships between AOPs, IOPs, and AOCs using the solution of the RTE (GORDON; BROWN; JACOBS, 1975). In the semi-analytical approach by direct modeling, the AOCs are related to the specific IOPs of each constituent, which are then related to the system's AOPs and the measured radiance. As in the scope of remote sensing the objective is obtaining the concentration of AOCs from the radiometric measurements, then inverse modeling is applied (Radiance →AOPs → IOPs → AOCs) (BARBOSA; NOVO; MARTINS, 2019). Simis et al. (2005), for example, addressed the spectral influence of Chl-a at 620 nm through adaptations of the Gons (1999) algorithm in a nested band semi-analytical algorithm. Compared to empirical and semi-empirical models, Simis algorithm presents the best performance in cyanobacteria-dominant eutrophic waters (RUIZ-VERDÚ et al., 2008; YAN; BAO; SHAO, 2018; RIDDICK et al., 2019).

Recently, another class of empirical models has been related to machine learning algorithms (GEWALI; MONTEIRO; SABER, 2018). Machine learning techniques can overcome some of the limitations of empirical, semi-empirical, and semi-analytical models by applying complex networks and structures to capture data-rich features available from multiple optical data (PYO et al., 2019). Several approaches, including deep neural network (CAO et al., 2020), convolutional neural network (PYO et al., 2019), mixture density network (O'SHEA et al., 2021), and random forest (BEGLIOMINI, 2022), have been

used to derive absorption coefficients, and Chl-a and PC concentrations (CAO et al., 2020).

Despite the advances in the PC bio-optical models, their application to optically complex waters is still challenging. In the presence of different phytoplankton groups, accessory chlorophyll pigments overlap with PC absorption at 620 nm. Chl-a, for example, presents an absorption shoulder around 623 nm (RUIZ-VERDÚ et al., 2008). The effect of these pigments in the PC absorption region present a strong influence in the PC estimates (DEKKER, 1993; MISHRA et al., 2009; SIMIS et al., 2005, 2007). Similarly, the absorption by other constituents (suspended non-algal particles, detritus, and CDOM) and water itself cannot be ignored as they can reach the same order of magnitude as phycocyanin absorption (SIMIS et al., 2007). This way, PC algorithms are hard to be adapted to the present satellite imaging sensor because other phytoplankton pigments mask the phycocyanin reflectance spectrum. Only those imagers that sense a narrow absorption feature near 620 nm are well suited for the direct estimation of PC.

Furthermore, compared to Chl-a, PC retrieval is less user-friendly for measurements or estimation due to the low specific absorption coefficient (a_{PC}^{*}) (YAN et al., 2019). Consequently, a higher concentration of PC is required to produce a R_{rs} signal equivalent to that of Chl-a. Finally, it is essential to consider the a_{PC}^{*} variability as a function of the cyanobacterial response to changes in environmental and lighting conditions (GROSSMAN et al., 1993). For example, the PC production is reduced when cyanobacteria are exposed to high illumination intensities (HOTOS, 2021). Once the phycobiliproteins main function is to harvest light beyond the Chl-a absorption spectrum, when the amount of energy captured by Chl-a is enough to maintain the cell's basal metabolism, it triggers the reduction of phycobilisomes (LONNEBORG et al., 1985).

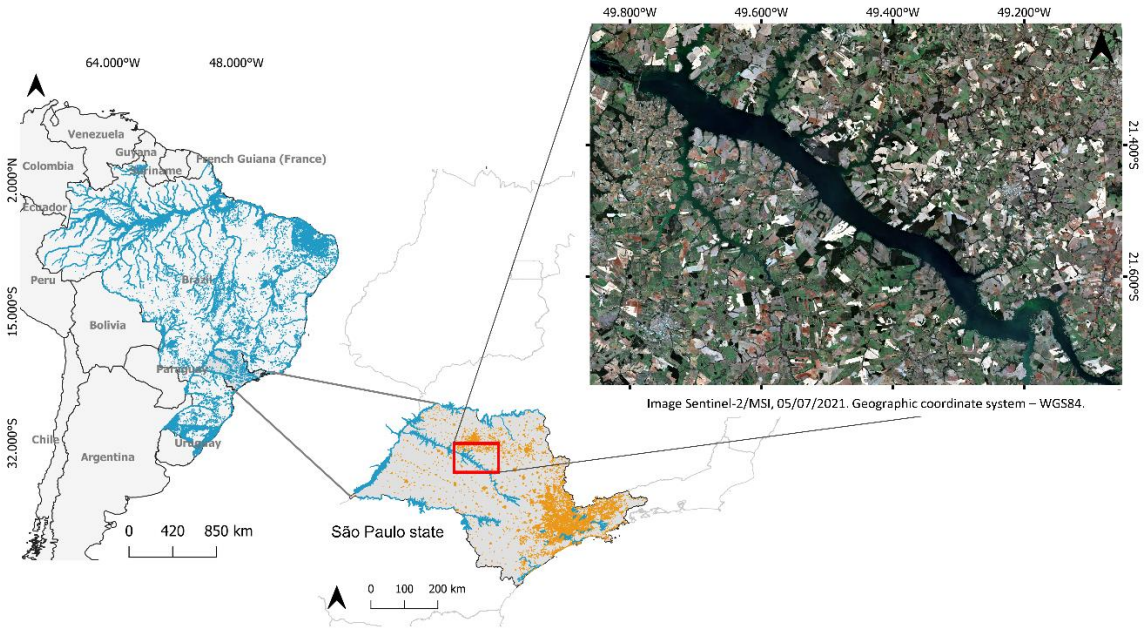
3 DATASET

3.1 Study area

Brazil is among the nine richest countries in freshwater, representing almost 12% of the world's available freshwater volume. Much of this reserved water is intended for electricity production, as the country's energy matrix is highly dependent on hydroelectricity (BARBOSA; NOVO; MARTINS, 2019). The main Brazilian hydroelectric reservoirs are in the Paraná River basin (57 reservoirs), which represents 6% of Brazilian water production. One of the main tributaries of the Paraná Basin is the Tietê River, which crosses the State of São Paulo. Considering its 1100 km of extension, the Tietê basin's Reservoirs' Cascade System (TCSR) includes six reservoirs, accumulating 29100 m^3 of water (BARBOSA; NOVO; MARTINS, 2019). The TCSR's waters are deeply affected by different non-natural contamination sources, such as discharges from pasture waste, wastewater from urban centers, and other agricultural activities developed in nearby areas, including sugarcane and citrus crops (BERNARDO et al., 2018).

The study area is Promissão Reservoir (PROM) (Figure 3.1), the fourth largest reservoir in the TCSR. Promissão Reservoir presents a flooded area of 560 km^2, 57,610 km^2 of drainage, and an average residence time of 134.1 days. Among its hydrological characteristics, the reservoir accumulates a useful volume of 2,128 km^3, and operates with a minimum and maximum water level of 379.7 m and 384 m, respectively (CESP, 1989).

Figure 3.1 - Study area – Promissão Reservoir.



Source: Author's elaboration.

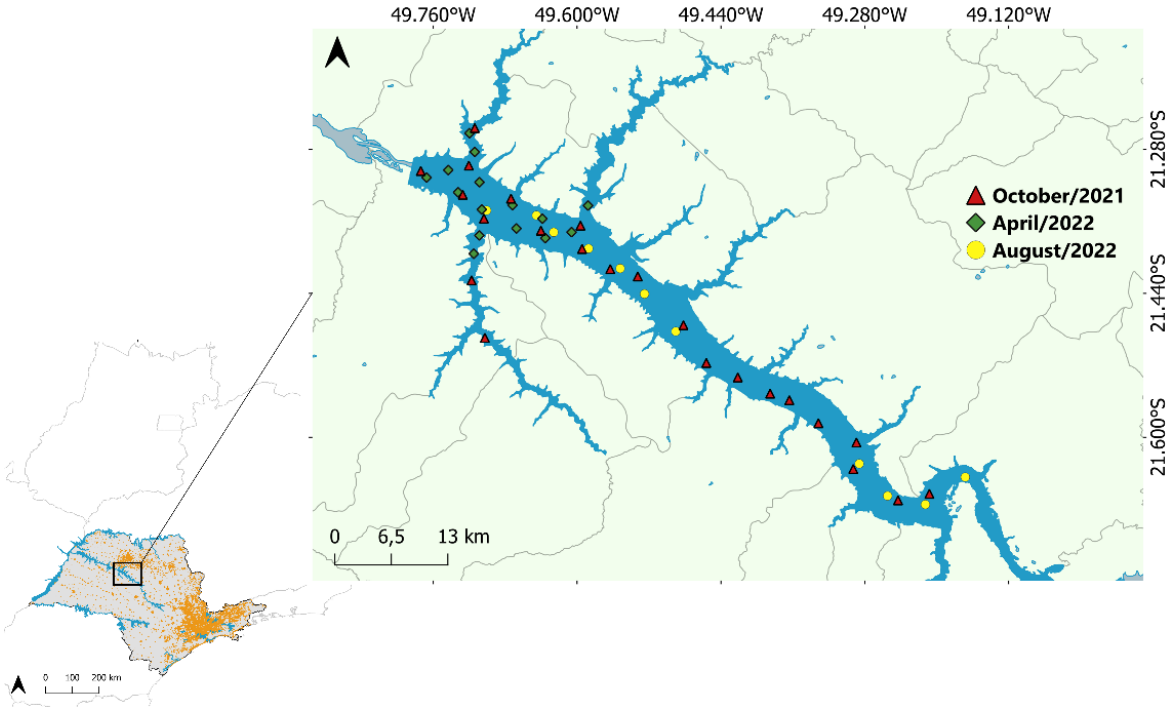
3.2 Field dataset

The entire dataset (N = 165 samples) consist of co-located *in-situ* R_{rs}, Chl-a, and PC measurements collected in two water bodies: Promissão and Billings Reservoir (BILL), both part of the TCSR. Although the object of study is the Promissão Reservoir, samples collected in the Billings Reservoir are also part of the input dataset in order to increase the variability in the pigment concentration, allowing a better generalization capacity. The dataset measured at the Billings Reservoir (here called as BIL) consists of 116 samples collected in 8 field campaigns between November 2020 and December 2021. The protocols used to measure radiometric and limnological data are the same as those used in Promissão Reservoir measurements, which will be detailed in the next sections. Details about the Billings dataset collection are described in Begliomini (2022).

The field dataset measured in the Promissão Reservoir was acquired in three field campaigns: 3rd – 9th October/2021 (PROM1), where 23 points were sampled (highlighted in the red circle in Figure 3.1); 11th April/2022 (PROM2), with 15 sampling points being collected, as indicated by the green symbols in

Figure 3.1; and 15th – 17th August/2022 (PROM3), where 11 samples were measured (yellow circle in Figure 3.1). The choice of periods for each campaign was based on meteorological conditions and the Sentinel-3 and PRISMA overpass, ensuring a match-up between *in-situ* measurements and orbital observations.

Figure 3.2 - Spatial distribution of the samples collected in the field campaigns carried out in the Promissão Reservoir.



Source: Author's elaboration.

3.2.1 Radiometric data

In-situ radiometric data were measured using three Trios-Ramses spectroradiometers, ranging from 400-900 nm, with a sampling interval of approximately 3.3 nm, positioned on top of the boat: two radiance radiometers with 7° Field-of-View (FOV), and one irradiance radiometer equipped with a cosine collector pointed up. The instruments were responsible for measuring the downward irradiance ($E_s(\lambda)$), the water-leaving radiance ($L_t(\theta,\phi,\lambda)$), and the sky radiance ($L_{sky}(\theta,\phi,\lambda)$). The geometry (azimuthal and zenith angles) of each sensor was pre-established following the protocol proposed by Mobley (1999) to minimize the specular reflection effects of direct and diffuse radiation.

The water color radiometry was defined in terms of the R_{rs}, which is calculated as the ratio of L_w (W.m^{-2}.sr^{-1}.nm^{-1}) and E_s (W.m^{-2}.nm^{-1}) (Section 2.2.1). However, L_w (Figure 2.1) cannot be directly measured due to the specular reflections on the water surface caused by diffuse solar irradiance (i.e., surface-reflected radiance) (JIANG; MATSUSHITA; YANG, 2020). To solve this problem, the L_t, which contains L_w plus the radiance reflected by the surface (L_r), is used to calculate R_{rs} through the following mathematical formulation (MOBLEY, 2015):

$$R_{rs}(\theta,\phi,\lambda) = \frac{L_w}{E_s} = \frac{L_t(\theta,\phi,\lambda) - L_r(\theta,\phi,\lambda)}{E_s} = \frac{L_t(\theta,\phi,\lambda) - \rho(\theta,\phi).L_{sky}(\theta',\phi,\lambda)}{E_s(\lambda)}, \quad (3.1)$$

where $\rho(\theta,\phi)$ is the water-surface reflected factor used to correct the surface radiance reflection or glint effect. The ρ values depend on the viewing geometry of the sensor, environmental conditions, and atmospheric conditions (MOBLEY, 1999).

During the radiometric measurements, variations in lightning and cloud cover result in variations in the intensity and quality of the measured radiance, leading to uncertainties in the model's application (MACIEL et al., 2019). In order to minimize this effect, for each sample point, approximately 150 spectra were measured, where the representative spectrum was chosen based on the following equation (MACIEL et al., 2019):

$$Dif_{Rrs(i)} = \sum_{\lambda=400}^{900} \left| R_{rs}(i,\lambda) - R_{rs_{median}}(\lambda) \right| \quad (3.2)$$

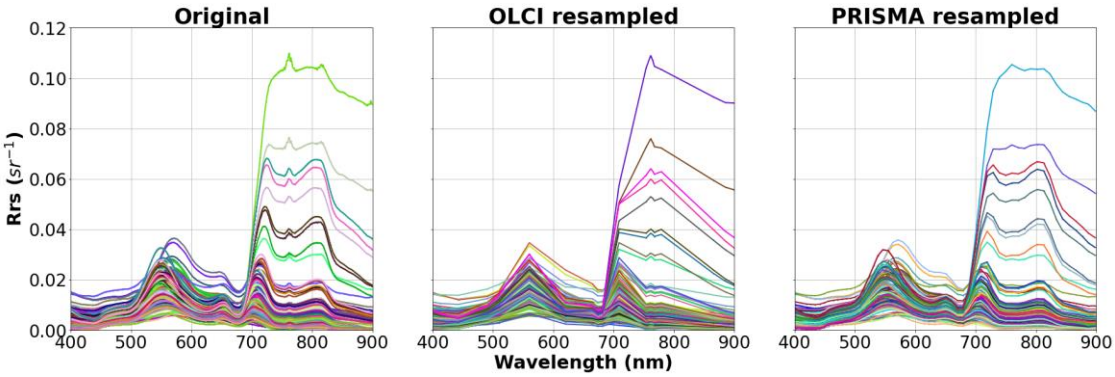
For each wavelength (λ), $R_{rs}(i,\lambda)$ is the sample R_{rs} and $R_{rs_{median}}(\lambda)$ is the median R_{rs} value of all samples. $Dif_{Rrs(i)}$ is the absolute spectral difference between an *i*-th sample and the median spectrum. The chosen spectrum is the one that reached the lowest value of $Dif_{Rrs(i,j)}$ in each sampling station.

To correct residual glint effects over the measured R_{rs} spectra, we applied the method of Kutser et al. (2013) with adaptations described by Cairo et al. (2020), in order to reduce the effect of spectra overcorrection, avoiding negative results (CAIRO et al., 2020). After the glint correction, each R_{rs} spectrum was used to simulate the spectral bands of OLCI sensor and PRISMA based on the spectral response function (SRF) of each sensor, applying the following equation (PELLOQUIN; NIEKE, 2012):

$$R_{rs_{sim}}(B_i) = \frac{\int_{\lambda_1}^{\lambda_2} SRF(\lambda).R_{rs_{m(\lambda)}} d\lambda}{\int_{\lambda_1}^{\lambda_2} SRF(\lambda)\, d\lambda}, \quad (3.3)$$

where $R_{rs_{sim}}(B_i)$ is the *i*-th simulated band for the target sensor and $R_{rs_{m(\lambda)}}$ is the *in-situ* measured R_{rs} for the wavelength range $[\lambda_1, \lambda_2]$ considered for each band. The original R_{rs} glint-corrected, and the respectively PRISMA and OLCI resampled R_{rs} are presented in Figure 3.3. The variability in the magnitude and shape of the R_{rs} spectra indicate waters dominated by phytoplankton, evidenced by the features near 620 nm and 665 nm driven by PC and Chl-a absorption, respectively. The reflectance peaks near 700 nm suggest material accumulation near the surface, masking the high absorption of pure water in the NIR.

Figure 3.3 – Original in-situ R_{rs} spectra, and OLCI and PRISMA resampled R_{rs}.



Author's elaboration.

3.2.2 Laboratory analysis

Water samples for limnological analyses (PC and Chl-a concentration, and phytoplankton absorption) were collected from the subsurface (≈ 0.2 m), 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. Then, the filters were placed into a 15 ml falcon tube and frozen at -80°C until analysis. The absorption by phytoplankton (a_{phy}) was measured according to Tassan and Ferrari (2002) method, using the Transmittance-Reflectance (T-R) technique. The T-R is based on the light transmittance and reflectance measurements of particles retained on the filters that are carried out using an integrating-sphere attached to a double-beam spectrophotometer. The data analysis is performed by a theoretical model that corrects for the effect of light backscattering by the particles. Based on the reflectance and transmittance values measured by the integrating sphere, the absorbance is calculated, from which the values of the absorption coefficients of suspended particles, detritus, and phytoplankton are obtained. Details about the protocol description can be found in Cairo (2015).

For Chl-a determination, the method described by APHA (1998) was applied. In the laboratory, the filters were homogenized, and the pigments were extracted using 90% acetone. After 12 hours at 4°C, the samples were centrifugated for

20 minutes (3000 rpm), and then read on a spectrophotometer. Subsequently, the Pheophytin-a correction was made, by acidifying the samples (HCl 0.1M) and reading it again in the spectrophotometer. Finally, the pigment concentration was calculated by Lorenzen's equation (Table 3.1) (APHA, 1998).

PC was quantified in the water samples following the method described by Sarada et al. (1999) and adapted by Horváth et al. (2013). Before carrying out the analysis, the filters were suspended in a phosphate buffer (100 mM and pH 7.2) and submitted to 3 freeze-thaw cycles (-80°C and 35°C). The samples were then sonicated (90 seconds at a frequency of 20 kHz) and centrifuged (30 minutes at 3000 rpm). PC concentration was calculated using Bennett and Bogorad (1973) equation according the spectrophotometric measurements (Table 3.1).

Table 3.1 - PC and Chla concentration ($mg.m^{-3}$), and PC:Chla ratio for each collected sample.

		Min	**Máx**	**Mean**	**Median**	**Std**
PROM1 (N = 23)	PC	0.33	136.39	10.01	3.00	27.20
	Chl-a	25.44	183.47	66.51	50.04	41.00
	PC:Chl-a	0.008	0.74	0.095	0.073	0.14
PROM2 (N = 15)	PC	1.36	65.89	26.98	18.16	19.23
	Chl-a	15.59	487.82	167.45	115.76	138.57
	PC:Chl-a	0.074	0.47	0.18	0.15	0.10
PROM3 (N = 11)	PC	5.50	123.61	22.80	11.49	32.74
	Chl-a	30.60	297.08	81.75	55.4	72.36
	PC:Chl-a	0.11	0.42	0.22	0.22	0.08
BIL (N = 116)	PC	0.1	301.81	20.74	2.92	46.78
	Chl-a	4.54	906.15	89.19	41.67	132.57
	PC:Chl-a	0.003	0.48	0.14	0.08	0.12

3.3 Satellite imagery data

3.3.1 Sentinel-3/OLCI

As part of the European Copernicus Programme, the OLCI sensor, on board of Sentinel-3A/3B, was designed to bring a wealth of ocean colour data at medium spatial resolution. Launched in 2016 (Sentinel-3A) and 2018 (Sentinel-3B), OLCI is a multispectral radiometer with 21 spectral bands (400–1020 nm) with

high signal-to-noise ratio (between 305 and 2188), and approximately 300 m of spatial resolution. The two satellites (3A and 3B) can provide a revisit time of fewer than two days at the equator, with higher overpass frequencies at higher latitudes. The push-broom imaging spectrometer of OLCI is tilted off-nadir in a westerly direction by 12.6°, in an attempt to mitigate the effects of sun glint. Within its 21 spectral bands, measurements in the band “Oa07” (central wavelength = 620 nm) allow the estimation of phycocyanin. OLCI products are divided into three main types: *(i)* Level-1B Product, which provides calibrated, ortho-geolocated and spatially re-gridded Top of Atmosphere (TOA) radiances; *(ii)* Level-2 Land Product containing land and atmospheric geophysical products; and *(iii)* Level-2 Water Product (L2-WFR) which provide water-leaving radiance, as well as water and atmospheric geophysical products (ESA, 2022).

In this study, we use L1B Full Resolution Product as input to the atmospheric correction models, and L2 Water-Full Resolution Product as the AC corrected water surface Reflectance. The L1B products were downloaded via Copernicus Online Data Access (CODA and CODArep for reprocessed data). Digital Numbers from the product were converted into TOA reflectance through the SNAP Sentinel-3 Toolbox (https://step.esa.int/main/download/snap-download/, accessed on November 10th 2022). L2-WFR baseline products were obtained from the CODA database hosted by EUMETSAT (coda.eumetsat.int).

3.3.2 PRISMA

PRISMA is an Earth Observation system funded by Italian Space Agency (ASI) and launched in March 2019. The satellite payload is composed of an imaging spectrometer able to capture images at 30 m spatial resolution in a continuum spectrum of 239 spectral bands between 400-2500 nm (66 in the Visible and Near-Infrared (VNIR), and 173 in the Short Wave Infrared (SWIR) spectrum) (GIARDINO et al., 2020). A panchromatic camera is also onboard of the satellite that provides a single band image at 5 m spatial resolution. The satellite revisit time in nadir-looking configuration is of 29 days. However, the system is capable of acquiring images distant 1000 km in a single pass (with a total rotation left to right side looking and vice versa), so that the temporal resolution

can be significantly improved (less than one week) (BRESCIANI et al., 2022). PRISMA acquires images on demand, in specific individual locations requested by the users (https://prisma.asi.it, accessed on 10 June 2022), in a “standard” mode, resulting in a 30x30 km scene and a “strip” mode, generating an image of 30 km width.

The image products are released with different levels of pre-processing. The Level 1 (L1) product is a Top-of-Atmosphere (TOA) radiance imagery organized in two radiometrically and geometrically calibrated hyperspectral and panchromatic radiances cubes. The Level 2B (L2B) contains geolocated and atmospherically corrected hypercube and panchromatic radiances images – bottom of atmosphere (BOA) radiance. Level 2C (L2C) contains geolocated and atmospherically corrected hypercube and panchromatic reflectance images, including aerosol optical thickness (AOT) and water vapor map. Level 2D (L2D) also contains similar images of L2C but geolocated and geocoded. More detailed information can be found in the PRISMA Product Specifications document (ASI, 2021).

L2C geolocated, and atmospherically corrected reflectance images, and L1 TOA radiance imagery products were considered in this study. L2C products were chosen over the updated ASI’s product (L2D) once the geolocation for the study area was displaced compared to a reference raster (Sentinel-2/MSI). The L1 product was used as input in different atmospheric correction methods, while L2C composed the set of surface reflectance products evaluated in this study. The images, distributed in a hierarchical data format (he5), were downloaded from the PRISMA portal and were first converted into a suitable format using the R package “prismaread” tool developed by the National Research Council of Italy (CNR). Further details on the prismaread tool can be accessed from https://irea-cnr-mi.github.io/prismaread (accessed on June 10th 2022).

3.3.3 Imagery pre-processing

3.3.3.1 Atmospheric correction

Although the high influence of the atmosphere on the signal is not a new concern, the correction of the atmospheric effects over continental water bodies still presents some challenges that impose limitations on the accuracy of atmospheric modeling and, consequently, on the estimation of the emergent reflectance of the water column (MOSES et al., 2017). Here, the satellite retrieved surface reflectance with *in-situ* data were compared to better understand how uncertainties in Atmospheric Correction (AC) models propagate in multispectral and hyperspectral satellite data. Following, the AC methods tested are presented:

(i) Standard PRISMA L2C processor: The Level 2 standard atmospheric correction processor is based on MODTRAN v6.0, using a multi-dimensional Look-up-Table (LUT) approach (BRAGA et al., 2022). This method uses the hyperspectral bands for deriving atmospheric parameters (e.g., water vapor and Aerosol Optical Depth (AOD)). The water vapor is retrieved pixel-by-pixel using the water's absorption features at NIR bands. The retrieval of PRISMA AOD is based on Dense Dark Vegetation (DDV) algorithm approach (OUAIDRARI; VERMOTE, 1999), exploiting the correlation between reflectance in the SWIR region, blue and red bands. An extended description of the algorithms used to generate PRISMA products is available on ASI (2021).

(ii) OLCI L2-WFR: The OLCI L2-Water contains the water leaving reflectance in 16 spectral bands, related to the R_{rs} dividing by π. The AC is applied by combining two approaches: the baseline AC algorithm (BAC), which involves the correction of glint and white caps effects, followed by the estimation of the aerosol type from the NIR bands at 779 and 865 nm. The Rayleigh contribution is determined from tabulated results of radiative transfer computations, corrected for local variations in atmospheric pressure. The second approach consists of an alternative AC, in which the atmospheric parameters and water-leaving reflectance are inverted using neural networks. The selection of the

BAC or alternative AC is made using a dedicated switch, configurable via auxiliary data (RENOSH et al., 2020).

(iii) 6SV: Second Simulation of a Satellite Signal in the Solar Spectrum (6SV) is an advanced radiative transfer code designed to simulate a specific atmosphere condition based on advance knowledge of atmospheric and illumination conditions, and the sensor used. The algorithm outputs the necessary parameters to apply the Radiative Transfer Equation (ETR) for estimating the surface reflectance (VERMOTE et al., 2006). For the validation of PRISMA L1 and OLCI L1 products, 6SV was applied using the Py6S Python programming language interface (WILSON, 2013). The aerosol and atmospheric profile were set as Continental and Tropical, respectively. The AOD (550) value and the geometry parameters were obtained from PRISMA metadata. The correction was made for each PRISMA and OLCI band using their respective spectral response functions.

(iv) ACOLITE: ACOLITE version 20221114.0 applies a dark spectrum fitting (DSF) scheme as the default setting to estimate the AOD and hence atmospheric path reflectance, transmittances, and spherical albedo (VANHELLEMONT, 2019). The DSF assumes: *(i)* a homogeneous atmosphere over a certain extent of an image, and *(ii)* that there are pixels within this subscene that contain near-zero water-leaving radiances in one band. The DSF approach explores the differences between the surface reflectance and the TOA reflectance of a target contained in the image and with known spectral response to estimate the remote sensing aerosol. For clear waters, the aerosol scattering is estimated based on the NIR bands, and SWIR bands for moderate and turbid water. The aerosol reflectance is retrieved at those bands and extrapolated to the VNIR bands based on the aerosol type, or ratio of Rayleigh corrected reflectance in these infrared bands (MARTINS et al., 2017). Despite being primarily designed for processing multispectral images, ACOLITE is now adapted to support processing PRISMA data, where the L1 and L2C data products are required as inputs. The ACOLITE/DSF processing is available in a GitHub code and in binary releases (https://github.com/acolite/acolite).

3.3.3.2 Performance assessment

Atmospheric correction models were assessed qualitatively by comparing *in-situ* R_{rs} measurements (resampled to PRISMA and OLCI bands) against satellite-derived surface R_{rs}, and quantitatively through the coefficient of determination (R^2), Bias, Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the Spectral Angle (SA) (Table 3.2). SA was used to determine the spectral similarity between satellite-derived surface R_{rs} and *in-situ* R_{rs} (lower SA indicates higher similarity), since SA is not sensitive to differences in the magnitude of spectrums.

Table 3.2 - Statistics error metrics, where y_i and x_i represent the AC-glint corrected R_{rs} and in-situ R_{rs}, respectively.

Coefficient of determination	$R^2 = \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}\right]^2$
Bias	$Bias = \frac{\sum_{i=1}^{n}(\frac{y_i - x_i}{x_i})}{n} \times 100$
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}}$
Mean Absolute Percentage Error (MAPE)	$MAPE = \frac{100}{n}\sum_{i=1}^{n}\left\lvert\frac{y_i - x_i}{x_i}\right\rvert$
Spectral Angle (SA)	$SA = \cos^{-1}\frac{\sum_{i=1}^{n} y_i x_i}{\sqrt{\sum_{i=1}^{n} y_i^2}\sqrt{\sum_{i=1}^{n} x_i^2}}$

4 CYANOBACTERIA MAPPING USING HYBRID ALGORITHMS AND SENTINEL-3 (OLCI) DATA IN BRAZILIAN RESERVOIRS

4.1 Introduction

CyHABs are the most common forms of harmful algal bloom in freshwater systems (COFFER et al., 2020). Under favorable growth conditions, they may form massive blooms that severely affect water quality for drinking water production (CODD; BELL; BROOKS, 1989). The World Health Organization reported significant toxic cyanobacterial blooms in inland water worldwide, which contributes to the deterioration of water quality and public health (CHORUS; BARTRAM, 1999). Because many inland waters are under eutrophication stress, regular monitoring of CyHABs and the early detection of potentially harmful blooms are important for the management of inland waters (RUIZ-VERDÚ et al., 2008).

All cyanobacteria contain at least a low concentration of the accessory photosynthetic pigment phycocyanin (AHN et al., 2007). Due to PC specific absorption peak in 620 nm, remote sensing of freshwater cyanobacteria has been largely focused on developing bio-optical algorithms to estimate de PC concentration (DEKKER, 1993; SCHALLES; YACOBI, 2000; SIMIS et al., 2005, 2007). These algorithms have provided reasonable cyanobacteria interpretation. However, the conventional optical algorithms present estimates with acceptable accuracy only for specific scenarios (OGASHAWARA et al., 2013; LI; LI; SONG, 2015; MATSUSHITA et al., 2015; SMITH; ROBERTSON LAIN; BERNARD, 2018). These models are sensitive to the optical complexity of turbid waters, in which the concentration and IOPs of phytoplankton, suspended matter, and CDOM have large variations (LE et al., 2011). As result of these effects, the application of common methodologies in optically dynamic systems reduces the general accuracy of the estimate and limits the temporal and spatial coverage of the model within the same water body.

To circumvent this limitation, specific thresholds have been used to switch between different algorithms through an approach known as hybrid algorithms (MATSUSHITA et al., 2015; SMITH; ROBERTSON LAIN; BERNARD, 2018). In

the hybrid model (HM), two or more bio-optical algorithms, with adequate accuracies for different concentration ranges, are combined (MATSUSHITA et al., 2015; SMITH; ROBERTSON LAIN; BERNARD, 2018; CAIRO, 2020). Gomez-Jakobsen et al. (2011), for example, proposed two normalized indices to separate Mediterranean Lakes into two types, where an algorithm for estimating Chl-a was applied to each classified type. Recently, Cairo et al. (2020) developed a hybrid algorithm for estimating Chl-a in the Ibitinga-SP reservoir, where a decision tree classifier was used to determine the best algorithm to be applied to each image pixel corresponding to different optical classes.

In this sense, we evaluate the applicability of a hybrid algorithm to mapping PC concentration in the Promissão Reservoir using the Sentinel-3/OLCI image data, comparing the performance with conventional optical models. We also provide an assessment of different atmospheric corrections models well consolidated in the literature. It is expected to improve the application of multispectral bio-optical algorithms to monitor CyHABs in optically complex waters.

4.2 Dataset

4.2.1 Field dataset

The *in-situ* dataset acquired from PROM1 (October/2021), PROM2 (April/2021), PROM3 (August/2022), and BIL (Billings Reservoir) field campaigns (Section 3.2) was used in this study (N = 163 samples). The *in-situ* dataset based on PROM1, PROM2, and BIL was used to calibrate the hybrid model, while PROM3 was applied to validate the final model. The sampling includes the *in-situ* R_{rs} resampled for OLCI bands (Section 3.2.1), Chl-a, and PC concentrations (Section 3.2.2) measurements (Table 3.1).

4.2.2 Sentinel-3/OLCI imagery data

Three OLCI cloud-free scenes (L1 TOA radiance, and L2-WFR water leaving radiance) of Promissão Reservoir acquired on October 3rd, 2021, October 8th, 2021, and August 15th, 2022 were considered in this study (Section 3.3.1). The images were co-located with 18 stations visited within ± 1 day with OLCI

overpass, and used to validate the atmospheric correction models, and the PC algorithms.

Table 4.1 shows the PC and Chl-a concentrations for the samples, in each field campaign, co-located with OLCI overpass. The L1 OLCI product image was atmospherically corrected using 6SV, and ACOLITE (Section 3.3.3.1). For all AC models tested, the surface reflectance image was divided by π in order to obtain the atmospherically corrected R_{rs}. Finally, the performance of each AC model was assessed considering the *in-situ* resampled R_{rs} (Section 3.3.3.3).

Table 4.1 - PC and Chl-a concentrations (mg.m^{-3}), and PC:Chla ratio for each collected sample co-located with OLCI overpass (October 3rd, 2021, October 8th, 2021, and August 15th, 2022).

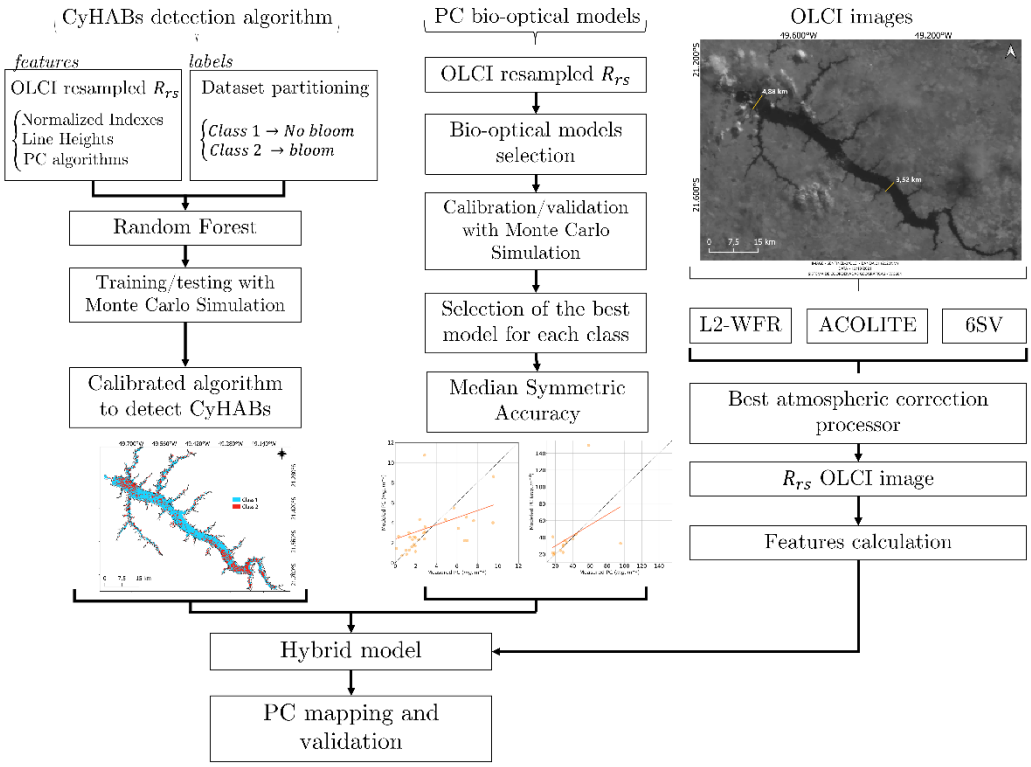
		Min	Máx	Mean	Median	Std
PROM1 (N = 11)	PC	0.32	6.93	2.97	2.52	1.90
	Chl-a	25.44	77.08	49.66	47.29	12.76
	PC:Chl-a	0.008	0.10	0.05	0.06	0.03
PROM3 (N = 7)	PC	5.50	123.61	26.98	6.83	40.39
	Chl-a	30.62	297.08	90.41	47.69	89.17
	PC:Chl-a	0.11	0.41	0.21	0.17	0.09

4.3 Methods

4.3.1 Hybrid model construction

The basic structure of the hybrid algorithm (Figure 4.1) is composed of two main parts: *(i)* CyHAB detection algorithm responsible for identifying pixels with cyanobacteria bloom presence. In this sense, two classes are defined: bloom presence and no bloom presence, and a machine learning algorithm is applied to classify the image. *(ii)* For each defined class a PC bio-optical algorithm is associated. Based on this structure, each pixel is initially classified in one of the optical classes, and then the respective bio-optical algorithm is applied, allowing the estimative of PC concentration.

Figure 4.1 - Block diagram representing the hybrid algorithm architecture.



Source: Author's elaboration.

4.3.1.1 CyHAB detection algorithm

4.3.1.1.1 Sample partitioning

In order to detect cyanobacteria bloom presence, a machine learning algorithm was used. First, the *in-situ* dataset was divided into two classes: low and high PC concentration, where the higher PC concentration class was related to the CyHABs presence (Bloom), and the lower PC concentration class was related to the absence of CyHABs (No Bloom) (SIMIS et al., 2007). The dataset division was based on the Spectral Angle Mapper (SAM) supervised classifier. SAM classification technique quantifies the similarity degree between different spectra using the angle difference of spectral curves. Starting from a set of reference spectra, SAM is able to classify sampled spectra according to their shape similarity (PAHLEVAN et al., 2021). Different studies have shown that R_{rs} classification is able to represent variations in the concentration of the OACs (SPYRAKOS et al., 2018; DA SILVA et al., 2021; PAHLEVAN et al., 2021).

Here, the SAM was applied considering two reference spectra representing the presence or not of CyHABs. The choice of the reference spectra was based on ensuring the best separability in the PC concentration samples.

4.3.1.1.2 Machine learning classifier

To identify image pixels with CyHABs presence, the supervised machine learning algorithm Random Forest (RF) was used, given its powerful generalization capacity and the good estimative with a small training dataset (BELGIU; DRĂGUT, 2016). The application of RF is based on ensemble learning training using the bagging approach, a technique that combines other classifiers to provide solutions to complex problems. In the bagging approach, each classifier in the ensemble is trained on a random subset of a training samples set selected with replacements, i.e., individual data points can be chosen more than once. After several data samples are generated, these weak models are trained independently, reducing the variance and yielding a more accurate estimate (BELGIU; DRĂGU, 2016). RF classifier yields reliable classifications using predictions derived from an ensemble of decision trees, selecting and ranking those variables with the greatest ability to discriminate classes. In the end, RF creates an uncorrelated forest of a decision tree. This is an important asset given the high dimensionality of remotely sensed data (KORTING et al., 2013).

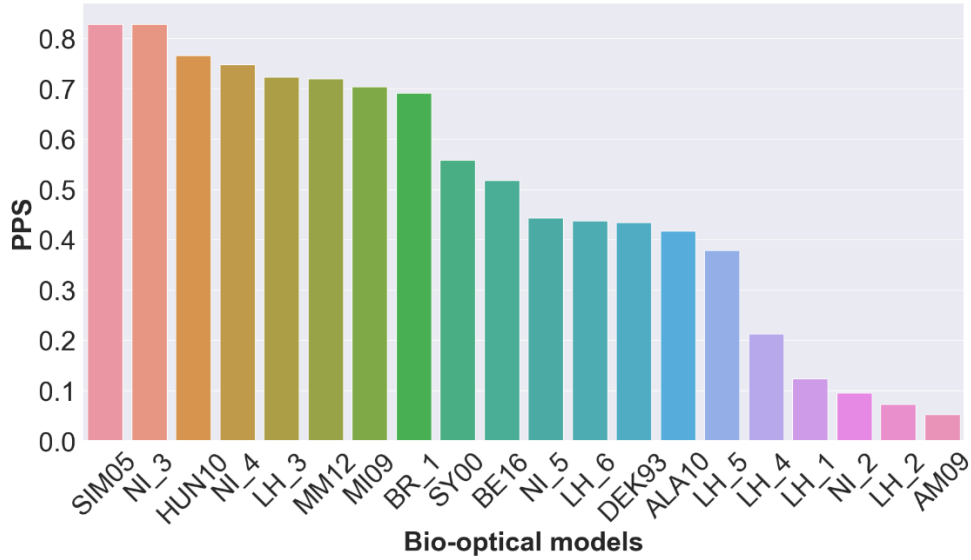
The optimal set of input variables is crucial for the model performance (BERHANE et al., 2018). In this study, several band ratios (BRs), line heights (LHs), and established algorithms related to PC estimative were calculated as potential explanatory variables (listed in Table 4.2). BRs were chosen as they capture relationships between separate bands, while LHs are relatively insensitive to uncertainties in R_{rs} (O'SHEA et al., 2021). The pure spectral bands were not used as an input feature to avoid creating a sensitive model to atmospheric-derived uncertainties.

Table 4.2 - Band math used as initial features selected as input to Random Forest. All equations used were adapted to OLCI bands. Highlighted lines in orange indicate the features selected as final input.

Name	Reference	Equation
DEK93	Dekker (1993)	PC ∝ 0.5*[(Rrs(560) + Rrs(665)) - (Rrs(620))]
SY00	Schalles & Yacobi (2000)	PC ∝ Rrs(665) / Rrs(620)
SIM05	Simis et al. (2005)	PC ∝ Rrs(708.75) / Rrs(620)
MI09	Mishra et al. (2009)	PC ∝ Rrs(708.75) / Rrs(665)
HUN10	Hunter et al. (2010)	Chla ∝ [1/Rrs(620) - 1/Rrs(709)] * Rrs(753)
ALA10	Alawadi et al. (2010)	PC ∝ (Rrs(865) - Rrs(665)) / (Rrs(442.5) + Rrs(510))
AM09	Amin et al. (2009)	PC ∝ (Rrs(681.25) - Rrs(665)) / (Rrs(681.5) + Rrs(665))
BE16	Beck et al. (2016)	PC ∝ Rrs(510) - [Rrs(665) + (Rrs(442.2) - Rrs(665))]
GI03	**Gitelson et al. (2003)**	PC ∝ [1/Rrs(673.5) - 1/Rrs(708.75)] * Rrs(761.25)
MM12	Mishra & Mishra (2012)	PC ∝ (Rrs(708.75) - Rrs(665) / (Rrs(708.5) + Rrs(665))
NI_1	**NI_1**	Chla ∝ [Rrs(764.37) - Rrs(665)] / [Rrs(764.37) + Rrs(665)]
NI_2	NI_2	Chla ∝ [Rrs(560) - Rrs(665)] / [Rrs(560) + Rrs(665)]
NI_3	NI_3	Chla ∝ [Rrs(708.75) - Rrs(620)] / [Rrs(708.75) + Rrs(620)]
NI_4	NI_4	Chla ∝ [Rrs(764.37) - Rrs(620)] / [Rrs(764.37) + Rrs(620)]
NI_5	NI_5	Chla ∝ [Rrs(560) - Rrs(620)] / [Rrs(560) + Rrs(620)]
LH_1	LH_1	Rrs(673.5) - (Rrs(681.25) - ((Rrs(665) - Rrs(681.25)) * ((681.25 - 673.5)/(681.25 - 665))))
LH_2	LH_2	Rrs(620) - (Rrs(665) - ((Rrs(560) - Rrs(665)) * ((665 - 620)/(665 - 560))))
LH_3	LH_3	Rrs(681.25) - (Rrs(708.75) - ((Rrs(665) - Rrs(708.75)) * ((708.75 - 681.25)/(708.75 - 665))))
LH_4	LH_4	Rrs(665) - (Rrs(681.25) - ((Rrs(620) - Rrs(681.25)) * ((681.25 -665)/(681.25 - 620))))
LH_5	LH_5	Rrs(708.75) - (Rrs(753.75) - ((Rrs(681.25) - Rrs(753.75)) * ((753.75 - 708.75)/(753.75 - 681.25))))
LH_6	LH_6	Rrs(665) - (Rrs(673.5) - ((Rrs(620) - Rrs(673.5)) * ((673.5 -665)/(673.5 - 620))))
BR_1	Band Ratio 1	Rrs(708.75)/Rrs(681.25)

Machine learning algorithms, in general, are sensitive to correlated features, with a sharp decrease in the performance (CHAN et al., 2022). Given the number of potential variables for the classification (Table 4.2), a feature selection procedure was applied to create a parsimonious set of input variables, minimizing the information redundancy and input model complexity. First, the features were ordered by their importance using the Predictive Power Score (PPS), which scores each variable based on their relevance to predict a label (KIM et al., 2022). PPSs are calculated based on the loss function of a single decision tree, which are created from the evaluated feature and the target label. The score varies from 0 to 1, where variables with no predictive power are penalized with 0, and a perfectly predictive feature achieves a unit score (Figure 4.2).

Figure 4.2 - Predictive Power Score (PPS) for all the math bands used as initial features (Table 4.2).



Source: Author's elaboration.

Then, a pairwise relationship analysis was applied for every pair of features to find the correlation between them. A threshold was chosen to the Pearson's Ranked Correlation Coefficient to exclude more correlated features, starting from the feature with the highest PPS in direction to the lowest PPS (Figure 4.2). In the subspace of threshold search (from 0.50 to 0.95, varying 0.05), the threshold of 0.70 was chosen as it was found to produce estimative with the best accuracy. The variables defined as optimal input to RF classifier are highlighted in orange in Table 4.2

In the construction of a machine learning model, the learning rate of the algorithm is controlled by the hyperparameters (JIN, 2022). Each algorithm is programmed with a default configuration, but fine adjustments of hyperparameters can increase the accuracy of the final predictions. To determine the structure of RF model, several hyperparameters, including the number of trees and the maximum depth, were tuned. The number of trees controls how many decision trees will be generated, and maximum depth limits the number of splits that each decision tree can do. In this way, a step-by-step tuning with a grid search strategy was used. This methodology generates all possible configurations of a set of hyperparameters and test each one in the

model, choosing the one that achieves the lowest mean accuracy metric. For instance, the number of trees selected was 300 (tests: 25, 50, 100, 150, 200, 300, 500, 1000), and the maximum depth was set to 10 (tests: 3, 5, 10, 15, 20, 25, 30, 35).

After the definition of the input features and the hyperparameters, the Monte Carlo Simulation approach was considered to calibrate the RF algorithm. The MC is an approach commonly used in models evaluation where only a limited samples are available (REMESAN; MATHEW, 2014). In MC, the calibration/validation is repeated *n* times, where at each iteration the entire dataset is randomly divided into two subsets: training and testing. (AUGUSTO-SILVA et al., 2014). In this study, the MC was applied with 1000 repetitions, using a random 75/25 training/testing split, which was selected in the dataset using the stratified approach. The stratified random sampling approach is done when the dataset is unbalanced, i.e., the number of observations in each class is highly different. For each MC trial, the accuracy (Equation 4.2), precision (Equation 4.3), and recall (Equation 4.4) were calculated (CAIRO et al., 2020). Precision is a measure of the result relevance, and recall is a classification sensitivity measure. High values for both performance metrics indicate precise results (high precision), with higher frequency for positive results (high sensitivity). At the end of the MC, the final statistical metrics were obtained through the mode for all rounds.

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \tag{4.2}$$

$$Precision = \frac{T_p}{T_p + F_p} \tag{4.3}$$

$$Recall = \frac{T_p}{T_p + F_n}, \tag{4.4}$$

where T_p, F_p, and F_n are the number of true positives, false positives, and false negatives, respectively.

4.3.1.2 Phycocyanin bio-optical models

A summary of the algorithms implemented for PC retrieval is provided in Table 4.3 . These algorithms were compiled based on a literature review and selected upon the criteria that they applicable or easily adapted to the OLCI bands, and were specifically developed for inland water. The models considered are function of remote sensing reflectance (R_{rs}), absorption coefficient by phycocyanin (a_{pc}), absorption coefficient by water (a_w), absorption coefficient by phytoplankton (a_{phy}), backscattering coefficient (b_b), absorption by chlorophyll-a (a_{chla}) and the specific absorption by phycocyanin (a^*_{pc}) at the respective wavelengths (λ).

Table 4.3 - Summary of the algorithms used to retrieve PC concentration.

Model	Equation
SY00 - Schalles and Yacobi (2000)	$PC \propto \frac{R_{rs}(665)}{R_{rs}(620)}$
MM14 – Mishra and Mishra (2014)	$\psi = 0{,}709\left[\frac{R_{rs}(560)}{R_{rs}(665)}\right]^{0{,}874}$ $PC \propto [R_{rs}^{-1}(620) - \psi R_{rs}^{-1}(665)] * R_{rs}(778{,}05)$
SIM05 – Simis et al. (2005)	$a_{chla}(665) = \left[\frac{R_{rs}(709)}{R_{rs}(665)} * (a_w(709) + b_b) - b_b - a_w(665)\right] * \gamma^{-1}$ $a_{pc}(620) = \left[\frac{R_{rs}(709)}{R_{rs}(620)} * (a_w(709) + b_b) - b_b - a_w(620)\right] * \delta^{-1} - (\varepsilon * a_{chla}(665))$ $[PC] = \frac{a_{pc}(620)}{a^*_{pc}(620)}$
LIU17 – Liu et al. (2017)	$PC \propto \left[\frac{1}{R_{rs}(620)} - \frac{0{,}4}{R_{rs}(560)} - \frac{0{,}6}{R_{rs}(709)}\right] * R_{rs}(754)$
OGA19 – Ogashawara and Li (2019)	$PC \propto \frac{\frac{R_{rs}(709)}{R_{rs}(620)} - \frac{R_{rs}(709)}{R_{rs}(665)}\varphi_1}{1 - (\varphi_1\varphi_2)}$

Among the models tested, we applied a parameterization in the SIM05 algorithm using *in-situ* data from the PROM1 field campaign (N = 23 samples). First, the correction factors δ and γ (Table 4.3) were retrieved from *in-situ* $R_{rs}(\lambda)$ and $a_{phy}(\lambda)$ spectra. Both factor were originally applied in order to minimize the errors derived from the model simplications. In this way, the

equations used to calculate $a_{chla}(665)$ and $a_{pc}(620)$ (Table 5) were initially applied considering $\delta = \gamma = 1$, obtaining an 'uncorrected' $a_{chla}(665)$ and $a_{pc}(620)$. The linear least-squares fit of these retrieved absorptions against $a_{phy}(\lambda)$ (for $\lambda = 665$ and $620\,nm$), yielded two slope values that were adopted for $\gamma = 0.14585$ and $\delta = 0.18055$. The conversion factor ε (Table 4.3) was originally retrieved in the Simis et al. (2005) work from the best fit of computed versus observed PC concentration. Our *in-situ* dataset was not used, in this case, to prevent ε from converging to an optimal value. Instead, O'Shea et al. (2021) applied a dataset of 939 samples to optimize the ε value by minimizing the median symmetric accuracy between the model estimates and the known *in-situ* concentration. The final value obtained ($\varepsilon = 0.251753$) was then applied here. Finally, the specific absorption coefficient for PC at 620 nm ($a_{pc}^{*}(620)$) was determined from the $a_{pc}(620)$ divided by the measured phycocyanin concentration.

For each defined optical class (Section 4.3.1.1.1), the algorithms listed in Table 4.3, except the SIM05 model, were calibrated/validated using BIL, PROM1 and PROM2 dataset, through MC simulation. Based on the sampling division (Section 4.3.1.1.1), 106 samples were used to calibrate the models for the No CyHABs class, and 46 samples for the CyHAb class. Here, we consider 10,000 iterations using 75/25 training/testing split. At each iteration, the testing dataset of *in-situ* PC concentration was compared with those estimated by the models by applying the linear least squares regression analysis. The goodness of fit was reported by the coefficient of determination (R^2) and the slope (*p*). Algorithm accuracy was quantified with measures of errors, including the median symmetric accuracy (ζ) (Table 4.4), which is equivalent to the MAPE; the symmetric signed percentage bias (β) (Table 4.4), and the median absolute error (MdAE) (Table 4.4). The first two norms (ζ and β) initially transform data into log space and then convert them back to linear space to assess the quality of the retrieved quantity. This transformation ensures that the metric is symmetric, i.e., switching the values of the predicted and observed value will

result in the same errors, unlike MAPE, which is asymmetric with respect to overforecasting and underforecasting (MORLEY; BRITO; WELLING, 2018).

Table 4.4 - Statistics error metrics, where $\boldsymbol{y_i}$ and $\boldsymbol{x_i}$ represent the modelled value and in-situ measured value, respectively.

Metric	Equation
Median symmetric accuracy	$\zeta = 100 \times (\exp\left(\frac{\sum_{i=1}^{n}\left\lvert \log_e \frac{y_i}{x_i}\right\rvert}{n}\right) - 1)$
Symmetric signed percentage bias	$\beta = 100 \times sgn\left(\frac{\sum_{i=1}^{n}\left\lvert \log_e \frac{y_i}{x_i}\right\rvert}{n}\right)\left(\exp\left(\left\lvert\frac{\sum_{i=1}^{n}\left\lvert log_e \frac{y_i}{x_i}\right\rvert}{n}\right\rvert\right) - 1\right)$
Median absolute error	$MdAE = median(\lvert y_i - x_i\rvert, \ldots, \lvert y_i - x_i\rvert)$

At the end of the MC iterative process, the determination of the adjustment coefficients (slope and intercept) of each semi-empirical model was determined based on the mode of ζ. Finally, the choice of the 'best algorithm' for each class was selected, considering the lowest ζ value. The ζ was chosen as the criteria since it presents a cost function linear and symmetric (MORLEY; BRITO; WELLING, 2018). In order to validate the hybrid algorithm and evaluate its applicability, the final selected algorithms for each optical class were also calibrated/validated considering the entire dataset (joining data from all classes).

4.3.2 Hybrid algorithm validation – in situ data

We applied a leave-one-out testing approach to estimate how well the hybrid-model is expected to be transferred to datasets not included within the training set. In leave-one-out testing, the training dataset consisted of the entire dataset except for one field campaign (PROM3, n = 11), which was used for the test (*in-situ* PC concentration, class, *in-situ* R_{rs} resampled to OLCI bands). First, the Random Forest model was applied in order to identify the CyHABs samples. The accuracy was evaluated considering the confusion matrix, and the statistical metrics of accuracy, precision, and recall (Equations 4.2, 4.3, and 4.4, respectively). Then, for each defined class, the selected bio-optical algorithm was applied to the sample, obtaining the PC concentration as a final result. To

assess the performance of the hybrid model, the estimated PC was compared to *in-situ* PC through the statistics metrics described in Table 4.4.

4.3.3 Hybrid algorithm application on satellite image

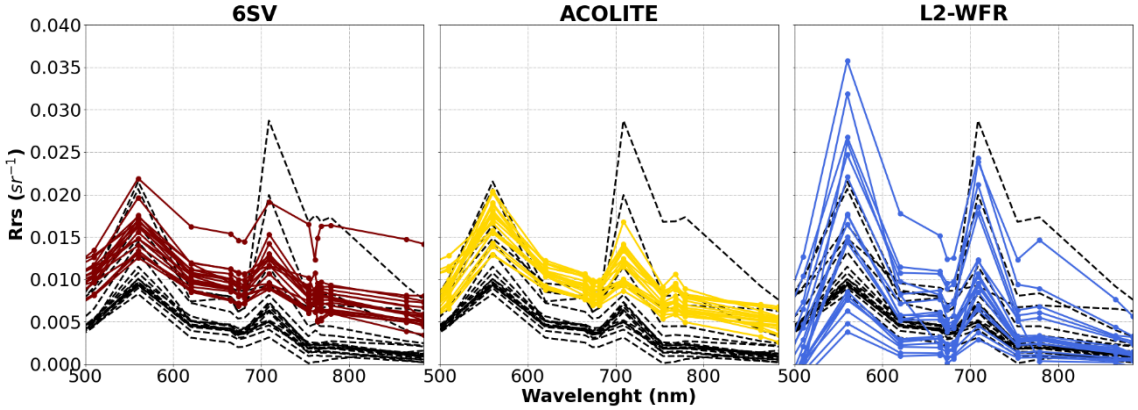
Based on the assessment of the atmospheric correction processors (Section 3.3.3.1 and Section 4.2.2), the hybrid model was applied over the AC-corrected R_{rs} images (October 3rd, 2021, October 8th, 2021, and August 15th, 2022) for each AC model considered, generating class maps, and PC concentration maps. The CyHABs detection algorithm and the final hybrid model validation were assessed considering the estimated class and PC concentration extracted from the pixels of each sampling station (latitude/longitude) of the co-located samples (Table 4.1). Each extracted pixel value was then compared with the *in-situ* PC concentration and its respective defined class. Confusion matrix, accuracy, precision, and recall (Equations 4.2, 4.3, and 4.4, respectively) were used to validate the CyHABs detection model, while the statistics metrics described in Table 4.4 were applied to assess the performance of the final hybrid model.

4.4 Results

4.4.1 Atmospheric correction modeling

Figure 4.3 shows the qualitative comparison between *in-situ* R_{rs} and the AC-derived R_{rs} for 6SV and ACOLITE models, and L2-WFR product. In order to avoid the commonly high uncertainties in R_{rs} within the blue region (BRAGA et al., 2021), a lower bond of 500 nm was chosen. Eighteen R_{rs} samples acquired in match-up condition with OLCI (± 1 day) were used as reference to evaluate the retrieved bottom-of-atmosphere reflectance. In general, all models removed most of the atmospheric effect, with the best R_{rs}-corrected shape achieved by ACOLITE (SA = 16.76°) and 6SV (SA = 18.76°) in comparison with *in-situ* measurements (Figure 4.4). Despite the best R_{rs}-corrected shape, 6SV and ACOLITE shows an offset across the spectra, indicating the presence of some residual glint effect. L2-WFR model overestimates *in-situ* R_{rs}, particularly in the green and red spectral region.

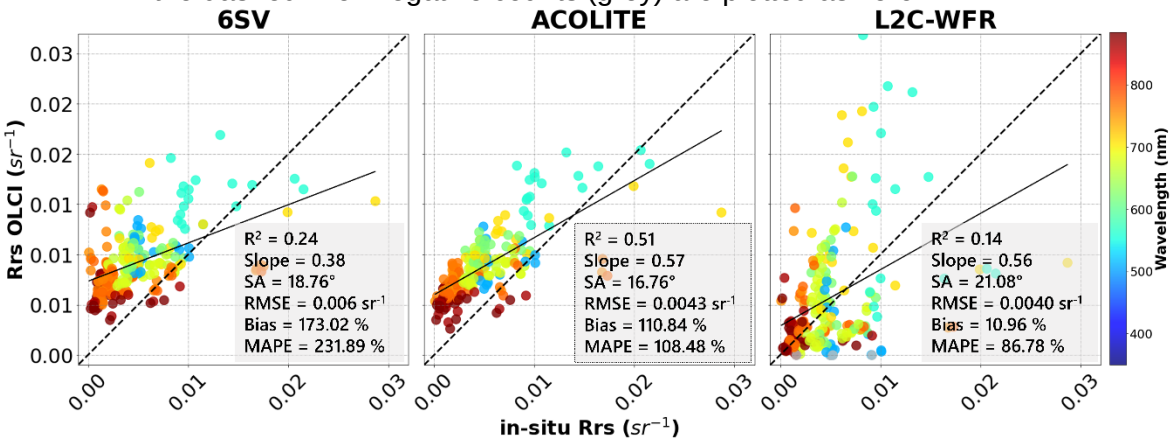
Figure 4.3 – Qualitative comparison of *in-situ* $\boldsymbol{R_{rs}}$ (dashed line) and OLCI AC-corrected $\boldsymbol{R_{rs}}$ (colored line) on October 3rd, 2021, October 10th, 2021, and August 15th, 2022 for each AC model tested.



Source: Author's elaboration.

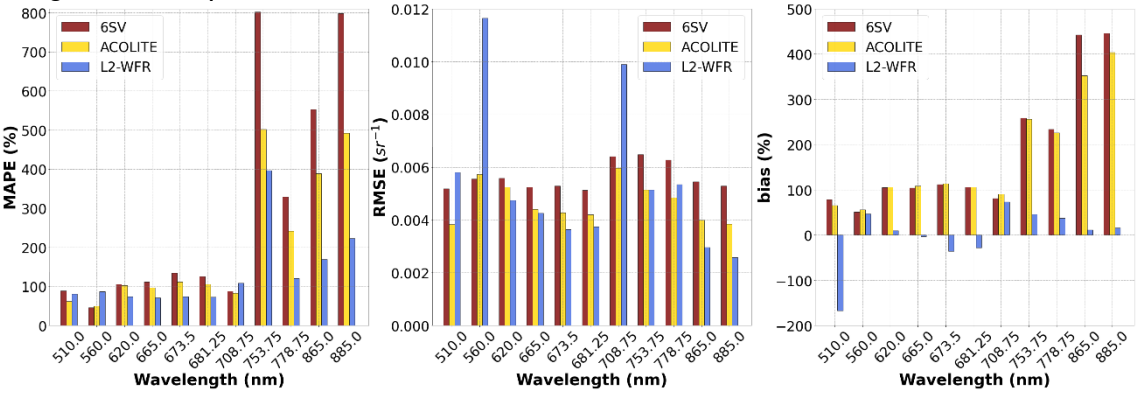
For a quantitative investigation into band-wise performance of AC models, scatter plots for the full matchup dataset, and the median of error metrics are shown in Figure 4.4, while Figure 4.5 demonstrates that the accuracy of atmospherically corrected R_{rs} is wavelength-dependent as MAPE, RMSE, and bias varies with wavelength across all AC algorithms. Among the models tested, LW-WFR presented the best performance in terms of error metrics (bias = 10.96% and MAPE = 86.78%), followed by ACOLITE (bias = 110.84% and MAPE = 108.48%). Even at longer wavelengths (from 700 nm onwards), where a tendency toward increased errors is observed, the L2-WFR model presents an improvement in the performance (Figure 4.5). Regarding bands around PC absorption feature (600 – 650 nm), which are typically used for retrieval of the pigment concentration, the best performances are found for L2-WFR, while ACOLITE and 6SV achieved nearly equal performance (Figure 4.5). Despite the significant advantage of L2-WFR, the standard OLCI L2 product presented two major errors at 560 and 708.75 nm (Figure 4.4 and Figure 4.5), and some negative R_{rs} values, represented by the grey dots in Figure 4.4. ACOLITE, on the other hand, showed the best linear fit, with and R^2 = 0.51, and slope = 0.57.

Figure 4.4 – Scatterplots and the median of error metrics of *in-situ* R_{rs} resampled for OLCI bands versus OLCI AC-corrected R_{rs} for each AC model tested. Black solid line is the best-fit linear regression. 1:1 line is represented by the dashed line. Negative counts (grey) are plotted as zero.



Source: Author's elaboration

Figure 4.5 – Spectral variation of MAPE, RMSE, and bias for each AC model tested.



Source: Author's elaboration.

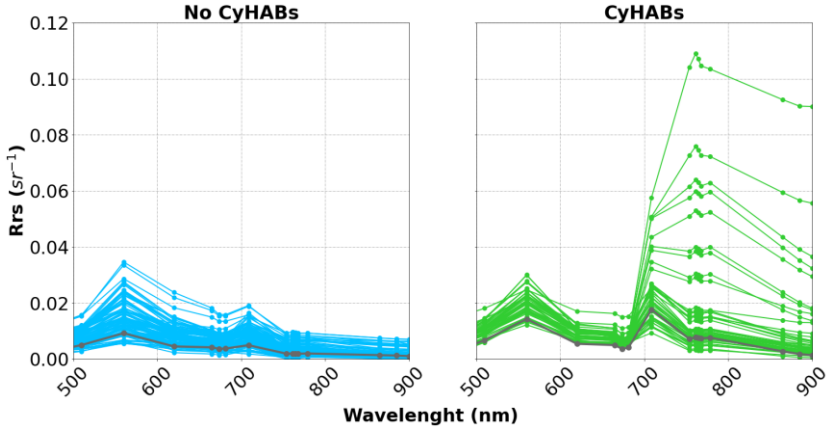
4.4.2 Calibration/validation of the hybrid algorithm

In order to obtain an PC algorithm suitable for optically complex waters, a hybrid model combining existing bio-optical models is proposed here. As the flowchart in Figure 4.1 details, the first step in building the hybrid model is the calibration of a CyHABs detection algorithm.

Figure 4.6 shows the result for the samples partitioning (N = 142 samples) into two classes using the SAM: *(i)* No CyHABs, and *(ii)* CyHABs, with the samples defined as the reference for each class highlighted in gray. Figure 4.7 highlight the separability of the classes in terms of PC concentration. The first class ('No

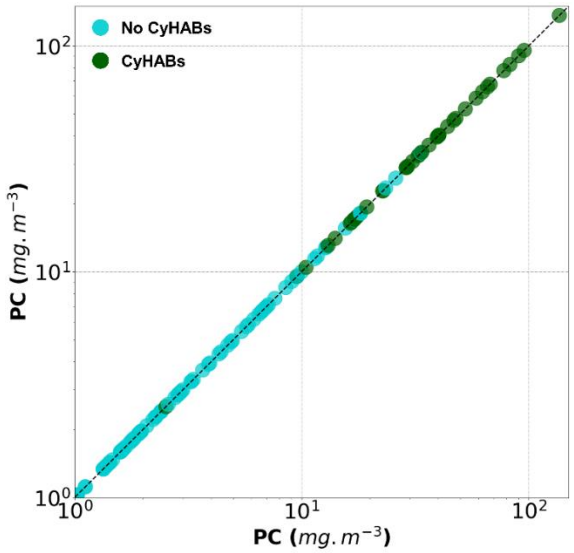
CyHABs', N = 106 samples) was related to the absence of cyanobacteria bloom, where the PC concentration ranged approximately between 0 – 14 $mg.m^{-3}$ (Figure 4.7). On the other hand, the second class ('CyHABs', N = 46 samples) was used here as a mask to identify cyanobacteria bloom, with a PC concentration range between 14.1 – 302 $mg.m^{-3}$. The input dataset used was unbalanced in terms of PC concentration, as the result of the second group was smaller than the first group. As the SAM algorithm established optical classes, the PC concentration ranges displayed small overlaps between classes related to the fuzzy nature of the water masses and the presence of the remaining optically active components. Assuming phytoplankton is the dominant constituent with regards to the casual IOPs, the magnitude of the 700 nm peak in the second-class spectra (Figure 4.6) is predominantly dependent on the specific backscattering coefficient (b_b) of the dominant phytoplankton species.

Figure 4.6 – *In-situ* R_{rs} resampled to OLCI bands classified in two classes representing the presence (green plot) or not (blue plot) of CyHABs, using the SAM supervised classifier. The grey line represents the reference spectrum of each class.



Source: Author's elaboration.

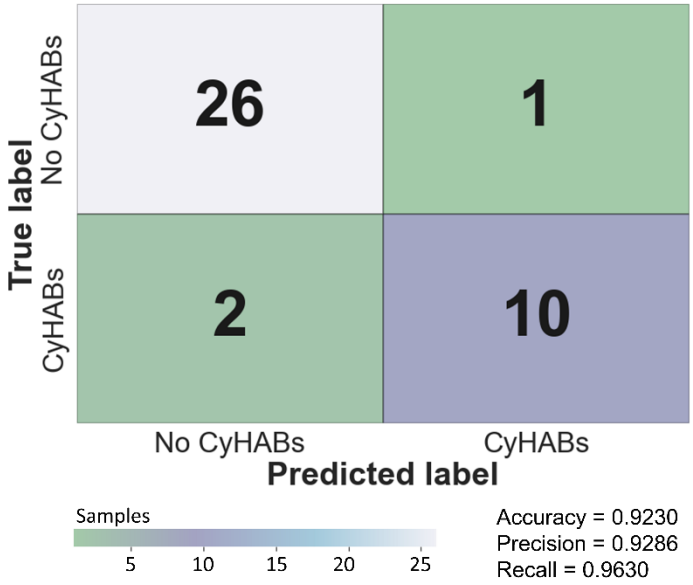
Figure 4.7 - Separability between classes.



Source: Author's elaboration.

The second part of the CyHABs detection algorithm construction was the calibration of the RF algorithm, which was then used as conditional structure in the hybrid model. The resulted dataset partitioning into the 'No CyHABs' and 'CyHABs' classes was used as input label in the machine learning model. Based on the features selection (Section 4.3.1.1.2, Table 4.2), the input dataset was divided into 75/25 training/testing using MC simulation, ensuring dataset balance. The final performance of the RF classifier was evaluated using mode accuracy, precision, and recall (Figure 4.8). In addition, as seen in Figure 4.8, the variation of false-positive rates versus true-positives rates was illustrated using the confusion matrix. In general, 39 samples (27 'No CyHABs', and 12 'CyHABs') were used to validate the classification, where only 1 sample was identified as false-positive and 2 samples as false-negative. Furthermore, the precision and recall obtained of 0.93 and 0.96 respectively, demonstrating the satisfying quality of the RF performance.

Figure 4.8 – Validation of the RF algorithm to identify CyHABs.



Source: Author's elaboration.

For each defined class the best performance bio-optical model was selected to estimate PC concentration. The performance of all selected algorithms listed in Table 4.3 and calibrated using the MC simulation as summarized in Table 4.5 for each defined group: No CyHABs, CyHABs, and All. The last group corresponds to the combination of the first two groups. Overall, all semi-empirical models were capable of estimating the PC concentration with good accuracy ($\zeta < 50\%$) for both classes. Higher error metrics, for all classes, was resulted from SIM05 and SY00 (Table 4.5). SIM05, the only semi-analytical model tested, presented a better performance for the 'No CyHABs' class ($\boldsymbol{\beta}$ = 30.52%). In fact, all bio-optical algorithms tested here had lower percentage errors (β) for the first class, in comparison with the 'CyHABs' class (Table 4.5). Considering the combination of both classes (All), lower retrieval uncertainties were achieved by LIU17 (ζ = 25.85%) and OGA19 (ζ = 24.32%). However, all models presented best performance when calibrated for specific concentrations range of PC.

For the 'No CyHABs' class, the best accuracy was achieved by OGA19 model (ζ = 28.86%). Considering bias and MdAE, however, lower errors were obtained from LIU17 and MM14 models. On the other hand, for the 'CyHABs' class,

LIU17 model showed the best accuracy (ζ = 8.44 %), and the lowest MdAE (2.06 mg.m^{-3}). In order to combine the best model's combination to estimate PC concentration, we selected one bio-optical algorithm for each class ('No CyHABs', and 'CyHABs') based on the accuracy. In this sense, the OGA19 was selected to represent the 'No CyHABs' class, while LIU17 was chosen for the 'CyHABs' class.

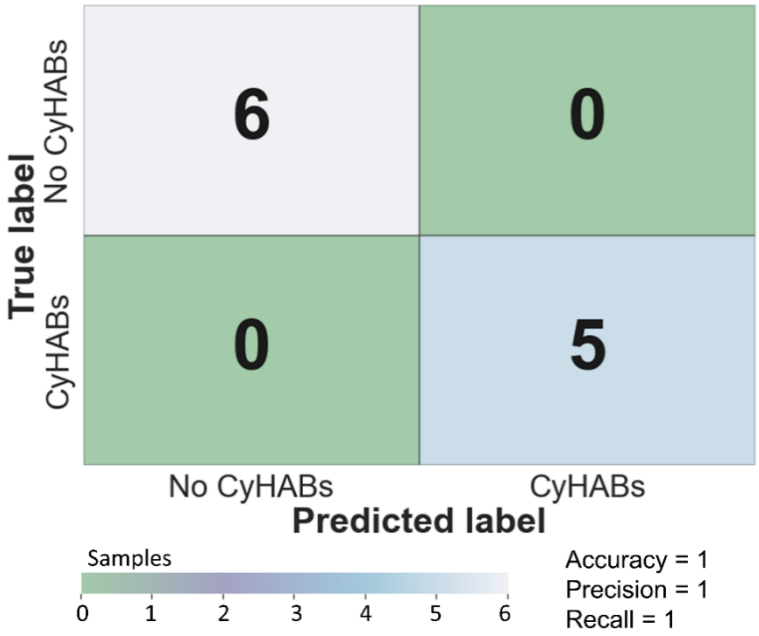
Table 4.5 – Performances of all PC estimation algorithms for each defined class ('No CyHABs', and 'CyHABs') as well as the entire dataset ('All'). For each class, the most accurate model is highlighted in green.

		SIM05	SY00	MM14	LIU17	OGA19
No CyHABs	**ζ**	77.9	39.56	34.4	29.21	28.86
	β	30.52	5.6	1.47	0.52	4.18
	MdAE	1.73	1.04	0.71	0.72	1.07
CyHABs	**ζ**	141.33	36.84	11.43	8.44	12.9
	β	141.33	7.17	1.83	4.47	1.69
	MdAE	16.9	11.51	5.61	2.06	4.38
All	**ζ**	99.89	80.9	35.63	25.85	24.32
	β	4.97	59.59	3.61	15.67	4.09
	MdAE	2.2	9.99	8.68	2.45	5.21

4.4.3 *In-situ* performance of the hybrid algorithm

In the leave-one-out analysis, the PROM3 field measurements (N = 11 *in-situ* samples resampled to OLCI bands) were used to assess the final performance of the hybrid model in a dataset not included in the model construction. Figure 4.9 reports the CyHABs detection algorithm achievement based on the confusion matrix and accuracy, precision, and recall metrics. Results show no classification error during the validation process (accuracy = precision = recall = 1), where 5 samples were detected as CyHABs spots.

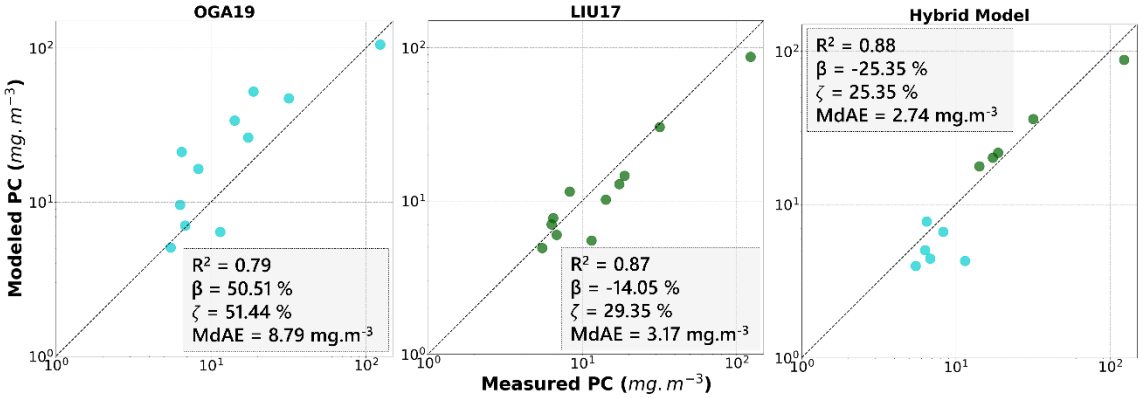
Figure 4.9 – *In-situ* validation of the CyHABs detection algorithm.



Source: Author's elaboration.

Figure 4.10 shows the performances of the hybrid algorithm as well as those of two selected algorithms (OGA19 and LIU17) calibrated for the entire dataset. It can be seen that OGA19 yielded better performance for low PC concentration, but performed poorly for higher PC concentration values (overestimation), with a general PC uncertainty of 51.44% and MdAE of 8.79 $mg.m^{-3}$. The LIU17 algorithm achieved good accuracy performance (ζ = 29.35%), generalizing better for higher PC values. In contrast, the retrieval uncertainties from the hybrid model were lower than the individual bio-optical models (ζ = 25.35%, and MdAE = 2.74 $mg.m^{-3}$). Overall, the scatterplot demonstrates that PC concentration values estimated by the hybrid algorithm were in good agreement with *in-situ* measured values, with all data points close to the 1:1 line.

Figure 4.10 – Performance assessment of Hybrid model, and OGA19 and LIU17 models using *in-situ* measured PC from the PROM3 field campaign (N = 11 samples). Reported metrics are R^2, median symmetric accuracy (ζ), symmetric signed percentage bias (β), and the median absolute error (MdAE). Cyan dots correspond to 'No CyHABs', and green dots do 'CyHABs' class.



Source: Author's elaboration.

4.4.4 Hybrid algorithm assessment on satellite observations

The first step for applying the HM to OLCI images was the detection of CyHABs pixels, based on the RF classifier. Then, the best-performing PC algorithm selected for each class was applied pixel by pixel according to its classification. Thus, an estimated PC concentration map is obtained. In this sense, the proposed CyHABs detection algorithm was first applied to three OLCI images atmospherically corrected using 6SV, ACOLITE, and L2-WFR. Each image was acquired in match-up condition with the field campaigns in the Promissão Reservoir on October 3rd, 2021, October 8th, 2021, and August 15th, 2022.

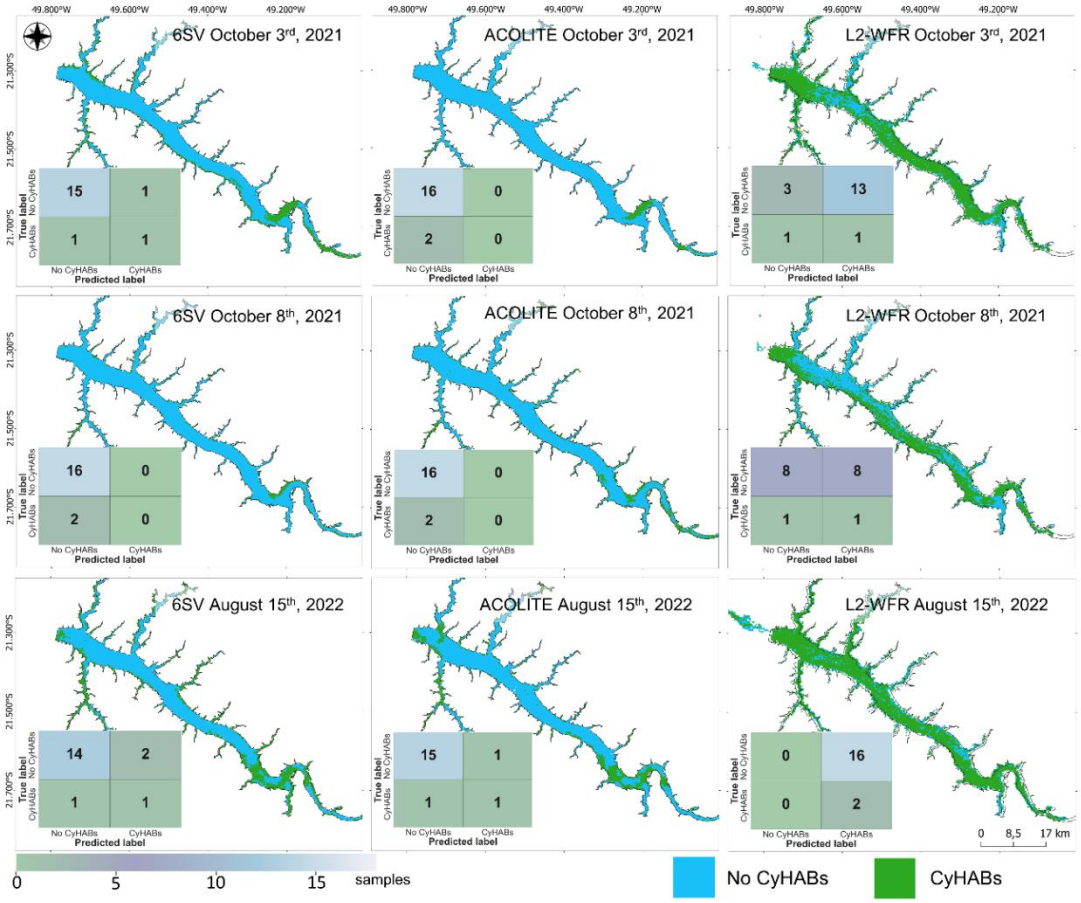
Figure 4.11 illustrates the final classification maps, with the CyHABs pixels mapped in green. For each image classification, match-up samples were extracted to assess the model's performance based on the confusion matrix (Figure 4.11), and the accuracy, precision, and recall (Table 4.6) in comparison with *in-situ* measurements.

Among the three AC models tested, the worst classification performance was achieved by L2-WFR corrected images (Figure 4.11). The application of the CyHABs detection model in the L2 OLCI product failed to correctly identify pixels with the presence of CyHABs, resulting in a large number of false

positives. For August 15th, for example, the model classified all samples as CyHABs, generating 16 false positives (accuracy = 0.11). On the other dates analyzed, the model was still able to generate high precision values related to true positives, but still with low accuracy values (Table 4.6).

6SV and ACOLITE presented better performance, with highly related results. However, the application of the RF classifier in both AC models also had particularly inaccurate results in identifying CyHABs pixels in terms of false negatives, which may be related to the unbalance of the considered samples. For October 8th, 2021, the algorithm did not detect any pixels with bloom, classifying the real CyHABs pixels as ‘No CyHABs’ (Figure 4.11). However, the good classification of the 'No CyHABs' class ensured good overall accuracy in both 6SV and ACOLITE images (0.89).

Figure 4.11 – CyHABs detection maps of Promissão Reservoir from atmospherically corrected OLCI images (AC processors: 6SV, ACOLITE, and L2-WFR) for three dates (October 3rd, 2021, October 8th, 2021, and August 15th, 2022). For each map, the confusion matrix is highlighted.



Source: Author's elaboration.

Table 4.6 – Assessment of the CyHABs detection model applied to OLCI images. For each AC processor and date, the best metric error value is highlighted in green.

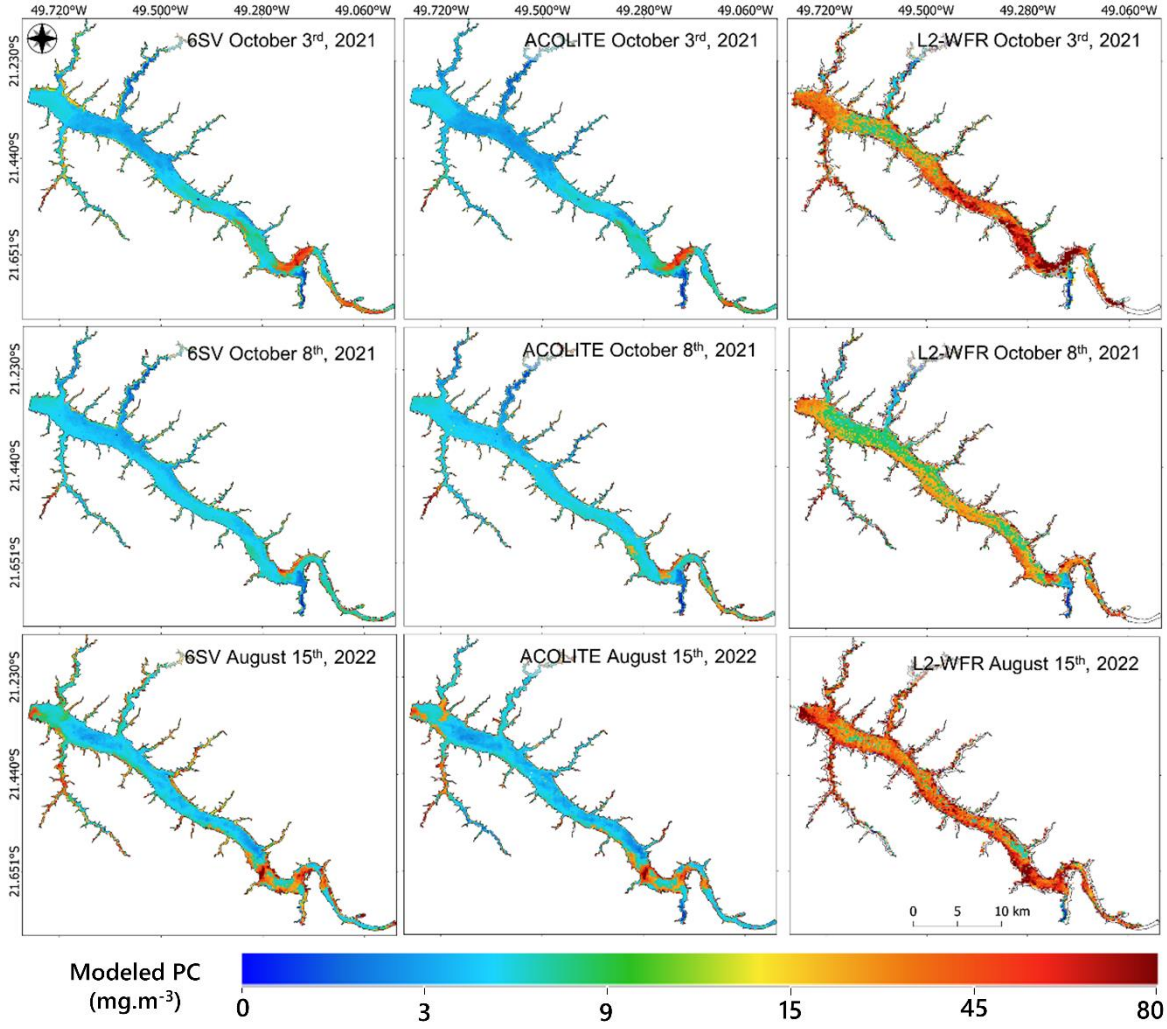
	October 3rd, 2021			**October 8th, 2021**			**August 15th, 2022**		
	6SV	ACOLITE	L2-WFR	6SV	ACOLITE	L2-WFR	6SV	ACOLITE	L2-WFR
Accuracy	0.89	0.89	0.22	0.89	0.89	0.50	0.83	0.89	0.11
Precision	0.94	0.89	0.75	0.89	0.89	0.89	0.93	0.94	0.00
Recall	0.94	1.00	0.19	1.00	1.00	0.50	0.88	0.94	0.00

Figure 4.12 presents the final PC maps for each AC processor and the date considered. Based on match-up samples with *in-situ* measurements, Figure 4.13 shows the performance of the HM, as well as the error metric of OGA19

and LIU17 calibrated for the entire dataset and applied to the OLCI images. The retrieved product maps produced from 6SV, ACOLITE, and L2-WFR agreed in terms of general distribution of pigment estimates, where higher PC concentration occurred in the south part of the reservoir (Figure 4.12). However, the magnitude of the estimates varies substantially, mainly with L2-WFR. For example, the scatterplots demonstrated similar distribution for 6SV and ACOLITE, with better accuracy achieved by ACOLITE (ζ = 64.38%). In general, the PC retrievals from both AC processors agreed well with the *in-situ* measurements ($MdAE_{6SV} = 2.06\, mg.m^{-3}$, $MdAE_{ACOLITE} = 1.95\, mg.m^{-3}$). The main difference among them occurred for higher PC concentration samples, where 6SV performed better (Figure 4.13). On the other hand, L2-WFR failed to generate an accurate estimate of the pigment concentration, significantly overestimating the low PC concentrations (ζ = 428%).

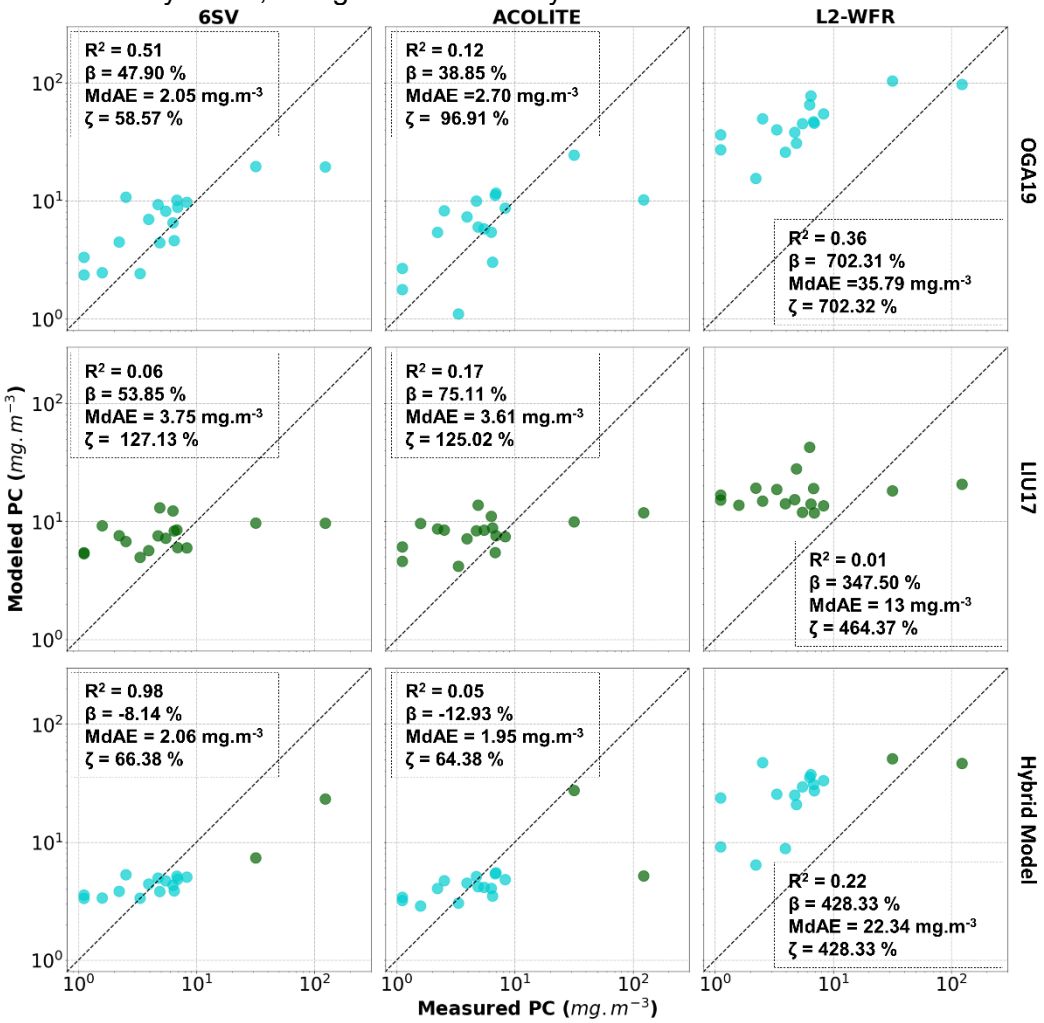
Comparing the hybrid performance with OGA19 and LIU17 models, the PC retrieval uncertainties from the HM were lower than both bio-optical models. For instance, considering ACOLITE AC-processor, OGA19 and LIU17 achieved an accuracy of 96.91% and 125.02%, respectively, while the HM was 64.38%. In terms of bias, the HM also presented better results (β = -12.93%) than OGA19 (38.85%) and LIU17 (75.11%), indicating that the hybrid approach can perform better than the single algorithm used for the entire dataset.

Figure 4.12 – PC products maps produced by the final hybrid model applied in three atmospherically corrected OLCI images (6SV, ACOLITE, and L2-WFR) for three dates (October 3rd, 2021, October 8th, 2021, and August 15th, 2022).



Source: Author's elaboration.

Figure 4.13 - Performance assessment of Hybrid model applied on atmospherically corrected OLCI images (6SV, ACOLITE, and L2-WFR) using co-located *in-situ* measured PC (N = 18 samples collected on October 3rd, 2021, October 8th, 2021, and August 15th, 2022). Reported metrics are R^2, median symmetric accuracy (ζ), symmetric signed percentage bias (β), and the median absolute error (MdAE). Cyan dots correspond to 'No CyHABs', and green dots do 'CyHABs' class.



Source: Author's elaboration.

4.5 Discussion

4.5.1 Atmospheric correction of OLCI data

Around 70-90% of the sensor-measured signal over water bodies comes from the atmosphere. In addition, other environmental factors, e.g., sunglint, wind speed, the geometry of the observations, and land adjacency effects, are added to the sum of the signal detected by the sensor. This high amount of

interferences has to be removed to obtain the water-leaving reflectance, the real value coming from the water surface. Atmospherically corrected OLCI data with three different AC processors are matched-up here to *in-situ* measured R_{rs} in order to test, analyze, and validate their performance. In general, all models tested were capable of accurately retrieving the R_{rs} spectra. With the exception of the spectral angle, general statistics showed a better performance achieved with the L2-WFR product, the only AC processor that applied glint correction. However, at shorter wavelengths, L2-WFR showed an increase of the relative errors in comparison with 6SV and ACOLITE, demonstrating a poor performance in modeling the atmosphere scattering effects. The negative R_{rs} values reported can mainly be related to the overestimation of aerosol radiance. The L2 OLCI product estimates the aerosol type and load from the NIR bands at 779 and 865 nm, which are highly affected by the presence of chlorophyll organisms (SIMIS et al., 2007). Thus, when determining the atmospheric parameters, a significant part of the water-leaving reflectance in the NIR bands was taken into account. As a result, the contribution of the atmosphere was overestimated and, the retrieved R_{rs} spectra were negative. In addition, the L2-WFR AC-corrected spectra showed a pronounced spectral feature in the green and red bands, presumably deviating from the *in-situ* R_{rs} spectra. In fact, the SA analysis showed that ACOLITE gave the most similar spectra in comparison with the *in-situ* measurements, and provided products with significantly less noise.

ACOLITE and 6SV presented similar results, with better error metrics achieved with ACOLITE. Both AC models overestimated the R_{rs} values through the whole spectrum. That offset might be a result of a residual glint component that was not corrected, and also due to AOD-derived errors related to the underestimation of the atmospheric path reflectance. Bassani et al. (2015) and Paulino et al. (2022) highlighted the impact on the reflectance accuracy when the AOD value used does not characterize the local aerosol microphysics. In addition, Vanhellemont (2019) reported about how processing a whole scene with the same aerosol model to estimate the path reflectance can be likely insufficient. In this sense, the lower performance of 6SV can be attributed to the

poor characterization of aerosols using a generalized model from the radiative transfer code. The practical application of 6SV includes standard aerosol types for land applications, such as biomass burning, urban, continental, desert, and maritime, which can directly impact the reflectance accuracy (BASSANI et al., 2015). In this study, the aerosol profile used was continental, but the mixture of urban and agricultural cultivation areas around the reservoir might impose a constraint for the standard aerosol use, limiting the radiative transfer simulation and the atmospheric effect removal. Furthermore, the AOD value used here was derived from PRISMA, which can be susceptive by several uncertainty sources (e.g., sensor calibration, cloud screen, aerosol models and surface properties), and may have increased the errors in 6SV outputs (LI et al., 2009).

ACOLITE, on the other hand, presents an image-based AOD estimate which retrieves the parameter using the DSF approach and requires a high SNR for the accurate quantification of the aerosol effects, which is well supplied by the OLCI SNR (ECK et al., 1999). Another major challenge is related to the spectral perturbation caused by light scattered into the sensor FOV by highly reflecting surfaces near the target water body, known as the adjacency effect (BULGARELLI; KISELEV; ZIBORDI, 2017). At wavelengths above 700 nm, the impact of the adjacency effect is particularly strong since neighboring land pixels show distinctly higher reflectance. ACOLITE, for instance, does not apply an adjacency correction approach, in which, considering the necessity of having dark pixels from shadowy land, adjacency pixels could contribute to the contamination of the reflectance of neighboring water pixels. Furthermore, the 12.6° camera tilt aboard OLCI leads to an Observation Zenith Angles (OZA's) near the edge of the swath reaching upwards of 60°. Increasing OZA from nadir to the edge of swath can potentially increase the amount of adjacency influence, with larger effects seen in shorter wavelengths due to increased Rayleight scattering (BULGARELLI; KISELEV; ZIBORDI, 2017). Thus, a correction of adjacent pixel appropriate for the water/land environment is essential for reliable spectra retrieval.

4.5.2 A method for the identification of cyano-dominant pixels

In the present study, a hybrid algorithm to estimate PC concentration was proposed, where a RF classifier was used as a mask to identify CyHABs pixels, switching among two bio-optical algorithms. The first step in this construction was the dataset partitioning based on the SAM. Both approaches use the R_{rs} spectra shapes to classify the samples, and could be directly used to mask the CyHABs pixels. However, these applications resulted in a very unbalanced dataset partitioning with high confusion between the classes in terms of PC concentration, making it impossible to calibrate the bio-optical models. In addition, SAM is based on the similarity of *in-situ* R_{rs} spectra shapes, and it is important to notice that the spectral range between 500 – 900 nm was used here to split the dataset. As reported by Silva et al. (2020), R_{rs} responds mostly to the backscattering and absorption properties of Chl-a, total suspended matter, and CDOM, resulting in different R_{rs} shapes and magnitudes. Thus, considering the application of SAM on OLCI spectra, the classifier may be sensitive to the influence of other components, or other groups of pigments that affect the spectra shape, and may not be related to the PC concentration. To better relate the SAM classification with the PC concentration, we applied the SAM iteratively, varying the reference spectra, and ensuring good separability between the data, as reported in Figure 4.7.

In order to minimize the diverse influences in the R_{rs} that affect the final classification, we used as input features to the RF classification math bands strictly related to the PC spectral feature. In addition, given the diversity of input datasets that could be used to improve classification, it is important to include only the most relevant datasets in order to reduce the computational burden without sacrificing the accuracy of the results. Filter methods are able to identify the most relevant features, removing redundant, noisy, and irrelevant variables. Here, we applied the PPS approach with Spearman's rank order correlation to determine pair-wise correlation between the features, reducing the error associated with the structural risk of including the entire feature vector in the analysis.

Among the AC processors tested, the CyHABs detection algorithm applied in the L2-WFR images achieved the worst performance, with a high number of false positives. The low accuracy observed can be strictly related to the input features used in the classifier. The AC analysis demonstrated that, in comparison with *in-situ* R_{rs}, 6SV and ACOLITE showed higher differences in magnitude, but a good similarity in the R_{rs} spectra shape. While L2-WFR removed the glint effect, but overestimate the R_{rs} in the green and red bands. Here, BR's, LH's, and multispectral algorithms were used as input for the RF classification model training. The selected features have shown to be able to avoid interference of R_{rs} intensity into the hybrid model conditional threshold and to not restrict the application of the hybrid model on the image to specific R_{rs} intensities.

Despite the efforts to choose the best feature set, another important point is to evaluate the sensitivity of the RF classification to sampling design and imbalanced training samples. Millard and Richardson (2015) reported that RF can be sensitive to spatial autocorrelation of the training classes and to the proportions of the different classes within the training samples. In general, RF classifier fails to cope with imbalanced training data and tends to favor the most representative classes (DALPONTE et al., 2013). In fact, the results showed a difficulty in correctly identifying pixels with CyHABs (false positives for L2-WFR, and false negatives for 6SV and ACOLITE), which can be mainly related to the lack of samples with higher PC concentration. Thus, our finding suggest that the RF classifier can serve as a good indicator to qualitatively classify waters for selecting the appropriate PC estimation algorithm for a given water pixel, but improvements can be applied, increasing the variability of PC concentration in the dataset.

4.5.3 A hybrid approach to estimate PC concentration

Each algorithm selected was developed based on different assumptions, and are applicable to waters where the assumptions were valid. Among the models tested, OGA19 achieved the best performance for the 'No CyHABs' class (ζ = 28.86%). OGA19 was specially designed to remove the contribution of Chl-a

absorption at 620 nm. The authors determined the PC contribution to the absorption at 665 nm and the Chl-a contribution to the absorption at 620 nm based on empirical relationships. This differs from the other algorithms where the absorption coefficient of phytoplankton at 665 nm is assumed to only depend on Chl-a, being used as a proxy for a_{chl-a}. It is well documented that, at low PC concentrations, the absorption by Chl-a can overlap the target wavebands for PC around 620-624 nm, leading to an overestimation of the estimates (RUIZ-VERDÚ et al., 2008).

In this sense, a good performance was also expected from the SIM05 and MM14 models, since both incorporate correction for Chl-a absorption. In fact, Ogashawara et al. (2013) assessed PC retrieval algorithms for a tropical reservoir. They concluded that the SIM05 and MM14 models had the best performances for the estimation of PC, being less sensitive to the interference of Chl-a. However, when compared to OGA19 model, the MM14 model was more sensitive to variations in Chl-a concentration. Regarding the SIM05, although the model showed good results for the 'No CyHABs' class (ζ = 77.9%), the PC retrieval uncertainties from SIM05 were higher than that of thesemi-empirical models tested (Table 4.5). This can be explained by the application of a constant value of a_{PC}^{*}, which may not be representative. It is well known that the specific absorption coefficient can vary across the lake or between phytoplankton communities due to differences in pigmentation, physiology, cell size, and pigment packaging effect (BRICAUD et al., 1995). This variability is even higher at low PC concentrations (< 10 mg.m^{-3}) (YACOBI et al., 2015), what may explain the poor model performance. In addition, SIM05 model assumes a linear relationship between the depth of the PC absorption feature at 620 nm and PC concentrations. However, several studies have shown that this relationship is no longer linear at low and high PC concentrations, (SIMIS et al., 2007).

For Cyano-dominant waters, the lowest uncertainties were achieved by LIU17, a four-band semi-analytical model, which isolates PC absorption from second-order variability caused by TSS and other phytoplankton pigment absorption. Similar to MM14 (a three-band algorithm), the model also accounts for the

variability in the backscattering that might be caused by the presence of non-phytoplankton particulates. In fact, the main difference between the models (LIU17 and MM14) is the presence of $\lambda_3 = 709\,nm$ that captures the influence of TSS in the PC absorption band. Furthermore, the LIU17 model was originally calibrated for turbid inland waters, and may be representative of the phytoplankton community in the Promissão Reservoir during the period of analysis. Despite being called semi-analytical by the authors, in its original form, the model is related to the PC concentration through a linear relationship, where the authors empirically define the slope and intercept values. Here, we calibrate these values from the PROM and BIL dataset. The entire set of characteristics makes the selected algorithms (OGA19 and LIU17) useful for different optical conditions, and the combination of these models has the potential to be an operational tool for effectively monitoring PC from satellite in dynamic waters.

Figure 4.12 demonstrates the final performance of the hybrid model for three different AC-processors. Figure 4.13 compares final PC estimates against the *in-situ* measurements taken during OLCI overpass. With the exception of L2-WFR corrected images, the majority of each scene was resolved with OGA19 algorithm, whilst only some specific cyano-dominant areas in the reservoir necessitate the use of the LIU17 algorithm. In response to the RF classification errors, the PC maps from L2-WFR do not correctly reproduce either spatial features of the PC distribution or quantitative estimates of their concentration. It is well known that imperfect corrections of the atmospheric effects typically result in further inaccuracies in the retrieved products depending on the sensitivity of the retrieval algorithm (MOSES et al., 2012). In fact, the final application of the hybrid model proved to be highly sensitive to variations in the spectrum shape, which makes the choice of the AC processor an extremely important step for the accuracy of the final PC estimate.

4.6 Conclusion

The optical properties of inland waters are very complex, with unique optical conditions. A single water body can also present different optical properties, both spatially and temporally (MOORE et al., 2014). These dynamic conditions

may prevent the application of a single global model to estimate optics components, making the hybrid approach a possible solution to deal with waters with unknown optical properties. We proposed here a hybrid algorithm to retrieve the PC concentration in inland water with a wide range of optical properties. Among five individual PC estimation algorithms, OGA19 and LIU17 models were selected for pixels with no presence and presence of CyHABs, respectively, to compose the hybrid model. A simple but robust method based on the RF algorithm for detecting cyano-dominant pixels was implemented to switch between the two algorithms.

To apply the final model to the OLCI images, three atmospheric correction models were initially tested: L2-WFR, 6SV, and ACOLITE. Although the OLCI sensor settings minimized the influence of the glint effect, an offset was observed on the 6SV (Bias = 173.02%) and ACOLITE (Bias = 110.84%) processors, possibly related to a residual glint effect. L2-WFR, on the other hand, was capable of removing the offset, achieving the best-observed results (Bias = 10.96%). However, higher errors were observed in the green and red bands, resulting in a significant shape difference when compared with the *in-situ* R_{rs} shape. More studies in different atmospheric conditions should be pursued to better validate the L2-WFR OLCI product.

The final application of the HM on OLCI images showed that the hybrid model was superior to the single algorithms when considering the entire range of optical properties, with an accuracy of 64.38% using ACOLITE. In general, the error sources include instrumental error measurement and measurement methodology, and the presence of pigments with overlapping light absorption features. In addition, as with all reflectance-based algorithms, the successful application of the hybrid model proved to be highly dependent on the accuracy of the applied atmospheric correction processor. In fact, the result analysis demonstrated that the final HM's application proved to be less sensitive to variations in magnitude, and more sensitive to variations in the spectra shape, i.e., neutral offset errors will possibly have a limited influence on the accuracy of PC estimates.

Our finding indicates that the proposed HM has the potential to be used as the initial step in the development of an operational tool for monitoring PC in waters with widely varying trophic conditions without the requirement of reparameterization. This technique thus offers a generally robust and adaptable approach to PC detection and facilitates the application of water type-appropriate algorithms over vast biomass ranges.

5 ASSESSMENT OF ESTIMATED PHYCOCYANIN CONCENTRATION FROM PRISMA AND SENTINEL-3 (OLCI) IN BRAZILIAN INLAND WATERS: A COMPARISON BETWEEN SEMI-ANALYTICAL AND MACHINE LEARNING ALGORITHMS

5.1 Introduction

Access to clean water is one of the 17 Sustainable Development Goals of the 2030 Agenda elaborated and agreed upon by the member countries of the United Nations, to which Brazil is a signatory (MIOLA; SCHILTZ, 2019). Despite the importance of continental aquatic systems to sustainable development, anthropogenic activities have been promoting the eutrophication of water bodies, leading to a significant increase in the frequency of potentially toxic CyHABs (AZEVEDO; VASCONCELOS, 2006). Regular and continuous CyHABs monitoring is necessary to assess, understand, and mitigate environmental changes, supporting informed decisions and evidence-based policies for the efficient use of our planet's resources (GIULIANI et al., 2017; LEHMAN et al., 2017). In this context, the use of remote sensing provides advantages to observe spatio-temporal dynamics, since it provides a synoptic and high temporal resolution view of a feature of interest, providing the basis for reliable and accountable scientific understanding and knowledge of the cyanobacterial dynamics (SHI et al., 2019; COFFER et al., 2021; SCHAEFFER et al., 2022).

An assertive indicator for the presence of CyHABs using Earth Observation data has been the PC, an auxiliary photosynthetic pigment with major presence in cyanobacteria (AHN et al., 2007). Due to PC-specific absorption features and by considering the challenging to retrieve water constituents in inland waters (AUGUSTO-SILVA et al., 2014; WATANABE et al., 2015; CAIRO et al., 2020), the remote sensing of freshwater cyanobacteria biomass has been largely focused on algorithms developed on hyperspectral reflectance (DEKKER, 1993; SCHALLES; YACOBI, 2000; KUTSER et al., 2004; SIMIS et al., 2007; LI et al., 2015). In particular, for the retrieval of PC, Simis et al. (2005) presented a nested band PC hyperspectral algorithm addressing the spectral influence of

Chlorophyll-a (Chl-a) in the PC absorption peak (620 nm), showing high performance in cyanobacteria-dominated temperate eutrophic waters (RUÍZ-VERDÚ et al., 2008; YAN et al., 2018; RIDDICK et al., 2019). More recently, machine learning (ML) algorithms have leveraged the prediction of PC from spaceborne imaging spectroscopy (O'SHEA et al., 2021) over a variety of inland water bodies.

However, when applied to hyperspectral satellite observations, the accuracy of the water quality parameters retrieval depends on the successful removal of the atmospheric contributions to the signal measured by the satellite sensor. The hyperspectral retrieved Remote sensing reflectance (R_{rs}) is strongly affected by several degrees of uncertainties related to a suboptimal SNR, the presence of sun-glint, and the adjacency effects (BRAGA et al., 2022). These influences require investigation and development of suitable correction models to minimize the uncertainties propagation on parameter estimates.

In this scenario, the use of PRISMA, a hyperspectral satellite mission of ASI, has been gaining attention for water quality mapping over the last years. Niroumand-Jadidi, et al. (2020); O'shea et al. (2021); Bresciani et al. (2022) demonstrated the use of PRISMA for retrieving water quality parameters and bottom properties in inland waters starting from R_{rs} data. The results highlighted the potential of PRISMA's fine and continuous spectral information offered in the retrieval of standard aquatic products, providing more detailed information about the OACs such as particle size, phytoplankton species, and functional types.

In this study, we evaluated the applicability and the accuracy of PRISMA hyperspectral imagery for mapping PC in Brazilian inland waters with concurrent field measurements and the state-of-the-art algorithms for retrieving both the R_{rs} and PC concentrations. Along with PRISMA Level 2 products, we assessed the performance of two atmospheric correction (AC) models. Then, as imagery data were affected by sunglint, three methods for mitigating the glint effects well consolidated in the literature were also tested. A semi-analytical model and two ML approaches to estimate PC concentrations were then tested

and compared to field measurements. Sentinel-3/OLCI has been added in the comparison as it is a frequent data source for inland water remote sensing. This work is expected to extend PRISMA products exploitation for mapping PC in continental waters in the tropics, allowing the improvement in the application of current and future satellite imaging spectroscopy missions for water quality monitoring in freshwater reservoirs.

5.2 Background

The main characteristic of spaceborne hyperspectral imagery is the availability of a full spectrum that enables the detection and identification of key biophysical properties (GIARDINO et al., 2019). In 2000, Hyperion, onboard National Aeronautics and Space Administration's Earth Observing-1 (NASA's EO-1) spacecraft, was the first imaging spectrometer to acquire image data from Earth in the 400-2500 nm spectral range at 30 m spatial resolution (PEARLMAN et al., 2003). In the next year, the Compact High-Resolution Imaging Spectrometer (CHRIS) was launched, acquiring 13 square km scenes in 18-62 user-selected VNIR spectral bands (400-1000 nm) (LAVIGNE et al., 2021). From 2009 to 2014, the Hyperspectral Imager for the Coastal Ocean (HICO) operated onboard the International Space Station (ISS), providing imagery focused on aquatic targets with a 90 m spatial resolution (BRAGA et al., 2013). Over the last decade, several hyperspectral missions have been launched, e.g., the Earth Sensing Imaging Spectrometer (DESIS) onboard the ISS (KRUTZ et al., 2019); the PRISMA sensor launched in 2019 by ASI (LOIZZO et al., 2018); and the German satellite mission DLR Environmental Mapping and Analysis Program (EnMAP) launched in 2022 (GUANTER et al., 2015). In the coming years, a new generation of hyperspectral missions is expected, including the Israeli/Italian Spaceborne Hyperspectral Applicative Land and Ocean Mission (SHALOM) (FEINGERSH; BEN-DOR, 2015), the Italian PRISMA Second Generation (FORMARO et al., 2021), the Copernicus Hyperspectral Imaging Mission for the Environment (CHIME), and the Plankton, Aerosol, Cloud, Ocean Ecosystem (PACE) developed for ocean to coastal water applications.

The capabilities of these spaceborne imaging spectrometers in aquatic environments have been demonstrated in several case studies (BRANDO; DEKKER, 2003; GIARDINO et al., 2007; IBRAHIM et al., 2018; O'SHEA et al., 2021; BRAGA et al., 2022). Niroumand-Jadidi et al. (2020) examined the potential of PRISMA level 2D images in retrieving standard water quality parameters (Chl-a, TSS and CDOM), comparing the estimates with Sentinel-2/MSI, which presents an important role in inland water applications. The authors observed a good agreement between the sensors. Similar findings were pointed out by Giardino et al. (2020), which demonstrated a high consistency between TOA radiance observed by PRISMA and Sentinel-2 over a wide range of water types. Bresciani et al. (2022) also explore the potential of hyperspectral observations presenting examples of water quality mapping based on PRISMA and DESIS imagery. The applications provided include the retrieval of Chl-a, TSM, CDOM, the mapping of fractional cover of bottom types, the assessment of phytoplankton pigment as well as of emergent macrophytes biomass.

5.3 Dataset

5.3.1 Field dataset

The field data considered in this study were gathered during two campaigns per-formed in October 2021 (PROM1) and April 2022 (PROM2), by covering 23 and 15 stations, respectively, and include both radiometric (R_{rs} resampled for PRISMA and OLCI bands) and phytoplankton (Chl-a and PC concentrations (Section 3.2.2) measurements (Table 3.1)).

5.3.2 Satellite imagery data

Three PRISMA and OLCI cloud free scenes of the Promissão Reservoir acquired on October 3rd, 2021, September 04th, 2021, and July 26th, 2022, were considered in this study. The images acquired on October 3rd, 2021 was co-located with four field stations visited within ± 1 day with PRISMA and OLCI overpass, and used to validate the atmospheric and glint correction models, and the PC algorithms. The images acquired on September 04th, 2021, and July 26th, 2022 were used to further apply the calibrated models to compare the PC estimates between hyperspectral and multispectral sensors.

The L1 OLCI product and PRISMA L1 images were atmospherically corrected using the L2-WRF Product, PRISMA L2C processor, 6SV, and ACOLITE (Section 3.3.3.1). For all AC models tested, the surface reflectance image was divided by π in order to obtain the atmospherically corrected R_{rs}.

5.3.2.1 Glint correction

In order to evaluate the glint influence on the reflectance spectrum, three different glint correction methods were applied in all PRISMA atmospherically corrected products to choose the best possible combination for the assessed sensors:

(i) Wang and Shi (2007) (WS07) method assumes that the reflectance values of SWIR bands come from the specular reflection in the water's surface (sun and sky glint), as the signal from this spectral region is considered negligible in natural inland waters (O'SHEA et al., 2020). Considering the 173 SWIR PRISMA channels (942 – 2496 nm), a SWIR range band centered near 1600 nm was considered as the reference bands for performing the glint correction. In principle, a band range centered near 2200 nm could also be selected as a reference. However, these bands are much noisier than 1600 nm (GAO et al., 2009). Therefore, in order to avoid the possible noise propagation to the glint-removed bands in the visible region, the 1600 nm SWIR band was selected as the reference band, where the average was considered and then subtracted of each band in the VNIR spectrum.

Hedley et al. (2005) (HED05) method assumes negligible water leaving signal in NIR part of the spectrum. Relative sun-glint intensity of the image is obtained based on the NIR brightness and the light in the visible band using a set of pixels, which could be homogeneous if not for the presence of glint. Establishing a linear relationship between the NIR band and each visible band allows the removal of the glint contribution. In case of hyperspectral data, there is a need to find the regression algorithm for each spectral band. HED05 method was used in a deglint processor implemented in the Sen2Coral toolbox available in the SNAP software (https://sen2coral.argans.co.uk/). Bands between 833 – 972 nm were considered as glint references.

Kutser et al. (2009) (KUT09) proposed an alternative glint removal procedure for hyperspectral imagery when SWIR data is not available. The method is based on the assumption that there is no spectral feature in the R_{rs} at 760 nm if it does not contain glint. Furthermore, it considers that the depth of oxygen absorption feature at 760 nm (called D) is proportional to the amount of glint in this pixel. The model assumes that pixels with D values close to zero do not contain glint, and pixels with the highest D value contain mainly glint. The glint spectrum is obtained by subtracting the pixel spectrum which D is close to zero from the spectrum with the highest D. In order to avoid land pixels or adjacency pixels, a water mask must be applied before the application of the model.

Atmospheric and glint correction models were assessed qualitatively by comparing in-situ R_{rs} measurements (resampled to PRISMA and OLCI bands) against satellite-derived surface R_{rs}, and quantitatively through the R^2, Bias, RMSE, MAPE and the SA (Section 3.3.3.3).

5.4 Phycocyanin and Chlorophyll-a modeling

Two approaches were used to retrieve PC and Chl-a concentrations from radiometric data: semi-analytical algorithms and ML models. They are outlined below.

5.4.1 Nested semi-analytical algorithm

The most widely used semi-analytical algorithm to estimate PC concentration was developed by Simis et al. (2005) (SIMIS05). The SIMIS05 algorithm assumes that PC dominates the absorption at 620 nm, with some spectral contribution of Chl-a (absorption peak at 665 nm). The algorithm attempts to isolate PC absorption by relating the reflectance at 620 nm with wavelengths without pigments absorption features (e.g., 709 nm and 779 nm) and removing the Chl-a spectral influence by empirically estimating the absorption of Chl-a at 620 nm from the 665 nm coefficient (SIMIS; PETERS; GONS, 2005):

$$a_{pc}(620) = \left[\frac{R_{rs}(709)}{R_{rs}(620)} * (a_w(709) + b_b) - b_b - a_w(620)\right] * \delta^{-1} - (\varepsilon * a_{chla}(665)), \qquad (5.1)$$

where a_w is the absorption by water ($a_w(620) = 0.281\, m^{-1}$); δ is a correction factor to relate the R_{rs} ratio to measured pigment absorption, and ε is a conversion factor that relates the absorption of Chl-a at 665 nm with its absorption at 620 nm. The backscattering (b_b) is assumed to be spectrally neutral and was derived by Simis et al. (2005) from R_{rs} at a single wavelength in the NIR. Here, we apply $b_b(700) = 0.1345$ estimated by *in-situ* measurements using HydroScat-6. Hydroscat-6 measures the scattering at a specific angle (140°), which is used to model the volumetric scattering function for estimating the backscattering coefficient (b_b). The Chl-a absorption at 665 nm was calculated by applying the Equation (5.2) developed by Gons (1999) (GON99):

$$a_{chla}(665) = \left[\frac{R_{rs}(709)}{R_{rs}(665)} * (a_w(709) + b_b) - b_b - a_w(665)\right] * \gamma^{-1} \qquad (5.2)$$

A correction factor (γ) was also introduced by Simis et al. (2005) to relate the R_{rs} ratio to measured Chl-a absorption. Finally, the PC and Chl-a concentrations were calculated by dividing the absorption coefficient ($a_{pc}(620)$; $a_{chla}(665)$) by the specific absorption coefficient ($a^*_{pc}(620)$; $a^*_{chla}(665)$):

$$[PC] = \frac{a_{pc}(620)}{a^*_{pc}(620)} \qquad (5.3)$$

$$[Chla] = \frac{a_{chla}(665)}{a^*_{chla}(665)} \qquad (5.4)$$

In order to achieve a better model performance, we reparametrized the SIMIS05 algorithm using *in-situ* data. First, the correction factors δ (Equation 5.1) and γ (Equation 5.2) were retrieved from *in-situ* $R_{rs}(\lambda)$ and $a_{phy}(\lambda)$ spectra. In this way, Equations 5.1 and 5.2 were initially applied considering $\delta = \gamma = 1$,

obtaining an 'uncorrected' $a_{chla}(665)$ and $a_{pc}(620)$. The linear least-squares fit of the uncalibrated-equations' derived absorptions against $a_{phy}(\lambda)$ measurements (for $\lambda = 665$ and $620\ nm$), yielded the slopes that were adopted as the new parameters ($\gamma = 0.14585$ and $\delta = 0.18055$). The conversion factor ε (Equation 5.1) was originally estimated in Simis et al. (2005) from the best fit of computed versus observed PC concentration. Our *in-situ* data were not used for calculating ε to avoid overfitting. Instead, O'Shea et al. (2021) applied a dataset of 939 samples to optimize the ε value by minimizing the median symmetric accuracy between the model estimates and the known *in-situ* concentration. The final value obtained ($\varepsilon = 0.251753$) was then applied here. Finally, the specific absorption coefficient for PC at 620 nm ($a^*_{pc}(620)$) and Chl-a at 665 nm ($a^*_{chla}(665)$) was determined from the absorption $a_{pc}(620)$ and $a_{chla}(665)$ divided by the measured PC and Chl-a concentration, respectively.

5.4.2 Machine learning models

O'Shea et al. (2021) calibrated a MDN for predicting PC based on *in-situ* radiometric measurements resampled for HICO and PRISMA sensors. The dataset consisted of 939 samples composed by *in-situ* R_{rs}, Chl-a, and PC measurements for water bodies worldwide. The MDN architecture uses as input features BRs, LHs, and multispectral PC algorithms based on their correlation with PC (500-550, 620, 650, and 710 nm) (all features used in the model are listed in O'Shea et al., 2021). The results obtained demonstrate the robustness of the model for estimating low PC concentrations (<10 mg m^{-3}) and the reduced impact of the uncertainties derived from orbital measurements of the R_{rs} on medium-to-high *in-situ* PC measurements (>10 mg m^{-3}). Considering the uniform accuracy for different concentration ranges and the broad number of water bodies used in the model calibration/validation, MDN was assumed to be a globally applicable algorithm. The entire *in-situ* R_{rs} (resampled to PRISMA bands) data and PRISMA images were used to predict PC and Chl-a using the MDN model. As the algorithm does not allow any extra training or parameters setting, MDN was applied in default mode as provided by the authors (https://github.com/STREAM-RS/STREAM-RS). The model took 26 PRISMA

bands (500-719 nm) as input and calculated PC and Chl-a concentration as final output.

Begliomini (2022) used radiometric *in-situ* R_{rs} resampled to PRISMA bands, and biological data (PC and Chl-a) measured in eight field campaigns in the Billings Reservoir – Brazil (115 samples) to calibrate/validate three different ML algorithms (Random Forest (RF), Extreme Gradient Boosting, and Support Vector Machine). The input features in the model calibration consisted of normalized indexes, LHs, and PC algorithms. A data-driven feature selection was applied using the Predictive Power Score to retain only relevant features for predicting PC. Then, the Pearson Coefficient was calculated for each pair of the filtered features, and a threshold was established to remove the most correlated layers. RF achieved the best performance among the assessed ML algorithms for orbital PC retrieval (MAE = 45%). Therefore, it was chosen to be evaluated in this study for the Promissão Reservoir. Following the application of MDN, the RF model also does not allow extra training. Thus, the entire dataset (*in-situ* R_{rs} resampled to PRISMA bands and PRISMA image) was applied in order to validate de PC estimates.

5.4.3 Performance indicators

The performances of each algorithm were assessed using *in-situ* R_{rs} resampled to PRISMA and OLCI bands, comparing them with satellite image R_{rs} within 1 day of *in-situ* PC and Chl-a measurements. The robustness of the relationship between the algorithm estimated values and measured values was evaluated by applying linear least squares regression analysis. The goodness of fit is reported by R^2 and the slope (*p*). Algorithm accuracy was quantified with measures of errors, including the median symmetric accuracy (ζ), which is equivalent to the median unsigned percentage error, and the MdAE. The first norm (ζ) initially transform data into log space and then convert them back to linear space to assess the quality of the retrieved quantity. This transformation ensures that the metric is symmetric, i.e., switch the values of the predicted and observed value will result in the same error (unlike MAPE). The interpretation of

the median symmetric accuracy is that 50% of the unsigned percentage errors are smaller than ζ (MORLEY; BRITO; WELLING, 2018).

Table 5.1 - Statistical error metrics, where $\boldsymbol{y_i}$ and $\boldsymbol{x_i}$ represent the modelled value and *in-situ* measured value, respectively.

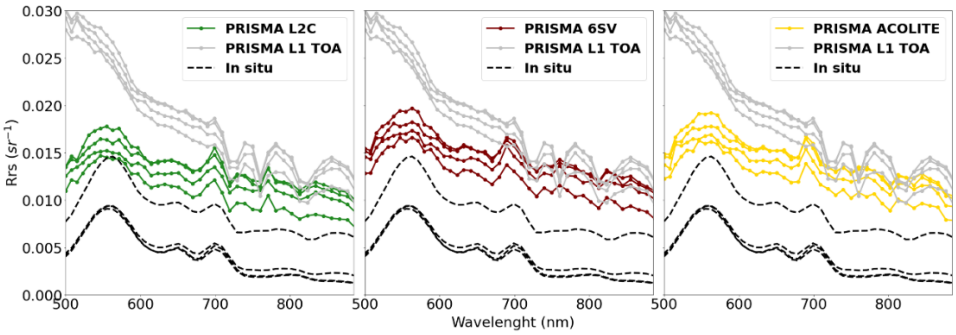
Median symmetric accuracy	$\zeta = 100 \times (\exp\left(\frac{\sum_{i=1}^{n}\left\lvert \log_e \frac{y_i}{x_i}\right\rvert}{n}\right) - 1)$
Median absolute error	$MdAE = median(\lvert y_i - x_i\rvert, \ldots, \lvert y_i - x_i\rvert)$

5.5 Results and discussion

5.5.1 Atmospheric and glint correction

Qualitative comparisons between TOA R_{rs}, AC-derived R_{rs} and *in-situ* R_{rs} are presented in Figure 5.1 with the aim of pointing out artefacts that may characterize either satellite or *in-situ* data. The lower bound (500 nm) was chosen to avoid commonly high uncertainties in R_{rs} within the blue region (BRAGA et al., 2022). Due to the presence of high glint in the PRISMA image for the date (October 3rd, 2021), an offset difference was highlighted in all AC models (Figure 5.1). Despite the offset, the three methods were capable of removing most of the atmospheric effects. Before AC, TOA R_{rs} spectrum presents an exponential behavior indicating dominance by Rayleigh and aerosol scattering effects. After the correction, the R_{rs} spectrum shows a shape close to that of the *in-situ* measurements. This can be mainly highlighted by the spectral angle of 20.21° by L2C, 19.9° by ACOLITE, and 20.22° by 6SV, which can indicate a good similarity between the spectral shapes of *in-situ* measurements and AC models. More significant differences between *in-situ* R_{rs} and AC-derived R_{rs} shapes occur around 700 nm, where a shift in the reflectance peak is observed in all models. This spectral shift is likely the result of inter-band calibration issues (LAVIGNE; RUDDICK, 2021), which also justifies the band-to-band variations in the PRISMA spectra.

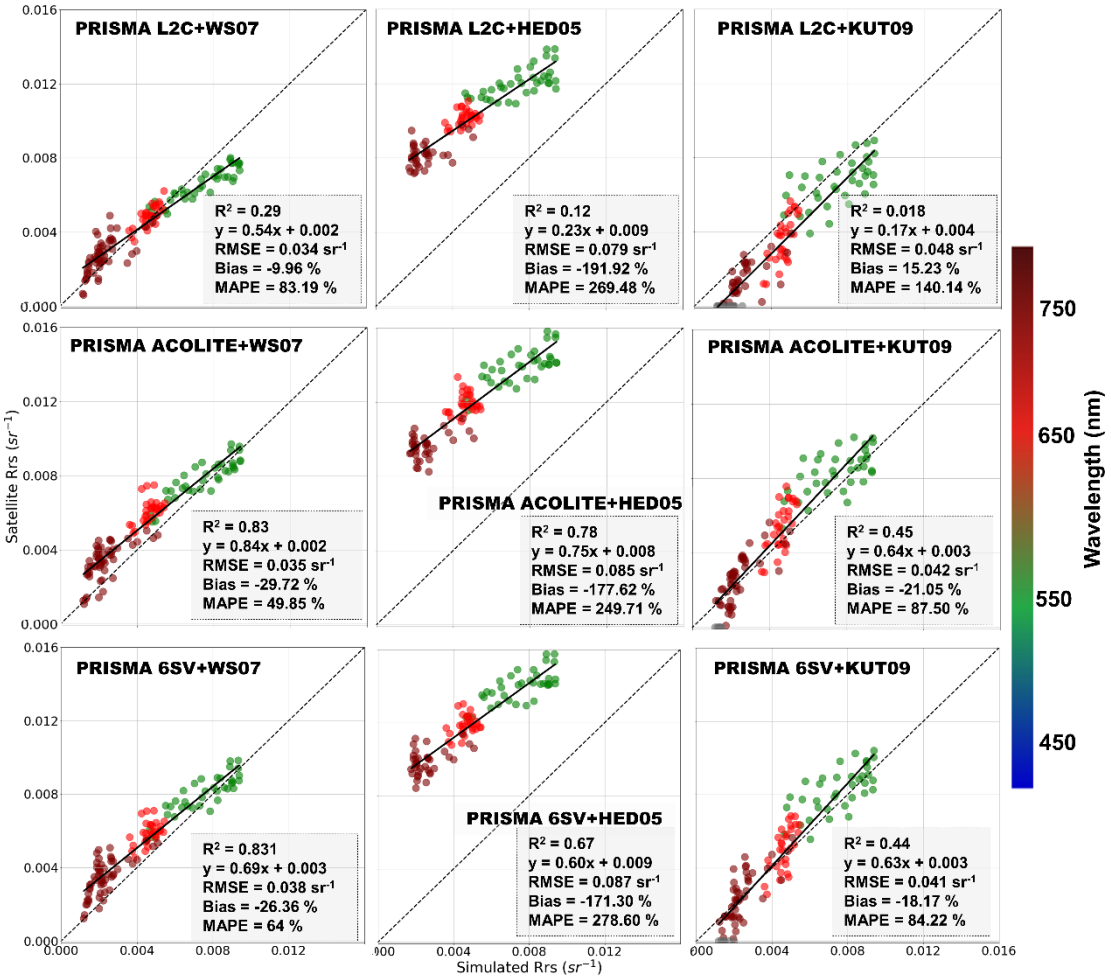
Figure 5.1 - Qualitative comparison of *in-situ* $\boldsymbol{R_{rs}}$ and PRISMA AC-corrected $\boldsymbol{R_{rs}}$ on October 3rd, 2021 for each atmospheric correction algorithm tested.



Author's elaboration.

For each AC method (L2C, ACOLITE, and 6SV), three glint correction approaches were evaluated (WS07, HED05, and KUT09). Figure 5.2 describes the algorithm combinations (atmospheric + glint correction methods) performances presenting the scatter plots between modeled and measured R_{rs}, and the mean Bias, RMSE and MAPE. These error metrics are also spectrally presented in Figure 5.3 , where the spectral range around PC absorption feature (588 – 679 nm) was highlighted. The best performance for all AC models was obtained in combination with WS07, which shows the potential of explore SWIR bands in the retrieval of the R_{rs}. In the PC absorption feature (rectangle in the Figure 5.3), the best result was obtained with L2C + WS07 combination. Since most bio-optical models that estimate PC apply R_{rs} mainly around 620 nm, PRISMA L2C product shows a high potential to deliver accurate R_{rs} products to assess PC concentration. Despite the good performance, hyperspectral sensors present low SNR, due to the trade-off between the narrow band-width of the spectral channels/spatial resolution and the required energy to illuminate detector elements (MOSES et al., 2017).

Figure 5.2 - Scatterplots of *in-situ* R_{rs} resampled for PRISMA bands versus estimated PRISMA R_{rs} using four atmospheric correction algorithms (L2C, ACOLITE, and 6SV) and the three glint methods (WS07, HED05, and KUT09) for the 500-885 nm spectral range. Black solid line is the best-fit linear regression. 1:1 line is represented by the dashed line. Negative counts (grey) are plotted as zero.

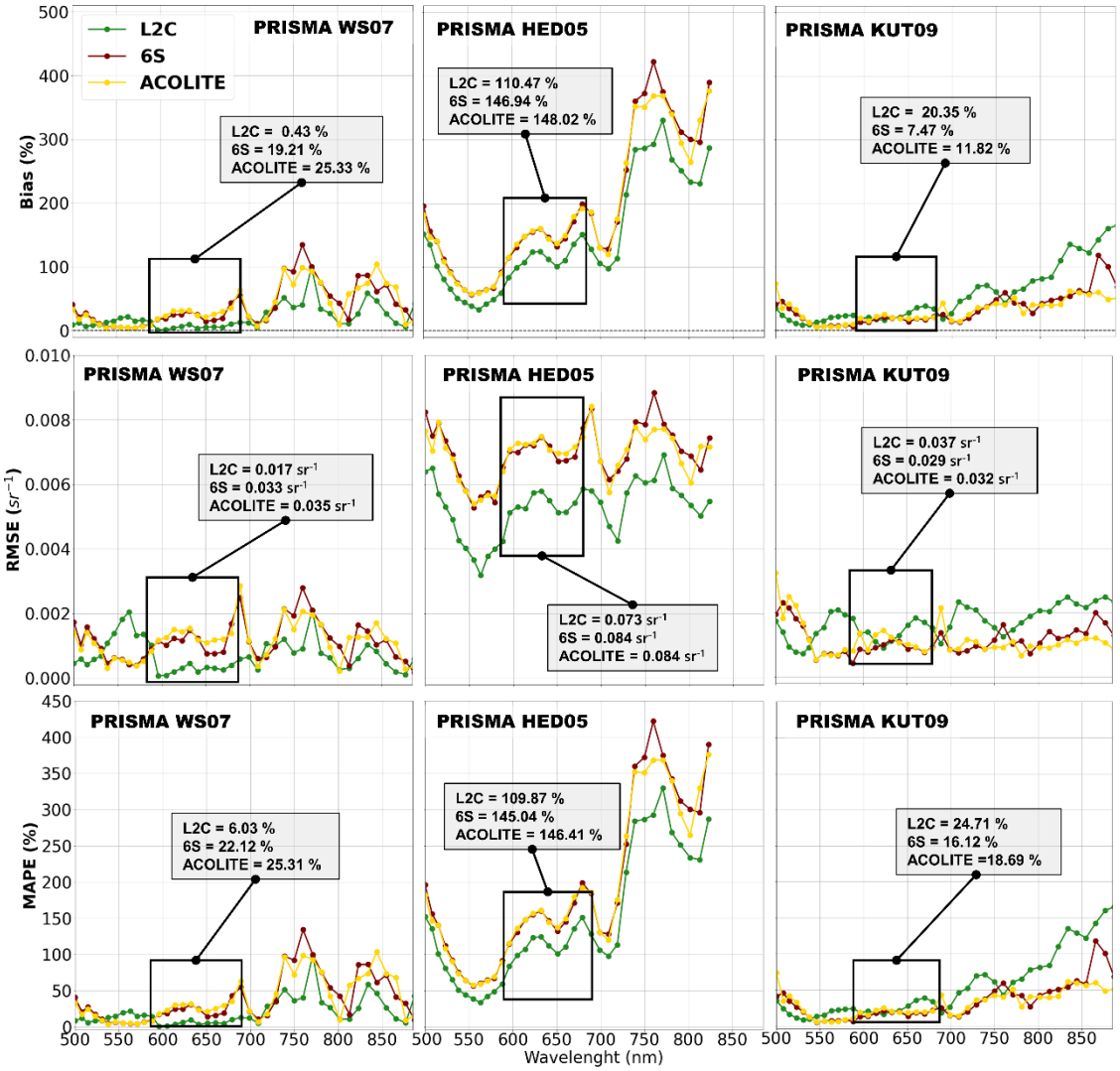


Author's elaboration.

Followed by WS07 model, KUT09 also presented a good agreement with the *in-situ* dataset (Figure 5.2), with better results achieved in combination with 6SV (Bias = -18.17%). In general, both glint correction models (WS07 and KUT09) showed higher errors in the NIR bands (700-813 nm), resulting in negative R_{rs} values for KUT09 model (Figure 5.3). These errors might traduce the reduced capacity of the assessed models to accurately predict the variations of the effects of atmospheric constituents in this range (FROUIN et al., 2019). Additionally, the glint overestimation observed with the KUT09 model might be a

result of the influence of other components besides glint. In fact, the application of HED5 and KUT09 models assumes that the brightness in the NIR region is composed only of sun-glint. However, this assumption is not valid for waters optically dominated by phytoplankton where high reflectance values are frequently observed in the NIR (KUTSER; VAHTMAE; PRAKS, 2009). In addition, the adjacency effects are often more expressive in NIR bands. The radiance increment led by adjacent pixels might also overestimate the glint effects in KUT09 model. Despite these further environmental effects over the NIR bands, HED05 model presented a higher overestimation of the R_{rs} for all spectral bands, suggesting an underestimation of the influence of the glint effect (Figure 5.2). HED05 model requires manual selecting the glinted areas through a visual inspection, which can be an important source of error if the selected area is not representative.

Figure 5.3 - Plots of the spectral statistic (bias, RMSE and MAPE) for the three glint correction methods (Wang and Shi (2007), Hedley (2005) and Kutser (2009)). For each model, bias, MAPE and RMSE around phycocyanin absorption feature (588 - 679 nm) is highlighted. For HED05 the reference bands used to estimate the glint effect are not corrected and, therefore, not considered in the results.

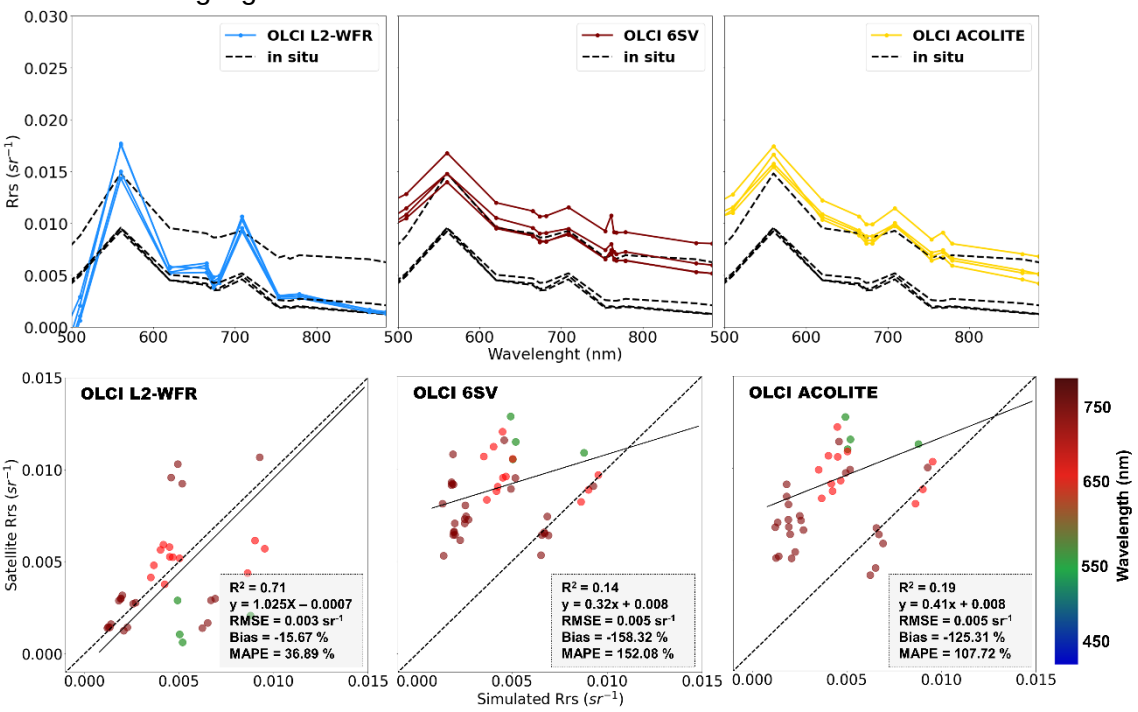


Author's elaboration.

The results of OLCI atmospherically corrected images are presented in Figure 5.4. Here, no glint correction model was applied to OLCI images. The spectral range of OLCI bands (400-1020 nm) does not allow the application of WS07 algorithm, which requires the availability of SWIR bands. The decametre-scale of OLCI sensor imposes challenges on the visual selection of the glint-affected areas in the HED05 model. Furthermore, the push-broom imaging spectrometer

of OLCI is tilted off-nadir in a westerly direction by 12.6°, in an attempt to mitigate the effects of glint (KRAVITZ et al., 2020). In general, all AC models tested were capable of preserving the *in-situ* R_{rs} spectral shape. Lowest error metrics were observed with L2-WFR model (Bias = -15.57%; MAPE = 36.89%). However, ACOLITE model achieved the highest spectral agreement with *in-situ* R_{rs} (SA = 25.83°). Despite that, a difference in the magnitude is highlighted, probably related to some remaining glint effect. This offset was also observed with 6SV, which also showed the worst performance (Bias = -158.32%; MAPE = 152.08%).

Figure 5.4 - First line: qualitative comparison between *in-situ* resampled $\boldsymbol{R_{rs}}$ and OLCI atmospheric-corrected $\boldsymbol{R_{rs}}$ (L2-WFR, 6SV and ACOLITE) for October 3rd, 2021. Second line: Scatterplots comparing *in-situ* resampled $\boldsymbol{R_{rs}}$ and atmospheric corrected OLCI $\boldsymbol{R_{rs}}$ (L2-WFR, 6SV and ACOLITE). Black solid line is the best-fit linear regression, and the dashed line represents the 1:1 line. Note the different axis scales used for the plots. For each atmospheric correction method, the mean Bias, MAPE and RMSE are highlighted.



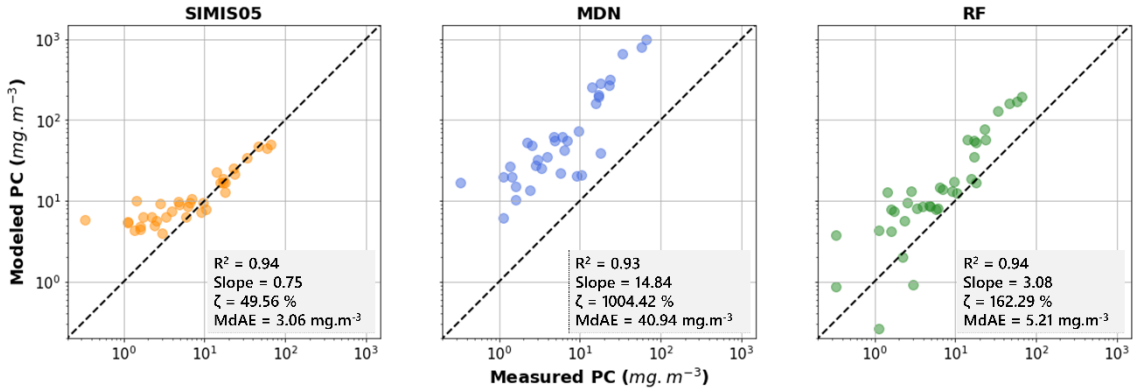
Author's elaboration.

5.5.2 Phycocyanin algorithm performance

The results of the modeled PC were compared to reference PC measurements and plotted in Figure 5.4 with associated error metrics. The dataset was log-transformed to reduce the data distribution's asymmetry (QIANG; XINDONG, 2006). Among the evaluated models to estimate PC concentration, the semi-analytical algorithm (SIMIS05) had the best fit ($R^2 = 0.94$) with the closest linear regression coefficient to the unit (slope = 0.75), and the lowest standard error (ζ = 49.56%, MdAE = 3.06 mg.m^{-3}). Despite the good performance, the SIMIS05 algorithm showed a tendency to overestimate PC at low concentration ranges. When considering only PC concentrations below 20 mg.m^{-3}, we observed a decrease in the accuracy metrics (ζ = 87.13% and MdAE = 3.15 mg.m^{-3}) which can be related to the high variation in the $a_{pc}^{*}(620)$ for this concentration range. In general, $a_{PC}^{*}(620)$ may vary in function of season, cell morphology, pigment compositions, changes in nutrient concentrations, and light conditions (SIMIS; KAUKO, 2012). As reported by Simis et al. (2007), the choice of $a_{PC}^{*}(620)$ highly affect PC concentration estimate. We considered an unique value of $a_{PC}^{*}(620)$ for the whole dataset range, what may have contributed to the increased errors in low pigment concentrations.

Furthermore, estimating low PC in diverse phytoplankton communities has been particularly challenging due to the presence of other optically relevant constituents, such as accessory pigments and CDOM. The most commonly influencing factor is the dominance of cyanobacteria over the entire phytoplankton biomass, which can be represented by the PC:Chl-a ratio. Li, Li and Song (2015) observed that for a PC:Chl-a < 0.5, the bio-optical models significantly increases the error estimates. Except for one sample, all other measurements from the Promissão Reservoir had PC:Chl-a < 0.5 (Table 3.1). The lack of spectral dominance of PC and the presence of other optically active constituents that absorbs electromagnetic radiation near 620 nm (e.g., Chl-a, Chl-b) might have reduced the accuracy of the parameters used for retrieving PC (e.g., $a_{pc}(620)$ in SIMIS05).

Figure 5.5 - Performance assessment of semi-analytical (Simis et al. (2005)), and machine learning algorithms (MDN and RF) using *in-situ* measured PC. The reported metrics are median symmetric accuracy (ζ), slope, regression coefficient (R2), and the median absolute error (MdAE).



Author's elaboration.

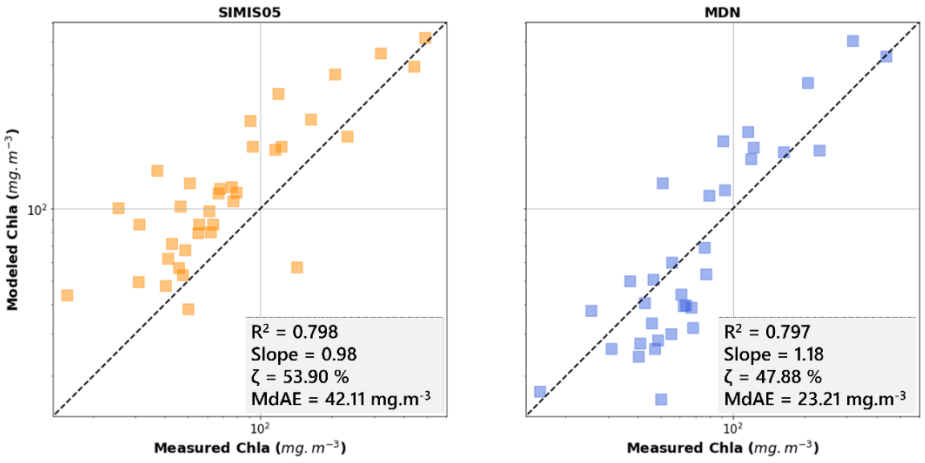
Both ML models (MDN and RF) highly overestimated PC values, increasing toward high pigment concentration. Despite that, the coefficients of determination were good for both models ($R^2_{MDN} = 0.93$ and $R^2_{RF} = 0.945$), where better accuracy was obtained with RF model (ζ = 162.29%, and MdAE = 5.21 mg.m^{-3}) compared to MDN (ζ = 1004.42%, and MdAE = 40.94 mg.m^{-3}). At low PC concentration (< 20 mg.m^{-3}), an improvement in both models was observed, mainly with the RF (ζ = 132.4%, and MdAE = 3.98 mg.m^{-3}). The RF model was calibrated based on a dataset from Billings Reservoir, which is part of the Tietê Cascade System (Figure 3.1). The dataset collected at the Billings Reservoir had a similar mean PC:Chl-a (0.14) when compared with Promissão (0.13). Those reduced values indicate a non-spectral dominance of PC, which increases the influence of Chl-a and other optically active constituents and might have increased the models' output errors.

MDN was calibrated using a wider range of water bodies, even though most samples were from few regions worldwide. The lack of samples from tropical inland waters might have reduced MDN's generalization capability. The training dataset also had a limited number of high PC concentrations (> 200 mg.m^{-3}), which might reduce the ability of MDN to predict accurately in this concentration range (O'SHEA et al., 2021). It ws already expected that MDN generalizes best on low PC (< 20 mg.m^{-3}), since it comprises the bulk of the training data.

However, the improvement observed for this concentration range was not significant (ζ = 923.46%, and MdAE = 28.57 mg.m^{-3} for PC < 20 mg.m^{-3}). The results suggest that the dataset used for calibrating MDN and RF was not able to capture the relationship between radiometric data and PC concentrations for the Promissão Reservoir. Therefore, the high variability in the performance of ML models outside the areas used for training should be considered when applied to new locations.

Considering Chl-a estimative (Figure 5.6), both algorithms tested (Gons, 1999 and MDN) performed with a good correlation with *in-situ* data ($R^2_{GONS} = 0.798$ and $R^2_{MDN} = 0.797$). Despite that, an overestimation was observed with the Gons model for all dataset (MdAE = 42.11 mg.m^{-3}). Unlike to the PC estimation, the MDN model performed well in the Chl-a estimation, at low and high pigment concentrations (ζ = 47.88%, and MdAE = 23.21 mg.m^{-3}).

Figure 5.6 - Performance assessment of semi-analytical (Simis et al (2005)), and machine learning algorithms (MDN) using *in-situ* measured Chl-a. The reported metrics are median symmetric accuracy (ζ), slope, regression coefficient (R2), and median absolute error (MdAE).



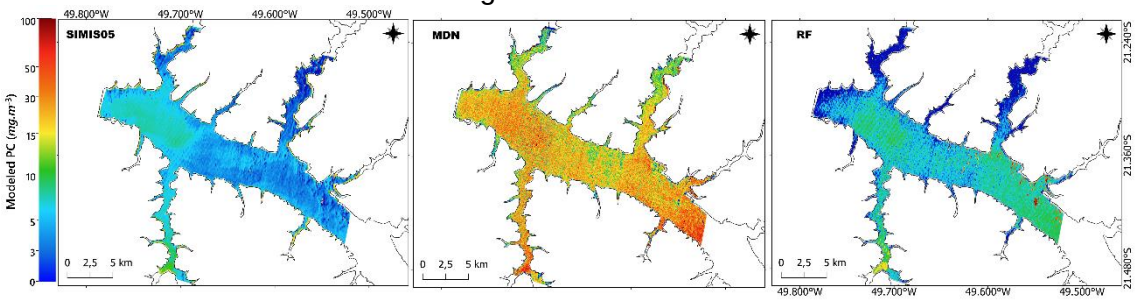
Author's elaboration.

5.5.3 Mapping PC: model assessment on satellite observations

Figure 5.7 shows PC concentration mapped with the three models applied to the PRISMA data corrected using L2C+WS07 combination. Comparisons

between the satellite-estimated PC against PC *in-situ* measurements are shown in Table 5.2. For PRISMA image, the spatial distribution of the predicted values was similar for all methods, where higher PC concentrations were observed in the main river (Figure 5.7). However, the magnitude of the estimates varies substantially. Following the *in-situ* results, all the models overestimated the PC concentrations (Table 5.2), specially MDN, which presented the worst MdAE (23.58 $mg.m^{-3}$), followed by RF (MdAE = 5.41 $mg.m^{-3}$). The SIMIS05 model achieved the best performance for PRISMA (MdAE = 3.18 $mg.m^{-3}$), which, based on the significant improvement in the PC estimates (Table 5.2), was used in the assessment of OLCI image, ensuring consistency between both satellites (Figure 5.8). The pigment concentration from the multispectral sensor is also shown in Table 5.2.

Figure 5.7 - PC product maps of Promissão Reservoir on October 3th, 2021 produced by Simis et al. (2005) model, and ML algorithms (MDN and RF) applied on PRISMA L2C+WS07 image.

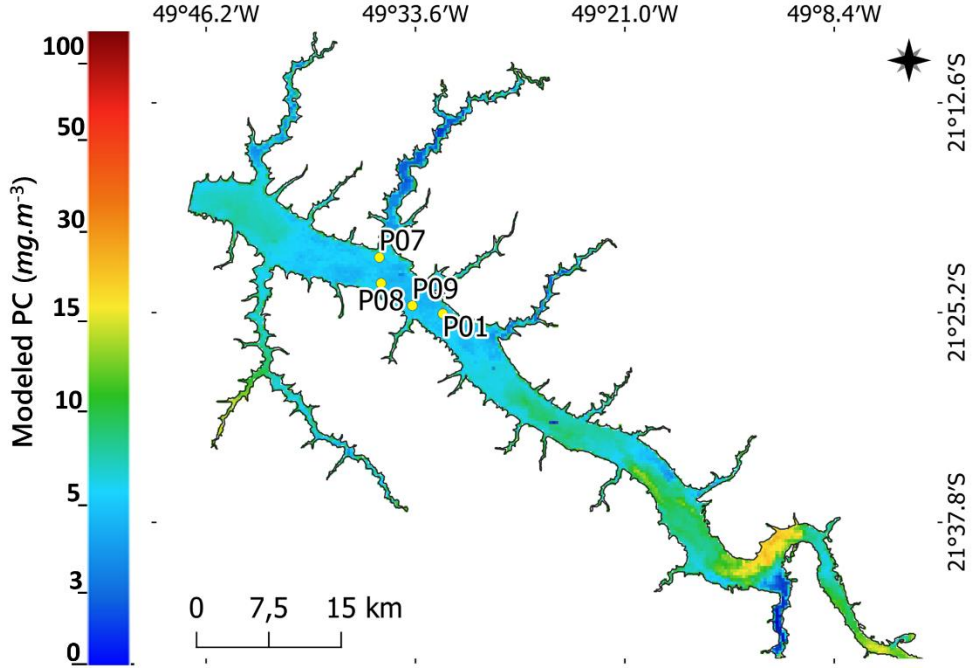


Author's elaboration.

The PC mapping based on the OLCI image (Figure 5.8) agreed to PRISMA PC using the SIMIS05 model, with a MdAE = 3.93 $mg.m^{-3}$. The PC estimated values (Table 5.2) indicated a good performance of the parameterized SIMIS05 model applied in a multispectral sensor with medium spatial resolution, even at low PC concentrations. In general, the behavior of the estimated PC from satellite imagery was expected, considering the previously *in-situ* analysis and the concentration range of the matchup samples (0.33 – 3.33 $mg.m^{-3}$). The observed errors might result from spatial and temporal misalignment between the *in-situ* measured and remotely estimated PC. Variations in the surface

cyanobacteria concentration can occur during the time window between field sampling and imaging (KUTSER, 2004). Furthermore, the vertical structure of the cyanobacterial bloom and dominant species can vary spatially and temporally, what affects the signal used for PC retrieval. Thus, spatial scale and temporal misalignment can result in large differences in the *in-situ* measured and remotely estimated PC.

Figure 5.8 - PC product map using SIMIS05 model in OLCI (AC corrected using ACOLITE) image on October 3rd, 2021, with associated *in-situ* matchup points (labels, which match with *in-situ* concentrations Table 5.2).



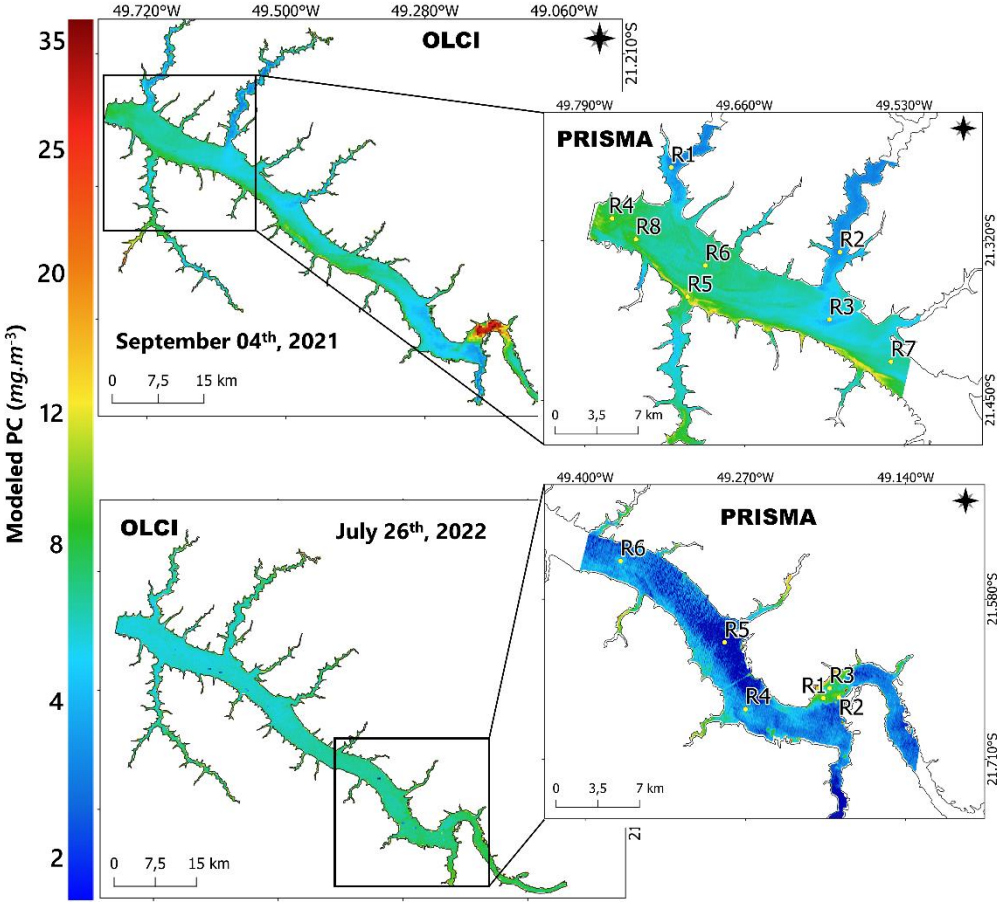
Author's elaboration.

Table 5.2 - *in-situ* measured PC (mg.m^{-3}) compared against retrieved PC (mg.m^{-3}) from PRISMA and OLCI images for October 3rd, 2021. The spatial distribution of the labeled stations is shown in Figure 5.8.

Station	*In-situ*	PRISMA			OLCI
		SIMIS05	MDN	RF	SIMIS05
P07	1.12	4.69	24.46	6.23	5.17
P08	1.12	3.89	38.68	5.09	4.93
P09	0.33	4.12	24.16	8.65	4.57
P01	3.33	4.07	24.97	9.05	4.69

Further comparisons between PRISMA and OLCI PC products maps are shown in Figure 5.9. For each considered date, Regions of Interest (ROI) were randomly selecteded, and the PC concentration estimate from each sensor was compared (Table 5.3). Despite the difference in the spatial resolution between OLCI and PRISMA, the visual inspection of the maps conveys an intuitively good agreement between the two sensors. The differences between the PC concentration from PRIMA and OLCI were, in general, in order of ~2 $mg.m^{-3}$ (Table 5.3). For both dates, higher differences in the pigment concentration occurred in pixels with small blooms spots (R5 on September 04th, 2021, and R2 on July 26th, 2022) that were not detected in the OLCI images. In fact, the 30 m spatial resolution of PRISMA allows for a finer spatial distribution of the pigment concentration, particularly along the tributaries rivers (reduced flooded areas) and discrete spots of blooms (high spatial frequency within small areas).

Figure 5.9 - Comparison between OLCI and PRISMA PC product maps using SIMIS05 model in Promissão Reservoir for (a) September 04th, 2021, and (b) July 26th, 2022.



Author's elaboration.

Table 5.3 - Comparison between SIMIS05 retrieved PC (mg.m^{-3}) from PRISMA and OLCI images. The spatial distribution of the labeled stations is shown in Figure 10. The station highlighted in green indicates the presence of bloom spots.

September 04th, 2021			**July 26th, 2022**		
ROI	PRISMA	OLCI	ROI	PRISMA	OLCI
R01	2.42	3.32	R01	7.15	5.45
R02	2.73	3.06	R02	25.52	7.42
R03	3.36	3.99	R03	4.20	5.56
R04	7.98	6.92	R04	4.76	6.14
R05	10.27	6.51	R05	0.85	6.52
R06	6.57	5.60	R06	3.52	5.91
R07	5.64	5.21	-	-	-
R08	8.17	6.64	-	-	-

5.6 Conclusions

Although different studies reveal the potential of PC as a proxy for cyanobacteria biomass, monitoring this pigment from space is limited by the availability of a spectral band centered near 620 nm. In this context, the hyperspectral sensors' capacity to measure across the full VNIR spectrum allows the optical constituents present in the water bodies to be better estimated (GIARDINO et al., 2019; DIERSSEN et al., 2021). However, the uncertainties in the hyperspectral reflectance resulted from instrument noise and the atmospheric correction process, can propagate to errors in the retrieved satellite products. These effects still require an amount of investigation to minimize the uncertainties propagation. In this study, we conducted a preliminary analysis of water reflectance derived from PRISMA hyperspectral sensor in a Brazilian reservoir. Above-water *in-situ* reflectances were used to evaluate different atmospheric and glint correction models. We also assessed different models (semi-analytical and ML) for PC estimate. Among the AC models tested in PRISMA and OLCI images, the L2C PRISMA product and L2-WFR product presented a better spectral similarity with the *in-situ* data in the presence of glint, respectively, for each sensor.

Regarding the application of glint correction in the PRISMA image, the best achievements, with AC algorithm, were derived from Wang and Shi (2007) model. PRISMA sensor provides measurements in the SWIR part of the spectrum, which could relieve some previous assumptions made for the NIR part of the spectrum. First, absorption of water molecules is more pronounced in the SWIR than in the NIR wavelength range by more than one order of magnitude, thereby reinforcing the assumption of 'black water' in the SWIR. Secondly, the atmosphere is more transparent with a weak amount of diffuse light (i.e., the lower contribution of aerosols and air molecules), making measurements more sensitive to the glint contribution at such wavelengths (HARMEL et al., 2018). As a result, 'deglinting' methods could be potentially improved based on the exploitation of the SWIR data. Despite the promising results, the match-up analysis presented in this study is very limited, requiring further studies to confirm our results and extend performance analysis to other

optical water types with a wider range of PC concentration. Furthermore, a robust method for retrieving atmospheric aerosol load for AC is still one of the biggest challenges of remote sensing in productive inland waters, and further studies are still needed to understand the impact of these uncertainties in pigment estimation.

The performances of two algorithm approaches (semi-analytical and ML) for retrieval PC were tested considering a dataset of pigment radiometric data collected in a Brazilian reservoir, and the effects of several potential sources of error were analyzed. These error sources include instrumental error, measurement methodology, and the presence of pigments with overlapping light absorption features. The nested semi-analytical model of Simis, Peters and Gons (2005) gave the best fit and accuracy for all concentration ranges, different from the ML models that significantly overestimated the pigment concentration. Our findings suggest that these errors could be markedly reduced by using site-specific adaptations of each model.

Finally, the comparisons between OLCI and PRISMA PC maps derived from Simis, Peters and Gons (2005) showed good agreement, with the advantage of PRISMA in detecting smaller spatial variations. Despite the spatial resolution of OLCI being one order of magnitude lower than PRISMA, its high revisit time always ensured a match-up, while providing 15 spectral bands potentially matching those of PRISMA. In fact, the synergic use of different sensors could improve the temporal analysis useful to understand, for example, cyanobacteria dynamics.

6 FINAL CONSIDERATIONS

The quantitative estimation of PC is essential for detecting and monitoring CyHABs in inland waters. Despite the advances in the development of bio-optical algorithms, the use of PC as a proxy for cyanobacteria biomass is still very error-prone, and it is important to understand the sources of these uncertainties. In this study, we provided detailed information about the PC retrieval from multispectral and hyperspectral data using different approaches. The research questions (*italic*) listed in section 1.2 are addressed below.

1. *Is the hybrid approach feasible to better estimate the space-time variation of PC concentration in an optically complex reservoir? What are the challenges in monitoring CyHABs using remote sensing?*

Our findings in Chapter 4 show that the combined use of different bio-optical models, calibrated for specific PC concentrations, presents distinct retrieval performance, with relative better results than the isolated application of these models. Despite the improved performance, it is important to highlight some limitations of the hybrid approach, mainly related to the detection of cyano-dominant waters. In general, the CyHABs detection model proved to be quite accurate when applied to *in-situ* data. However, when applied to orbital images, the model showed to be highly sensitive to the uncertainties in R_{rs} from the atmospheric correction. It is well known that imperfect corrections of the atmospheric effects typically result in further inaccuracies in the retrieved products. Here, the AC analysis demonstrated some limitations in the processors, mainly related to the residual glint effect and correct aerosol estimates, which directly affected the final PC estimates. In this sense, a robust method for retrieving atmospheric aerosol load for atmospheric correction was still one of the big challenges to assess the final water quality parameter accurately.

Additionally, despite the large dataset used in the model calibration (N = 153 samples), most of the samples were related to the absence of CyHABs. The unbalanced dataset can directly affect the random forest performance, reducing the ability of the model to learn the relationship between the input features and

the respective classes. Regarding the match-up samples with OLCI overpass, only two samples were classified as CyHABs, which limits a robust conclusion about the model accuracy. In fact, the lower values of PC:Chl-a ratio showed in Table 3.1 induce to consider the absence of Cyanobacteria species. However, taxonomical analysis in the Promissão Reservoir shows the dominant presence of cyanobacteria. In this way, an explanation might be related to the low extraction efficiency from the PC samples, which is an underestimation of the PC concentration. Despite the several limitations, the final HM presented good results in combination with 6SV and ACOLITE AC processors, overcoming previous limitations in the bio-optical model applicability.

2. *What are the gains of using hyperspectral data over multispectral data for predicting PC concentration?*

In Chapter 5, the results showed that the PC reflectance signature, mainly located at 620 nm, when compared to that of Chl-a, is both weaker and more difficult to sense directly with most sensors due to the interfering reflectance of Chl-a and other OACs (BECK et al., 2017). This interference makes the width and spacing of spectral bands for each imager important concerning avoiding other spectral components. In this sense, the OLCI spectral band located at 620 nm gives an advantage to the sensor, which combined with its high temporal resolution, allows the sensor to be used as a continuous CyHABs monitoring tool. Despite the spectral advantage, a potentially large source of error with medium-resolution sensors such as OLCI (300 m) is the influence of horizontal heterogeneity. Chl-a, for example, can vary by two orders of magnitude within a 300x300 m pixel (KUTSER, 2004). This influence can be visualized in Figure 5.8, where the PC concentration in the point R02 was underestimated using OLCI.

The hyperspectral PRISMA sensor, on the other hand, offers a significant advantage in terms of spectral and spatial resolution, which covers the application of orbital remote sensing to small water bodies. Furthermore, our results demonstrated that PRISMA data achieved the most accurate results for PC modeling, with a mean absolute error of 3.18 $mg.m^{-3}$. 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.

3. *Is Machine Learning approach able to predict PC more accurately than the semi-analytical models? What are the difficulties in developing a global model for estimating PC concentration?*

In Chapter 5, we compared the performance of the semi-analytical model developed by Simis et al. (2005), and two machine learning models (MDN and RF). The PC retrieval uncertainties were low with the application of the semi-analytical model, followed by the RF. The MDN had the worst prediction, highly overestimating the PC concentrations. The differences observed between the models are directly related to the dataset used for calibration, and show the difficulty in developing a global model for PC estimation. Among the ML models, the RF was initially calibrated in a reservoir that is part of the same system of the Promissão reservoir, where both presented similar bio-optical characteristics. The MDN model, on the other hand, despite having been calibrated with a large dataset, was not able to correctly estimate the relationship between the R_{rs} and the PC concentration.

Despite the results indicating the best performance for the semi-analytical model, the combination of ML models with hyperspectral data deserves attention as it allows exploring non-linear relationships between different spectral features and pigment concentration. However, the definition of a global PC model still has great challenges due to the numerous environmental influences in the cyanobacteria pigment production, and other optical components which directly affect the spectral features, making it difficult to generalize the estimates.

REFERENCES

AGENZIA SPAZIALE ITALIANA (ASI). **PRISMA products specification document**. 2021. Available from: http://prisma.asi.it/missionselect/docs/PRISMA%20Product%20Specifications_Is2_3.pdf. Accesss in: 24 Oct 2022.

AHN, C. Y.; JOUNG, S. H.; YOON, S. K.; OH, H. M. Alternative alert system for cyanobacterial bloom, using phycocyanin as a level determinant. **Journal of Microbiology**, v.45, n.2, p. 98-104, 2007.

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

AUGUSTO-SILVA, P.B.; OGASHAWARA, I.; BARBOSA, C.C.; DE CARVALHO, L.A.S.; JORGE, D.S.F.; FORNARI, C.I.; STECH, J. Analysis of MERIS reflectance algorithms for estimating chlorophyll-a concentration in a Brazilian reservoir. **Remote Sensing**, v. 6, n. 12, p. 11689-11707, 2014.

AZEVEDO, S. M. F. O.; VASCONCELOS, V. M. Toxinas de cianobactérias: causas e consequências para a saúde pública. In: ZAGATTO, P. A.; BERTOLETTI, E. (Ed.). **Ecotoxicologia aquática:** príncipios e aplicações. [S.l.]: Rima, 2006. p.433-452.

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: INPE, 2019. 178p. ISBN 978-85-17-0009-9.

BASSANI, C.; MANZO, C.; BRAGA, F.; BRESCIANI, M.; GIARDINO, C.; ALBEROTANZA, L. The impact of the microphysical properties of aerosol on the atmospheric correction of hyperspectral data in coastal waters. **Atmospheric Measurements Technique**, v. 8, p. 1593–1604, 2015.

BEGLIOMINI, F.N. **Cyanobacteria monitoring on urban reservoir using hyperspectral orbital remote sensing data and machine learning**. 2022. 104 p. Dissertation (Master in Remote Sensing) – Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2022.

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.

BENNETT, A.; BOGORAD, L. Complementary chromatic adaptation in a filamentous blue-green alga. **Journal of Cell Biology**, v. 58, n. 3, p. 419-435, 1973.

BERHANE, T.M.; LANE, C.R.; WU, Q.; AUTREY, B.C.; ANENKHONOV, O.A.; CHEPINOGA, V.V.; LIU, H. Decision-tree, rule-based, and random forest classification of high-resolution multispectral imagery for wetland mapping and inventory. **Remote Sensing**, v. 10, 2018. Available from: https://doi.org/10.3390/rs10040580.

BERNARDO, N.; ALCÂNTARA, E.; WATANABE, F.; RODRIGUES, T.; CARMO, A.; GOMES, A.; ANDRADE, C. Glint removal assessment to estimate the remote sensing reflectance in inland waters with widely differing optical Properties. **Remote Sensing**, v. 10, n. 10, 2018.

BORTOLLI, S.; PINTO, E. Cyanotoxins: general characteristics, history, legislation and methods of analysis. In: POMPEO, M. et al. **Ecology of reservoirs and interfaces.** São Paulo: Institute of Bioscience, University of São Paulo, 2015. p.321-339.

BRAGA, F.; FABBRETTO, A.; VANHELLEMONT, Q.; BRESCIANI, M.; GIARDINO, C.; SCARPA, G. M.; MANFÈ, G.; CONCHA, J. A.; BRANDO, V. E. Assessment of PRISMA water reflectance using autonomous hyperspectral radiometry. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 192, p. 99–114, 2022.

BRANDO, V.E.; DEKKER, A.G. Satellite hyperspectral remote sensing for estimating estuarine and coastal water quality. **IEEE Transaction 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. 8, 2022.

BRICAUD, A.; BABIN, M.; MOREL, A.; CLAUSTRE, H. Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: analysis and parameterization. **Journal of Geophysical Research: Oceans**, v. 100, p. 13321-13332, 1995.

BRIVIO, P.A.; GIARDINO, C.; ZILIOLI, E. Determination of chlorophyll concentration changes in Lake Garda using an image-based radiative transfer code for Landsat TM images**. International Journal of Remote Sensing**, v. 22, n. 2-3, p. 487 – 502, 2001.

BULGARELLI, B.; KISELEV, V.; ZIBORDI, G. Adjacency effects in satellite radiometric products from coastal waters: a theoretical analysis for the northern Adriatic Sea. **Applied Optics**, v. 56, p. 854-869, 2017.

CAIRO, C.T. **Caracterização temporal das propriedades bio-ópticas do reservatório de Ibitinga/SP**. Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2015.

CAIRO, C.; BARBOSA, C.; LOBO, F.; NOVO, E.; CARLOS, F.; MACIEL, D.; JNIOR, R. F.; SILVA, E.; CURTARELLI, V. Hybrid chlorophyll-a algorithm for assessing trophic states of a tropical Brazilian reservoir based on msi/sentinel-2 data. **Remote Sensing**, v. 12, n. 1, 2020.

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.

CARMICHAEL, W. W. Health effects of toxin-producing cyanobacteria: the CyanoHABs. **Human and Ecological Risk Assessment: An International Journal**, v. 7, n. 5, p. 1393–1407, 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.

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.

CHORUS, I.; BARTRAM, J. **Toxic cyanobacteria in water:** a guide to their public health consequences, monitoring, and management. London: E&FN Spon, 1999. 416p.

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.

COFFER, M. M.; SCHAEFFER, B. A.; DARLING, J. A.; URQUHART, E. A.; SALLS, W. B. Quantifying national and regional cyanobacterial occurrence in US lakes using satellite remote sensing. **Ecological Indicators**, v. 111, 2020.

COFFER, M. M.; SCHAEFFER, B. A.; FOREMAN, K.; PORTEOUS, A.; LOFTIN, K. A.; STUMPF, R. P.; DARLING, J. A. Assessing cyanobacterial frequency and abundance at surface waters near drinking water intakes across the United States. **Water Research**, v. 201, n. 117377, 2021;

COMPANHIA ENERGÉTICA DE SÃO PAULO (CESP)**. Caracterização limnológica do reservatório da UHE de Mário Lopes Leão, Promissão, SP**. São Paulo: CESP, 1989.

DA SILVA, E. F. F.; NOVO, E. M. L. M.; LOBO, F. L.; BARBOSA, C. C. F.; NOERNBERG, M. A.; ROTTA, L. H. S.; CAIRO, C. T.; MACIEL, D. A.; FLORES JÚNIOR, R. Optical water types found in Brazilian waters. **Limnology**, v. 22, n. 1, p. 57–68, 2021.

DALPONTE, M.; ORKA, H.O.; GOBAKKEN, T.; GIANELLE, D.; NAESSET, E. Tree species classification in boreal forests with hyperspectral data**. IEEE Transactions on Geoscience and Remote Sensing**, v. 51, p. 2632-2645, 2013.

DEKKER, A. G. **Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing**. 1993. Thesis (PhD – Research and Graduation internal) – Vrije Universiteit Amsterdam, Amsterdam, 1993.

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, 2021.

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

EUROPEAN SPACE AGENCY (ESA). **Sentinel-online: product types information**. 2022. Available from: https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-3-olci/product-types. Accessed in: Nov. 22 2022.

FEINGERSH, T.; BEN-DOR, E. Shalom–a commercial hyperspectral space mission. **Optical Payloads for Space Missions**, p. 247–263, 2015.

FORMARO, R.; LONGO, F.; VARACALLI, G.; FASANO, L.; PULCINO, V. ASI roadmap in technology and programmes for earth advanced monitoring and assessment of hazards. In: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM IGARSS, 2021, Belgium. **Proceedings…** 2021. p. 1875–1878.

ECK, T.F.; HOLBEN, B.N.; REID, J.S.; DUBOVIK, O.; SMIRNOV, A.; O'NEILL, N.T.; SLUTSKER, I.; KINNE, S. Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. **Journal of Geophysical Research**, v. 104, p. 31333–31349, 1999.

ELLIOT, J. A. Is the future blue-green? a review of the current model predictions of how climate change could affect pelagic freshwater cyanobacteria. **Water Research**, v.46, n.5, p.1364-1371, 2012.

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, p. 2017-224, 2009.

GELETEU, T. T. Lake eutrophication: control of phytoplankton overgrowth and invasive aquatic weeds. **Lakes & Reservoirs,** v. 28, 2023.

GEWALI, U. B.; MONTEIRO, S. T.; SABER, E. Machine learning based hyperspectral image analysis: a survey. **arXiv:1802.08701**, 2018.

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, ahievements and perspectives. **Surveys in Geophysics**, v. 40, p. 401-429, 2019.

GIARDINO, C.; BRANDO, V.E.; DEKKER, A.G.; STROMBECK, 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. et al. Imaging spectrometry of inland and coastal waters: state of the art, achievements and perspectives. **Survey Geophysics**, v. 40, p. 401-429, 2019.

GIARDINO, C.; BRESCIANI, M.; BRAGA, F.; FABBRETTO, A.; GHIRARDI, N.; PEPE, M.; GIANINETTO, M.; COLOMBO, R.; COGLIATI, S.; GHEBREHIWOT, S.; LAANEN, M.; PETERS, S.; SCHROEDER, T.; CONCHA, J. A.; BRANDO, V. E. First evaluation of prisma level 1 data for water applications. **Sensors** (Switzerland), v. 20, n. 16, p. 1–16, 2020.

GIULIANI, G.; DAO, H.; DE BONO, A.; CHATENOUX, B.; ALLENBACH, K.; DE LABORIE, P.; RODILA, D.; ALEXANDRIS, N.; PEDUZZI, P. Live Monitoring of Earth Surface (LiMES): a framework for monitoring environmental changes from Earth Observations. **Remote Sensing of Environment**, v. 202, p. 222–233, 2017.

GHOLIZADEH, M.; MELESSE, A.; REDDI, L.A. Comprehensive review on water quality parameters estimation using remote sensing techniques. **Sensors**, v. 16, n. 1298, 2016.

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.

GOMEZ-JAKOBSEN, F.; MERCADO, J.M.; COSTÉS, D.; RAMÍREZ, T.; SALLES, S.; YEBRA, L. A new regional algorithm for estimating chlorophyll-a in the Alboran Sean (Medterranean Sea) from MODIS-Aqua satellite imagery. **International Journal of Remote Sensing**, v. 37, p. 1431-1444, 2016. Available from: https://doi.org/10.1080/01431161.2016.1154223.

GONS, H. J. Optical teledetection of chlorophyllain turbid inland waters. **Environmental Science Technology**, v.33, n.7, p.1127-1132, 1999.

GORDON, H. R.; BROWN, O.B.; JACOBS, M.M. Computed relationships between the inherent and apparent optical properties of a flat homogeneous ocean. **Applied Optics**, v. 14, p. 417–427, 1975.

GORDON, H.R.; WANG, M. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. **Applied Optics**, v. 33, p. 443-452, 1994.

GUANTER, L.; KAUFMANN, H.; et al. The EnMAP spaceborne imaging spectroscopy mission for earth observation. **Remote Sensing**, v. 7, n. 7, p. 8830–8857, 2015.

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, p. 3479-3490, 2011.

HARMEL, T.; CHAMI, M.; TORMOS, T.; REYNAUD, N.; DANIS, P.A. Sunglint correction of the Multi-Spectral Instrument (MSI)-SENTINEL-2 imagery over inland and sea waters from SWIR bands. **Remote Sensing of Environment**, v. 204, p. 308–321, 2018.

HEDLEY, J.D.; HARBONE, A.R.; MUMBY, P.J. Simple and robust removal of sun glint for mapping shallow-water benthos. **International Journal of Remote Sensing**, v. 26, n. 10, p. 2107-2112, 2005.

HORNE, A.J.; GOLDMAN, C.R. **Limnology**. 2.ed. New York: McGraw Hill, 1994.

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. 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.; 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.

IBRAHIM, A.et al. Atmospheric correction for hyperspectral ocean color retrieval with application to the Hyperspectral Imager for the Coastal Ocean (HICO). **Remote Sensing of Environment,** v. 204, p. 60-75, 2018.

JEFFREY, S.W.; WRIGHT, S.W.; ZAPATA, M. Chlorophylls and carotenoids. In: ROY, S.; LLEWELLYN, C.A.; EGELAND, E.S.; JOHSEN, G. (Ed.). **Phytoplankton pigments:** characterization, chemotaxonomy and application in oceanography. New York, USA: Cambridge University Press, 2011. ISBN 978-1-107-00066-7.

JIANG, D.; MATSUSHITA, B.; YANG, W. A simple and effective method for removing residual reflected skylight in above-water remote sensing reflectance measurements**. ISPRS Journal of Photogrammetry and Remote Sensing**, v. 165, p. 16–27, 2020.

JIN, H. **Hyperparameter importance for machine learning algorithms**. 2022. Available from: <https://arxiv.org/abs/2201.05132v1>.

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.

KIMES, D.S.; KNYAZUKHIN, Y.; PRIVETTE, J.L.; ABUELGASIM, A.A.; GAO, F. Inversion methods for physically-based models. **Remote Sensing Reviews**, v.18, p. 381-439, 2000. Available from: https://doi.org/10.1080/02757250009532396.

KIRK, J. T. O. **Light and photosynthesis in aquatic ecosystems**. [S.l: s.n.], 2011. 665 p. ISBN 9780521151757.

KRUTZ, D. et al. The instrument design of the DLR earth sensing imaging spectrometer (DESIS). **Sensors**, v. 19, n. 7, 2019.

KORTING, T.S.; GARCIA, L.M.; CÂMARA, G. GeoDMA: geographic data mining analysts. **Computers & Geoscience**, v. 57, p. 133-145, 2013.

KUTSER, T. Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. **Limnology and Oceanography**, v. 49, n. 6, p. 2179-2189, 2004.

KUTSER, T.; VAHTMAE, E.; PRAKS, J. A sun glint correction method for hyperspectral imagery containing areas with non-negligible water leaving NIR signal. **Remote Sensing of Environment**, v. 113, p. 2267-2274, 2009.

KUTSER, T.; VAHTMAE, E.; PAAVEL, B.; KAUER, T. Removing glint effects from field radiometry data measured in optically complex coastal and inland waters. **Remote Sensing of Environment**, v. 113, p. 85-89, 2013.

LAVIGNE, H.; RUDDICK, K. Inter-band calibration for hyperspectral water remote sensing: Demonstration for Chris-Proba, In: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM IGARSS, 2021, Brussels. **Proceedings…** 2021. p. 7916–7919.

LAVIGNE, H.; VANHELLEMONT, Q.; RUDDICK, K.; DOGLIOTTI, A. New processor and reference dataset for hyperspectral CHRIS-PROBA images over coastal and inland waters. In: IEEE INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM IGARSS, 2021, Brussels. **Proceedings…** 2021. p. 7916–7919. Available from: https://doi.org/10.1109/IGARSS47720.2021.

LE, C.; LI, Y.; ZHA, Y.; WANG, Q.; ZHANG, H.; YIN, B. Remote sensing of phycocyanin pigment in highly turbid inland waters in Lake Taihu, China. **International Journal of Remote Sensing**, v. 32, n. 23, p. 8253–8269, 2011.

LEE, R.E. **Phycology**. 4.ed. Colorado: Cambridge University Pres, 2008. 561p. ISBN 978-0-521-68277-0.

LEHMANN, A.; CHAPLIN-KRAMER, R.; LACAYO, M.; GIULIANI, G.; THAU, D.; KOY, K.; GOLDBERG, G.; SHARP, R., JR. Lifting the information barriers to address sustainability challenges with data from physical geography and Earth observation. **Sustainability**, v. 9, n. 858, 2017.

LI, L.; LI, L.; SHI, K.; LI, Z.; SONG, K. A semi-analytical algorithm for remote estimation of phycocyanin in inland waters. **Science of the Total Environment**, v. 435–436, p. 141–150, 2012.

LI, L.; LI, L.; SONG, K. Remote sensing of freshwater cyanobacteria: an extended IOP Inversion Model of Inland Waters (IIMIW) for partitioning absorption coefficient and estimating phycocyanin. **Remote Sensing of Environment**, v. 157, p. 9–23, 2015.

LI, Z.; ZHAO, X.; KAHN, R.; MISHCHENKO, M.; REMER, L.; LEE, K.-H.; WANG, M.; LASZLO, I.; NAKAJIMA, T.; MARING, H. Uncertainties in satellite remote sensing of aerosols and impact on monitoring its long-term trend: a review and perspective. **Annales Geophysicae**, v. 27, p. 2755–2770, 2009.

LIU, G.; SIMIS, S.; 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. **IEE Transactions on Geoscience and Remote Sensing,** v. 56, n. 3, p. 1374-1385, 2017.

LOBO, F.L.; JORGE, D.S. Processamento de dados, modelagem e mapeamento de parâmetros bio-ópticos. In: BARBOSA, C.C.F.; NOVO, E.; MARTINS, V.S. (Ed.). **Introdução ao sensoriamento remoto de sistemas**

aquáticos: Princípios e aplicações. São José dos Campos: INPE, 2019. ISBN 978-85-17-0009-9.

LOIZZO, R.; GUARINI, R.; LONGO, F.; SCOPA, T.; FORMARO, R.; FACCHINETTI, C.; VARACALLI, G. PRISMA: The italian hyperspectral mission. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM ON OBSERVING, UNDERSTANDING AND FORECASTING THE DYNAMICS OF OUR PLANET (IGARSS), 2018, Valencia – Spain. **Proceedings…** 2018.

LONNEBORG, A.; LIND, L.K.; KALLA, S.R.; GUSTAFSSON, P.; OQUIST, G. Acclimation process in light-harvesting system of the cyanobacterium *Anacystis nidulans* following a light shift from white to red light. **Plant Physiology**, v. 78, p. 110-114, 1985.

LUNETTA, R.S.; SCHAEFFER, B.A.; STUMPF, R.P.; KEITH, D.; JACOBS, S.A.; MURPHY, M.S. Evaluation of cyanobacteria cell count detection derived from MERIS imagery across the eastern USA. **Remote Sensing of Environment**, v. 157, p. 24-34, 2014.

MACIEL, D.; NOVO, E.; DE CARVALHO, L. S.; BARBOSA, C.; 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, 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.

MATSUSHITA, B.; YANG, W.; YU, G.; OYAMA, Y.; YOSHIMURA, K.; FUKUSHIMA, T. A hybrid algorithm for estimating the chlorophyll-a concentration across different trophic states in Asian inland waters. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 102, p. 28–37, 2015.

MATTHEWS, M. W. Eutrophication and cyanobacterial blooms in South African inland waters: 10years of MERIS observations. **Remote Sensing of Environment**, v. 155, p. 161–177, 2014.

MEREL, S. et al. State of knowledge and concerns on cyanobacterial blooms and cyanotoxins. **Environment International**, v.59, p.303-327, 2013.

MILLARD, K.; RICHARDSON, M. On the importance of training data sample selection in random forest image classification: a case study in Peatland Ecosystem Mapping. **Remote Sensing**, v. 7, 2015.

MIOLA, A.; SCHILTZ, F. Measuring sustainable development goals performance: how to monitor policy action in the 2030 Agenda implementation? **Ecological Economics**, v. 164, 2019.

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

MISHRA, S.; MISHRA, D.R.; SCHLUCHTER, W.M. A novel algorithm for predicting phycocyanin concentrations in cyanobacteria: a proximal

hyperspectral remote sensing approach. **Remote Sensing**, v.1, p.758-775, 2009.

MOBLEY, C. D. **Light and water:** radiative Trasnfer in natural waters. [S.l.]: Academic Press, 1994. 592 p.

MOBLEY, C. D. Estimation of the remote-sensing reflectance from above-surface measurements. **Applied optics**, v. 38, n. 36, 7442–7455, 1999.

MOBLEY, C. D. Polarized reflectance and transmittance properties of windblown sea surfaces. **Applied Optics**, v. 54, n. 15, p. 4828–4849, 2015.

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.

MOSES, J.W.; STECKX, S.; MONTES, M.J.; KEUKELAERE, L.D.; KNAEPS, E. Atmospheric correction for inland waters. In: MISHRA, D.R.; OGASHAWARA, I.; GITELSON, A.A. (Ed.). **Bio-optical modeling and remote sensing of inland waters**. United States: Elsevier, 2017. 334p. ISBN 978-0-12-804644-9.

MUELLER, J. L.; AUSTIN, R. W. **SeaWiFS Technical Report Series**. Washington: NASA, 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, n. 23, 2020.

OGASHAWARA, I. Determination of phycocyanin from space: a bibliometric analysis. **Remote Sensing**, n. 1, p. 1–16, 2020.

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, n. 10, p. 4774–4798, 2013.

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

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.

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

OUAIDRARI, H.; VERMOTE, E.F. Operational atmospheric correction of Landsat TM data**. Remote Sensing of Environment**, v. 70, p. 4–15, 1999.

OWODUNNI, A.A.; ISMAIL, S.; KURNIAWAN, S.B.; AHMAD, A.; IMRON, M.; ABDULLAH, S.R.S.; A review on revolutionary technique for phosphate removal

in wastewater using green coagulant. **Journal of Water Process Engineering,** v. 54, 2023.

PAERL, H.M.; HUISMAN, J. Blooms like it hot. **Science**, v. 320, n. 57, 2008.

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.

PAERL, H.W.; OTTEN, T.G. Harmful cyanobacterial blooms: causes, consequences, and controls. **Environmental Microbiology**, v.65, n.4, p.995-1010, 2013.

PAHLEVAN, N.; MANGIN, A.; BALASUBRAMANIAN, S. V.; SMITH, B.; ALIKAS, K.; ARAI, K.; BARBOSA, C.; BÉLANGER, S.; BINDING, C.; BRESCIANI, M.; GIARDINO, C.; GURLIN, D.; FAN, Y.; HARMEL, T.; HUNTER, P.; ISHIKAZA, J.; KRATZER, S.; LEHMANN, M. K.; LIGI, M.; MA, R.; MARTIN-LAUZER, F. R.; OLMANSON, L.; OPPELT, N.; PAN, Y.; PETERS, S.; REYNAUD, N.; SANDER DE CARVALHO, L. A.; SIMIS, S.; SPYRAKOS, E.; STEINMETZ, F.; STELZER, K.; STERCKX, S.; TORMOS, T.; TYLER, A.; VANHELLEMONT, Q.; WARREN, M. ACIX-Aqua: A global assessment of atmospheric correction methods for Landsat-8 and Sentinel-2 over lakes, rivers, and coastal waters. **Remote Sensing of Environment**, v. 258, 1 jun. 2021.

PAULINO, R.; MARTINS, V.S.; NOVO, E.M.L.; 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. 1829, 2022.

PEARLMAN, J.S.; BARRY, P.S.; SEGAL, C.C.; SHEPANSKI, J.; BEISO, D.; CARMAN, S.L. Hyperion, a space-based imaging spectrometer. **IEEE Transactions on Geosciences and Remote Sensing**, v. 41, n.6, p. 1160–1173, 2003.

PELLOQUIN, C.; NIEKE, J. **Sentinel-3 OLCI and SLSTR simulated spectral response function**.[S.l.]: ESA, 2012.

PITOIS, S.; JACKSON, M.; WOOD, B. Problems associated with the presence of cyanobacteria in recreational and drinking waters. **International Journal of Environmental Health Research**, v.10, p.203-218, 2000.

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, 2019.

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.

RAMESAN, R.; MATHEW, J. Model data selection and data pre-processing approaches: hydrological data driven modeling. In: BLASIUS, B.; SOLOMATINE, D.P. (Ed.). **Earth systems data and models.** [S.l.]: Springer, 2014.

RANDOLF, K.; WILSON, J.; TEDESCO, L.; LI, L.; PASCUAL, D.L.; SOYEUX, E. Hyperspectral remote sensing of cyanobacteria in turbid productive water using optically active pigments, chlorophyll a and phycocyanin. **Remote Sensing of Environment**, v. 112, p. 4009-419, 2008.

RAY, A.K.; CHATURVEDI, R.; KUMAR, A. Proteomic vidences for microcystin-RR-induced toxicological alteration in mice liver. **Scientific Reports,** v. 8, 2018.

RENOSH, P. R.; DOXARAN, D.; DE KEUKELAERE, L.; GOSSN, J. I. Evaluation of atmospheric correction algorithms for sentinel-2-MSI and sentinel-3-OLCI in highly turbid estuarine waters. **Remote Sensing**, v. 12, n. 8, 2020.

RIDDICK, C. A. L.; HUNTER, P. D.; DOMÍNGUEZ GÓMEZ, J. A.; MARTINEZ-VICENTE, V.; PRÉSING, M.; HORVÁTH, H.; KOVÁCS, A. W.; VÖRÖS, L.; ZSIGMOND, E.; TYLER, A. N. Optimal cyanobacterial pigment retrieval from ocean colour sensors in a highly turbid, optically complex lake. **Remote Sensing**, v. 11, n. 13, 2019.

ROESLER, C. S. Theoretical and experimental approaches to improve the accuracy of particulate absorption coefficients derived from the quantitative filter technique, **Limnology and Oceanography**, v. 43, n. 7, p. 1649-1660, 2003.

ROYER, A.; VINCENT, P.; BONN, F. Evaluation and correction of viewing angle effects on satellite measurements of bidirectional reflectance. **Photogrammetric Engineering and Remote Sensing**, v. 51, n. 12, p. 1899–1914, 1985.

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.

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.

SCHAEFFER, B. A.; URQUHART, E.; COFFER, M.; SALLS, W.; STUMPF, R. P.; LOFTIN, K. A.; WERDELL, P. J. Satellites quantify the spatial extent of cyanobacterial blooms across the United States at multiple scales. **Ecological Indicators**, v. 140, e108990, 2022.

SCHALLES, J.; YACOBI, Y. Z. **Remote detection and seasonal patterns of phycocyanin, carotenoid and chlorophyll pigments in eutrophic waters**. 2000.Available from: <https://www.researchgate.net/publication/285636096>.

SCHINDLER, D. W. Recent advances in the understanding and management of eutrophication**. Limnology and Oceanography**, v. 51, n. 1 II, p. 356–363, 2006.

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.

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. **Science Direct**, v. 106, p. 414–427, 2007.

SMITH, M. E.; ROBERTSON LAIN, L.; BERNARD, S. An optimized Chlorophyll a switching algorithm for MERIS and OLCI in phytoplankton-dominated waters. **Remote Sensing of Environment**, v. 215, p. 217–227, 2018.

SONG, K.; LI, L.; LI, S.; TEDESCO, L.; HALL, B.; LI, Z. Hyperspectral retrieval of phycocyanin in potable water sources using genetic algorithm–partial least squares (GA–PLS) modeling. **International Journal of Applied Earth Observation and Geoinformation**, v. 18, p. 368–385, 2012.

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, 1 mar. 2018.

STEINMETZ, F.; RAMON, D. Sentinel-2 MSI and Sentinel-3 OLCI consistent ocean colour products using POLYMER. In: SPIE ASIA-PACIFIC REMOTE SENSING, 2018, Honolulu, Hawaii. **Proceedings…** 2018. Available from: https://doi.org/10.1117/12.2500232.

STRAMSKI, D.; REYNOLDS, R. A.; KACZMAREK, S.; UITZ, J.; ZHENG, G. Correction of pathlength amplification in the filter-pad technique for measurements of particulate absorption coefficient in the visible spectral region. **Applied Optics**, v. 54, n. 22, p. 6763-6782, 2015.

STUMPF, R. P.; DAVIS, T. W.; WYNNE, T. T.; GRAHAM, J. L.; LOFTIN, K. A.; JOHENGEN, T. H.; GOSSIAUX, D.; PALLADINO, D.; BURTNER, A. Challenges for mapping cyanotoxin patterns from remote sensing of cyanobacteria. **Harmful Algae**, v. 54, p. 160-173, 2016.

SUN ET AL. Estimating phycocyanin pigment concentration in productive inland waters using Landsat measurements: a case study in Lake Dianchi. **Journal of Geophysical Research: Ocean**s, v. 115, n. 4, 2010.

TASSAN, S.; FERRARI, G.M. A sensitivity analysis of the 'Transmittance-Reflectance' method for measuring light absorption by aquatic particles. **Journal of Plankton Research**, v. 24, n. 8, p. 757-774, 2002.

TONG, Y.; ZHANG, W.; WANG, X.; COUTURE, R. M.; LARSSEN, T.; ZHAO, Y.; LI, J.; LIANG, H.; LIU, X.; BU, X.; HE, W.; ZHANG, Q.; LIN, Y. Decline in Chinese lake phosphorus concentration accompanied by shift in sources since 2006. **Nature Geoscience**, v. 10, n. 7, p. 507–511, 2017.

TUNDISI, J. G. Recursos hídricos no futuro: problemas e soluções. **Estudos Avançados**, v.22, n.63, 2008.

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.

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

VEERABADHRAN, M. et al. State-of-the-art review on the ecotoxicology, health hazards, and economic loss of the impact of microcystin and their ultrastructural cellular changes. **Aquatic Toxicology,** v. Pre-proof, 2023. Available from: https://doi.org/10.1016/j.aquatox.2023.106417.

VERMOTE, E.; HERMAN, M.; MORCRETTE, J.-J.; TANRÉ, D.; DEUZÉ, J. L.; HERMAN, M.; MORCRETTE, J. J.; KOTCHENOVA, S. Y. Second simulation of a satellite signal in the solar spectrum-vector (6SV): an overview. **IEEE Transactions on Geoscience and Remote Sensing**, v. 35, p. 675-586, 1997.

XIA, R.; ZHANG, Y.; CRITTO, A.; WU, J.; FAN, J.; ZHENG, Z.; ZHANG, Y. The potential impacts of climate change factors on freshwater eutrophication: implications for research and countermeasures of water management in China. **Sustainability** (Switzerland), v. 8, n. 3, 2016.

WANG, M.; SHI, W. The NIR-SWIR combined atmospheric correction approach for MODIS ocean color data processing. **Optics Express**, v. 15, p. 15722–1573, 2007.

WATANABE, F.S.Y.; ALCÂNTARA, E.; RODRIGUES, T.W.P.; IMAI, N. N.; BARBOSA, C. C. F.; ROTTA, L. H. S. Estimation of chlorophyll-a concentration and the trophic state of the Barra Bonita Hydroelectric Reservoir using OLI/Landsat-8 images. **International Journal of Environmental Research and Public Health**, v. 12, p. 10391-10417, 2015.

WILSON, R. T. Py6S: A Python interface to the 6S radiative transfer model. **Computers and Geosciences**, v. 51, p. 166-171, 2013.

WYNNE, T.T.; TOMLINSON, M.C.; WARNER, R.A.; TESTER, P.A.; DYBLE, J.; FAHNENSTEIL, G.L. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. **International Journal of Remote Sensing,** v. 29, p. 3665-3672, 2008.

YACOBI, Y. Z.; KÖHLER, J.; LEUNERT, F.; GITELSON, A. Phycocyanin-specific absorption coefficient: eliminating the effect of chlorophylls absorption. **Limnology and Oceanography: Methods**, v. 13, n. 4, p. 157–168, 2015.

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.

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