



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA,
INOVAÇÕES E COMUNICAÇÕES



sid.inpe.br/mtc-m21c/2020/03.03.16.25-TDI

USING OPTICAL WATER TYPES FOR SATELLITE MONITORING OF BRAZILIAN INLAND WATERS

Edson Filisbino Freire da Silva

Master's Dissertation of the Graduate Course in Remote Sensing, guided by Drs. Evlyn Márcia Leão de Moraes Novo, and Felipe de Lucia Lobo, approved in February 27, 2020.

URL of the original document:

<<http://urlib.net/8JMKD3MGP3W34R/4247NRS>>

INPE
São José dos Campos
2020

PUBLISHED BY:

Instituto Nacional de Pesquisas Espaciais - INPE
Gabinete do Diretor (GBDIR)
Serviço de Informação e Documentação (SESID)
CEP 12.227-010
São José dos Campos - SP - Brasil
Tel.:(012) 3208-6923/7348
E-mail: pubtc@inpe.br

**BOARD OF PUBLISHING AND PRESERVATION OF INPE
INTELLECTUAL PRODUCTION - CEPPII (PORTARIA Nº
176/2018/SEI-INPE):****Chairperson:**

Dra. Marley Cavalcante de Lima Moscati - Centro de Previsão de Tempo e Estudos
Climáticos (CGCPT)

Members:

Dra. Carina Barros Mello - Coordenação de Laboratórios Associados (COCTE)
Dr. Alisson Dal Lago - Coordenação-Geral de Ciências Espaciais e Atmosféricas
(CGCEA)
Dr. Evandro Albiach Branco - Centro de Ciência do Sistema Terrestre (COCST)
Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia e Tecnologia
Espacial (CGETE)
Dr. Hermann Johann Heinrich Kux - Coordenação-Geral de Observação da Terra
(CGOBT)
Dra. Ieda Del Arco Sanches - Conselho de Pós-Graduação - (CPG)
Sílvia Castro Marcelino - Serviço de Informação e Documentação (SESID)

DIGITAL LIBRARY:

Dr. Gerald Jean Francis Banon
Clayton Martins Pereira - Serviço de Informação e Documentação (SESID)

DOCUMENT REVIEW:

Simone Angélica Del Ducca Barbedo - Serviço de Informação e Documentação
(SESID)
André Luis Dias Fernandes - Serviço de Informação e Documentação (SESID)

ELECTRONIC EDITING:

Ivone Martins - Serviço de Informação e Documentação (SESID)
Cauê Silva Fróes - Serviço de Informação e Documentação (SESID)



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA,
INOVAÇÕES E COMUNICAÇÕES



sid.inpe.br/mtc-m21c/2020/03.03.16.25-TDI

USING OPTICAL WATER TYPES FOR SATELLITE MONITORING OF BRAZILIAN INLAND WATERS

Edson Filisbino Freire da Silva

Master's Dissertation of the Graduate Course in Remote Sensing, guided by Drs. Evlyn Márcia Leão de Moraes Novo, and Felipe de Lucia Lobo, approved in February 27, 2020.

URL of the original document:

<<http://urlib.net/8JMKD3MGP3W34R/4247NRS>>

INPE
São José dos Campos
2020

Cataloging in Publication Data

Silva, Edson Filisbino Freire da.
Si38u Using optical water types for satellite monitoring of brazilian inland waters / Edson Filisbino Freire da Silva. – São José dos Campos : INPE, 2020.
xxiv + 117 p. ; (sid.inpe.br/mtc-m21c/2020/03.03.16.25-TDI)

Dissertation (Master in Remote Sensing) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2020.

Guiding : Drs. Evlyn Márcia Leão de Moraes Novo, and Felipe de Lucia Lobo.

1. Remote sensing. 2. Optical water types. 3. Inland waters.
I.Title.

CDU 528.8:628.1.03



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

Aluno (a): **Edson Filisbino Freire da Silva**

Título: "USING OPTICAL WATER TYPES FOR SATELLITE MONITORING OF BRAZILIAN INLAND WATERS"

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de **Mestre** em **Sensoriamento Remoto**

Dr. Cláudio Clemente Faria Barbosa



Presidente / INPE / S.J.Campos - SP

() Participação por Video - Conferência

Aprovado () Reprovado

Dra. Evlyn Márcia Leão de Moraes Novo



Orientador(a) / INPE / S.J.Campos - SP

() Participação por Video - Conferência

Aprovado () Reprovado

Dr. Felipe de Lucia Lobo



Orientador(a) / UFPEL / Pelotas - RS

() Participação por Video - Conferência

Aprovado () Reprovado

Dr. Mauricio Almeida Noernberg



Convidado(a) / UFPR / Paraná - PR

() Participação por Video - Conferência

Aprovado () Reprovado

Dra. Maycira Costa



Convidado(a) / UVic / Victoria - CAN

Participação por Video - Conferência

Aprovado () Reprovado

Este trabalho foi aprovado por:

() maioria simples

unanimidade

São José dos Campos, 27 de fevereiro de 2020

“When we read, another person thinks for us: we merely repeat his mental process. In learning to write, the pupil goes over with his pen what the teacher has outlined in pencil: so in reading; the greater part of the work of thought is already done for us. This is why it relieves us to take up a book after being occupied with our own thoughts. And in reading, the mind is, in fact, only the playground of another’s thoughts. So it comes about that if anyone spends almost the whole day in reading, and by way of relaxation devotes the intervals to some thoughtless pastime, he gradually loses the capacity for thinking; just as the man who always rides, at last forgets how to walk. This is the case with many learned persons: they have read themselves stupid.”

Arthur Schopenhauer

ACKNOWLEDGEMENTS

Agradeço primeiramente aos meus orientadores, Dra. Evlyn Márcia Leão de Moraes Novo e Dr. Felipe de Lucia Lobo, pela aprendizagem e orientação no desenvolvimento deste trabalho. Suas orientações contribuíram muito para o desenvolvimento do meu conhecimento e na conclusão deste trabalho.

Agradeço ao Dr. Claudio Clemente Faria Barbosa pela sua orientação extraoficial, o qual contribuiu muito no desenvolvimento e finalização deste trabalho.

Agradeço a todo pessoal do LabISA, principalmente a Carolline Tressmann Cairo, Daniel Andrade Maciel, Rogério Flores Júnior, Victor Pedroso Curtarelli e Felipe Menino Carlos. Todos me ajudaram muito esclarecendo dúvidas durante a pesquisa.

Agradeço ao Programa de Pós-Graduação do Instituto Nacional de Pesquisas Espaciais (INPE), aos professores do SERE pelo conhecimento, e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo auxílio financeiro.

ABSTRACT

One of the key issues of monitoring inland water quality is spatial-temporal sampling because water quality can rapidly change due to natural and anthropogenic influence. Remote sensing of inland waters is a reliable tool for monitoring water quality in large areas and time-series. However, the traditional method of calibrating bio-optical algorithms for limnological parameters (e.g., Chlorophyll-a (Chl-a), Total Suspended Matter (TSM), and Colored Dissolved Organic Matter (CDOM)) is limited to bio-optical characteristics of the study sites used for algorithm calibration. Consequently, bio-optical algorithms are not suitable for monitoring inland waters on a macro-scale level. On the other hand, monitoring Optical Water Types (OWT) has shown a macro-scale application, while those OWTs also represent changes in Chl-a, TSM, and CDOM concentrations. Thus, monitoring Brazilian OWTs could be a useful tool for water management on a wide scale. The objective of this study is to create a method for monitoring the water quality of Brazilian inland waters using OWTs. The study is described in three chapters; the chapter 3 assesses the uncertainties related to the merging of spectra measurements obtained under different protocols of computing remote sensing reflectance (R_{rs}); the chapter 4 describes the identification of Brazilian OWTs using hyperspectral *in situ* R_{rs} , which was acquired for water bodies encompassing a wide range of optical characteristics in Brazil; the chapter 5 describes the training of classification algorithms for detecting the OWTs using satellite sensors. In the chapter 3, it is shown that R_{rs} computed on Kutser's method is lower than that of Mobley's in all water types, with bias reaching up to -100%. Both methods allow satisfactory calibration of bio-optical algorithms when they are used apart, but there is a significant accuracy reduction when both methods are mixed in the same database. Furthermore, almost half of the samples are labeled with different clusters depending on the R_{rs} method. Hence, merge both methods for calibrating bio-optical algorithms is viable when a validation dataset is used, but spectral clustering should be avoided. In the chapter 4, a total of eight OWTs are computed based on R_{rs} shape and magnitude, which represent different optical and limnological characteristics of Brazilian waters. The OWT 1 represents transparent waters with low TSM, Chl-a, and CDOM concentrations; the OWT 2 represents transparent waters with moderate CDOM and TSM; OWT 4 is characterized by waters with algae bloom in aquatic system with moderate TSM concentration; OWT 5 is characterized by waters with algae bloom in low TSM concentration; OWT 6 is composed by waters with severe algae bloom density; OWT 7 is characterized by waters with the highest CDOM concentration; OWT 8 is waters with high TSM concentration; OWT 9 is waters with the highest scattering and TSM concentration. In the chapter 5, classification algorithms are trained for detecting the OWTs in satellite images of Sentinel-2 MSI, Landsat-8 OLI, and Landsat-7 ETM+. Sentinel-2 MSI has the best spectral resolution for classifying OWTs and exhibited satisfactory accuracy (Recall from 0.77 to 0.99) in satellite images. On the other hand, Landsat-8 OLI and Landsat-7 ETM+ classifications are profoundly affected by the overestimation of near-infrared bands, causing weak accuracy water bodies characterized by algae blooms (OWTs 4, 5, and 6). In

conclusion, the proposed have many applications, such as i) support of sampling design and survey campaigns; ii) detection of water quality anomalies caused by abrupt changes such in sediment loading and onset of algal blooms; iii) it could also be used for a census of Brazilian surface waters and provide reliable data in a macroscale level; last, iv) It could be used for improving the accuracy and the scope of semi-analytical algorithms based on R_{rs} by using the OWT in the calibration and validation process.

Keywords: Remote sensing. Optical Water Types. Inland Waters.

USO DE TIPOS ÓPTICOS DE ÁGUA PARA O MONITORAMENTO DAS ÁGUAS INTERIORES BRASILEIRAS POR SATÉLITES

RESUMO

Um dos fatores limitantes para se monitorar a qualidade das águas interiores é a cobertura espaço-temporal de amostragens, visto que a qualidade da água pode mudar rapidamente por causas naturais ou antropogênicas. O sensoriamento remoto de águas interiores é uma excelente ferramenta para monitorar com maior frequência grandes regiões. Entretanto, o método tradicional de calibração de algoritmos bio-ópticos para a estimativa de parâmetros como concentração de Material Total Suspenso (TSM), clorofila-a (Chl-a), e matéria orgânica colorida dissolvida (CDOM) tem sua aplicação limitada às regiões para as quais estes foram calibrados, e, portanto, não podem ser aplicados em macro escala. Por outro lado, o monitoramento de tipos ópticos de água (OWT) possui aplicação em macro escala pois estes também representam alterações em TSM, Chl-a, e CDOM. Sendo assim, o uso de OWTs para o monitoramento de grandes regiões, como o território brasileiro, pode ser vantajoso. Essa vantagem fundamenta o objetivo deste estudo, de criar um método para o monitoramento das águas brasileiras usando OWTs. Este estudo encontra-se descrito em três capítulos; o capítulo 3 investiga as incertezas geradas quando diferentes métodos de correção de glint para o cálculo da reflectância de sensoriamento remoto (R_{rs}) são misturados para formar uma única base de dados; o capítulo 4 descreve como as OWTs de sistemas aquáticos interiores brasileiros foram geradas a partir de medidas de R_{rs} hiperespectral representativas de um amplo range de características ópticas dos corpos de água do Brasil; o capítulo 5 descreve o processo de treinamento de algoritmos classificadores desenvolvidos para detectar OWTs definidas por dados hiper espectrais aplicados a sensores multiespectrais. Nos resultados do capítulo 3, a R_{rs} calculada utilizando o método do Kutser é mais baixa que a R_{rs} calculada pelo método do Mobley em todos tipos de água, sendo que a R_{rs} do Kutser pode subestimar em até -100% a R_{rs} corrigida por Mobley. Ambos os métodos permitem calibrar algoritmos bio-ópticos quando são usados separados, mas quando são combinados em uma única base de dados, pode haver uma queda significativa na acurácia dos algoritmos. Além disso, o processo de classificação (clustering) de espectros de R_{rs} , aproximadamente metade das amostras são agrupadas em classes distintas em função do método de correção de glint utilizado. Portanto, ambos os métodos de correção de glint podem ser combinados para calibração de algoritmos bio-ópticos, desde que uma base de dados de avaliação seja utilizada. Por outro lado, deve se evitar clusterizar espectros combinando os dois métodos. No capítulo 4, um total de oito OWTs foram obtidas utilizando a forma e magnitude da R_{rs} , cujas propriedades ópticas e limnológicas são distintas. A OWT 1 compreende águas transparentes com baixa concentração de TSM, Chl-a e CDOM; a OWT 2 compreende águas transparentes com moderada concentração de CDOM e TSM; a OWT 4 inclui águas com floração

de algas em ambientes com concentração moderada de TSM; a OWT 5 compreende águas com floração de algas em ambientes com baixa concentração de TSM; a OWT 6 inclui águas com florações de algas com alta densidade; a OWT 7 compreende águas com elevadas concentrações de CDOM; a OWT 8 são águas com alta concentração de TSM; e a OWT 9 são águas com alto retro espalhamento e as mais altas concentrações de TSM. No capítulo 5, algoritmos classificadores foram treinados para detectarem as OWTs em imagens dos sensores Sentinel-2 MSI, Landsat-8 OLI e Landsat-7 ETM+. Sentinel-2 MSI mostrou a melhor capacidade espectral para classificar as OWTs. Nas imagens de satélite, o desempenho dos classificadores é muito sensível a correção atmosférica, sendo o Sentinel-2 MSI o que apresenta o melhor desempenho entre as imagens orbitais. Por outro lado, as classificações das imagens obtidas por Landsat-8 OLI e Landsat-7 ETM+ foram significativamente afetadas pela super estimativa de suas bandas no infravermelho próximo, o que levou à redução da acurácia de classificação de OWTs relacionadas a floração de algas (OWTs 4, 5 e 6). Concluindo, a utilização do método proposto neste estudo para o monitoramento das águas interiores brasileiras pode: i) fornecer subsídios para o delineamento amostral e para o planejamento de campanhas de campo; ii) permitir detecção de anomalias em mudanças abruptas do ambiente, como alto aporte de sedimentos e a floração de algas; iii) o método pode também ser utilizado para um cadastro das águas superficiais brasileiras, informação essencial para determinar um nível de referência da qualidade da água contra qual medir impactos antropogênicos e naturais em um nível de macro escala.; iv) melhorar a acurácia de algoritmos semi-analíticos baseados em R_{rs} , usando as OWTs durante o processo de calibração e validação.

Keywords: Sensoriamento Remoto. Tipos Ópticos de Água. Águas Interiores.

LIST OF FIGURES

	<u>Pg.</u>
Figure 2.1 – Decision tree example.....	20
Figure 2.2 - Linear support vector machine example.	21
Figure 3.1. Study area and sampling stations.	25
Figure 3.2 - The mean (solid line) and samples (shaded lines) of R_{rs} (Mobley) (red color) and R_{rs} (Kutser) (blue color) for the limnological classes A, B, C, and D. ...	34
Figure 3.3 - The mean (dashed line) and standard deviation (shaded area) of glint removed by Mobley’s (red color) and Kutser’s (blue color) methods for classes A, B, C, and D.....	35
Figure 3.4 - The divergences between Kutser’s and Mobley’s methods for classes A, B, C, and D.	36
Figure 3.5 - Scatter plots between simulated bands /indexes with TSM and Chl-a.....	39
Figure 3.6 - The mean (solid lines) and standard deviation (shaded area) of R_{rs} (shape) for a clustering interaction using the Mobley R_{rs} (shape) in the labeling procedure.	42
Figure 4.1 - Study areas and sampling stations located in the Brazilian waters.....	48
Figure 4.2 - The mean (solid line) and SD (shaded area) of R_{rs} (black and left y-axis) and R_{rs} (shape) (blue and right y-axis) of all OWTs.	53
Figure 4.3 - Limnological characteristics of each OWT. The measurements are Chl-a concentration, TSM concentration, CDOM concentration, and Z_{sd}	56
Figure 4.4 - The absorption (a), scattering (b), and attenuation (c) coefficients of all OWTs.....	57
Figure 5.1 - Study design flowchart.	69
Figure 5.2 - Spectral R_{rs} of Brazilian OWTs.....	76
Figure 5.3 - Satellite and <i>in situ</i> match-ups of Sentinel-2 MSI bands, corresponding to data measured on the same day.	80
Figure 5.4. Satellite and <i>in situ</i> match-ups of Landsat-8 OLI bands corresponding to data measured on the same day.	81

Figure 5.5 - Satellite and <i>in situ</i> match-ups of Landsat-7 ETM+ bands, corresponding to data measured on the same day.	82
Figure 5.6 - OWT classification in the Funil reservoir in 2019-08-10	87
Figure 5.7 - OWT classification in the Curuai lake in 2018-08-01.....	88
Figure 5.8. OWT classification in the lower Amazon basin in 2017-09-04.....	89
Figure A.1 - How accuracy and number of targets detected (n) change as the probability threshold increases, for each OWT in the MSI SVC algorithm.....	108
Figure A.2 - How accuracy and number of targets detected (n) change as the probability threshold increases, for each OWT in the MSI RF algorithm.	109
Figure A.3 - How accuracy and number of targets detected (n) change as the probability threshold increases, for each OWT in the OLI SVC algorithm.	110
Figure A.4 - How accuracy and number of targets detected (n) change as the probability threshold increases, for each OWT in the OLI RF algorithm.....	111
Figure A.5 - How accuracy and number of targets detected (n) change as the probability threshold increases, for each OWT in the ETM+ SVC algorithm. .	112
Figure A.6 - How accuracy and number of targets detected (n) change as the probability threshold increases, for each OWT in the ETM+ RF algorithm.	113

LIST OF TABLES

	<u>Pg.</u>
Table 3.1 - The mean and standard deviation (SD) of Chl-a, TSM, and DOC concentrations for classes A, B, C, and D.	33
Table 3.2 - The mean statistical results of calibrated algorithms for TSM in the Curuai Lake, after 1000 interactions.....	37
Table 3.3 - The mean statistical results of calibrated algorithms for Chl-a in the Ibitinga reservoir, after 1000 interactions.	40
Table 3.4 - Confusion matrix between Kutser and Mobley labeled data for the same clustering interaction.....	42
Table 4.1 - The median (med), interquartile range (IQR), and the number of samples (n) of Chl-a concentration, TSM concentration, CDOM concentration, and Z_{sd} , of all OWTs.	55
Table 4.2 - The number of representative samples of each OWT per region. ...	58
Table 5.1 - Technical specifications summary of Sentinel-2 MSI, Landsat-8 OLI, and Landsat-7 ETM+.	68
Table 5.2 - Parameters specifications of SVC and RF training algorithms and inputs for classification.	71
Table 5.3 - The accuracy and detection rate of all classification algorithms.	77
Table 5.4 - Performance of the classifications algorithms for each OWT using the <i>in situ</i> simulated sensors.	78
Table 5.5 - Statistical results of the match-up between each sensor image and <i>in situ</i> data. The match-ups correspond to R_{rs} samples measured concurrent to satellite overpass.....	79
Table 5.6 - Performance of the classifications models for each OWT in the satellite and <i>in situ</i> match-ups.	83
Table 5.7 - Performance of the classifications models for each OWT using the <i>in situ</i> simulated sensors with uncertainty estimated from the satellite and <i>in situ</i> match-ups.....	85
Table 5.8 - Total OWTs areas estimated in Funil Reservoir, Curuai Lake, and the Lower Amazon Basin.	90

Table A.1 – SVC and RF probability thresholds for the MSI, OLI, and ETM+ sensors..... 107

LIST OF ABBREVIATIONS

2B	Two Bands Ratio
3B	Three Bands Ratio
6S	Second Simulation Of Satellite Signal In The Solar Spectrum
6SV	6S Vector Version
AOP	Apparent Optical Properties
Chl-a	Chlorophyll-a
DOC	Dissolved Organic Carbon
ECP	Estuarine Complex Of Paranaguá
EMR	Electromagnetic Radiation
FDA	Functional Data Analysis
FN	False Negatives
FP	False Positives
IOP	Inherent Optical Properties
IQR	Interquartile Range
k	Number Of Clusters
MAPE	Mean Absolute Percentage Error
MARD	Mean Absolute Relative Difference
NAP	Non-Algae Particles
NDCI	Normalized Chlorophyll Index
NIR	Near-Infrared
OAC	Optically Active Constituents
OWT	Optical Water Type
P	Probability
Phy	Algae Particles
RF	Random Forest
RMSE	Root Means Squared Error
SAM	Spectral Angle Mapper
SD	Standard Deviation
SSE	Sum Squared Error
SVC	SVM Classifier

SVM Support Vector Machine
SWIR Short Wave Infrared
TP True Positives
TSM Total Suspended Matter

LIST OF SYMBOLS

ΔB_n	Bandwidth (nm).
Δ_M^2	Mahalanobis distance
Δ_T^2	Theoretical threshold of Δ_M^2
ΔV	Water volume (m ³).
Δr	Thickness (m).
$\Delta\Omega$	Solid angle (sr).
λ	Wavelength (nm).
Φ_i	Incident radiant power (W).
Φ_a	Absorbed radiant power (W).
Φ_s	Scattered radiant power (W).
Φ_t	Transmitted radiant power (W).
ψ	Scattering angle (°).
θ	Nadir angle (°).
ϕ	Azimuth angle (°).
Σ	Covariance matrix
$ \Sigma $	Matrix determinant of Σ
A	Absorptance.
a	Absorption coefficient (m ⁻¹).
a_{CDOM}	CDOM absorption coefficient (m ⁻¹).
a_{NAP}	NAP absorption coefficient (m ⁻¹).
a_{phy}	Phy absorption coefficient (m ⁻¹).
ang	The angle between two spectrums.
B	Scatterance.
b	Scattering coefficient (m ⁻¹).
b_c	Specific scattering coefficient (m ⁻¹ μg ⁻¹ L).
b_{phy}	Phy scattering coefficient (m ⁻¹).
b_{NAP}	NAP scattering coefficient (m ⁻¹).
β	Volume scattering function (m ⁻¹).
b_f	Forward scattering coefficient (m ⁻¹).
b_b	Backscattering coefficient (m ⁻¹).

Bn_{sensor}	A specific band according to sensor design.
E_s	Downwelling irradiance incident in the water surface ($W m^{-2}$).
L_s	Downwelling sky radiance ($W m^{-2} sr^{-1}$).
L_t	Total upwelling radiance leaving the water ($W m^{-2} sr^{-1}$).
ρ	Proportionality factor that relates how much of L_s was reflected towards the sensor direction.
$P (y=1 f)$	Probability estimated using a fitted sigmoid function from the SVM output values.
R	Irradiance reflectance just below the surface.
R^2	Coefficient of determination
$RF_{Bn}(\lambda)$	Response function of a specific band.
R_{rs}	Remote sensing reflectance (sr^{-1}).
$R_{rs}(Bn)(\text{standardized})$	R_{rs} of bands Bn standardized by the sum of all bands.
$R_{rs}(\text{Glnt})$	R_{rs} computed without glint correction (sr^{-1}).
$R_{rs}(\text{Kutser})$	R_{rs} computed using Kutser's method (sr^{-1}).
$R_{rs}(\text{Mobley})$	R_{rs} computed using Mobley's method (sr^{-1}).
$R_{rs}(\text{shape})$	R_{rs} normalized R_{rs} by its integer (finite approximation per 1 nm).
$R_{rs}(\text{pow function})$	R_{rs} computed by an adjusted power function using all wavelengths between 350 – 380 nm and 890 – 900nm (sr^{-1}).
$R_{rs}(\text{shape})$	Remote sensing reflectance normalized by its integer.
S	The spectral slope between two wavelengths (λ and λ_0).
Z_{sd}	Secchi depth (m).

CONTENTS

	<u>Pg.</u>
1 INTRODUCTION	1
1.1 Hypothesis	3
1.2 Objective	3
2 THEORY	5
2.1 Optically active constituents	5
2.2 Optical properties of inland waters	7
2.2.1 Inherent and apparent optical properties	7
2.2.2 Optical properties of the OACs	10
2.3 Optical classification of natural waters	14
2.4 Classification algorithms	17
2.4.1 Clustering algorithms	17
2.4.2 Supervised algorithms	18
3 INCONSISTENCIES BETWEEN KUTSER'S AND MOBLEY'S GLINT REMOVAL METHODS AND HOW THEY AFFECT ALGORITHM CALIBRATION AND SPECTRAL CLUSTERING: ASSESSMENT FOR INLAND WATERS	23
3.1 Introduction	23
3.2 Methods	24
3.2.1 Study area and sampling stations	24
3.2.2 Radiometric measurements and deglint	26
3.2.3 Limnological classification	27
3.2.4 Estimation of Kutser and Mobley inconsistencies	28
3.2.5 Algorithms calibration	29
3.2.6 Spectral clustering	32
3.3 Results and discussion	33
3.3.1 The limnological and radiometric characteristics of limnological classes	33
3.3.2 Inconsistencies between Kutser's and Mobley's deglint methods	34
3.3.3 The effect of merging Kutser's and Mobley's methods in the algorithms calibration	36

3.3.4	The effect of merging Kutser’s and Mobley’s methods in the spectral clustering.....	40
3.4	Conclusion.....	43
4	BRAZILIAN OPTICAL WATER TYPES	45
4.1	Introduction	45
4.2	Material and methods	47
4.2.1	Study area and survey campaigns	47
4.2.2	AOPs, IOPs, and limnological measurements.....	49
4.2.3	Establishment of the optical water types	51
4.3	Results and discussion	52
4.3.1	Clusters and adjustments to OWTs.....	52
4.3.2	The OWTs.....	54
4.3.3	The sources of OWTs	60
4.3.4	The Brazilian OWTs from a global perspective	63
4.4	Conclusion	64
5	SATELLITE DETECTION OF BRAZILIAN OPTICAL WATER TYPES	65
5.1	Introduction	65
5.2	Material and methods	66
5.2.1	<i>In situ</i> radiometric data	66
5.2.2	Satellite images.....	67
5.2.3	Classification algorithms.....	69
5.2.4	Accuracy assessment	73
5.3	Results and discussion	75
5.3.1	Classification using <i>in situ</i> data	75
5.3.2	Classification using satellite images.....	79
5.3.2.1	<i>Atmospheric correction assessment</i>	79
5.3.2.2	<i>In situ and satellite classification match-ups</i>	82
5.3.2.3	<i>Image analyses of OWTs</i>	86
5.4	Conclusion	91
6	FINAL CONSIDERATIONS	93
	REFERENCES.....	95

APPENDIX A - HOW THE ACCURACY CLASSIFICATION IMPROVES USING PROBABILITY ESTIMATES AND THE PROBABILITY THRESHOLDS FOR ALL CLASSIFICATION MODELS AND OWTS.....	107
A.1 Probability thresholds.....	107
A.2 Increasing accuracy using probability	107

1 INTRODUCTION

Inland waters comprise many environments such as lakes, ponds, wetlands, bogs, artificial reservoirs, rivers, and estuaries. Those environments are essential for human activities, providing resources for energy production, farming, aquaculture, navigation, and recreation, urban and industrial activities. As a consequence, those human uses cause environmental impacts on water quality, such as increased turbidity, eutrophication level, and frequency of toxic algal blooms. The increased degradation of inland waters causes losses of ecosystem services, human health hazards, socioeconomic impacts, and adverse effects on the aquatic biota (JØRGENSEN et al., 2012). For that reason, studies and monitoring programs of water quality are needed for giving scientific support to the sustainable use of water resources and the development of ecosystem restoration programs.

Various studies on water quality are based on satellite remote sensing, which, compared to *in situ* data collection has the advantage of frequent data acquisition over vast areas of the Earth's surface (CAIRO et al., 2020; DOGLIOTTI et al., 2016; MACIEL et al., 2019). However, these studies are mostly regional and time-specific. Estimate essential information (e.g., total suspended matter (TSM), chlorophyll-a (Chl-a), and colored dissolved organic matter (CDOM) concentrations) from remote sensing requires calibration of bio-optical algorithms, and each calibration tends to be specific for an exclusive range of optically active constituents (OACs) composition, size distribution, and concentration. Regarding inland waters, each waterbody, depending on its drainage basin features and land-use history, has different ranges of concentration and biogeochemistry, demanding specific calibrations. Likewise, even analytical and semi-analytical bio-optical algorithms are also time-specific since OACs type and range are in constant change. Hence, the monitoring of vast areas encompassing different water bodies requires a considerable effort in the calibration process. This effort involves surveys, human resources, and financing, limiting the monitoring capability of large areas and the application of satellite time-series in the study of highly dynamic environments.

Alternatively, several studies have shown that approaches relying on optical water types (OWT) can be applied for monitoring different water bodies and in studies relying on the application of satellite time-series (DEYONG et al., 2013; HUANG et al., 2014; LIU et al., 2013; LUBAC; LOISEL, 2007; REINART et al., 2003; SHI et al., 2013a; SPYRAKOS et al., 2018; SUN et al., 2014; VANTREPOTTE et al., 2012). This approach consists of using radiometric measurements (e.g., remote sensing reflectance (R_{rs})) for classifying water masses in a way that each class represents specific ranges of CDOM, Chl-a, and TSM concentration. Then, the classification provides a qualitative knowledge about the waterbody properties such as water masses with different degrees of clarity and turbidity (CHEN et al., 2004 ; LOBO et al., 2012), presence of algae blooms, and rich in dissolved organic carbon (DOC) concentration (LOBO et al., 2012). Furthermore, the clustering of OWTs may also improve the performance of Chl-a algorithms (CAIRO et al., 2020; SHI et al., 2013b) and TSM concentration (VANTREPOTTE et al., 2012). These classes may have similar OACs ranges, then, allowing for a more accurate algorithm calibration for each OWT. Therefore, for monitoring vast regions composed of several water bodies and for time series applications, this approach could be an alternative to the search for regional algorithms.

Brazil is the 5th larger country in the world and covers 47% of the South American continent ($8.5 \times 10^6 \text{ km}^2$). Brazil has surface water availability of $78600 \text{ m}^3 \text{ s}^{-1}$, corresponding to 12% of the surface fresh water available in the world (BRASIL, 2012). These waters are subjected to quality degradation, such as eutrophication and an increase in algae bloom caused by crop and animal production systems (SHIGAKI et al., 2006), and increase of turbidity and TSM concentration caused by dam failures (HATJE et al., 2017). Many studies have been carried out in specific waterbodies, but there is a considerable gap in the assessment of time changes in the level of the Brazilian inland water degradation. The elimination of that gap at the national level is relying only on *in situ* methods and on the calibration of regional bio-optical algorithms for each waterbody, which is a difficult task. Thus, the development of an optical classification of Brazilian inland

waters and a method for detecting them based on orbital sensors could bridge that spatial-temporal gap of water quality data.

1.1 Hypothesis

OWTs can be used for satellite monitoring of Brazilian inland waters.

1.2 Objective

The objective of this study is to create a method for monitoring the Brazilian inland waters using OWTs. Thus, three specific objectives are set and comprise three chapters of this study. Chapter 3 assess if a database composed of Kutser's and Mobley's glint removal methods (KUTSER et al., 2013; MOBLEY, 1999) can be used in the same database for establishing the Brazilian OWTs. Chapter 4 establishes the Brazilian OWTs based on *in situ* R_{rs} . Chapter 5 train classification algorithms for detecting the Brazilian OWTs using different orbital sensors. Based on these objectives, the following research questions are addressed in this dissertation:

- Can the Kutser's and Mobley's glint removal be used in the same database for computing OWT? (Chapter 3)
- What are the Brazilian OWTs? (Chapter 4)
- How to detect the Brazilian OWTs in satellite images? (Chapter 5)

2 THEORY

2.1 Optically active constituents

The main Optically Active Constituents (OACs) are the CDOM, algae particles (Phy), and non-algae particles (NAP). Each component interacts with electromagnetic radiation (EMR) and makes it possible to detect them by remote sensing. Besides, monitoring the OACs is necessary because they also have an essential role in the water quality of inland waters (KIRK, 2011). This section describes the composition of each OAC and its influence in inland waters.

CDOM is the fraction of dissolved organic matter which interacts with EMR. This fraction is derived from the dissolved humic substances, which are a heterogeneous group of compounds consisting of polymers of aromatic rings formed due to the decomposition of carbon material by microbial activity. The CDOM varies in size, solubility, and molecular weight and is classified into 2 types: (1) humic acids, which are the precipitate fraction resulting from its acidification, and (2) fulvic acid, which is the remaining material. Both fractions are chemically similar, but humic acid molecules are larger than those of fulvic acid (KIRK, 2011).

The CDOM role in aquatic environments is associated with photobleaching, pollutants, and absorbing hazardous EMR. Photobleaching is the process through which CDOM is degraded by EMR, affecting nutrient availability and the production of CO₂ (BUSHAW et al., 1996). Pollutants can strongly bind with CDOM and thereby have its transport and toxicity affected (SANTOS et al., 2008). CDOM absorbs ultraviolet radiation that can damage phytoplankton (WILLIAMSON et al., 2001). On the other hand, CDOM increases light attenuation and limits the EMR available for photosynthesis (EVANS et al., 2005).

Phy is composed by the pigments and cell structure of phytoplankton, which are living organisms comprising of diverse phylum, from prokaryotic cells such as Cyanophyta to eukaryotic cells such as Cryptophyta. Most phytoplankton is autotrophic and dependent on their pigments for absorbing EMR and producing energy and organic matter. For example, blue-green algae that have Chl-a, phycocyanin, and phycoerythrin pigments. Moreover, phytoplankton size also

differs into pico (0.2 – 2 μm), nano (2 – 20 μm), micro (20 – 200 μm), meso (0.2 – 2 mm), and macro (> 2 mm) (REYNOLDS, 2007).

For evaluating the water quality of inland waters, a standard measurement related to Phy is the Chl-a concentration. This measurement is easily obtained by standalone and survey devices, as well as laboratory analysis. The Chl-a concentration is used as a proxy of the phytoplankton biomass, although the relationship between Chl-a concentration and phytoplankton biomass differs in different species (REYNOLDS, 2007). Monitoring Chl-a concentration can aid the identification of algae blooms, whose occurrence is related to a fast increase of organic matter in the waterbody, leading to hypoxia and mortality of heterotrophic fauna. Moreover, a toxic algae bloom can also contaminate water supplies and impact animal and human health. Furthermore, changes in the Chl-a concentration can indicate sewers and agricultural contaminants (LIBES, 2009).

NAP consists of mineral particles derived from land erosion, resuspension of bottom sediments, organic components like bacteria, dead cells, and the fragment of cells. Thus, NAP varies in size, composition, and concentration. NAP size can be expressed as particle size distribution in the water column, the composition by its refraction index, and concentration by mass in a volume of water (KIRK, 2011). TSM is the limnological measurement related to NAP. Although, TSM can also include part of Phy depending on the laboratory method.

The TSM influence on the aquatic environment is related to turbidity and transport of adsorbed compounds. The increase of TSM concentration increases the turbidity, and so, light attenuation. Higher light attenuation means less EMR available for photosynthesis; thus, a reduced amount of organic matter and energy available to the heterotrophic fauna (REYNOLDS, 2007). Also, TSM can adsorb some compounds enabling the distribution of nutrients in the waterbody. On the other hand, particles also help the spread of pollution (HATJE et al., 2017).

2.2 Optical properties of inland waters

The optical properties of water are divided into inherent optical properties (IOPs) and apparent optical properties (AOPs). IOPs are dependent only upon the medium, and therefore are independent of the ambient light field and environment conditions. On the other hand, AOPs are dependent on the medium, geometric structure of the ambient light field, and environment conditions (MOBLEY, 1994). Both IOPs and AOPs comprise various parameters. The next sections describe only the IOPs and AOPs variables relevant to this project and the optical properties of each OAC.

2.2.1 Inherent and apparent optical properties

This topic represents a summary of IOPs and AOPs described by MOBLEY, (1994). Considering a small volume of water (ΔV) and thickness (Δr) illuminated by a collimated beam of monochromatic light of spectral radiant power (Φ), $W\text{ nm}^{-1}$, some part of incident radiant power (Φ_i) is absorbed in the water (Φ_a), while another part of the incident radiant power is scattered out of the beam at all directions (Φ_s), and the remaining radiant power is transmitted through the volume with no change in the direction (Φ_t). Disregarding the inelastic scattering (fluorescence), all the incident radiant power in a volume of water follows the conservative radiative transfer defined by the Kirchhoff's law:

$$\Phi_i(\lambda) = \Phi_a(\lambda) + \Phi_s(\lambda) + \Phi_t(\lambda) \quad (2.1)$$

Hence, the spectral absorptance $A(\lambda)$ is the fraction of incident power that is absorbed within the volume:

$$A(\lambda) = \frac{\Phi_a(\lambda)}{\Phi_i(\lambda)} \quad (2.2)$$

Likewise, the spectral scatterance $B(\lambda)$ is the fraction part of the incident power that is scattered out of the beam:

$$B(\lambda) = \frac{\Phi_s(\lambda)}{\Phi_i(\lambda)} \quad (2.3)$$

Thus, the IOPs can be defined using these variables. The spectral absorption coefficient $a(\lambda)$ is defined as:

$$a(\lambda) = \lim_{\Delta r \rightarrow 0} \frac{A(\lambda)}{\Delta r} \quad (\text{m}^{-1}) \quad (2.4)$$

So the spectral scattering coefficient $b(\lambda)$ is defined as:

$$b(\lambda) = \lim_{\Delta r \rightarrow 0} \frac{B(\lambda)}{\Delta r} \quad (\text{m}^{-1}) \quad (2.5)$$

The scattering direction in the water column can also be taken into account. The volume scattering function $\beta(\psi; \lambda)$ is the fraction of incident radiant power scattered out the beam into a solid angle ($\Delta\Omega$) centered on a specific scattering angle ψ . The $\beta(\psi; \lambda)$ is defined as:

$$\beta(\psi; \lambda) = \lim_{\Delta r \rightarrow 0} \lim_{\Delta\Omega \rightarrow 0} \frac{B(\psi; \lambda)}{(\Delta r \Delta\Omega)} \quad (\text{m}^{-1}\text{sr}^{-1}) \quad (2.6)$$

Integrating $\beta(\psi; \lambda)$ overall directions gives $b(\lambda)$:

$$b(\lambda) = 2\pi \int_0^{\pi} \beta(\psi; \lambda) \sin \psi \, d\psi \quad (\text{m}^{-1}) \quad (2.7)$$

This integration is often divided into forwarding scattering b_f ($0 \leq \psi \leq \pi/2$) and backscattering b_b ($\pi/2 \leq \psi \leq \pi$), where 0 is the downward zenith direction. The forward and backscattering coefficients are, respectively:

$$b_f(\lambda) = 2\pi \int_0^{\pi/2} \beta(\psi; \lambda) \sin \psi \, d\psi \quad (\text{m}^{-1}) \quad (2.8)$$

$$b_b(\lambda) = 2\pi \int_{\pi/2}^{\pi} \beta(\psi; \lambda) \sin \psi \, d\psi \quad (\text{m}^{-1}) \quad (2.9)$$

The IOPs are represented by Eq. 2.4 to Eq. 2.9, and they are useful for comparing the optical properties of different water bodies; however, obtaining them by orbital sensors requires sophisticated modeling. On the other hand, the foremost AOPs utilized in inland waters, the R_{rs} , is easily obtained from orbital sensors. The R_{rs} is defined by the following equation:

$$R_{rs}(\theta, \phi; \lambda) = \frac{L_w(\theta, \phi; \lambda)}{E_d(\lambda)} \quad (\text{sr}^{-1}) \quad (2.10)$$

Where $L_w(\theta, \phi; \lambda)$ is the water-leaving radiance into the specific zenith (θ) and azimuth (ϕ) angles, which is measured at the orbital sensor, and E_d is the downwelling irradiance incident in the water surface. R_{rs} is a measure of how much the downwelling EMR incident onto the water surface is returned in the direction (θ, ϕ), so it can be detected by a radiometer pointed in the opposite direction and operated in orbital sensors. Furthermore, R_{rs} is related to the IOPs by the following equations:

$$R = \alpha \frac{b_b}{a} \quad (2.11)$$

$$R_{rs} = \frac{t_- t_+}{n^2} \frac{R}{Q(1 - \gamma R)} \quad (\text{sr}^{-1}) \quad (2.12)$$

Where R is the irradiance reflectance just below the surface, α is the proportionality constant that depends on the radiance distribution (e. g., solar zenith angle, diffuse sky lighting, and sea state), t_- is the radiance transmittance from below to above the surface, t_+ is the irradiance transmittance from above to below the surface, n is the refractive index of water, γ is the water-to-air internal reflection coefficient, and Q is the ratio of E_u (Irradiance ascendant below the surface) and L_u (radiance ascendant below the surface) (LEE et al., 1998).

2.2.2 Optical properties of the OACs

The primary influence of CDOM is its absorption, while scattering is not significant. CDOM absorbs very little in the red, but there is a fast increase in the absorption with decreasing wavelength, reaching its maximum in the ultraviolet (MOBLEY, 1994). The CDOM spectral absorption coefficient (a_{CDOM}) can be described by the model (BRICAUD et al., 1981):

$$a_{\text{CDOM}}(\lambda) = a_{\text{CDOM}}(\lambda_0) \exp(-S(\lambda - \lambda_0)) \quad (m^{-1}) \quad (2.13)$$

Where S is the slope and determines how fast the absorption reduces with increasing wavelength. The value of $a_{\text{CDOM}}(\lambda)$ depends on the CDOM concentration, while the value of S depends on the relative proportions of humic and fulvic acids. Consequently, $a_{\text{CDOM}}(\lambda)$ is different for each region, for example, in the Estonian aquatic systems $a_{\text{CDOM}}(400 \text{ nm})$ varies from 0 to 85 m^{-1} (OGASHAWARA et al., 2017).

The CDOM composition affects the slope and spectral shape of $a(\lambda)$. In the Sweden lakes, Wünsch et al. (2018) report different spectral shapes in response to CDOM molecule size. For smaller molecules (mainly composed of fulvic acids), the slope is higher than that of larger molecules (mainly composed of humic acids). Thus, the slope of absorption can be used for inferring the CDOM composition. Likewise, Silva (2018) uses the a_{CDOM} slope to infer the CDOM molecular weight, and further the CDOM source.

CDOM also has fluorescence properties once excited by EMR in the ultraviolet region, emitting broadband in the blue area. Chen et al. (2017) show the spectral response of Hangzhou Bay and Qiandao lake after exciting it with an EMR at 405 nm. An emission peak at 508 nm is observed and is highly correlated with the CDOM absorption coefficient at 440 nm ($R^2 = 0.9$; $R^2 = 0.87$). Moreover, the peak position of CDOM fluorescence can also change according to its molecular size. After excited by EMR at 275 nm, the emission peak changes to longer wavelengths as the molecule's size increases ($R^2 = 0.53$) (WÜNSCH et al., 2018).

The optical properties of Phy depend on its concentration, pigments, and cell structure of phytoplankton. Pigments have a variety of absorption peaks along the visible spectrum. For pigments dissolved in solvents with no influence of cell structure, Chl-a absorption peaks are at 432 nm and 662 nm; chlorophyll-b absorptions peaks are at 458 nm and 646 nm; chlorophyll-c absorption peaks at 444 nm and 630 nm; fucoxanthin absorption peak is at 450 nm; and β -Carotene absorption peak is at 452 nm (BIDIGARE et al., 1990). Each species of phytoplankton has different intracell concentration and relative proportion of pigments, thus, each species can have different specific absorption spectrums $a_{\text{phy}}(\lambda)$.

Cell size and concentration of phytoplankton can also affect absorption along the visible spectrum. Morel & Bricaud (1981) analyzed the influence of cell size on the absorption coefficient of phy (intracellular self-shading), where for the same Chl-a concentration, they observed a flattening process at 430 nm absorption peak when the diameter of the cell increases. Ciotti et al. (2002) observe the same pattern for natural phytoplankton communities; the absorption of picoplankton

was more flattened in the blue region than that of the microplankton. Hence, for the same Chl-a concentration, absorption is reduced by the self-shading effect, which is increased by a larger particle size (DUYSENS, 1956).

Phytoplankton can also contribute significantly to the scattering, which varies from among species. Species with mineralized cell walls such as diatoms and coccolithophores have higher scattering than those with no mineralized cell walls (GANF et al., 1989). Furthermore, species can be compared using the specific scattering coefficient (b_c). Species such as *Scenedemus bijuga*, *Chlamydomonas* sp., *Nostoc* sp. and *Anabaena oscillarioides* have b_c (550 nm) of 0.107, 0.044, 0.113 and 0.139, respectively (Davies-Colley et al., 1986).

Phytoplankton also expresses fluorescence emission when excited, with spectral peak varying with species. Johnsen & Sakshaug (2007) analysis the fluorescence excitation in 33 marine phytoplankton species divided into pigment groups. They observe that spectral fluorescence excitation differs under low and high light conditions but also in different pigment groups.

The NAP optical characteristics are mainly on the absorption and scattering but no considerable in fluorescence. The NAP absorption is low in the red and increases towards blue, decreasing again in the ultraviolet. Its absorption has a similar shape to CDOM absorption, caused by humic particulate or humic substance adsorbed in the mineral particles. Thus, NAP absorption follows an exponential equation similar to CDOM absorption:

$$a_{\text{NAP}}(\lambda) = a_{\text{NAP}}(\lambda_0)\exp(-S * (\lambda - \lambda_0)) \quad (\text{m}^{-1}) \quad (2.14)$$

Where $a_{\text{NAP}}(\lambda)$ is the NAP absorption coefficient, and S is the slope. The slope is typically smaller than that of CDOM. Studies in coastal waters report values of $0.0123 \pm 0.0013 \text{ nm}^{-1}$ (KIRK, 2011).

The scattering in aquatic environments is highly modified by NAP concentration, size, and refraction index. Simulated results of oceanic waters by Roesler & Boss (2008) at 555 nm demonstrate the influence of each parameter. The

concentration increases the scattering coefficient with small changes in the $\beta(\psi; \lambda)$ shape. Different refractive index show, not only distinctive $b(\lambda)$ but also changes in the shape of $\beta(\psi; \lambda)$. The particle size distribution (Junge value) is positively correlated with the scattering coefficient; the higher the Junge value (proportion of small particles is more present than larger particles), the higher is the scattering. Also, Junge values change the shape of $\beta(\psi; \lambda)$, which small Junge values has a higher slope and shoulders next to 60° and 110° .

For all NAP particle sizes in the water column, smaller particles are responsible for most of the scattering. Peng et al. (2009) analyzed the influence of particle size in the scattering and backscattering of Superior Lake waters. From particle sizes from 0.5 to 10 μm , particle sizes up to 2.05 μm are responsible for 50% of the total scattering; and particle sizes up to 2.93 μm are responsible for 50% of total backscattering. Thus, their study demonstrates that between 0.5 and 10 μm , smaller particles scatter more EMR than bigger particles.

The $a(\lambda)$ and $b(\lambda)$ of water bodies combine CDOM, Phy, and NAP. Thus, the absorption and scattering coefficient can be defined as:

$$a(\lambda) = a_{\text{CDOM}}(\lambda) + a_{\text{phy}}(\lambda) + a_{\text{NAP}}(\lambda) \text{ (m}^{-1}\text{)} \quad (2.15)$$

$$b(\lambda) = b_{\text{phy}}(\lambda) + b_{\text{NAP}}(\lambda) \text{ (m}^{-1}\text{)} \quad (2.16)$$

Where $a_{\text{CDOM}}(\lambda)$, $a_{\text{phy}}(\lambda)$ and $a_{\text{NAP}}(\lambda)$ are the absorption coefficient of CDOM, Phy, and NAP, respectively. And $b_{\text{phy}}(\lambda)$ and $b_{\text{NAP}}(\lambda)$ scattering coefficient of Phy and NAP, respectively. The spectral $R_{\text{rs}}(\lambda)$ is influenced by both $a(\lambda)$ and $b(\lambda)$, and the ambient light field (see eq. 2.11 and 2.12). Furthermore, the IOPs are influenced by CDOM concentration and size, phytoplankton concentration, species and size, and NAP concentration. Thus, the R_{rs} spectral shape and magnitude are influenced by all those parameters.

The contribution of each OAC may differ in different water bodies. For example, Sun et al. (2010) compare the influence of Phy and NAP in the scattering and absorption in the Taihu Lake. For all visible spectral range, Phy has a small influence in b , contributing less than 10% to the total scattering; in contrast, NAP contributes up to 90 %. For the same spectral range, Phy varied from 0.3 to 3.9 % of total b_b , while NAP contributed up to 100%. Considering the absorption, Phy and NAP have a similar contribution to the total absorption, both ranging from 20 to 70 % in the visible spectral range.

As CDOM, phy, and NAP change absorption and scattering; consequently, they also change the reflectance. Lobo et al. (2012) show the reflectance spectra for different TSS, Chl-a, and DOC concentrations. Clear waters with a low concentration of all limnological parameters exhibit low reflectance on the NIR region, increasing towards lower wavelengths, with the highest peak at 550 nm. Waters rich in DOC and low Chl-a and TSM concentration have a flat shape in the visible spectrum and low reflectance in the NIR region. High TSM concentration increases reflectance in the most of the spectrum, for instance, from the mean of 13 mg l⁻¹ to mean of 580 mg l⁻¹, the reflectance at 600 nm increases from 4% to 12%. This reflectance increase shows the influence of $b_{\text{NAP}}(\lambda)$ in high concentrations. Furthermore, waters with high Chl-a concentration have a reflectance peak at 700 nm and increases with the concentration. This peak position also changes from 680 to 715 nm when Chl-a concentration rises (GITELSON, 1992).

2.3 Optical classification of natural waters

A well-known optical classification of natural waters is the case 1 and case 2 waters (MOREL; PRIEUR, 1977). Case 1 water has phytoplankton concentration much higher than other water components, having Chl-a and carotenoids as the main absorbers. In contrast, case 2 are waters that have inorganic particles as the main absorbers, with pigments playing a less critical role. Both classes have distinct spectral shapes of absorption and reflectance. In case 1, the spectral absorption curve is V-shape due to the absorption of Chl-a and carotenoid,

centered at 440 and 670 nm, respectively. In case 2, there is no absorption peak due to the steady increase of absorption towards shorter wavelengths, leading to a flattering shape.

Recently, optical classification of natural waters has been carried out in numerous studies with two main purposes: (1) to understand the environment dynamics based on a categorical approach and (2) to optimize the performance of bio-optical algorithms for estimating Chl-a and TSM concentration. In the categorical approach, water masses are clustered into classes according to their optical properties, which may have different limnological parameters such as Chl-a or TSM concentration. For example, Reinart et al. (2003) classify lakes and coastal waters of Estonia and south Finland in five bio-optical classes using optical and limnological parameters. Each class has significant differences in Secchi depth (Z_{sd}), Chl-a, and TSM concentration, which are well correlated with the trophic state of Estonian lakes. Thus, those water classes are named Clear, Moderate, Turbid, Very Turbid, and Brown according to their optical and limnological characteristics. This approach categorizes the waters and simplifies the interpretation, which differs from the use of algorithms for retrieving quantitative results.

Optimizing algorithms involves the classification of OWTs in order to adjust specific algorithms for each class. Optically similar classes tend to have comparable OACs. As a result, algorithm calibration for each optical class becomes more accurate than those developed for a mixture of optically distinct water types. For example, Vantrepotte et al. (2012) show that the optical classification improved the relationship between TSM concentration and R_{rs} at 670 nm. They have obtained three classes and by comparing the mean absolute relative difference MARD (%) in each one of them, before and after tuning the algorithm reported the following results: Classes 1, 2 improved their MARD from 57%, 80% to 42%, 47% , respectively, and class 3 had a small degradation from 38% to 42% highly compensating the overall classification accuracy . Likewise, for optimizing Chl-a algorithms, Shi et al. (2013) optically have classified waters of Taihu, Chaohu, and Dianchi lakes, and Three Gorges and Three Indiana reservoirs. They show that by adjusting a three-band factor algorithm for each

class, the overall results improved from $R^2 = 0.81$ (Chl-a = $0.81x + 6.62$) to $R^2 = 0.95$ (Chl-a = $0.95x + 1.01$).

Regarding the use of optical properties for classifying water masses with different limnological parameters, Spyrakos et al. (2018) show that OWTs derived from different regions have different TSM and Chl-a concentration. Thirteen different spectrums of R_{rs} represented different mean and variance concentration of Chl-a and TSM. Although some classes have different spectrum shapes but no differences in the Chl-a and TSM concentration, those differences are the result of other optical parameters not assessed in the study, such as composition and cell size of phytoplankton or CDOM concentration. In summary, the R_{rs} spectrum shape enables the retrieval of OWTs with different limnological concentrations of optically contrasting waters types around the globe and can be used for monitoring inland waters. For this reason, normalized R_{rs} is suitable for establishing OWTs in diverse environments.

After establishing optical water classes using *in situ* data, it is necessary to train an algorithm for retrieving the same classes from orbital sensors. For example, Lobo et al. (2012) generated ten reference spectrum of R_{rs} to classify Amazon water types. First, they have classified sample data using limnological parameters retrieving a reference spectral R_{rs} for each class. Then, they have used the spectral angle mapper (SAM) for detecting those classes using Hyperion images. Although the detection of optical classes defined by hyperspectral data may suggest that hyperspectral orbital sensors are necessary, the multispectral sensors can also be used for detecting them. Lobo et al. (2012) have used MERIS images for detecting the hyperspectral classes. In fact, using ground truth data, MERIS results outperformed Hyperion's, with accuracies of 67% and 48%, respectively. The authors highlighted that low and spectrally variable Hyperion SNR had a more significant impact on the classification performance than the smaller spectral resolution of MERIS.

Retrieving optical classes from different water systems is also possible. For example, combining data from the eastern English Channel, the southern North Sea, and French Guiana, Vantrepotte et al. (2012) have used *in situ* R_{rs} and

SeaWiFS images for classifying those waters. First, the OWTs are established using normalized hyperspectral R_{rs} by its integer. Second, a novelty detection based on Mahalanobis distance and SeaWiFS simulated bands is applied for detecting those classes in SeaWiFS images. Likewise, Shi et al. (2013) classify Taihu, Chaohu and Dianchi lakes, and Three Gorges and Three Indiana reservoirs. They also use hyperspectral *in situ* R_{rs} for retrieving optical classes, which are then applied to MERIS images. Although both authors detect OWTs in different environments, they do not report the accuracy of their classification results. For this reason, the accuracy of detecting hyperspectral classes in different water systems is not well-known.

2.4 Classification algorithms

The classification algorithms can be unsupervised (clustering) and supervised. Clustering algorithms are used when the classes are not known, and the objective is only grouping data into classes with similar characteristics. Supervised algorithms are used when the classes are known, and the objective is to classify new datasets. Two unsupervised algorithms often used in optical classification of waters are the k-means (SPYRAKOS et al., 2018; ZHANG et al., 2015) and the hierarchical (CHEN et al., 2004; LUBAC; LOISEL, 2007; SHI et al., 2013b; VANTREPOTTE et al., 2012) methods. The supervised algorithms reported in the literature are SAM (LOBO et al., 2012), decision trees (HUANG et al., 2014; LIU et al., 2013; SHI et al., 2013a; SHI et al., 2013b) and novelty detection using Mahalanobis distance (MÉLIN et al., 2011; VANTREPOTTE et al., 2012). This section describes the most used algorithms for the optical classification of waters.

2.4.1 Clustering algorithms

The k-means algorithm uses as input the desired number of clusters (k), the tolerance, and the objective function. The algorithm composed of 3 steps: (1) random or chosen initial centroids (with the number of clusters) in defining n -dimension plan; (2) each sample position in this Euclidean plan is associated with the nearest centroid using a similarity criterion (e.g., Euclidean distance); (3) the

mean of each group is calculated and updated as the new centroid. Then, the distance of the previous and new centroids is calculated, if the distance is lower than a chosen tolerance, the clustering ends, or if the distance is higher than the tolerance, the steps two a three are repeated. In the case that initial centroids are random, each time that k-means algorithm is executed, different clusters are created. Thus, k-means are executed several times and the best clustering is computed by using an objective function, such as the sum squared error (SSE). After executing k-means n times, the centroids converge to the lower SSE (TAN et al., 2008).

The hierarchic agglomerative algorithm is also used for clustering optical data. This algorithm starts with each sample being one individual group; then, the most similar groups are merged according to a distance measurement. This distance is obtained from a distance matrix. When two groups are merged, the distance matrix is updated, and two groups are merged again. This process continues until the desired number of clusters or a threshold of distance is reached. For comparing the distance of clusters composed of two or more samples, the minimum distance, maximum distance, centroids distance, or the Ward linkage method can be used. The Ward method is the most used agglomerative clustering approach and uses SSE as the distance measurement. Ward method combines all possible groups; then, the outcome combination with minimum increase in SSE is selected to be merged. However, the agglomerative methods are computationally expensive and sensitive to noisy data.

2.4.2 Supervised algorithms

The spectral angle mapper (SAM) measures the spectral similarity between a set of spectra and a reference spectrum. This algorithm calculates the angle between two spectra, treating them as vectors in a space with dimensionality equal to the number of bands. The lines connecting each spectrum point and the origin forms an angle between them. Notice that the angle between them is the same regardless of their length; thus, this comparison is not influenced by the R_{rs} magnitude and the angle is a function of the spectral shape. The computation

consists of taking the arccosine of the dot product of the spectrum (KRUSE et al., 1993). The equation can be written as:

$$\alpha_{ng} = \cos^{-1} \left(\frac{\sum_{i=1}^{nb} t_i \text{ref}_i}{(\sum_{i=1}^{nb} t_i^2)^{1/2} (\sum_{i=1}^{nb} \text{ref}_i^2)^{1/2}} \right) \quad (2.17)$$

Where α_{ng} is the angle between the target spectrum (t) and the reference spectrum (ref) calculated for a given number of bands (nb). The closest α_{ng} is from zero, the highest is the similarity between them. Thus, spectrum references and a defined threshold of α_{ng} can be used for classifying different spectrums.

Novelty detection based on Mahalanobis distance assumes that the subsets of R_{rs} of each OWT are described by a multivariate log-normal distribution with a mean (μ) and covariance matrix (Σ) (BISHOP, 1994). The probability density function (P) of a target to a given OWT is computed using the following equation:

$$P(x) = \frac{1}{2\pi^{d/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} \Delta_M^2 \right] \quad (2.18)$$

where x is the $\log(R_{rs})$ of a sample; d is the dimension (e.g., number of bands or wavelengths) of x ; the $|\Sigma|$ is the matrix determinant of Σ ; and the Δ_M^2 is the Mahalanobis distance between a sample and a reference R_{rs} of the respective OWT, which is computed using the equation:

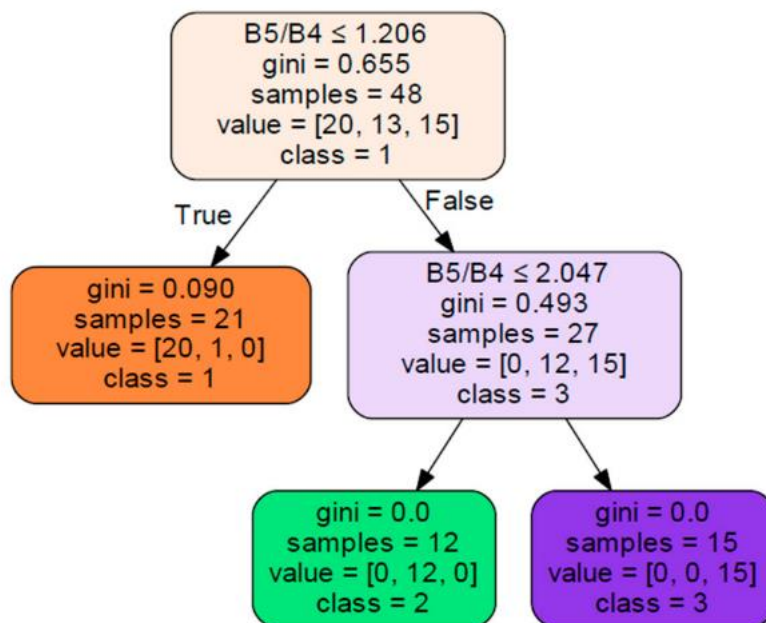
$$\Delta_M^2 = (x - \mu)^T \Sigma^{-1} (x - \mu) \quad (2.19)$$

where T is the transpose, and Σ^{-1} is the inverse matrix. Further, the Δ_M^2 has to be compared to a theoretical threshold (Δ_T^2), where Δ_M^2 lower than Δ_T^2 , the sample x statistically belongs to the respective OWT. The Δ_T^2 may be computed from the

Chi-square distribution that represents a given percentage of the data distribution for a degree of freedom corresponding to the dimension d (VANTREPOTTE et al., 2012).

A decision tree algorithm splits the data based on Boolean rules, where those rules are trained using the input data. All rules follow a logical path, building the decision tree classifier (Figure 2.1). When a set of decision trees are built throughout bootstrap subsamples, the model is called random forest (RF). Generally, RF models tend to be more robust since they are less sensitive to overfitting when no meaningful rules are generated in the decision tree, which leads to misclassifications in the independent data (TAN et al., 2008).

Figure 2.1 – Decision tree example.

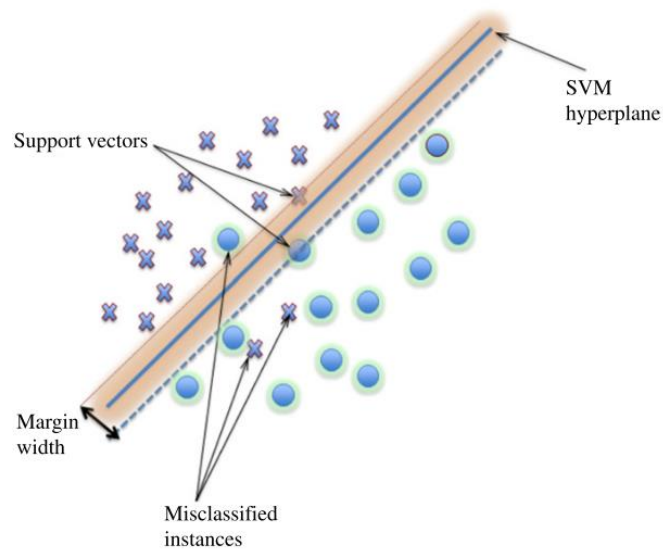


Source: Cairo et al. (2020).

Support Vector Machine (SVM) Classifier (SVC) is a classification algorithm widely used in machine learning problems, and its application in remote sensing has been widely used (MOUNTRAKIS et al., 2011). In its simplest form, the SVC is a binary classifier based on a margin hyperplane that maximizes the margin

limit of decision between classes in a high dimension space (e.g., number of bands), minimizing the misclassifications (Figure 2.2). The margin hyperplane is computed using support vectors (closest samples between two classes) (TAN et al., 2008). Although this machine learning algorithm has not been used for classifying OWTs, its approach using hyperplanes with a maximized limit could be favorable for classifying OWTs.

Figure 2.2 - Linear support vector machine example.



Source: Mountrakis et al. (2011).

3 INCONSISTENCIES BETWEEN KUTSER'S AND MOBLEY'S GLINT REMOVAL METHODS AND HOW THEY AFFECT ALGORITHM CALIBRATION AND SPECTRAL CLUSTERING: ASSESSMENT FOR INLAND WATERS¹

3.1 Introduction

Monitoring the quality of inland waters using remote sensing relies on *in situ* measurements of R_{rs} and OACs, such as Chl-a, TSM, and CDOM concentration. The measurements of *in situ* R_{rs} are used for calibrating/validating algorithms and for clustering waters types optically, which is then applied to orbital sensors images for the effective monitoring of broad areas and time series (LOBO et al., 2012, 2015; MACIEL et al., 2019; MILLER; MCKEE, 2004; PETUS et al., 2010; SHI et al., 2013b; SPYRAKOS et al., 2018; VANTREPOTTE et al., 2012). The above-water R_{rs} measurements are affected by sun and sky glint onto the water's surface, which increases the R_{rs} intensity (MOBLEY, 1999). Consequently, glint can cause a weak relationship between the concentration of OACs and R_{rs} , which may result in weak calibrated algorithms. For this reason, removing glint is a critical part of computing R_{rs} .

Mobley (1999) and Kutser et al. (2013) are well-established glint removal methods (called here as deglint). Mobley's method measures the downwelling sky radiance (L_s) and estimates how much of L_s is reflected as glint to the sensor direction using the factor ρ . This factor was computed using a radiative transfer model and relies on wind speed and sensor/sun nadir and azimuth angles. Despite the correction, Mobley (1999) mentions that some residual sun glint may still persist after deglinting. Kutser's method assumes that R_{rs} of the water body is insignificant in the ranges 350 - 380 nm and 890 - 900 nm, being any signal in those spectral ranges assumed as glint; then, a power function is computed using these spectral ranges which are then subtracted from the R_{rs} (380 - 890 nm). Kutser's method may lead to overcorrection when the assumption of insignificant

¹ This chapter will be submitted to a indexed scientific journal.

R_{rs} (350 - 380 nm; 890 - 900 nm) is not valid, such as in water bodies with high TSM or Chl-a concentrations. Therefore, the R_{rs} intensity and shape computed using each of those methods may not be comparable since Mobley's method tends to undercorrect R_{rs} , while Kutser's method tends to overcorrect R_{rs} .

A few studies have compared the performance of methods for computing *in situ* R_{rs} using different approaches. Bernardo et al. (2018) evaluated four deglint methods by comparing the estimated R_{rs} to simulated R_{rs} . Zibordi et al. (2012) assessed five systems/methods for computing *in situ* R_{rs} by comparing them to a single reference measurement. Garaba and Zielinski (2013) compared differences in five deglint methods in spectral sign (negative/positive) and intensity values. Although these studies have provided valuable results on how different methods produce different *in situ* R_{rs} , there has not been any attempt to assess the effect of merging R_{rs} computed by different measuring methods on calibrating/validating algorithms and on spectral clustering. Sharing R_{rs} data from different sources has increased in the remote sensing of inland waters, and merging R_{rs} from different measuring methods for calibrating/validating algorithms and optically clustering waters is likely to occur.

This chapter investigates the inconsistencies between the $R_{rs}(\lambda)$ derived from Kutser's and Mobley's deglint methods, and how those inconsistencies affect the calibration of TSM/Chl-a algorithms and the spectral clustering. Section 3.3.1 gives an overview of limnological and radiometric data. Section 3.3.2 analyses the inconsistencies between Kutser's and Mobley's methods. Section 3.3.3 analyses two case studies for TSM/Chl-a algorithms calibration using merged data. Issues regarding optical clustering are discussed in section 3.3.4.

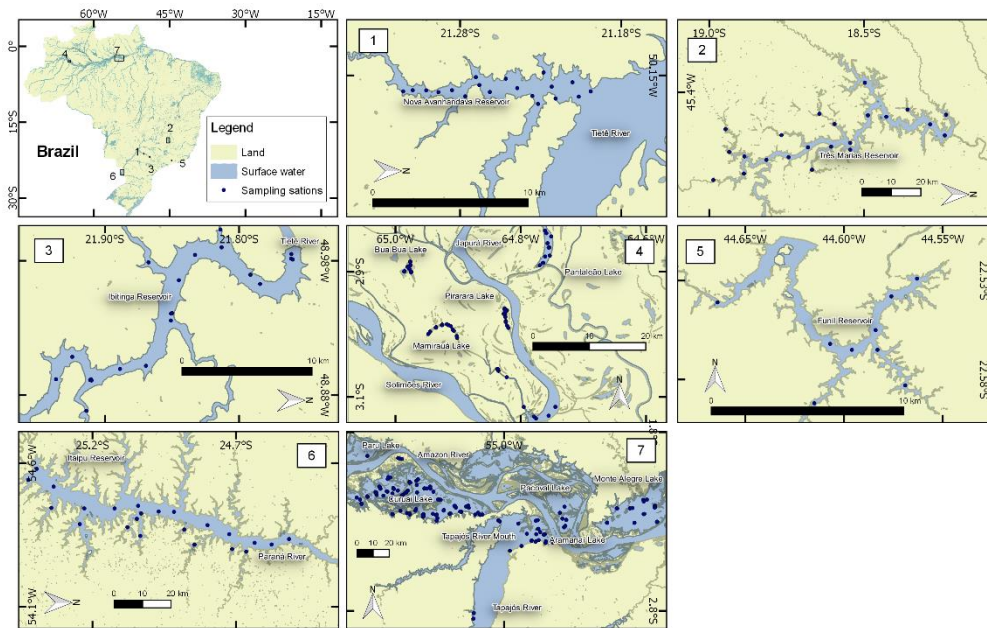
3.2 Methods

3.2.1 Study area and sampling stations

Radiometric, Chl-a, TSM, and dissolved organic carbon (DOC) concentration measurements were carried out in 394 stations in diverse reservoirs, rivers, and lakes in Brazil (Figure 3.1). The reservoirs are Funil (n = 9), Ibitinga (n = 32),

Itaipu (n = 21), Nova Avanhandava (n = 19), and Três Marias (n = 22). The lakes are Bua (n = 15), Pantaleão (n = 12), Mamirauá (n = 57), Pirarara (n = 18), Paru (n = 3), Curuai (n = 120), Pacoval (n = 2), Aramaí (n = 8), Monte Alegre (n = 23); and the rivers are Japurá (n = 5) and Tapajós (n = 28).

Figure 3.1. Study area and sampling stations.



The localization of (1) Nova Avanhandava reservoir; (2) Três Marias reservoir; (3) Ibitinga reservoir; (4) Bua Bua lake, Pantaleão lake, Mamirauá lake, Pirarara lake, and Japurá river; (5) Funil reservoir; (6) Itaipu reservoir; (7) Paru lake, Curuai lake, Tapajós river, Pacoval lake, Aramaí lake, and Monte Alegre lake.

Source: The author.

All data except for Nova Avanhandava were accessed from the LabISA (www.dpi.inpe.br/labisa/) database. All measurements of Chl-a, TSM, and DOC concentration from LabISA followed the Nush (1980), Wetzel and Likens (1991), and APHA (1985) methods, respectively. Survey campaign details can be found in Ferreira (2014) for Três Marias; Jorge (2018) for Pantaleão, Bua Bua, Mamirauá, Pirarara, and Japurá; Rotta et al. (2016) for Nova Avanhandava; de Carvalho et al. (2017) and D. Maciel (2019) for Paru, Curuai, Tapajós, Pacoval,

Aramanaí, and Monte Alegre; Barbosa et al. (2016) for Funil reservoir. For Itaipu reservoir, there is no published study yet.

3.2.2 Radiometric measurements and deglint

The instrument employed was three intercalibrated TriOS-RAMSES spectroradiometers operating from 320 to 950 nm (3 nm of spectral resolution). Concurrently, the irradiance incident onto the water surface ($E_s(\lambda)$) was measured above the water by a cosine sensor, the total upwelling radiance leaving the water ($L_t(\lambda)$) was measured approximately 40° nadir angle (θ) and 135° azimuthal angle (ϕ) from the sun, and the downwelling sky radiance ($L_s(\lambda)$) was measured an approximately 130° θ and 135° ϕ from the sun.

Numerous n replicate radiometric measurements were taken for each sampling station. First, all replicates were resampled to 1 nm of spectral resolution, because L_t , L_s , and E_s slightly differ in the wavelengths measured. Further, R_{rs} equation and deglint were done for each concomitant replicate and later represented by the median $R_{rs}(\lambda)$ for each station.

Two different datasets composed of each deglint method were produced using all samples. The methods used were Mobley (1999) and Kutser et al. (2013). The Eq. (1) computed the R_{rs} using Mobley method:

$$R_{rs(Mobley)}(\lambda, \theta, \Phi) = \frac{L_t(\lambda, \theta, \Phi) - (\rho L_s(\lambda, \theta, \Phi))}{E_s(\lambda)} \quad (3.1)$$

Where $R_{rs(Mobley)}$ is the R_{rs} computed using the Mobley's method, and p is the proportionality factor that relates how much of L_s was reflected towards the sensor direction. For all samples, the $p = 0.028$ was used because the value is adequate for wind speed lower than 5 m s^{-1} at the 40° θ and 135° ϕ (MOBLEY, 1999), which is the conditions of all sampling stations used here

Beforehand calculating the R_{rs} using Kutser's method, the R_{rs} with no glint correction was calculated by:

$$R_{rs(Glint)}(\lambda, \theta, \Phi) = \frac{Lt(\lambda, \theta, \Phi)}{Es(\lambda)} \quad (3.2)$$

Where $R_{rs(Glint)}$ is the R_{rs} without glint correction.

Furthermore, the R_{rs} using Kutser's method was calculated using the equation:

$$R_{rs(Kutser)}(\lambda, \theta, \Phi) = R_{rs(Glint)}(\lambda, \theta, \Phi) - R_{rs(pow\ function)}(\lambda, \theta, \Phi) \quad (3.3)$$

where $R_{rs(Kutser)}$ is the R_{rs} computed using the Kutser's method, and $R_{rs(pow\ function)}$ is the R_{rs} computed by an adjusted power function using all wavelengths between 350 – 380 nm and 890 – 900nm. The software program used to compute the $R_{rs(pow\ function)}$ was the curve fit module from the SciPy ecosystem v1.3 (www.scipy.org).

Last, for evaluating the changes in the R_{rs} shape, each replicate for all methods was normalized by its integer from 400 to 800 nm by:

$$R_{rs(shape)}(\lambda, \theta, \Phi) = \frac{R_{rs}(\lambda, \theta, \Phi)}{\int_{400}^{800} R_{rs}(\lambda, \theta, \Phi) d\lambda} \quad (3.4)$$

Where $R_{rs(shape)}$ is the normalized R_{rs} by its integer (finite approximation per 1 nm). Further, the $R_{rs(shape)}(\lambda)$ median for each station was computed.

3.2.3 Limnological classification

The dataset was split into four limnological classes using different ranges of TSM and Chl-a. The purpose of splitting the data into limnological classes is related to the assumption of insignificant R_{rs} in the ultraviolet and near infra-red (NIR) assumed by Kutser's method, which may not hold with increasing TSM or Chl-a.

Consequently, the dissimilarities between Kutser's and Mobley's methods may increase as TSM and Chl-a increases.

The following thresholds were used in the classification: Class A with TSM and Chl-a are lower than 10 g m⁻³ / mg m⁻³, respectively; Class B with TSM or Chl-a between 10 and 20 g m⁻³ / mg m⁻³, and the other one lower than 20 g m⁻³ / mg m⁻³; Class C with TSM or Chl-a between 20 and 40 g m⁻³ / mg m⁻³, and the other one lower than 40 g m⁻³ / mg m⁻³; Class D with TSM or Chl-a between 40 and 100 g m⁻³ / mg m⁻³, and the other one lower than 100 g m⁻³ / mg m⁻³. For example, a sampling station with TSM = 5 g m⁻³ and Chl-a = 5 mg m⁻³ is class A, a sampling station with TSM = 5 g m⁻³ and Chl-a = 15 mg m⁻³ is class B, a sampling station with TSM = 25 g m⁻³ and Chl-a = 5 mg m⁻³ is class C, and a sampling station of TSM = 45 g m⁻³ and Chl-a = 15 mg m⁻³ is class D. Several TSM and Chl-a thresholds were tested, and we chose these particular thresholds since TSM and Chl-a showed significant differences (Mann Whitney U test, $\alpha = 0.01$; (MANN, H.B., WHITNEY, 1947)) between all classes with a representative number of samples for each one.

3.2.4 Estimation of Kutser and Mobley inconsistencies

For comparing Kutser and Mobley R_{rs} , four statistical analyses were employed for each wavelength from 400 nm to 800 nm, the bias (sr⁻¹), normalized bias (%), Pearson correlation (R), and Wilcoxon signed-rank test (WILCOXON, 1945). The bias (sr⁻¹) was computed using the equation:

$$\text{bias}(\text{sr}^{-1})(\lambda) = R_{rs(\text{Kutser})}(\lambda) - R_{rs(\text{Mobley})}(\lambda) \quad (3.5)$$

And normalized bias (%) using the following equation:

$$\text{bias}(\%) (\lambda) = \frac{R_{rs(\text{Kutser})}(\lambda) - R_{rs(\text{Mobley})}(\lambda)}{R_{rs(\text{Mobley})}(\lambda)} \times 100 \quad (3.6)$$

The bias (sr^{-1}) and normalized bias (%) denotes how much of R_{rs} (Kutser) underestimates or overestimates R_{rs} (Mobley) in steradian and percentage, respectively.

The Wilcoxon test was performed between the associated samples of Kutser's and Mobley's methods, for each wavelength using their replicates. The number of replicates varied for each sample with a minimum of 20. Wilcoxon test evaluated if the median of the spectra for each wavelength corrected by both Kutser's and Mobley's methods showed significant differences ($\alpha = 0.01$). The selection of the Wilcoxon test instead of Student's t-test relates to the fact that most samples showed no normal distribution and outliers. The software program used to compute the Wilcoxon signed-rank test was the Wilcoxon module from the SciPy ecosystem v1.3 (www.scipy.org).

For each limnological class, the Pearson correlation between the Kutser and Mobley concomitant samples was computed.

3.2.5 Algorithms calibration

The calibration performance of TSM and Chl-a algorithms were assessed for three datasets: Kutser deglinted spectra, Mobley deglinted spectra, and Merged deglinted spectra. For this purpose, Curuai and Ibitinga were utilized as case studies for TSM and Chl-a algorithms, respectively. At Curuai lake, D. Maciel (MACIEL et al., 2019) calibrated TSM algorithms and concluded that Band 4 (655 nm; $R_{rs}(\text{B4}_{oli})$) of Landsat-8 OLI and Bands 6 (740 nm; $R_{rs}(\text{B6}_{msi})$) and 7 (783 nm; $R_{rs}(\text{B7}_{msi})$) of Sentinel-2 MSI produced satisfactory results ($R^2 = 0.91$; MAPE = from 30 to 32 %). For Ibitinga reservoir, Cairo et al. (CAIRO et al., 2019b) calibrated Chl-a algorithms and reported that the normalized chlorophyll index (NDCI) (MISHRA; MISHRA, 2012), two bands Ratio (2B), and 3 bands ratio (3B) (GITELSON et al., 2003) had satisfactory performance ($R^2 =$ from 0.78 to 0.98; MAPE = from 22 to 28 %). The NDCI, 2B and 3B ratios were computed using simulated Sentinel-2 MSI bands 3 (560 nm; $R_{rs}(\text{B3}_{msi})$), 4 (665 nm; $R_{rs}(\text{B4}_{msi})$), 5 (705 nm; $R_{rs}(\text{B5}_{msi})$) and $R_{rs}(\text{B6}_{msi})$. Therefore, those bands and indexes were

used to evaluate the effect of merging spectra submitted to both Mobley's and Kutser's deglint method in algorithm calibration.

First, the Landsat-8 OLI and Sentinel-2 MSI bands were simulated using the equation:

$$R_{rs}(B_n) = \frac{\int_{\Delta B_n} R_{rs}(\lambda) RF_{B_n}(\lambda) d\lambda}{\Delta B_n} \quad (3.7)$$

Where B_n is a specific band according to sensor design, $R_{rs}(B_{n_{sensor}})$ is the R_{rs} for each band, ΔB_n is the bandwidth, and $RF_{B_n}(\lambda)$ is the response function for each sensor band. The response function for each band was accessed at the Goddard Space Flight Center (<https://landsat.gsfc.nasa.gov>) and at the European Space Agency (<https://earth.esa.int>) websites. The simulated bands were $R_{rs}(B4_{oli})$, $R_{rs}(B3_{msi})$, $R_{rs}(B4_{msi})$, $R_{rs}(B5_{msi})$, $R_{rs}(B6_{msi})$, and $R_{rs}(B7_{msi})$ using both R_{rs} (Kutser) and R_{rs} (Mobley). Further, NDCI was calculated by the equation:

$$NDCI = \frac{R_{rs}(B5_{msi}) - R_{rs}(B4_{msi})}{R_{rs}(B5_{msi}) + R_{rs}(B4_{msi})} \quad (3.8)$$

The 2B was calculated by the equation:

$$2B = \frac{R_{rs}(B5_{msi})}{R_{rs}(B3_{msi})} \quad (3.9)$$

And the 3B was calculated by the equation:

$$3B = \left(\frac{1}{R_{rs}(B4_{msi})} - \frac{1}{R_{rs}(B5_{msi})} \right) R_{rs}(B6_{msi}) \quad (3.10)$$

The bands and indexes were then used for calibrating TSM and Chl-a algorithms, separately for Mobley, Kutser, and the Merged dataset of half of each method. The calibration procedures for the Mobley, Kutser, and Merged datasets were done 1000 times. At each interaction, the Merged dataset was randomly mixed, and all samples were split into train and test, containing 70% and 30% of samples, respectively. The same training and test samples were utilized for Mobley, Kutser, and Merged datasets at each interaction.

The calibration procedures were empirical using linear and non-linear least squared methods to fit a linear, power, and polynomial (2^o order) functions. The power function was not computed for Chl-a because NDCI showed negative values, which is unfeasible for the algorithm adjustment. The software program used to adjust the empirical models was the line regress and curve fit modules from the SciPy ecosystem v1.3 (www.scipy.org).

The calibrated algorithms were evaluated in terms of coefficient of determination (R²), root means squared error (RMSE), and the mean absolute percentage error (MAPE). All statistical measurements were taken in the test dataset. The RMSE was calculated by:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - J_i)^2} \quad (3.11)$$

And the MAPE by the equation:

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^n \left| \frac{Y_i - J_i}{Y_i} \right| \quad (3.12)$$

Where n is the number of predictions, Y is the observed data, and J is the predicted data by the algorithm.

3.2.6 Spectral clustering

In order to understand the effect of merging Kutser's and Mobley's methods in the spectral clustering, the k-means algorithm (MACQUEEN, 1967) was applied, using the R_{rs} (shape) from 400 to 800 nm as input. The purpose of using all wavelengths is to assess the divergence between both methods in the Euclidean space with equal weight to all wavelengths. Consequently, any dimensionality reduction, such as principal component analysis and functional data analysis, was discarded. Although being aware that the use of all wavelengths may not hold the assumption of independence among variables, but as dimensionality reduction is dataset dependent, it would prevent trough comparisons between Kutser and Mobley datasets. The software program used to compute the k-means algorithm was the KMeans module from the Scikit-learn v0.21 (www.scikit-learn.org).

The k-means algorithm may be outlined in two steps: (1) compute the centroids for each cluster, and (2) label each sample to the nearest centroid. Here, the Mobley dataset was used for defining the clusters' centroids. Then, the Mobley and Kutser samples were labeled to the nearest clusters' centroid using Euclidean distance. Considering the Mobley labeled dataset as True values, a confusion matrix can be generated between both methods. This method represents a viable alternative to assess the influence of changing the deglint method in the spectral clustering process.

Eight clusters were chosen using the gap statistic (TIBSHIRANI et al., 2001) after running several k-means in the Mobley dataset. The k-means algorithm was computed using random seeds in 1000 interactions, and the balanced accuracy for each confusion matrix was computed. Furthermore, one interaction was chosen to demonstrate how the optical clusters differed when changing from Mobley to Kutser R_{rs} (shape).

3.3 Results and discussion

3.3.1 The limnological and radiometric characteristics of limnological classes

The mean Chl-a and TSM between limnological classes showed significant differences, while for DOC, there were no significant differences among classes (Table 3.1). The TSM and Chl-a mean and standard deviation increased from classes A to D, and consequently, in the classes with higher concentration, Kutser's method should overcorrect R_{rs} . Hence, the four limnological classes are distinct groups, and they are suitable for comparing inconsistencies between Kutser and Mobley inconsistencies throughout the different ranges of TSM and Chl-a.

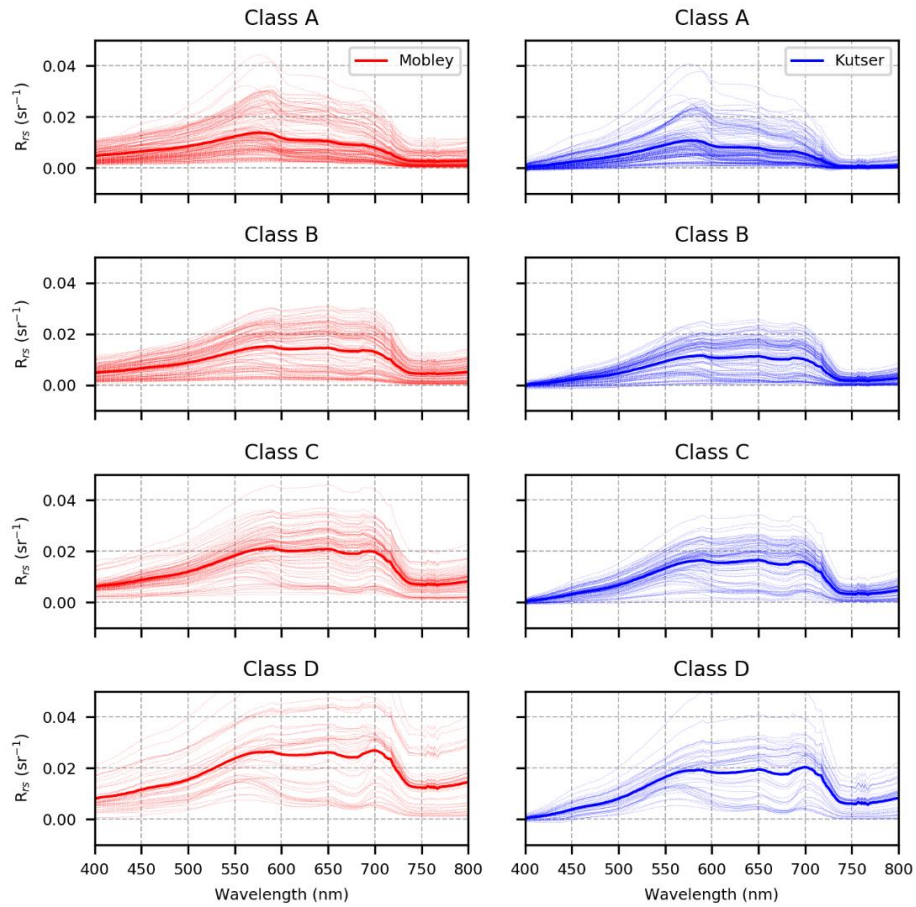
Table 3.1 - The mean and standard deviation (SD) of Chl-a, TSM, and DOC concentrations for classes A, B, C, and D.

Class	n	Chl-a (mg m^{-3})		TSM (g m^{-3})		DOC (g m^{-3})	
		Mean	SD	Mean	SD	Mean	SD
A	147	4.7	2.9	4.1	2.4	3.5	1.8
B	112	10.7	5.2	10.4	5.4	5.3	2.3
C	89	18.6	11.4	19.9	10.3	5.7	3.8
D	46	35.2	24.7	39.9	27.9	4.6	2.3

Source: The author.

The four limnological classes showed distinct R_{rs} intensity and spectral shape (Figure 3.2). The means of R_{rs} (Mobley) and R_{rs} (Kutser) increased from class A to D as a result of higher TSM, mainly in the NIR, where the mean R_{rs} increased from 0 to 0.02. On the other hand, the R_{rs} (Kutser) nearby to 400 nm did not show significant changes from class A to D. Furthermore, Chl-a features diagnostic (absorption at 665 nm and the red edge next to 700 nm) appeared in classes C and D, caused by the higher Chl-a concentration in those classes. As expected, Kutser's method computed lower R_{rs} than that of Mobley's in all limnological classes, since Kutser's method may overcorrect (KUTSER et al., 2013) and Mobley may overestimate R_{rs} (MOBLEY, 1999).

Figure 3.2 - The mean (solid line) and samples (shaded lines) of R_{rs} (Mobley) (red color) and R_{rs} (Kutser) (blue color) for the limnological classes A, B, C, and D.



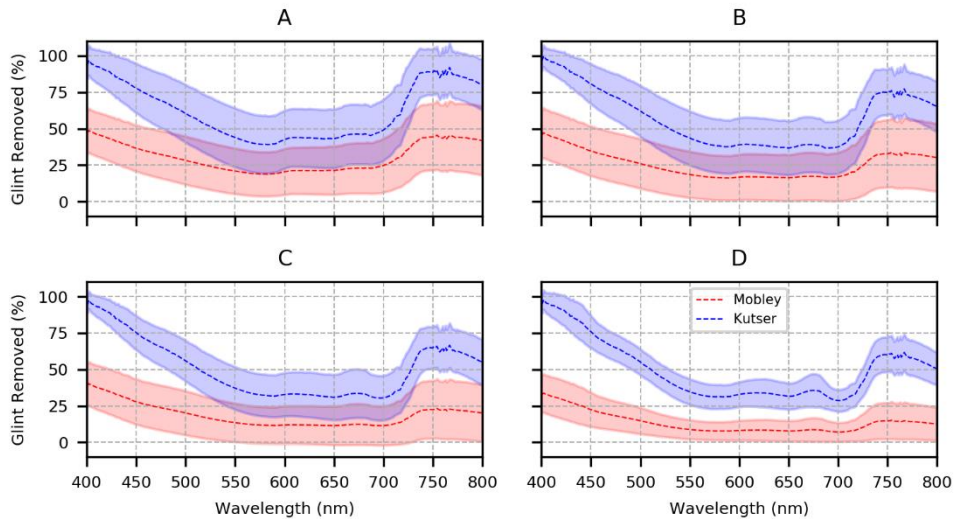
Source: The author.

3.3.2 Inconsistencies between Kutser's and Mobley's deglint methods

Kutser's and Mobley's methods significantly differed in the proportion of R_{rs} removed as glint signal (Figure 3.2). Kutser's method reduced R_{rs} (Glint) from 15 to 100%, even higher than 100% causing negative R_{rs} ($n = 125$). On the other hand, Mobley's method reduced between 0 and 80% of R_{rs} . For all limnological classes, the mean of the glint signal removed by Kutser's method was higher than that of Mobley's throughout the spectrum, particularly in the blue and NIR regions. The higher significant difference in the blue region was associated with the higher slope of the R_{rs} (pow function) when compared to that of the ρ_{L_s} . Overall, results show that even in lower concentrations (Class A) where overcorrection of Kutser's

method is not expected, the mean of R_{rs} removed was higher than that of Mobley's.

Figure 3.3 - The mean (dashed line) and standard deviation (shaded area) of glint removed by Mobley's (red color) and Kutser's (blue color) methods for classes A, B, C, and D.

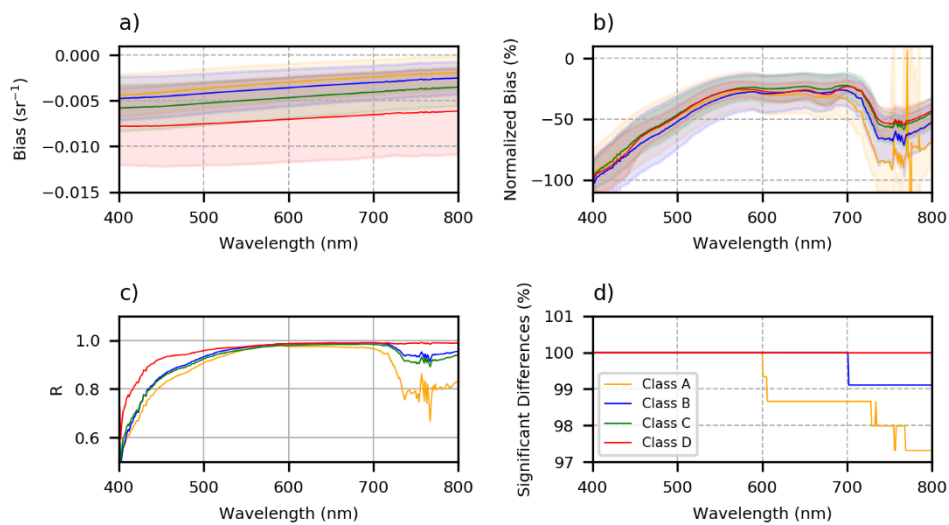


Source: The author.

In all samples analyzed, the R_{rs} (Kutser) and R_{rs} (Mobley) showed significant inconsistencies (Figure 3.4). The mean bias between both methods increased from class A to D, but the mean normalized bias was steady in all limnological classes except in the range between 700 nm and 800 nm. Furthermore, Kutser's and Mobley's methods showed significant differences throughout the spectrum in all samples of classes C and D, and in most samples of classes A and B. Nevertheless, there was a significant positive correlation between both methods in all limnological classes. The steady correlations and normalized bias between 500 and 710 nm throughout the classes suggest that the inconsistencies between both methods do not change in different TSM and Chl-a concentrations investigated here. On the other hand, for measurements above 710 nm, the inconsistencies between both methods reduce as the TSM/Chl-a concentration increases. The most remarkable result emerging from the data is that Kutser's method always underestimates Mobley's in the entire range of TSM and Chl-a,

with normalized bias varying from -20% to -100%. This result is much higher with respect to those reported in comparisons of other methods of measuring *in situ* R_{rs} , which the highest absolute relative percentage difference of 21.2 % between the TACCS-S (Tethered Attenuation Coefficient Chain Sensor) and WiSPER (Wire-Stabilized Profiling Environmental Radiometer) (ZIBORDI et al., 2012).

Figure 3.4 - The divergences between Kutser's and Mobley's methods for classes A, B, C, and D.



The a) is the mean (solid lines) and SD (shaded area) bias (sr^{-1}); b) is the normalized mean and SD bias (%); c) is the Pearson correlation; d) is the percentage of samples for each class that showed significant differences between R_{rs} (Kutser) and R_{rs} (Mobley).

Source: The author.

3.3.3 The effect of merging Kutser's and Mobley's methods in the algorithms calibration

The significance of merging R_{rs} (Kutser) and R_{rs} (Mobley) for calibration of TSM and Chl-a algorithms was evaluated using data from the Curuai Lake and Ibitinga Reservoir, respectively. The significance of differences between the metrics (R^2 , RMSE, and MAPE) describing the changes was statistically assessed with the Wilcoxon signed-rank test using the results of 1000 interactions. Therefore, all comparisons made between datasets represent changes in the R^2 , RMSE, and MAPE medians.

The simulated bands of Sentinel-2 MSI ($R_{rs}(B6_{msi})$, and $R_{rs}(B7_{msi})$) and Landsat-8 OLI $R_{rs}(B4_{oli})$, R_{rs} (Kutser) and R_{rs} (Mobley) showed satisfactory performance for estimating TSM in the Curuai Lake (Table 3.2). The application of Mobley's method allowed the retrieval of slightly more accurate estimates than that of Kutser's using $R_{rs}(B4_{oli})$, while Kutser's method surpasses Mobley's using $R_{rs}(B6_{msi})$ and $R_{rs}(B7_{msi})$. Mobley $R_{rs}(B4_{oli})$ showed statistically significant higher R^2 and lower RMSE than that of Kutser's method, although MAPE exhibited no differences. On the other hand, Kutser $R_{rs}(B6_{msi})$ and $R_{rs}(B7_{msi})$ presented statistically significant better performance than Mobley's method regarding R^2 , RMSE, and MAPE. This pattern occurred in the linear, power, and polynomial adjustments, suggesting that neither of the methods affects the relationship between R_{rs} and TSM. Moreover, the linear relationship between all the bands and TSM (Figure 3.5: 1, 2, and 3 subplots) explains the small difference in the linear, power, and polynomial adjustments.

Table 3.2 - The mean statistical results of calibrated algorithms for TSM in the Curuai Lake, after 1000 interactions.

Band	Dataset	Linear			Power			Polynomial		
		R^2	RMSE ($g\ m^{-3}$)	MAPE (%)	R^2	RMSE ($g\ m^{-3}$)	MAPE (%)	R^2	RMSE ($g\ m^{-3}$)	MAPE (%)
$R_{rs}(B4_{oli})$	Mobley	0.66	10.3	60	0.7	10.4	53	0.68	11.3	54
	Kutser	0.61	11.3	60	0.65	12	52	0.63	13.1	53
	Merged	0.56	11.9	61	0.58	11.9	57	0.58	12.5	58
$R_{rs}(B6_{msi})$	Mobley	0.76	8.5	54	0.78	8.1	51	0.78	8.2	53
	Kutser	0.85	6.8	40	0.87	6.3	37	0.87	6.4	38
	Merged	0.55	11.9	62	0.55	12.1	62	0.55	12.2	62
$R_{rs}(B7_{msi})$	Mobley	0.77	8.3	53	0.78	8	51	0.78	8.1	52
	Kutser	0.85	6.7	40	0.87	6.3	37	0.87	6.4	38
	Merged	0.58	11.5	61	0.58	11.7	60	0.58	11.9	60

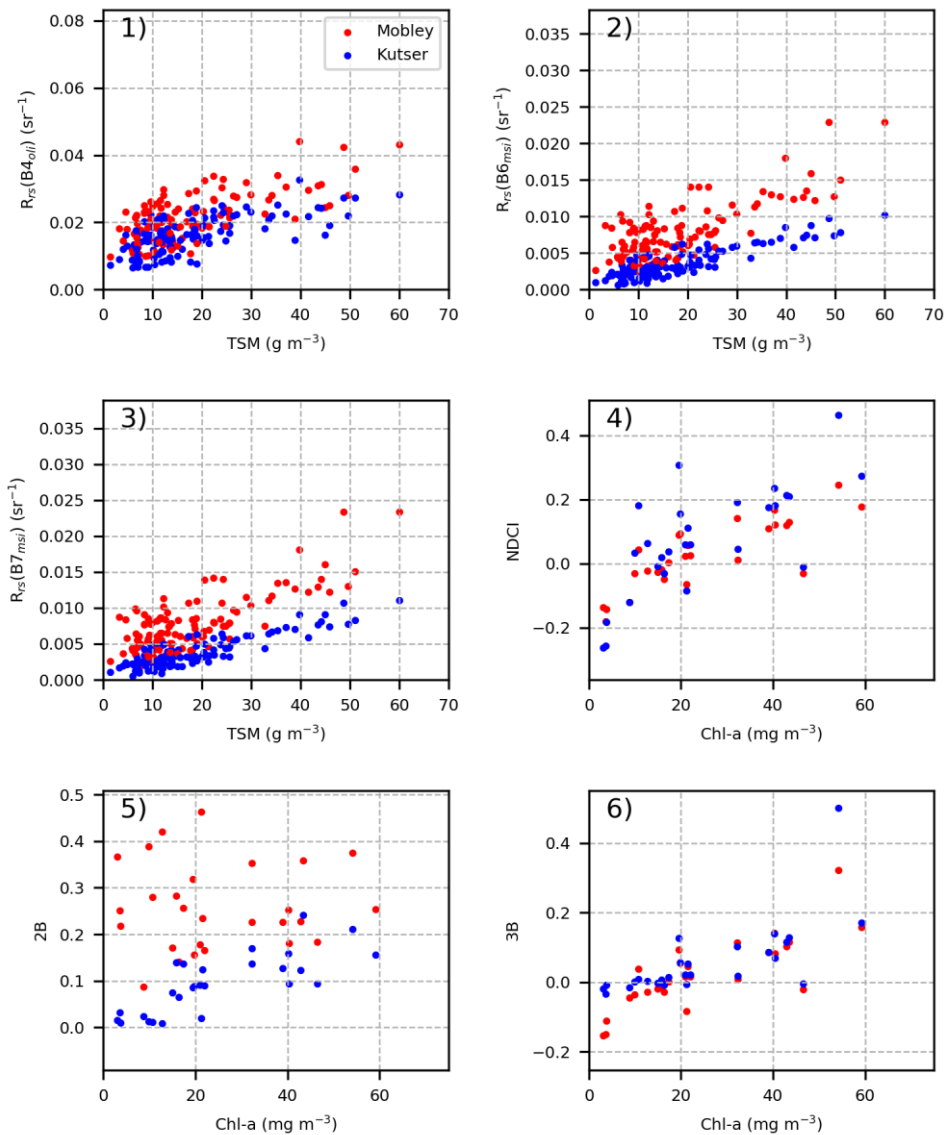
The columns are the linear, power, and polynomial adjustments, and their respective R^2 , RMSE, and MAPE. The rows are the simulated bands and the datasets used for simulating the bands.

Source: The author.

The performance of all bands and adjustments for estimating TSM in the Curuai Lake showed weaker performance with the Merged dataset in relation to that when either Kutser or Mobley datasets were used. The changes of R^2 , RMSE, and MAPE between the Merged and Kutser or Mobley were statistically different except the RMSE of OB4 power and polynomial adjustments. The decrease in performance varied from $R^2 = 0.05$, $RMSE = 0.06 \text{ gm}^3$, and $MAPE = 1\%$ ($R_{rs}(B4_{oli})$ linear model) to $R^2 = 0.23$, $RMSE = 4 \text{ gm}^3$, and $MAPE = 11\%$ ($R_{rs}(B6_{msi})$ power model). The $R_{rs}(B4_{oli})$ algorithms exhibited the lowest decreases, while $R_{rs}(B6_{msi})$ and $R_{rs}(B7_{msi})$ algorithms presented the highest drop in performance. As reported earlier (section 3.3.2), the normalized bias means between Kutser's and Mobley's methods at 655 nm where $R_{rs}(B4_{oli})$ is centered has been -25%, while the normalized bias means of 740 nm and 783 nm where $R_{rs}(B6_{msi})$ and $R_{rs}(B7_{msi})$ are centered, has varied from -50% to -90%. As a result, the higher difference of $R_{rs}(B6_{msi})$ and $R_{rs}(B7_{msi})$ produced a more dispersed distribution than that of $R_{rs}(B4_{oli})$ regarding their relationship with TSM (Figure 3.5: 1, 2, and 3 subplots), reducing their algorithm performances.

Chl-a calibrated algorithms using NDCI, 2B, and 3B indexes in the Ibitinga Reservoir exhibited acceptable performance using $R_{rs}(Kutser)$ and $R_{rs}(Mobley)$ in most cases (Table 3.3). For the NDCI and 3B, both deglint methods exhibited adequate performance, but Mobley's method showed statistically significant better results than those of Kutser's linear and polynomial adjustments. Surprisingly, the Mobley 2B showed poor performance, while Kutser 2B produced consistent results. The Mobley 2B and Chl-a relationship showed a dispersed distribution (Figure 3.5, subplot 5), while Kutser 2B and Chl-a relationship were linear. The reason for this rather unexpected result is not apparent, but it seems that Mobley uncertainties in the $R_{rs}(B3_{msi})$ (560 nm) could have a more considerable influence than Kutser uncertainties for the 2B model. Presumably, the $R_{rs}(B6_{msi})$ (740 nm) should not be the reason for Mobley 2B's poor performance, since $R_{rs}(B6_{msi})$ was also employed in the 3B and exhibited acceptable results.

Figure 3.5 - Scatter plots between simulated bands/indexes with TSM and Chl-a



Subplots from 1 to 3 are for the Curuai Lake. Subplots from 4 to 6 are for Ibitinga Reservoir. The red dots are using the Mobley dataset, and the blue dots are the Kutser dataset.

Source: The author.

Although the significant reduction of algorithms performance using the Merged dataset in the Curuai Lake, the Merged dataset of Ibitinga showed a minor reduction in the performance of NDCI and 3B algorithms. The linear and polynomial adjustments of the Merged dataset using NDCI exhibited higher

MAPE than using individual deglint methods, but R^2 and RMSE did not show significant changes compared to Kutser's. Considering 3B, the Merged dataset did not present significant changes except the MAPE of the Polynomial model. The minimum reduction in model performance when using NDCI and 3B was a consequence of a less dispersed distribution regarding their relationship with Chl-a for Mobley and Kutser datasets (Figure 3.5; subplots 4 and 6). The less dispersed relationship occurred because the equations of NDCI and 3B take into account the relative proportion between bands, which diminished the effect of the bias between Kutser's and Mobley's methods. Therefore, calibrating algorithms using NDCI and 2B are less sensitive by the bias between Kutser's and Mobley's methods than using individual bands.

Table 3.3 - The mean statistical results of calibrated algorithms for Chl-a in the Ibitinga reservoir, after 1000 interactions.

Index	Dataset	Linear			Polynomial		
		R^2	RMSE (g m ³)	MAPE (%)	R^2	RMSE (g m ³)	MAPE (%)
NDCI	Mobley	0.66	9.5	42	0.65	9.8	47
	Kutser	0.54	11.4	51	0.52	12.2	54
	Merged	0.55	11.3	55	0.52	12.4	60
2B	Mobley	0.08	17	133	0.08	17	137
	Kutser	0.6	10.7	57	0.58	11.3	58
	Merged	0.16	16.5	126	0.21	16.1	112
3B	Mobley	0.65	10	44	0.61	12.4	52
	Kutser	0.59	15.4	80	0.54	20.2	65
	Merged	0.6	12.1	66	0.55	18.7	78

The columns are the linear and polynomial adjustments, and their respective R^2 , RMSE, and MAPE. The rows are the indexes and the datasets used for each index.

Source: The author.

3.3.4 The effect of merging Kutser's and Mobley's methods in the spectral clustering

The Kutser R_{rs} (shape) was strongly modified by negative R_{rs} values, and for this reason, 125 samples were removed when labeling Kutser and Mobley R_{rs} (shape)

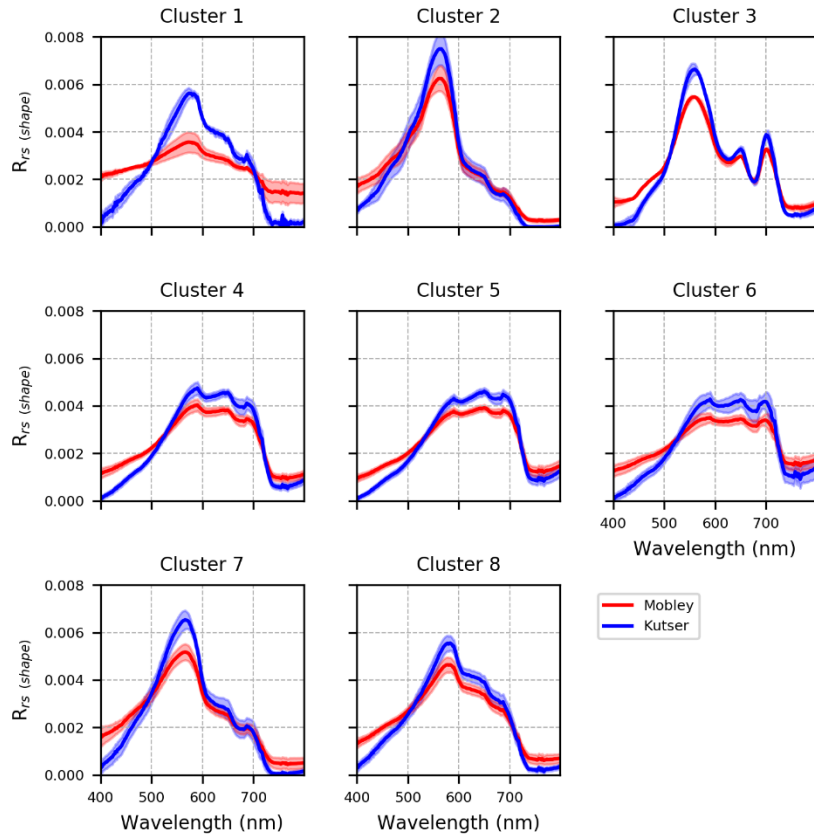
to the cluster centroids. Thus, in all 1000 interactions, the confusion matrixes and the computed balanced accuracy were not influenced by unusual R_{rs} shapes.

After running a thousand k-means interactions with random seeds, swapping from Mobley to Kutser dataset in the labeling step modified a large number of samples. The balanced accuracy mean and standard deviation were 0.55 and 0.2, respectively, which means that changing $R_{rs (shape)}$ from Mobley to Kutser assigned almost half (cluster weighted) samples to different clusters.

To illustrate the change of spectral shape from Mobley to Kutser dataset, a random clustering interaction was chosen. For the eight clusters mean, the Kutser $R_{rs (shape)}$ was lower than Mobley $R_{rs (shape)}$ from 400 to 500 nm and from 700 to 800 nm (Figure 3.6). On the other hand, Mobley $R_{rs (shape)}$ was lower than Kutser from 500 nm to 700 nm. This difference was a result of the higher removal of $R_{rs (Glint)}$ from Kutser than Mobley in the blue and NIR region (see section 3.3.2). Those differences were accentuated in the $R_{rs (shape)}$ equation, which reduces the influence of R_{rs} intensity but increases the influence of relative proportion among wavelengths.

Changing from Mobley to Kutser in the k-means labeling step packed most samples to a few clusters (Table 3.4). The most remarkable result was that cluster 1 samples computed by Kutser were nearer to cluster 8 centroid than cluster 1 centroid, which lead to Kutser not labeling any sample as cluster 1. This change occurred because Kutser was not able to compute flat spectral shapes, as Mobley did, caused by the significant removal of $R_{rs (Glint)}$ in the blue and NIR regions. For the remaining clusters, most samples of cluster 7 changed to cluster 2, and clusters 4 and 6 changed to cluster 5. In those cases, the samples were packed between similar clusters, which presumably should be more sensitive to changes in the $R_{rs (shape)}$ when changing deglint methods. Therefore, merging both methods led to two outcomes: (1) Kutser produced a different spectral shape in cases of flat spectral shape computed using Mobley, and (2) Kutser led samples of similar clusters to pack into a single one.

Figure 3.6 - The mean (solid lines) and standard deviation (shaded area) of $R_{rs}(\text{shape})$ for a clustering interaction using the Mobley $R_{rs}(\text{shape})$ in the labeling procedure.



The red is the Mobley $R_{rs}(\text{shape})$, and the blue is the Kutser $R_{rs}(\text{shape})$ for the same samples.

Source: The author.

Table 3.4 - Confusion matrix between Kutser and Mobley labeled data for the same clustering interaction.

		Kutser dataset								
		Clusters	1	2	3	4	5	6	7	8
Mobley dataset	1	0	0	0	0	0	0	0	0	4
	2	0	18	0	0	0	0	0	0	0
	3	0	0	2	0	0	0	0	0	0
	4	0	0	0	18	29	0	0	0	0
	5	0	0	0	0	92	0	0	0	0
	6	0	0	0	2	46	0	0	0	1
	7	0	28	1	0	0	0	0	5	0
	8	0	0	1	3	0	0	0	0	28

Source: The author.

The results show that when Kutser's and Mobley's R_{rs} (shape) are merged in the spectral clustering, the samples are labeled to different clusters depending on the method, which can cause unexpected results. Such unexpected results may be erroneous representative samples of optical clusters and retrieve poor results for algorithms based in those clusters. Assuming that the spectral shape is a result of the OACs, and therefore samples among one cluster have more similar OACs than other clusters, label an R_{rs} (shape) sample using Kutser to the cluster centroid computed by Mobley could produce a cluster with heterogeneous OACs. Furthermore, calibrate algorithms based in a non-homogeneous cluster may not show better results than the generic approach.

3.4 Conclusion

We have demonstrated that R_{rs} estimated using Kutser's and Mobley's methods are significantly different from 400 to 800 nm in a wide range of TSM/Chl-a conditions, where Kutser's method underestimated R_{rs} computed using Mobley's method in all cases, mainly in the blue and NIR region. We have also shown that both methods are useful for calibrating TSM/Chl-a algorithms, but their particular performance varies with bands and assessment indexes. The strong point of our study lies in demonstrating the effect of merging R_{rs} computed using Kutser's and Mobley's methods in calibrating TSM/Chl-a algorithms and spectral clustering. The use of datasets subject to both deglint methods reduced the performance of both, TSM and Chl-a algorithms, and labeled samples to different clusters.

Our study could help researchers that are handling large databases with different deglint methods for calibrating TSM/Chl-a algorithms and spectral clustering. However, our results are limited between Kutser's and Mobley's methods. Since studies that focus on validation between *in situ* methods for computing R_{rs} are quite rare, further studies need to be carried out to establish whether different methods are comparable in terms of algorithm calibration and spectral clustering.

4 BRAZILIAN OPTICAL WATER TYPES²

4.1 Introduction

The Brazilian waters are composed of distinct aquatic environments due to its continental territory and diversified environments, which results in waters with high variability of TSM, Chl-a, and CDOM concentrations. In the Amazon basin, rivers and floodplain lakes exhibit black waters that are rich in dissolved organic carbon, white waters with high TSM concentration, and clear waters with low TSM, Chl-a, and CDOM concentrations (SIOLI, 1968). Some tropical reservoirs present a high level of eutrophication due to severe algae blooms and high Chl-a concentration (AUGUSTO-SILVA et al., 2014; BARBOSA et al., 2016; CAIRO et al., 2016; LONDE, 2008). On the other hand, other reservoirs are oligotrophic with high water transparency and low TSM, Chl-a, and CDOM concentrations (BARBOSA et al., 2014; CURTARELLI et al., 2019; FERREIRA, 2014; ROTTA et al., 2019). Coastal waters area is susceptible to harmful algal bloom of (ALVES; MAFRA, 2018), which may reach the inner shelf of estuaries (NOERNBERG et al., 2017). This comprehensive range of environments and limnological properties responds to the occurrence of numerous OWTs with distinct R_{rs} shapes and magnitudes.

R_{rs} responds mostly to the backscattering and absorption properties of Chl-a, TSM, and CDOM (KIRK, 2011). The R_{rs} at 430 nm and 675 nm are reduced by Chl-a absorption peaks (MOREL; BRICAUD, 1981). In conditions of high Chl-a concentration, an R_{rs} peak next to 700 nm is created as a result of increased particulate scattering offset by dramatically increasing water absorption at wavelengths larger than 700 nm, and the adjacent Chl-a absorption band at 675 nm (MARK W. MATTHEWS, 2017). TSM derived from algae and non-algae particles intensify the backscattering (SUN et al., 2010) and, consequently, the R_{rs} intensity throughout all spectrum from 400 to 750 nm. Where CDOM has a high absorption influence (MILLER et al., 2002), R_{rs} magnitude is reduced in the blue range. Despite the fact that R_{rs} is an apparent optical property (AOP) and

² This chapter will be submitted to a indexed scientific journal.

modified by the light field geometry and environmental conditions, it has been widely used for defining OWT with distinct Chl-a, TSM, and CDOM concentrations (MOORE et al., 2014; SPYRAKOS et al., 2018; VANTREPOTTE et al., 2012).

Since OWTs based on R_{rs} discriminate changes in Chl-a, TSM, and CDOM concentrations, this approach has been applied for describing aquatic environments and detecting water changes in time-series. Spyrakos et al. (2018) investigated the major OWTs of inland and coastal waters around the world based on R_{rs} shape, discussing their limnological and the inherent optical properties (IOP). Uudeberg et al. (2019) studied the OWTs of lakes from the boreal regions and coastal areas using key features of R_{rs} such as wavelength maximum, slopes, and amplitude. Further, they developed remote sensing methods for detecting those boreal OWTs on Sentinel-2 MSI and Sentinel-3 OLCI images. Vantrepotte et al. (2012) analyzed the OWTs of different coastal regions and developed a remote sensing scheme for detecting them using SeaWiFS images. Then, the SeaWiFS provided substantial spatial-temporal data for the OWTs.

Furthermore, OWTs has been used for improving remote sensing algorithms by various researches, who established OWTs and then calibrated specific algorithms for each of them (CAIRO et al., 2020; LE et al., 2011; MOORE et al., 2014; SHI et al., 2013b). In those studies, each OWT had identical features (e.g., Chl-a range or absorption dominated by CDOM), which improved the performance of their algorithms. For example, the OWTs of Cairo et al. (2020) exhibited a substantial difference in Chl-a ranges, where the bio-optical algorithms were adjusted for each OWT and consequently Chl-a range. When the algorithms were adjusted to specific ranges of Chl-a, the overall estimation of Chl-a surpassed that of estimated for a unique algorithm.

Many studies defined OWTs of a variety of regions, but none of them have addressed Brazilian waters. Even though OWTs using global data have already been produced (SPYRAKOS et al., 2018), including Brazilian waters, those OWTs were based only in the R_{rs} shape. The waters of the Amazon basin show

distinct variability in the R_{rs} intensity where black waters with high CDOM concentration are associated with low R_{rs} . While white waters with high TSM concentration are related to high R_{rs} (JORGE et al., 2017). Despite the importance of using R_{rs} shape rather than R_{rs} magnitude in coastal waters (VANTREPOTTE et al., 2012), the removal of R_{rs} intensity from the analysis would decrease the representativeness of OWTs of Brazilian waters. Therefore, a study that focuses on Brazilian waters that combine the R_{rs} shape and magnitude could improve the present characterization of the Brazilian OWTs provided by the global data. Such OWTs could support remote sensing applications such as detecting algal blooms and high sediment loadings and also improve algorithms for estimating Chl-a, TSM, and CDOM concentrations.

Therefore, the objective of this study focuses on characterizing the OWTs of Brazilian waters. First, we compute the OWTs using R_{rs} shape and intensity. Second, their optical and limnological characteristics are exhibited and evaluated. Last, we discuss the OWTs sources.

4.2 Material and methods

4.2.1 Study area and survey campaigns

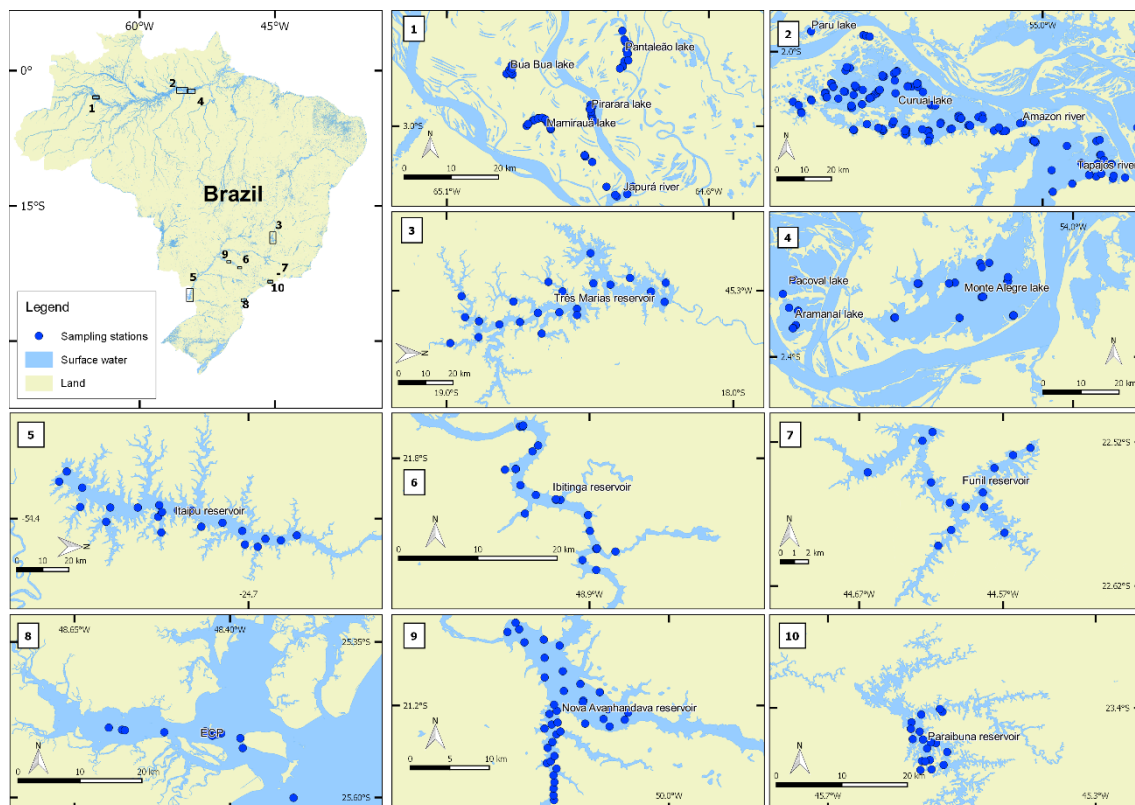
The survey campaigns were carried out in 19 water bodies in Brazil with a total of 483 sample stations, where radiometric measurements and limnological samples were taken from 2013 to 2018 (Figure 4.1). All R_{rs} measurements of those sample stations follow the Mobley (1999), while data from Kutser (2013) was not merged due to the uncertainty increase in defining the OWTs. The water bodies consist of several lakes, rivers, reservoirs, and a subtropical estuary, encompassing several different optical and limnological conditions that occur in the Brazilian waters.

Mamirauá (n = 71), Pantaleão (n = 12), Bua Bua (n = 15), and Pirarara (n = 13) lakes are located in the Solimões floodplain at the confluence of Japurá (n = 4) and Solimões rivers (Figure 4.1; 1). Survey campaigns were carried out in the rising and receding water periods. In the rising period, high input of organic matter

occurs in Bua Bua and Mamirauá lakes, while high sediment loading occurs in Pirarara and Pantaleão lakes (JORGE et al., 2017).

Paru (n = 3), Curuai (n = 125), Pacoval (n = 2), Aramanaí (n = 8), and Monte Alegre (n = 23) lakes and Amazon (n = 1) and Tapajós (n = 29) rivers are located in the lower basin of Amazon river (Figure 4.1 ; 2 & 4) (de Carvalho et al., 2017; Lobo et al., 2015; Maciel et al., 2019). In the Amazon region, the waters may vary among black waters rich in dissolved organic carbon, white waters with high TSM concentration, and transparent waters with low TSM, Chl-a, and CDOM concentrations (SIOLI, 1968).

Figure 4.1 - Study areas and sampling stations located in the Brazilian waters.



The (1) are Bua Bua, Pirarara, Pantaleão, Mamirauá, and Japurá river; (2) are Paru, Curuai, Amazon and Tapajós rivers; (3) is Três Marias reservoir, (4) are Aramanaí, Pacoval, and Monte Alegre, (5) is Itaipu reservoir, (6) is Ibitinga reservoir, (7) is Funil reservoir, (8) is the ECP, (9) is the Nova Avanhandava reservoir, and (10) is Paraibuna reservoir.

Source: The author.

Três Marias reservoir (n = 22) is located in the Minas Gerais state (Figure 4.1; 3) has been classified as having transparent waters with low TSM, Chl-a, and CDOM concentrations (FERREIRA, 2014). Itaipu reservoir (n = 23) is located in Paraná state at the Paraguay border (Figure 4.1; 5), having waters varying from high transparency to turbid waters (RIBEIRO FILHO et al., 2011). Paraibuna reservoir (n = 18) is located in the São Paulo state (Figure 4.1; 10), with waters also characterized as very transparent (CURTARELLI et al., 2019).

Ibitinga and Nova Avanhandava reservoirs are located in the Tietê river cascade system (Figure 4.1; 6 & 9), which is one of the most industrialized and densely populated drainage basins in Brazil. Survey campaigns in Ibitinga reservoir (n = 38) were conducted under oligotrophic to cyanobacterial blooms conditions (CAIRO et al., 2016, 2019b). For Nova Avanhandava (n = 40), the bio-optical conditions were characterized as transparent waters in all campaigns (RODRIGUES et al., 2015; ROTTA et al., 2016).

Funil reservoir (n = 14) is located in the Rio de Janeiro state (Figure 4.1; 7) and is characterized by eutrophic waters (BARBOSA et al., 2016). The Estuarine Complex of Paranaguá (ECP) (n = 13) is located on the Paraná state coast (Figure 4.1; 8), where low turbid waters were found (DA SILVA et al., 2019).

4.2.2 AOPs, IOPs, and limnological measurements

The above-water radiometric measurements were taken using a Hand-Held 2 VNIR for the Paraibuna reservoir and the ECP. While three intercalibrated TriOS-RAMSES spectroradiometers were used for the remaining water bodies. The method of Mobley (1999) was used for computing the R_{rs} . The total water-leaving radiance ($L_t(\lambda, \theta, \phi)$), the downwelling sky radiance ($L_s(\lambda, \theta, \phi)$), and the total irradiance incident onto the water surface ($E_s(\lambda)$) were measured. The wavelength (λ) interval used in this study ranged from 400 to 800 nm, where the spectral resolution of 3 nm of both devices was resampled to 1 nm. Taking the sun as a reference, the nadir (θ) and azimuth (ϕ) angles of L_t were 40° and 135° , respectively, while the nadir and azimuth of L_s were 130° and 135° , respectively. Then, the R_{rs} was computed using the equation:

$$R_{rs}(\lambda) = \frac{L_t(\lambda) - \rho L_s(\lambda)}{E_s(\lambda)} \quad (\text{sr}^{-1}) \quad (4.1)$$

Where ρ is a factor that relates how much of L_s is reflected in the water-air interface towards the sensor direction, and it is a function of wind speed, solar zenith angle, and sensor zenith and azimuth angle. The $\rho = 0.028$ was utilized here since it is suitable for the geometry of acquisition used here and at wind speeds lower than 5 m s^{-1} in blue sky conditions or at any wind speed in overcast skies (MOBLEY, 1999). Thus, the ρL_s accounts for the glint in the R_{rs} measurement that must be removed.

In situ absorption and attenuation coefficients (m^{-1}) measurements were carried out with a 15 cm and 25 cm pathlength AC-S attenuation-absorption meter with the spectral range from 400 to 750 nm and spectral resolution of 3.5 nm. The influence of temperature and water was removed following manual protocols (WET LABS, 2009), while scattering correction was done using the method of Kirk (1992) using the scattering constant fraction of 0.18. The 15 cm AC-S was used in very turbid waters while 25 cm AC-S was used in more transparent waters. As the AC-S profiles were taken in different depths depending on the survey campaign, we standardized using the median from 30 to 80 cm depth. Further, the measurements were resampled to 1 nm of spectral resolution, and scattering coefficient (m^{-1}) was computed by subtracting absorption from attenuation.

The limnological parameters measured were Chl-a (mg m^{-3}), TSM (g m^{-3}), CDOM (m^{-1}), and Secchi depth (Z_{sd}) (m). For CDOM, the absorption coefficient at 440 nm was used as a proxy to its concentration. The limnological measurements were taken using different methods, and details for each method are in the studies mentioned in section 4.2.1.

It is worth mentioning that some R_{rs} measurements do not have concomitant limnological, absorption, attenuation, and scattering measurements. The Japurá river and Itaipu reservoir do not have CDOM data. The AC-S measurements were

not taken in ECP, Funil, Itaipu, Pacoval, Pantaleão, Paraibuna, Amazon river, and Japurá river. Nevertheless, this should not influence the analysis because all OWTs exhibited representative limnological and AC-S measurements for characterizing them.

4.2.3 Establishment of the optical water types

For establishing the OWTs based on their shape and magnitude, first, the $R_{rs}(\lambda)$ shape was emphasized by normalizing the $R_{rs}(\lambda)$ using its integer (VANTREPOTTE et al., 2012), utilizing the following equation:

$$R_{rs\ (shape)}(\lambda) = \frac{R_{rs}(\lambda)}{\int_{400}^{800} R_{rs}(\lambda) d\lambda} \quad (4.2)$$

Where $R_{rs\ (shape)}$ is the normalized R_{rs} by its integer computed by finite approximation using 1 nm of spectral resolution, from 400 to 800 nm.

Furthermore, the Functional Data Analysis (FDA) method (RAMSAY, 2006) was used in each sample of $R_{rs\ (shape)}$. In this method, spectral $R_{rs\ (shape)}$ is viewed as a function of the wavelength that varies according to its basis coefficients. The use of basis coefficients for clustering has been proposed by Spyrakos et al. (2018) who have shown their following advantages of the $R_{rs\ (shape)}$ values: (1) reduction of data dimensionality supporting the assumption of independence among variables (FRALEY, 1998); (2) data noise reduction; and (3) capture of key $R_{rs\ (shape)}$ features by the basis functions. Since the number of basis functions that estimate the $R_{rs\ (shape)}$ controls the degree of flexibility, numbers varying from 15 to 35 were tested in the dataset. In other words, a small number of basis leads to a smoother $R_{rs\ (shape)}$ than numerous basis. On the other hand, a small number of basis can fail to capture the $R_{rs\ (shape)}$ curves. A total of 25 cubic basis functions equally spaced from 400 to 800 nm resulted in a reliable estimate of $R_{rs\ (shape)}$ while removing data noise. All functions related to FDA are implemented in the R-Project (RAMSAY et al., 2018).

The k-means algorithms (MACQUEEN, 1967) was used for clustering the basis functions. First, the Gap statistic (TIBSHIRANI et al., 2001) was used for choosing the optimal number of clusters (k). As Gap has not stabilized with increase k in the present study dataset, the number of k was set at the point where the Gap curve reaches the horizontal asymptote. For defining the initial seeds, the k-means++ algorithm (ARTHUR; VASSILVITSKII, 2007) was used, since it speeds up the convergence and improves the accuracy of the k-means algorithm. Furthermore, the k-means algorithm was fed with the initial seeds, and the convergence tolerance was set to 0.001. The implemented version of k-means utilized was the KMeans module from the Scikit-learn v0.21 (PEDREGOSA et al., 2011).

4.3 Results and discussion

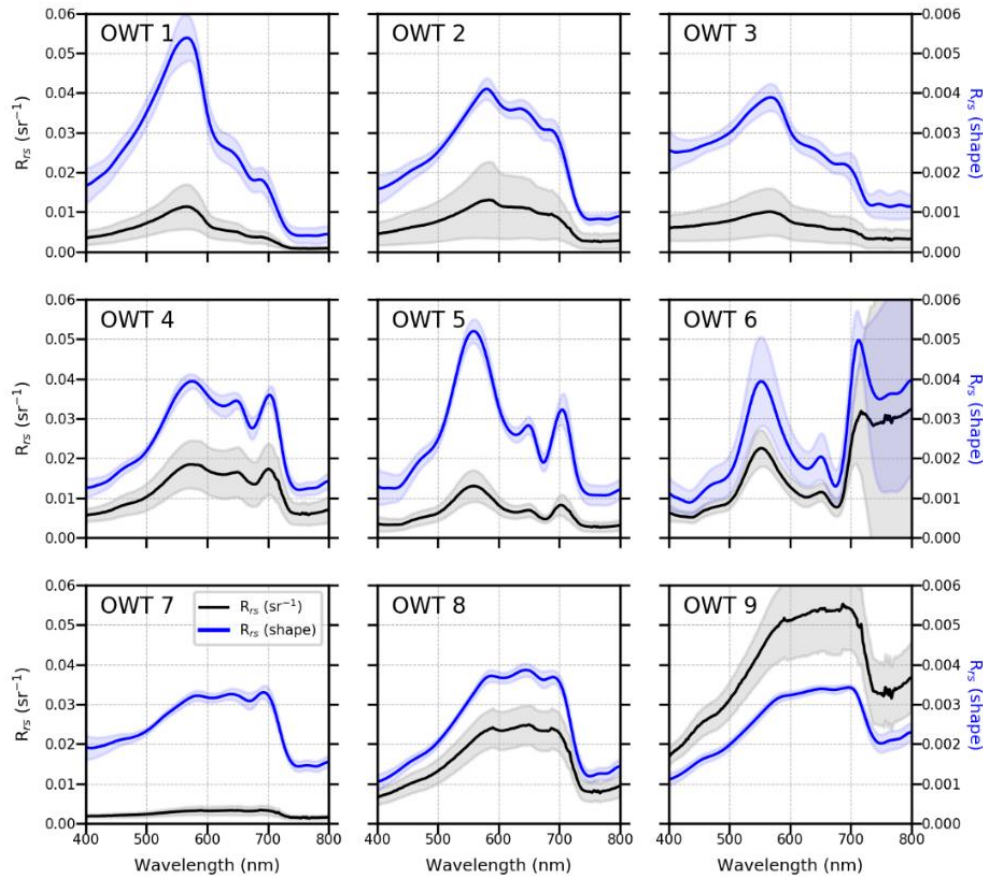
4.3.1 Clusters and adjustments to OWTs

A total of 16 clusters was originated using the k-means algorithm. However, many clusters showed very similar $R_{rs}(\text{shape})$ and no difference in TSM, Chl-a, or CDOM concentrations. To avoid redundancy among the OWTs, the number of clusters was reduced to 7 by merging the most similar clusters due to the Euclidean similarity of $R_{rs}(\text{shape})$. Thus, those 7 clusters could show a consistent difference in $R_{rs}(\text{shape})$ and also in their TSM, Chl-a, and CDOM concentrations.

Furthermore, one cluster showed high variability in $R_{rs}(\lambda)$, TSM, and CDOM, but very similar $R_{rs}(\text{shape})(\lambda)$. Since this high variability represents distinct optical and limnological properties but has an interrelated spectral shape, this cluster was separated using the $R_{rs}(650)$. The $R_{rs}(650)$ lower than 0.01 is the OWT 7 (Figure 4.2), the $R_{rs}(650)$ between 0.01 and 0.038 is the OWT 8, and the $R_{rs}(650)$ higher than 0.038 is the OWT 9. Those thresholds at 650 nm were set because they provided the best way for splitting OWTs 7, 8, and 9. As shown in Figure 4.2, the OWTs 7, 8, and 9 have some similarities in $R_{rs}(\text{shape})$, but their R_{rs} is what differs them mostly. In other words, the $R_{rs}(\text{shape})$ could provide a suitable clustering of OWTs at first, but the $R_{rs}(\text{signal intensity})$ should not be entirely

removed from the establishment of OWTs. Thus, the nine OWTs defined here were based on $R_{rs}(\text{shape})$ and R_{rs} .

Figure 4.2 - The mean (solid line) and SD (shaded area) of R_{rs} (black and left y-axis) and $R_{rs}(\text{shape})$ (blue and right y-axis) of all OWTs.



Source: The author.

To assure representative samples of each OWT regarding their $R_{rs}(\text{shape})(\lambda)$ and $R_{rs}(650)$, the similarity among samples was computed using the silhouette (ROUSSEEUW, 1987). The silhouette measures if a sample is more related to samples without their OWTs than to those within their OWTs. In practice, a sample silhouette lower than 0 means that this sample is more related to samples outside its OWT than within its OWT. First, the silhouette of all samples was computed using the basis coefficients in the Euclidean space. Here, the OWTs 7, 8, and 9 were considered as a single OWT since they have similar $R_{rs}(\text{shape})(\lambda)$.

Further, the silhouette was calculated among OWTs 7, 8, and 9 using the R_{rs} (650) as input. A total of 105 samples showed a silhouette lower than 0 and were removed, resulting in 378 samples representing all OWTs. This large number of samples occurred because some R_{rs} (shape) samples were transient among OWTs, which have resulted in negative silhouettes. Since the OWTs are based in k-means, and there is no fuzziness among them, the removal of those transient samples was essential for retrieving cohesive OWTs.

4.3.2 The OWTs

The OWT 1 is a very transparent water type with low Chl-a, TSM, and CDOM concentration (Figure 4.3; Table 4.1). The R_{rs} (shape) has a high slope increasing from 400 to 570 nm (Figure 4.2) that further decreases to 800 nm, where a weak absorption Chl-a feature at 675 nm is visible. The median R_{rs} is relatively low, reaching up to 0.01 sr^{-1} due to the low scattering coefficient (Figure 4.4). The Chl-a, TSM, and CDOM concentrations are low. This limnological condition results in very transparent waters with low attenuation caused by low absorption and scattering coefficients. The OWT 1 was found in ECP, Funil, Ibitinga, Nova Avanhandava, Paraibuna, Tapajós, and Três Marias (Table 4.2).

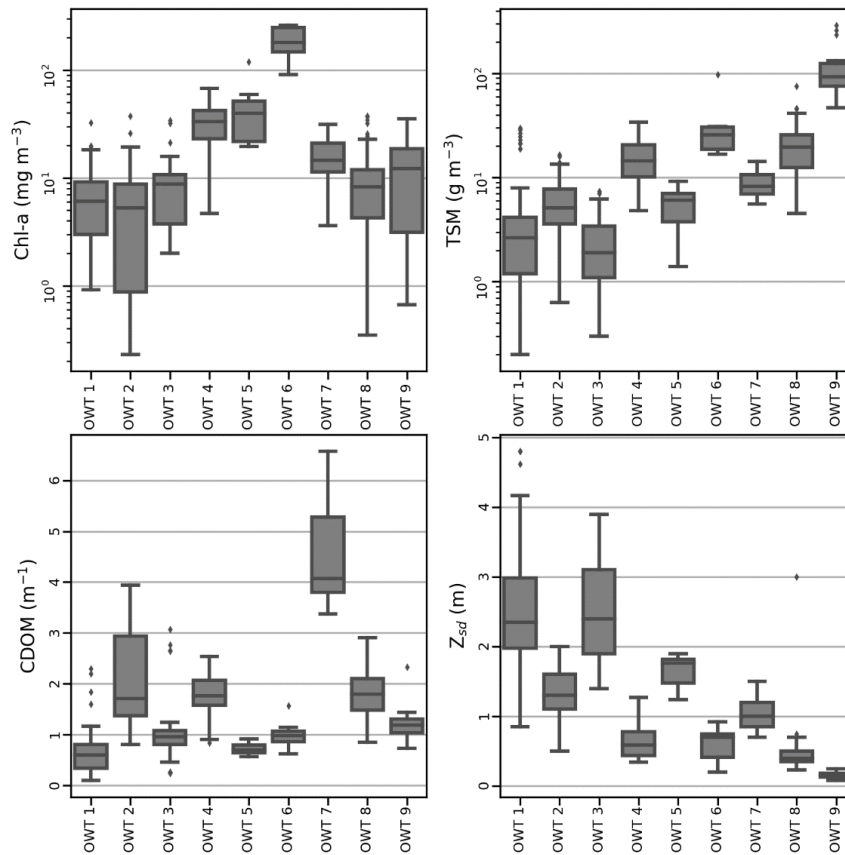
The OWT 2 is moderately transparent water with high CDOM concentration. The R_{rs} (shape) has a high slope increasing from 400 nm to 570 nm, caused by the high slope of the absorption coefficient found in those waters (Figure 4.4). Further, the R_{rs} (shape) slightly decreases towards 700 nm and significantly drops in the near infra-red (NIR) region. Two absorption features are evident at 610 nm and 675 nm. The mean of R_{rs} reaches up to 0.012 at 570 nm. The water transparency is moderate, where the scattering coefficient dominates attenuation at most wavelengths, while the absorption coefficient contributes more than the scattering coefficient to attenuation close to 400 nm. The Chl-a and TSM are low while the CDOM is high and shows high variability. The OWT 2 was found in Bua Bua, Curuai, Ibitinga, Itaipu, Mamirauá, Monte Alegre, Pantaleão, Pirarara, Tapajós, and Três Marias.

Table 4.1 - The median (med), interquartile range (IQR), and the number of samples (n) of Chl-a concentration, TSM concentration, CDOM concentration, and Z_{sd} , of all OWTs.

OWT	Chl-a (mg m^{-3})	TSM (g m^{-3})	CDOM (m^{-1})	Z_{sd} (m)
1	med: 6.04	2.64	0.6	2.35
	QRT: 6.18	2.94	0.46	1.0
	n: (77)	(76)	(55)	(52)
2	5.28	5.14	1.71	1.3
	7.93	4.1	1.56	0.5
	(75)	(75)	(59)	(75)
3	8.72	1.9	0.96	2.4
	6.89	2.3	0.28	1.20
	(26)	(25)	(26)	(25)
4	33.1	14.38	1.76	0.59
	19.1	10.5	0.48	0.34
	(28)	(28)	(24)	(29)
5	39.6	6.1	0.7	1.76
	29.7	3.25	0.16	0.34
	(14)	(14)	(14)	(12)
6	180.4	25.75	0.98	0.7
	101	11.8	0.21	0.33
	(7)	(7)	(7)	(6)
7	14.62	8.2	4.07	1.0
	9.6	3.67	1.4	0.35
	(15)	(15)	(14)	(15)
8	8.26	19.55	1.8	0.4
	7.67	13.0	0.62	0.15
	(102)	(102)	(88)	(102)
9	12.13	93.25	1.18	0.16
	15.6	49.1	0.27	0.06
	(14)	(14)	(14)	(14)

Source: The author.

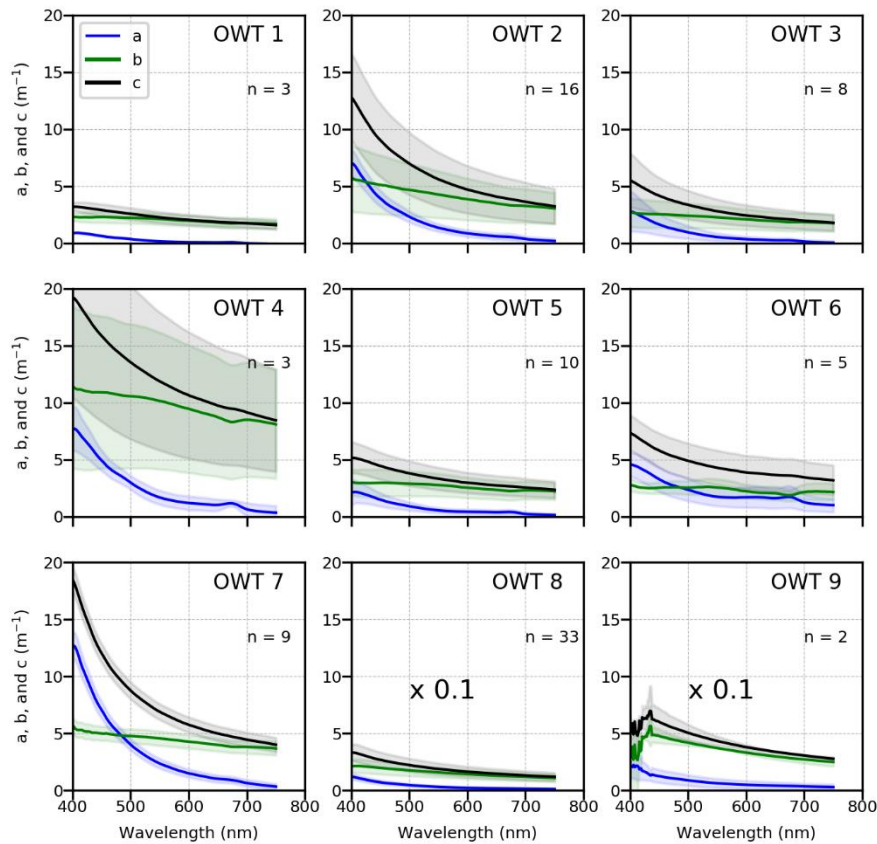
Figure 4.3 - Limnological characteristics of each OWT. The measurements are Chl-a concentration, TSM concentration, CDOM concentration, and Z_{sd} .



Source: The author.

OWT 3 is a very transparent water type with the influence of residual glint. The R_{rs} (shape) has a low slope increasing from 400 to 570 nm, and a low slope decreasing from 570 to 800 nm. The TSM and Chl-a concentrations are low, while CDOM showed a moderate concentration. The water transparency is similar to OWT 1, which both show the highest transparency among all OWTs, caused by the low TSM concentration, which favors low scattering and absorption coefficients. Thus, OWT 3 is artificial since it is a result of erroneous measurements instead of changes in the OACs. The OWT 3 was found in Funil, Ibitinga, Mamirauá, Nova Avanhandava, Tapajós, and Três Marias.

Figure 4.4 - The absorption (a), scattering (b), and attenuation (c) coefficients of all OWTs.



The values of OWT 8 and 9 are multiplied by 0.1 to preserve homogeneity in the y scale among subplots.

Source: The author.

In the OWT 3, the $R_{rs(\text{shape})}$ and R_{rs} are strongly influenced by residual glint, which is supported by three observations. First, the slope from 400 to 550 nm of $R_{rs(\text{shape})}$ of OWT 3 should be higher than OWT 1 since OWT 3 has a higher CDOM concentration and higher absorption coefficient in the same range. The lower slope of OWT 3 than OWT 1 may be caused by the residual glint in the blue range. Second, the mean R_{rs} (NIR) is very high, taking into account the TSM concentration similar to OWT 1. Also, the mean R_{rs} (NIR) is noisy, which was also checked in the individual samples. Even the higher scattering coefficient of OWT 3 than OWT 1 should not justify the high signal in the NIR range. Thus, the OWT 3 is more dependent on the glint removal method than a response of the Chl-a, TSM, and CDOM influence.

Table 4.2 - The number of representative samples of each OWT per region.

Region	OWTS								
	1	2	3	4	5	6	7	8	9
Amazon river	-	-	-	-	-	-	-	-	1
Aramanaí lake	-	-	-	-	-	-	-	6	-
Bua Bua lake	-	15	-	-	-	-	-	-	-
Curuai lake	-	5	-	21	-	2	-	54	11
ECP	11	-	-	-	-	-	-	-	-
Funil reservoir	9	-	1	-	2	1	-	-	-
Ibitinga reservoir	1	1	12	2	12	5	-	-	-
Itaipu reservoir	-	13	-	-	-	-	-	-	-
Japurá river	-	-	-	-	-	-	-	4	-
Mamirauá lake	-	13	5	5	-	-	15	17	-
Monte Alegre lake	-	3	-	1	-	-	-	12	-
Nova Avanhandava reservoir	27	-	4	-	-	-	-	-	-
Pacoval lake	-	-	-	-	-	-	-	2	-
Pantaleão lake	-	6	-	-	-	-	-	-	-
Paraibuna reservoir	5	-	-	-	-	-	-	-	-
Paru lake	-	-	-	-	-	-	-	2	-
Pirarara lake	-	9	-	-	-	-	-	6	-
Tapajós river	6	10	4	-	-	-	-	-	-
Três Marias reservoir	20	1	1	-	-	-	-	-	-

Source: The author.

The OWT 4 and 5 are high algae bloom that have comparable Chl-a concentrations but occurs in two specific circumstances: the OWT 4 occurs under higher TSM concentration from terrestrial input while OWT 5 occurs in lower TSM that mainly comes from the algae bloom. This difference is supported by three observations: (1) the OWT 4 has a higher TSM/Chl-a ratio than OWT 5; (2) the OWT 4 mainly occurs in Curuai lake where terrestrial input of TSM from Amazon river and sediment resuspension induced by wind is significant (BOURGOIN et al., 2007), and OWT 5 mainly occurs in Ibitinga reservoir where TSM that comes from its tributaries rapidly settles down due its low flow velocity (CAIRO et al., 2016); and (3) correlation between Chl-a and TSM for OWT 4 ($r = 0.48$) is lower than that of OWT 5 ($r = 0.89$), what indicates that the primary source of TSM in OWT 5 is related to phytoplankton.

Regarding the OWT 4, the $R_{rs (shape)}$ has a moderate slope increasing from 400 to 570 nm, a very low decrease from 570 to 700 nm where two absorption features at 620 and 675 nm are visible. Further, the $R_{rs (shape)}$ decreases in the NIR region. The mean R_{rs} is moderate and reaches up to 0.02, where the peaks occur in 570

and 700 nm. The Z_{sd} in those water is low caused by the high attenuation. The Chl-a and TSM concentrations are high with moderate variability, where the CDOM concentration is moderate. The OWT 4 was found in Curuai, Ibitinga, Mamirauá, and Monte Alegre.

Considering the OWT 5, the $R_{rs (shape)}$ has a high slope from 400 to 550 nm showing the Chl-a absorption feature at 440 nm. Further, the $R_{rs (shape)}$ decreases from 550 to 700 nm with a perceptible absorption feature at 620 and 675 nm and decreases in the NIR region. The mean R_{rs} reaches up to 0.012, and the two main peaks are at 550 and 700 nm. The Chl-a concentration is high, while CDOM and TSM concentrations are low. The OWT 5 was found in Funil and Ibitinga.

The OWT 6 is extreme algae bloom with a vegetation-like spectrum. The $R_{rs (shape)}$ significantly increases from 440 to 550 nm, where an absorption feature is visible at 430 - 440 nm. Further, the $R_{rs (shape)}$ decreases from 550 to 675 nm, where two great absorption features are visible at 630 nm and 675 nm. The $R_{rs (shape)}$ highly increase from 685 to 710 nm, characterizing the red-edge feature found in high Chl-a concentrations. The Chl-a concentration is the highest among all OWT caused by the intense algae bloom with a moderate TSM concentration and low CDOM. The water transparency is low caused by significant attenuation, where absorption and scattering contribute almost equally to attenuation in the Chl-a absorption peaks. The OWT 6 was found in Curuai, Funil, and Ibitinga.

The OWT 7 are waters rich in CDOM with the lowest R_{rs} of all OWTs, with the R_{rs} mean close to 0.005 nm in all wavelengths. The $R_{rs (shape)}$ has a low slope increasing from 400 to 570 nm. Furthermore, $R_{rs (shape)}$ changes to flat like shape from 570 to 700nm, and significant drops in the NIR region. A moderate absorption feature is noticeable at 675 nm. The water transparency is very low with is corroborated by the high attenuation, where absorption contributes more than scattering to attenuation in the blue range while scattering dominates in longer wavelengths. The CDOM concentration is the highest of all OWTs and shows moderate variation. The Chl-a median is moderate, and the TSM median is low. The OWT 7 was exclusively found in Mamirauá lake.

The OWT 8 is high scattering waters with high TSM concentration, being the second-highest TSM concentration that comes mainly from terrestrial input. The $R_{rs (shape)}$ has a moderate slope from 400 to 570 nm as a function of the absorption coefficient, which changes to flat shape from 570 to 700 nm, and significantly drops in the NIR region. Two moderate absorption features are visible at 600 nm and 675 nm. The median R_{rs} reaches up 0.025 from 570 to 700 nm as a result of the second-highest scattering coefficient of all OWTs. Furthermore, the high scattering also causes very low water transparency where the median Z_{sd} is 0.4 m. Regarding the limnological conditions, the TSM concentration is the second-highest among all of those OWTs that are not from an algae bloom. The OWT 8 was found in Curuai, Pacoval, Japurá, Mamirauá, Monte Alegre, Aramanaí, Paru, and Pirarara.

The OWT 9 is high scattering waters that show the lowest water transparency, the highest TSM concentration, and the highest R_{rs} magnitude. The $R_{rs (shape)}$ has a moderate slope increasing from 400 to 700 nm where a weak Chl-a absorption feature is visible at 675 nm, and further, the $R_{rs (shape)}$ decreases in the NIR region. This progressive increase to 700 nm is related to the higher absorption in the lower wavelengths while scattering changes lesser throughout the spectrum. The mean R_{rs} is the highest of all OWTs, reaching up to 0.055, which is evidenced by the high scattering coefficient. The Chl-a and CDOM concentrations are moderate, while TSM is the highest of all OWTs, with the median reaching up to 93.25 g m^{-3} . This high TSM concentration causes the lowest water transparency with a median Z_{sd} of 0.16 and the highest attenuation of all OWTs. The OWT 9 was found in Curuai and Amazon river.

4.3.3 The sources of OWTs

In the forest of Amazon floodplains, the OWT 2 and 7 originate from resuspension of organic carbon in podzol soils, which characterize them with high dissolved organic carbon concentrations (SIOLI, 1968) and consequently a high CDOM. The OWTs 8 and 9 come from rivers that originate in the Andean regions (e. g., Madeira and Solimões rivers), where the high erosion is conditioned by soil type,

terrain shape, and the amount of rain (SIOLI, 1951). In the rising period with the increase of water level, the waters of OWTs 8 and 9 flows into the floodplain lakes. During the high water level period in the Tapajós river, the TSM is more diluted in the water (LOBO et al., 2015), causing low TSM concentrations and the occurrence of OWT 1. In Curuai lake, the sedimentation of TSM that increases the light availability and nutrient input from cattle range is an example of algae bloom causes (BARBOSA, 2007) and the occurrence of OWTs 4 and 6.

An example of the OWT variation related to the Amazon basin is observed in the lakes located at the confluence of Japurá and Solimões rivers. In the rising water period, the resuspension of organic matter starts to increase, and OWT 2 takes place at Bua Bua and Mamirauá. Increasing the input organic matter, CDOM increases, and OWT 2 found in Mamirauá changes to OWT 7. Further, the OWT 8 coming from the Japurá river should also inundate Mamirauá lake. Thus, OWT 8 is also found in Mamirauá, while at Bua Bua, the OWT does not change. This example is limited by a minimal period of water sampling but illustrates how the OWT may change in de Amazon floodplains.

Ibitinga reservoir drainage basin is surrounded by sugar cane plantations that increase the availability of phosphorus and nitrogen, which further cause algae bloom (TUNDISI et al., 2008). The intensity of the bloom may vary from OTW 5 in low TSM or OWT 4 in high TSM concentration to OWT 6 when the bloom density significantly increases. High CDOM concentration may occur by the runoff of organic matter from reservoir banks near the Jacaré-Pepira river (CAIRO et al., 2016), and consequently, cause OWT 2 to occurs in Ibitinga reservoir. Since the occurrence of algae bloom varies throughout the year (Cairo et al., 2016), when algae bloom is not occurring, the OWT 1 should occur.

In Nova Avanhandava, the input of TSM that comes from rivers is low even in the rainy period, and Chl-a concentration may increase up to 38.59 mg m^{-3} in the dry period (RODRIGUES et al., 2017). The OWT 1 occurs in the dry period in Nova Avanhandava in areas with low Chl-a concentration.

In Funil reservoir, OWTs 5 and 6 also occur but probably caused by the accumulation of nutrients that comes from São Paulo state sewage (AUGUSTO-

SILVA et al., 2014). Further, OWT 1 may occur when algae bloom is not occurring.

Três Marias is located in the upstream São Francisco basin, where its main tributaries have its sources located in a National park at Canastra ridge, a very conserved and low degraded area. All those characteristics contribute to low input of TSM and CDOM and no occurrence of algae bloom, causing the predominance of OWT 1 in the Três Marias reservoir. In the Paraibuna reservoir, even in high precipitation periods, the input of TSM and CDOM is low, and no algae bloom occurs (CURTARELLI et al., 2019), causing the dominance of OWT 1. In the Itaipu reservoir, we did not find any study related to CDOM that would explain the OWT 2 occurrence.

The OWTs 1 samples show TSM lower than 10 g m^{-3} except for waters in the ECP, which comprises all outliers of TSM concentration of OWT 1. The medians of the TSM concentration of OWT 1 are 2.64 g m^{-3} . On the other hand, the TSM concentration of ECP in those OWTs varies from 18.8 to 29.9 g m^{-3} . Since all other outliers that occur in all limnological concentrations are not specific to a region or survey campaign, the TSM from ECP should have a specific meaning. Due June 2016 when the ECP data was collected (DA SILVA et al., 2019), there are evidences that a massive bloom of *Dinophysis acuminata* followed by a bloom of a *Noctiluca scintillans* occurred next to the ECP in the southern Brazilian coast (ALVES; MAFRA, 2018; NOERNBERG et al., 2017; SOBRINHO et al., 2018). Since *Noctiluca scintillans* are mainly composed of organic matter and have large cells that can reach up to 2 mm (SOBRINHO et al., 2018), it could reduce the refractive index and change particle size distribution of TSM in the ECP. Particles that have a low refractive index and a low change particle size distribution have lower scattering than that's of high refractive index and high change particle size (ROESLER; BOSS, 2008). Therefore, increase TSM concentration while the refractive index and change particle size distribution should have a minor change in the scattering and, consequently, the R_{rs} . Therefore, it should explain the reason for ECP with a higher TSM concentration shows the OWT 1 that is mainly characterized as low TSM concentration in other environments.

4.3.4 The Brazilian OWTs from a global perspective

Comparing the $R_{rs}(\text{shape})$ of OWTs computed here to global OWTs found in inland waters (SPYRAKOS et al., 2018), the Brazilian OWTs showed some similarities and particularities. The OWT 6 is comparable to their OWT 1 that represents hypereutrophic waters with the scum of cyanobacterial bloom and vegetation-like R_{rs} . The OWT 5 is also comparable to their OWT 8 that are productive waters with cyanobacteria abundance and with R_{rs} peak close to 700 nm. On the other hand, their OWT 10 that is related to CDOM rich waters differs from OWT 7 that is the CDOM rich waters in Brazil. The $R_{rs}(700)$ of OWT 10 can reach up to 0.05, while the $R_{rs}(\lambda)$ of OWT 7 is lower than 0.01. Furthermore, the OWTs 8 and 9 are analogous in R_{rs} shape to their OWT 5, which are sediment-laden waters, but their OWT 5 presented a wider variety of $R_{rs}(\lambda)$ than that's of OWTs 8 and 9. Those divergences occurred because the R_{rs} magnitude was utilized for separating the OWTs 7, 8, and 9, while the global OWTs were based exclusively on the R_{rs} shape.

The OWT 13 that represents clear blue waters (SPYRAKOS et al., 2018) was not registered in this study. The OWT 13 has the R_{rs} in the blue region higher than that of green and red with TSM and Chl-a concentration medians lower than 1 g m^{-3} and 1 mg m^{-3} , respectively. Such waters could occur in the coastal waters, but the ECP $R_{rs}(\text{shape})$ did not show any samples related to such OWT. Considering the other water bodies studied, the probability of OWT 13 occurrence is very low to the extent of our knowledge.

Comparing the OWTs identified in this study with those from different regions, $R_{rs}(\text{shape})$ may be similar, but the ranges of Chl-a, TSM, and CDOM concentrations can differ. For example, the $R_{rs}(\text{shape})$ of class 4 of turbid productive waters of inland waters in China (HUANG et al., 2014) is analogous to our OWT 4 but with higher R_{rs} and Chl-a concentration. Moreover, the $R_{rs}(\text{shape})$ of OWTs 2, 3, and 5 of Europe and China waters (ELEVELD et al., 2017) are similar to our OWTs 1, 5, and 4, respectively. However, their OWT 2 has higher TSM and CDOM concentrations than our OWT 1; their OWT 3 has higher TSM and CDOM and lower Chl-a concentrations than our OWT 5; their OWT 5 has higher TSM and

lower CDOM concentration than our OWT 4. In such cases, we hypothesize that the presence of other factors affecting water components, scattering and absorption coefficients and consequently, R_{rs} explain those divergences. Among them, parameters such as particle Junge distribution and refractive index (ROESLER; BOSS, 2008), CDOM molecular size (WÜNSCH et al., 2018), phytoplankton cell size and packaging effect (MOREL; BRICAUD, 1981) and species (JOHNSEN; SAKSHAUG, 2007). A better understanding would require absorption coefficients of CDOM, non-algae particles, and algae particles. Unfortunately, we do not have representative data for such comparisons.

4.4 Conclusion

This study characterizes 8 OWTs found in Brazilian waters using a rich dataset that represents different environmental conditions. The OWTs differ in R_{rs} shape and intensity, resulting in different combinations of Chl-a, TSM, and CDOM concentrations. Summarizing the OWTs that are presented here: the OWT 1 is transparent waters with low TSM, Chl-a, and CDOM concentrations; the OWT 2 is transparent waters dominated for CDOM; OWT 4 is high algae bloom in moderate TSM concentration; OWT 5 is high algae bloom in low TSM concentration; OWT 6 represents intensive algae bloom; the OWT 7 represents waters with the highest CDOM concentration; OWT 8 is water with high TSM concentration; OWT 9 are high scattering waters with the highest TSM concentration;

From this point forward, the OWTs defined here will support remote sensing applications of detecting spatial patterns and abrupt changes in waters quality, such as detecting OWTs related to algae bloom or high sediment loadings. Moreover, the OWTs could also be used for boosting algorithms of TSM, Chl-a, and CDOM concentrations, where specific algorithms may be calibrated for each OWT.

5 SATELLITE DETECTION OF BRAZILIAN OPTICAL WATER TYPES³

5.1 Introduction

A previous study identified 8 OWTs in Brazilian waters based on hyperspectral R_{rs} intensity and shape (R_{rs} (shape)) with different characteristics of Chl-a, TSM, and CDOM concentrations. OWT 1 is transparent waters with low Chl-a, TSM, and CDOM concentration. OWT 2 is moderate transparent waters with high CDOM concentration and moderate TSM. OWT 4 is waters with algae bloom that occurs in turbid waters. OWT 5 is waters with algae bloom that occurs in transparent waters. OWT 6 is waters with the highest algae bloom density. OWT 7 is waters dominated by CDOM and shows the highest CDOM level. OWT 8 is turbid waters dominated by TSM. OWT 9 is very turbid waters with the highest TSM concentration. Notice that OWT 3 is an artifact caused by erroneous measurements, and for this reason, was not considered in this chapter. Those OWTs were characterized in different environments in Brazil. However, this characterization was limited to *in situ* data, which may not hold well representative spatial-temporal variations of each water body.

To overcome the spatial-temporal gap of *in situ* data, a couple of studies have proposed using satellite remote sensing for detecting. Vantrepotte et al. (2012) defined OWTs using *in situ* R_{rs} in coastal waters and then trained a method for detecting them in SeaWiFS images. Uudeberg et al. (2019) evaluated the capability of detecting OWTs of boreal lakes and coastal waters using Sentinel-2 MSI and Sentinel-3 OLCI. Shi et al. (2013) tuned Chl-a bio-optical algorithms for specific OWTs, and further developed a decision tree method for detecting those OWTs in MERIS images.

Three challenges arise when detecting hyperspectral OWTs based on *in situ* data in remote sensing images. First, most available orbital sensors are multispectral, and thus the spectral resolution is limited to few bands. Second, the uncertainty between *in situ* and satellite R_{rs} caused by atmospheric correction, signal to noise

³ This chapter will be submitted to a indexed scientific journal.

ratio (SNR), the difference of spatial resolution, and adjacent effects (JORGE et al., 2017; MARTINS et al., 2017). Those uncertainties can lead to misclassification and have a significant influence on accuracy. For example, a low SNR can misclassify OWTs more than a degraded spectral resolution (LOBO et al., 2012). Likewise, misclassification can also depend on atmospheric correction (UUDEBERG et al., 2019). Third, the spatial-temporal coverage of remote sensing images surpasses that of *in situ* data. Consequently, those images should cover OWTs not used for training the classification algorithms. Traditional classification algorithms are not capable of novelty detection, and they will force any unknown OWT to a known OWT. Hence, any classification algorithm for OWTs based on *in situ* hyperspectral data should address all those issues.

The objective of this study is to develop a classification algorithm for detecting the Brazilian OWTs by remote sensing images. For this purpose, Sentinel-2 MSI, Landsat-8 OLI, and Landsat-7 ETM+ were tested. First, the classification algorithm of Support Vector Machines Classifier (SVC) and Random Forest (RF) were trained and validated using *in situ* simulated bands. Second, the classification algorithms were evaluated in satellite images. Last, the performance of each classification algorithm in each sensor was compared in three study areas.

5.2 Material and methods

5.2.1 *In situ* radiometric data

The above-water radiometric measurements were taken using a Hand-Held 2 VNIR and three intercalibrated TriOS-RAMSES. The Mobley (1999) method was used for computing R_{rs} . The total water-leaving radiance ($L_t(\lambda, \theta, \phi)$), the downwelling sky radiance ($L_s(\lambda, \theta, \phi)$), and the total irradiance incident onto the water surface ($E_s(\lambda)$) were measured. The wavelength (λ) range assessed here were from 400 to 800 nm, where the spectral resolution of 3 nm of both devices was resampled to 1 nm. Considering the sun as a reference, the nadir (θ) and azimuth (ϕ) angles of L_t were 40° and 135° , respectively, while the nadir and

azimuth of L_s were 130° and 135° , respectively. Then, the R_{rs} was computed using the equation:

$$R_{rs}(\lambda) = \frac{L_t(\lambda) - \rho L_s(\lambda)}{E_s(\lambda)} \quad (\text{sr}^{-1}) \quad (5.1)$$

Where ρ is a factor that relates how much of L_s is reflected in the water-air interface towards the sensor direction, and it is a function of wind speed, solar zenith angle, and sensor zenith and azimuth angle. The $\rho = 0.028$ was applied since it is suitable for the acquisition geometry used here and at wind speeds lower than 5 m s^{-1} in blue sky conditions or at any wind speed in overcast skies (MOBLEY, 1999).

5.2.2 Satellite images

The Sentinel-2 MSI, Landsat-8 OLI, and Landsat-7 ETM+ data were utilized to classify the OWTs. Those sensors were chosen because they have suitable spatial resolution and a considerable temporal coverage for monitoring inland waters (Table 5.1). Even though those sensors were not designed for water applications, they have demonstrated reliable applications in aquatic environments (CAIRO et al., 2019a; DA SILVA et al., 2019; LOBO et al., 2015; MACIEL et al., 2019; MILLER et al., 2011; MISHRA; GARG, 2011; RODRIGUES et al., 2017). We are aware that sensors such as Sentinel-3 OLCI have bands tuned for water environments. However, their 300 m spatial resolution would only be useful for a small number of large water bodies, limiting the operational use.

The Sentinel-2 MSI and Landsat-8 OLI images were accessed at level 1C reflectance at the top of the atmosphere, and Landsat-7 ETM+ images were accessed at level 2 surface reflectance. All images were downloaded from the United States Geological Survey earth explorer website (<<https://earthexplorer.usgs.gov/>>). The Landsat-7 ETM+ images were atmospherically corrected using the Landsat Ecosystem Disturbance Adaptive Processing System (MASEK et al., 2006), which is based on the 6S (Second Simulation of Satellite Signal in the Solar Spectrum) radiative transfer code (VERMOTE et al., 1997). For the Landsat-8 OLI and Sentinel-2 MSI images, the

atmospheric correction was performed using the software AtmosPy (CARLOS et al., 2019), which is based on the 6S vector version (6SV) and takes into account the radiation polarization (KOTCHENOVA et al., 2006; KOTEHENOVA; VERMOTE, 2007). The AtmosPy was not used in the Landsat-7 ETM+ images because the current version only works with Landsat-8 OLI, Sentinel-2 MSI, and Sentinel-3 OLCI.

Table 5.1 - Technical specifications summary of Sentinel-2 MSI, Landsat-8 OLI, and Landsat-7 ETM+.

Sensor	Band	Bandwidth (nm)	Spatial resolution (m)	Radiometric resolution (bits)	Revisit time (days)	Temporal coverage (years)
Sentinel-2 MSI	B2	458 – 523	10	12	2 - 5	2015 - ongoing
	B3	543 – 578	10			
	B4	650 – 680	10			
	B5	698 – 713	20			
	B6	733 – 748	20			
Landsat-8 OLI	B2	452 – 512	30	12	16	2013 - ongoing
	B3	533 – 590				
	B4	636 – 673				
	B5	851 – 879				
Landsat-7 ETM+	B1	450 – 520	30	8	16	1999 – 2020
	B2	520 – 600				
	B3	630 – 690				
	B4	770 – 900				

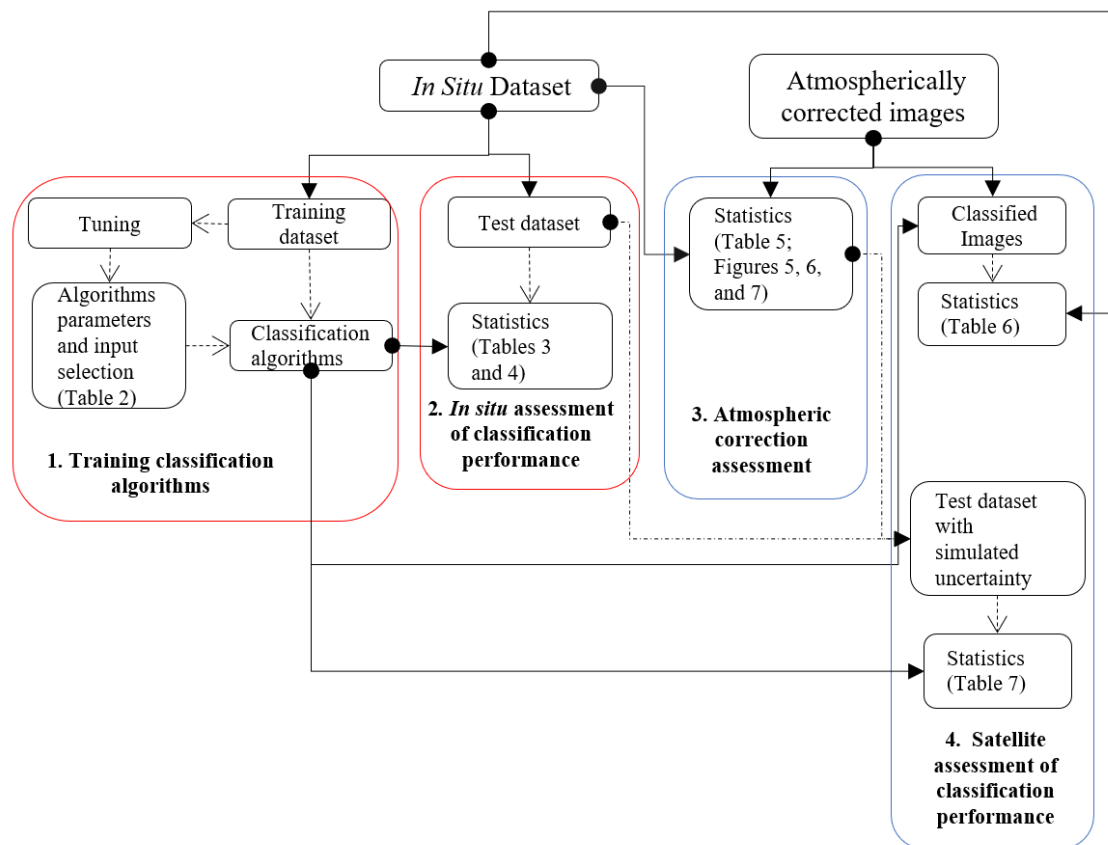
Source: The author.

Furthermore, surface reflectance of all images was converted to R_{rs} by dividing each pixel value by π . For glint removal, short wave infrared (SWIR) R_{rs} was subtracted from visible and NIR bands of each sensor. This simple correction has shown satisfactory performance in a wide range of environments (CAIRO et al., 2020; LOBO et al., 2015; MACIEL et al., 2019). For Sentinel-2 MSI, Landsat-8 OLI, and Landsat-7 ETM+, the SWIR bands used for glint removal were B12 (2130 nm), B7 (2200 nm), and B7 (2220 nm), respectively. Each image utilized was visually inspected for residual glint or artifacts. Last, the Sentinel-2 MSI bands with 10 m spatial resolution were resampled to 20 m using the mean values of a 2x2 window.

5.2.3 Classification algorithms

The dataset was split into a training (70%) and a test (30%) datasets a thousand times weighted by class, where the dataset samples of each split were not repeated (Figure 5.1). The training dataset was used for training the classification algorithms, while the test dataset was used for assessing the performance. At each interaction, the balanced accuracy was computed, and the classification algorithm of each sensor that showed the median balanced accuracy was chosen for further applying in the remote sensing images.

Figure 5.1 - Study design flowchart.



Source: The author.

OWT classification for each sensor was carried out using Support Vector Machine (SVM) Classifier (SVC) and Random Forest (RF) algorithms. The SVC is a binary classification algorithm that creates a margin hyperplane in an n-dimensional

space (number of bands) that separates two classes (OWTs) and maximizes this margin to the nearest support vectors (samples) of both OWTs. The hyperplane is defined by a kernel function, which can be linear, polynomial, radial basis function (RBF), and sigmoid. For multiclass classifications, such as the required here, the one-vs-one approach can be used. In this approach, the algorithm defines hyperplanes for all possible combinations of OWTs; then, the OWT receiving the majority of votes in all combinations is classified (TAN et al., 2008).

RF algorithm generates several decision trees by using bootstrap subsamples of the training dataset. Each decision tree is a series of nodes (Boolean rules using the input bands) that best split the OWTs. Then, the OWT with the majority votes of all decision trees is classified. (TAN et al., 2008).

There are infinite possible inputs (bands used) and adjusted parameters for SVC and RF algorithms. For example, the SVC algorithm requires penalty and kernel, while the RF algorithm requires the number of decision trees. The inputs and parameters of each algorithm were tuned using exploratory data analysis, where n-possible combinations were tested and evaluated using accuracy estimates. The chosen parameters and bands for each sensor (Table 5.2) were those with the best accuracy.

SVC and RF algorithms have not the capacity of novelty detection and will classify any target (sample or pixel in the image) as an OWT. Consequently, both algorithms could classify unknown OWTs as a known OWT and wrongly estimate OWTs in further applications. This issue requires pos processing using novelty detection, which statistically measures the probability of a sample belonging to the given OWT. For the SVC algorithm, the Platt (1999) method was used for probability estimates. In this method, the probability is estimated using a sigmoid function:

$$P(y = 1|f) = \frac{1}{1 + \exp(Af + B)} \quad (5.2)$$

Where $P(y = 1|f)$ is the probability, f is SVM output that corresponds to an uncalibrated value (which varies from -1 to 1 between the support vectors), A and B are the parameters fitted using the maximum likelihood of a training set. In multi-class classification, the probability estimates are pairwise among OWTs. Then, the pairwise probability was further converted to unique values using the Wu et al. (2004) method.

Table 5.2 - Parameters specifications of SVC and RF training algorithms and inputs for classification.

Training Algorithm	Algorithm Specifications	Sensor	The input of shape classification (OWTs 1, 2, 4, 5, 6, 789)	The input of intensity classification (OWTs 7, 8, and 9)
SVC	Penalty: 120, Kernel: RBF, multiclass approach: one-vs-one, class weight: balanced, probability estimates: (PLATT, 1999)	MSI	Standardized B2, B3, B4, B5, and B6 bands	B3
		OLI	Standardized B2, B3, B4, and B5 bands	B3
		ETM+	Standardized B1, B2, B3, and B4 bands	B2
RF	Number of decision trees: 100, class weight: balanced, probability estimates: number of decision trees that classified the same class	MSI	Standardized B2, B3, B4, B5, and B6 bands	B3
		OLI	Standardized B2, B3, B4, and B5 bands	B3
		ETM+	Standardized B1, B2, B3, and B4 bands	B2

Source: The Author.

Regarding novelty detection using the RF algorithm, the rate of decision trees that estimated an OWT over the total of decision trees was used for computing probability. In other words, a rate of 70 decision trees of 100 trees represents a probability of 70%. The SVC and RF training algorithms and novelty detection utilized are implemented in Python programming language on the Scikit-learn v0.21 (PEDREGOSA et al., 2011).

Probability thresholds must be set for each classification algorithm and OWT, in such a way that targets having probability higher than the threshold are classified, while targets having probability lower than the threshold are considered as a

novel and are not classified. For the SVC algorithm, the threshold was set where accuracy was 1 per OWT with a maximum of samples classified in the training dataset. Concerning the RF algorithm, the threshold was set to the minimum value of each OWT in the training dataset. More details are in appendix A.

Simulated bands of each sensor were used for training the classification algorithms, which were computed by the equation:

$$R_{rs}(Bn_{\text{sensor}}) = \frac{\int_{\Delta B_n} R_{rs}(\lambda) RF_{B_n}(\lambda) d\lambda}{\Delta B_n} \quad (5.3)$$

Where $R_{rs}(Bn_{\text{sensor}})$ is the R_{rs} of band Bn , ΔB_n is the bandwidth, and $RF_{B_n}(\lambda)$ is the response function for each sensor band. The response function for each band was accessed at the Goddard Space Flight Center (<https://landsat.gsfc.nasa.gov>) and the European Space Agency (<https://earth.esa.int>) websites.

Since most OWTs are based on the spectral shape, the brightness influence was removed, and the $R_{rs}(Bn_{\text{sensor}})$ were standardized as follows:

$$R_{rs}(Bn_{\text{sensor}}) \text{ (standardized)} = \frac{R_{rs}(Bn_{\text{sensor}})}{\sum bn} \quad (5.4)$$

Where $R_{rs}(Bn_{\text{sensor}})$ (standardized) is the R_{rs} of bands Bn standardized by the sum of all bands bn . The bands utilized in bn of Sentinel-2 MSI were $R_{rs}(B2_{\text{msi}})$, $R_{rs}(B3_{\text{msi}})$, $R_{rs}(B4_{\text{msi}})$, $R_{rs}(B5_{\text{msi}})$, and $R_{rs}(B6_{\text{msi}})$. The bands utilized in bn of Landsat-8 OLI were $R_{rs}(B2_{\text{oli}})$, $R_{rs}(B3_{\text{oli}})$, $R_{rs}(B4_{\text{oli}})$, and $R_{rs}(B5_{\text{oli}})$. The bands utilized in bn of Landsat-7 ETM+ were $R_{rs}(B1_{\text{etm+}})$, $R_{rs}(B2_{\text{etm+}})$, $R_{rs}(B3_{\text{etm+}})$, and $R_{rs}(B4_{\text{etm+}})$.

The classification algorithms were trained in two steps. First, a classification algorithm was trained to classify the OWTs by their shape rather than their intensity. In this step, the $R_{rs}(Bn_{\text{sensor}})$ (standardized) was used as input, and

OWTs 7, 8, and 9 were included in a unique class since their shapes are similar. Second, an additional classification algorithm using the same algorithm specifications was trained to classify OWTs 7, 8, and 9. In this step, the $R_{rs}(B_{n_{\text{sensor}}})$ was used for training the classification algorithm. The probability estimates for class OWT 7, 8, and 9 were computed as the mean probability of the two classification steps.

5.2.4 Accuracy assessment

The atmospheric correction was assessed by comparing the *in situ* $R_{rs}(B_{n_{\text{sensor}}})$ to the satellite $R_{rs}(B_{n_{\text{sensor}}})$ computed as the median of a 3x3 window, for image acquisition concurrent with *in situ* measurements. The coefficient of determination (R^2), the mean absolute percentage error (MAPE), and the mean and standard deviation (SD) bias were utilized for evaluating the atmospheric correction performance according to equations:

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{total}}} \quad (5.5)$$

$$\text{MAPE (\%)} = \frac{100}{n} \sum_{i=1}^n \left| \frac{R_{rs}(B_{n_{\text{sensor}}})_{(\text{sat})} - R_{rs}(B_{n_{\text{sensor}}})_{(\text{in situ})}}{R_{rs}(B_{n_{\text{sensor}}})_{(\text{in situ})}} \right| \quad (5.6)$$

$$\text{bias (\%)} = \frac{R_{rs}(B_{n_{\text{sensor}}})_{(\text{sat})} - R_{rs}(B_{n_{\text{sensor}}})_{(\text{in situ})}}{R_{rs}(B_{n_{\text{sensor}}})_{(\text{in situ})}} \times 100 \quad (5.7)$$

Where SS_{res} is the sum square of residuals, SS_{total} is the total sum of squares, n is the number of samples, $R_{rs}(B_{n_{\text{sensor}}})_{(\text{sat})}$ is the R_{rs} of band B_n in the satellite image, and $R_{rs}(B_{n_{\text{sensor}}})_{(\text{in situ})}$ is the R_{rs} of band B_n simulated *in situ*.

The performance of classification algorithms was assessed with *in situ* and satellite data. Three metrics were used: balanced accuracy, Precision, Recall, and detection rate. Those metrics are defined as:

$$\text{Precision}_j = \frac{TP_j}{TP_j + FP_j} \quad (5.8)$$

$$\text{Recall}_j = \frac{TP_j}{TP_j + FN_j} \quad (5.9)$$

$$\text{Balanced Accuracy} = \frac{1}{n_{\text{OWTs}}} \sum_{j=\text{OWT } 1}^{n_{\text{OWTs}}} \text{Recall}_j \quad (5.10)$$

$$\text{detection rate}_j = \frac{\text{classified Targets}_j}{\text{classified Targets}_j + \text{unclassified Targets}_j} \quad (5.11)$$

Where j denotes an OWT, TP is the true positives, FP is the false positives, FN is the false negatives, and n_{OWTs} is the total number of OWTs, classified targets are number of reference samples that were correctly or wrongly classified, and unclassified targets are the reference samples that were not detected due to low probability. Balanced accuracy was used instead of accuracy for the assessment to avoid bias since there are different sample sizes among OWTs. Thus, accuracy would be biased to the OWT that had more samples.

For evaluating the performance of the algorithms on the satellite image, *in situ* OWTs were compared to classified images with acquired within one day of *in situ* acquisition. A 3x3 window of OWTs was extracted at each match-up, and the OWT mode was computed. However, the main issue of evaluating the classification in satellite using *in situ* match-ups is that not all OWTs were detected, and consequently, the classification assessment for all OWTs is limited.

So as to guarantee a better understanding of the classification performance, it is necessary to fill this gap. Thus, the uncertainty satellite was simulated in the *in situ* test dataset, based on the following assumptions made: (1) the mean and SD $R_{rs}(B_{n\text{sensor}})$ match-up bias have a Gaussian distribution; and (2) the computed uncertainties are equivalent in all OWTs. Then, a random uncertainty for each sample and $R_{rs}(B_{n\text{sensor}})$ was generated using a Gaussian random function, which was further added to the test dataset. This process was repeated a thousand times, and the mean was computed.

5.3 Results and discussion

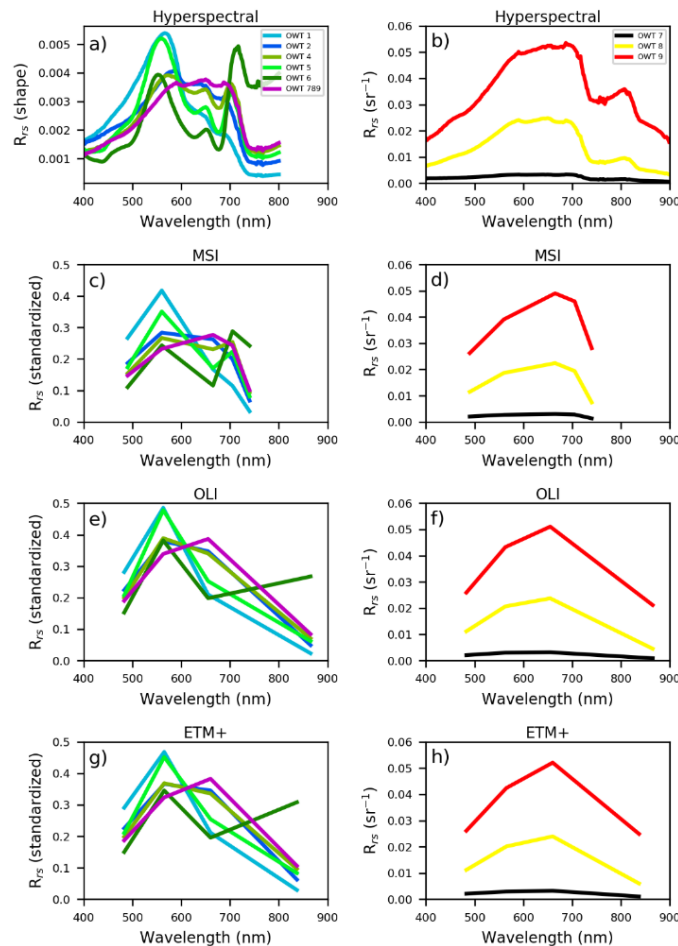
5.3.1 Classification using *in situ* data

The degradation of spectral resolution from hyperspectral to multispectral preserved significant differences of R_{rs} (standardized) and R_{rs} among most of OWTs (Figure 5.2). All simulated sensors could well discern OWTs 7, 8, and 9 using R_{rs} , with the highest differences occurring in the range from 500 to 700 nm. Considering the remaining OWTs, MSI bands exhibited distinctive R_{rs} (standardized) for each OWT, while Landsat sensors (OLI and ETM+) bands exhibited a moderate overlap between OWT 2 and 4. Furthermore, it is also worth noticing that MSI bands cover a narrow wavelength range, which was used for defining the OWTs (400 – 800 nm), whereas OLI and ETM+ bands cover a wider wavelength range and have their NIR bands above 800 nm. The effect of those differences among MSI bands, OLI bands, and ETM+ bands in the classification are presented and discussed then.

MSI, OLI, and ETM+ simulated sensors showed satisfactory performance for classifying the OWTs (Table 5.3), with accuracy ranging from 0.9 to 1 and detection rate varying from 81% to 97%. MSI was more accurate and showed a higher detection rate than ETM+ and OLI, except for the detection rate of OLI RF. One reason why MSI was more accurate than Landsat sensors is that MSI bands are located from 400 to 800 nm, which is the same OWTs range used for their definition, while OLI and ETM+ have the NIR band exceeding 800 nm. A second reason is that the MSI NIR bands are located at the red edges, key spectral

regions for characterizing algae bloom dominated water types, as OWTs 4, 5, and 6. Thus, the accuracy of each algorithm should have different performance depending on the OWT.

Figure 5.2 - Spectral R_{rs} of Brazilian OWTs.



The left column (a, c, e, and g) are the mean R_{rs} (shape) for hyperspectral data and mean R_{rs} (standardized) for multiband data for the OWTs that differ in shape, while the right column (b, d, f, and h) is the mean R_{rs} (sr^{-1}) for OWTs that differs in R_{rs} . The a) and b) are hyperspectral, c) and d) are MSI simulated bands, e) and f) are OLI simulated bands, and g) and h) are ETM+ simulated bands.

Source: The author.

Considering the classification performances at each OWT, the minimum values of Precision and Recall (0.67) occurred in the Landsat ETM+ and OLI classifications (Table 5.4). MSI showed the best results, where RF correctly

classified all samples, and the SVC algorithm only misclassified one OWT 7 sample as OWT 4. OLI classification misclassified samples among OWT 2 and 4 due to its overlap in OLI bands, causing their low Precision and Recall. Likewise, only ETM+ RF confused OWTs 2 and 4 due to its overlap bands. Regarding the remaining OWTs misclassifications, there was no pattern found as between OWT 2 and 4. Notice that those results were not computed using unclassified samples, which were evaluated using the detection rate presented next.

Table 5.3 - The accuracy and detection rate of all classification algorithms.

		Accuracy	Detection rate (%)
MSI	SVC	0.97	93
	RF	1	91
OLI	SVC	0.9	80
	RF	0.95	97
ETM+	SVC	0.95	81
	RF	0.98	88

Source: The author.

The lowest detection rates occurred to water types related to algae blooms (OWT 4, 5, and 6), what was similar for all sensors (MSI, OLI, and ETM+). The detection rates of OWT 4, 5, and 6 ranged from 33% to 89%, 25% to 100%, and 50% to 100%, respectively. Those OWTs have the largest R_{rs} (standardized) variation combined with a low number of samples. Consequently, the chance of a sample in the test dataset shows a lower probability than that set as the threshold for novelty detection should be higher. For this reason, the detection rate for those OWTs needs to be further evaluated using a larger dataset. More discussion about the detection rate in those OWTs will also be shown in the image analysis of OWTs (section 5.3.2.3), where the influence of those low detection rates can be elucidated.

Thus far, the spectral capability of MSI, OLI, and ETM+ in detecting OWTs based on hyperspectral was assessed. All sensors showed satisfactory performance, but MSI outperformed OLI and ETM+. Nevertheless, those accuracy estimates are limited to the *in situ* data where there are no uncertainties from atmospheric correction, SNR, the difference of spatial resolution, and adjacent effects. As

stated by Lobo et al., (2012) and Udeberg et al., (2019), SNR and atmospheric correction plays a vital role in classification and may have more impact than spectral resolution. The next section assesses the issues regarding classification in satellite images.

Table 5.4 - Performance of the classifications algorithms for each OWT using the *in situ* simulated sensors.

Sensor	Classification Model	OWTs								
		1	2	4	5	6	7	8	9	
MSI	SVC	Precision	1	1	0.88	1	1	1	1	1
		Recall	1	1	1	1	1	0.75	1	1
		Detection rate (%)	100	96	78	50	50	80	100	100
	RF	Precision	1	1	1	1	1	1	1	1
		Recall	1	1	1	1	1	1	1	1
		Detection rate (%)	95	96	56	75	100	80	97	100
OLI	SVC	Precision	1	0.88	1	1	1	1	0.97	1
		Recall	1	1	0.67	1	1	0.8	1	0.75
		Detection rate (%)	84	65	33	100	100	100	94	100
	RF	Precision	1	0.88	1	1	1	1	1	1
		Recall	1	1	0.63	1	1	1	1	1
		Detection rate (%)	100	96	89	100	50	100	100	100
ETM+	SVC	Precision	1	1	1	0.67	1	0.83	0.96	1
		Recall	0.8	1	1	1	1	1	0.96	0.75
		Detection rate (%)	90	65	44	100	100	100	74	100
	RF	Precision	1	0.96	1	1	1	1	1	1
		Recall	1	1	0.8	1	1	1	1	1
		Detection rate (%)	100	96	56	25	100	80	90	100
n		19	23	9	4	2	5	31	4	

Source: The author.

5.3.2 Classification using satellite images

5.3.2.1 Atmospheric correction assessment

Sentinel-2 MSI showed good agreement between *in situ* and satellite $R_{rs}(B3_{msi})$, $R_{rs}(B4_{msi})$, and $R_{rs}(B5_{msi})$, with a reliable R^2 , MAPE, and bias (Table 5.5, Figure 5.3). On the other hand, *in situ* and satellite $R_{rs}(B2_{msi})$ and $R_{rs}(B6_{msi})$ exhibited the most unsatisfactory results. The $R_{rs}(B2_{msi})_{(sat)}$ moderately overestimated $R_{rs}(B2_{msi})_{(in\ situ)}$, with a MAPE and mean bias of 27%, while $R_{rs}(B6_{msi})_{(sat)}$ also overestimated $R_{rs}(B6_{msi})_{(in\ situ)}$ with MAPE and mean bias of 37%.

Table 5.5 - Statistical results of the match-up between each sensor image and *in situ* data. The match-ups correspond to R_{rs} samples measured concurrent to satellite overpass.

Sensor	Band	R^2	MAPE (%)	Mean bias (%)	Std bias (%)	N
MSI	B2	0.93	27	27	15	8
	B3	0.94	12	9	13	
	B4	0.99	8	2	10	
	B5	0.99	14	9	17	
	B6	0.98	37	37	31	
	OLI	Blue	0.61	22	-3	
Green		0.74	14	0	18	
Red		0.85	13	-4	17	
NIR		0.3	74	65	88	
ETM+	Blue	0.81	18	-4	23	12
	Green	0.88	14	-9	17	
	Red	0.95	20	-7	27	
	NIR	0.75	221	215	255	

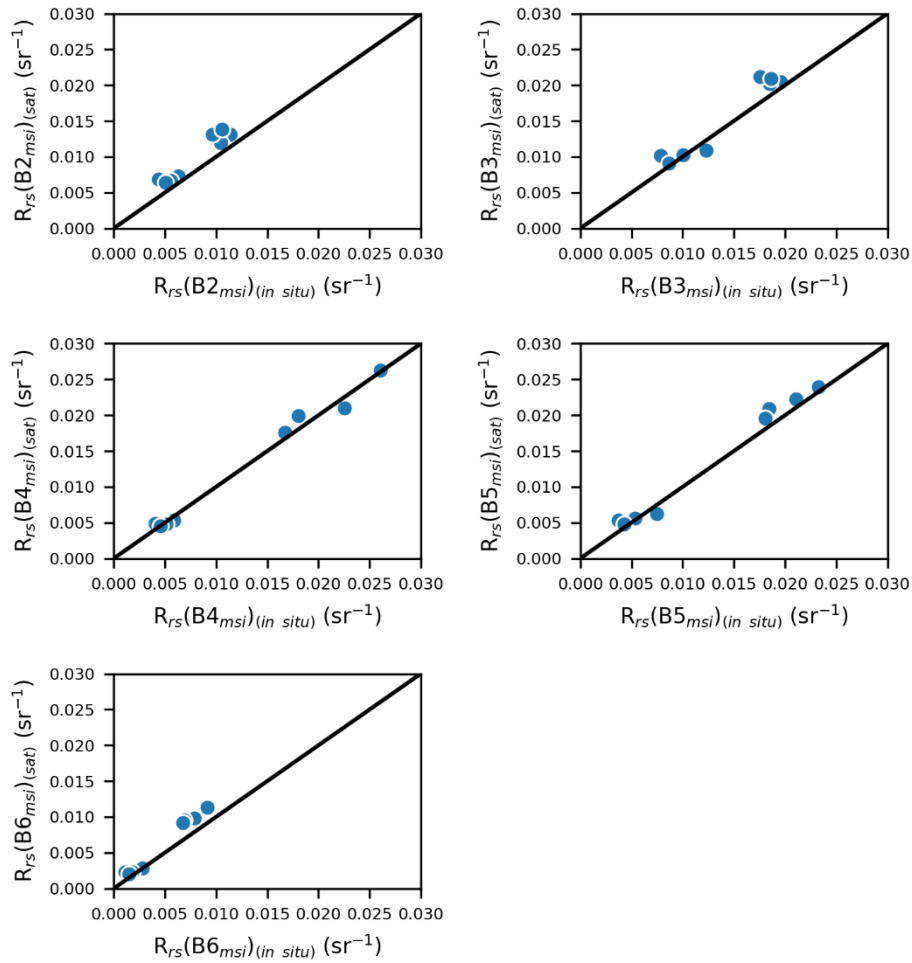
Source: The author.

Regarding Landsat-8 OLI match-ups, the satellite $R_{rs}(B2_{oli})$, $R_{rs}(B3_{oli})$, and $R_{rs}(B4_{oli})$ retrieved R_{rs} with satisfactory performance (Figure 5.4), where R^2 varied from 0.61 to 0.85, MAPE ranged from 13% to 22%, with a mean bias nearby 0. On the other hand, the $R_{rs}(B5_{oli})_{(sat)}$ overestimated $R_{rs}(B5_{oli})_{(in\ situ)}$ with a mean bias of 65%, an SD bias of 88%, and a MAPE of 74%.

The match-ups of Landsat-7 ETM+ exhibited consistent agreement between satellite and *in situ* $R_{rs}(B1_{etm+})$, $R_{rs}(B2_{etm+})$, and $R_{rs}(B3_{etm+})$ (Figure 5.5), with an

R^2 ranging from 0.81 to 0.95, a MAPE ranging from 18% to 20%, and a mean bias nearby 0. In contrast, the $R_{rs}(B4_{etm+})_{(sat)}$ strongly overestimated $R_{rs}(B4_{etm+})_{(in situ)}$, showing a mean bias of 215%, an SD bias of 255%, and a MAPE of 221%.

Figure 5.3 - Satellite and *in situ* match-ups of Sentinel-2 MSI bands, corresponding to data measured on the same day.

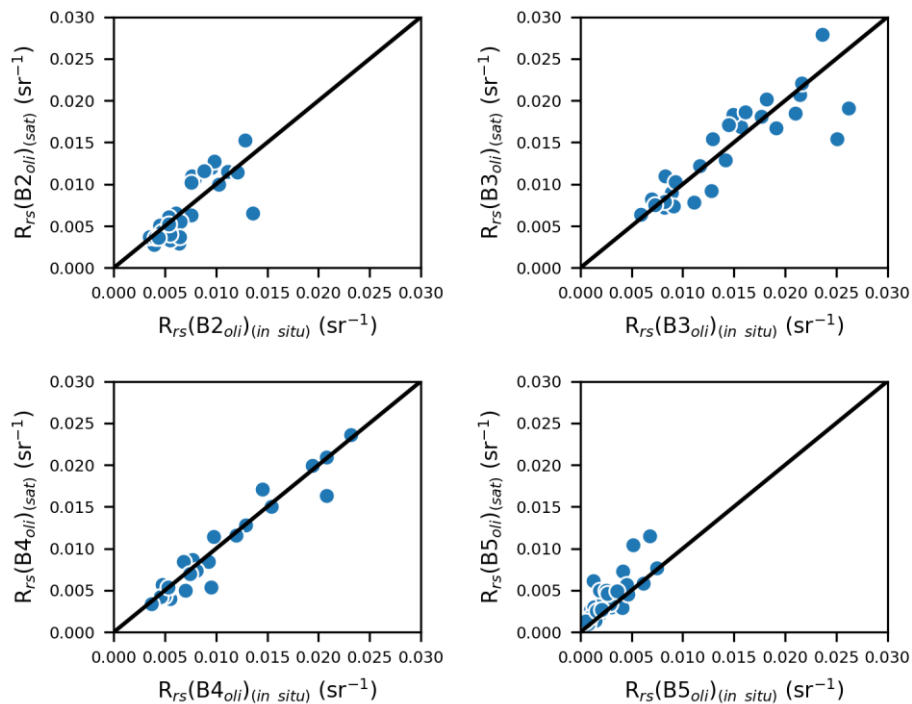


Source: The author.

The use of 6S and 6SV transfer radiative code for atmospheric correction provided satisfactory R_{rs} retrieval in the visible bands of all sensors. In the NIR range, the $R_{rs}(B_{nsensor})_{(sat)}$ of all sensors tended to overestimate $R_{rs}(B_{nsensor})_{(in situ)}$. The higher error in the NIR bands was also found in different studies using 6SV in aquatic environments (CAIRO et al., 2020; MACIEL et al., 2019; MARTINS et

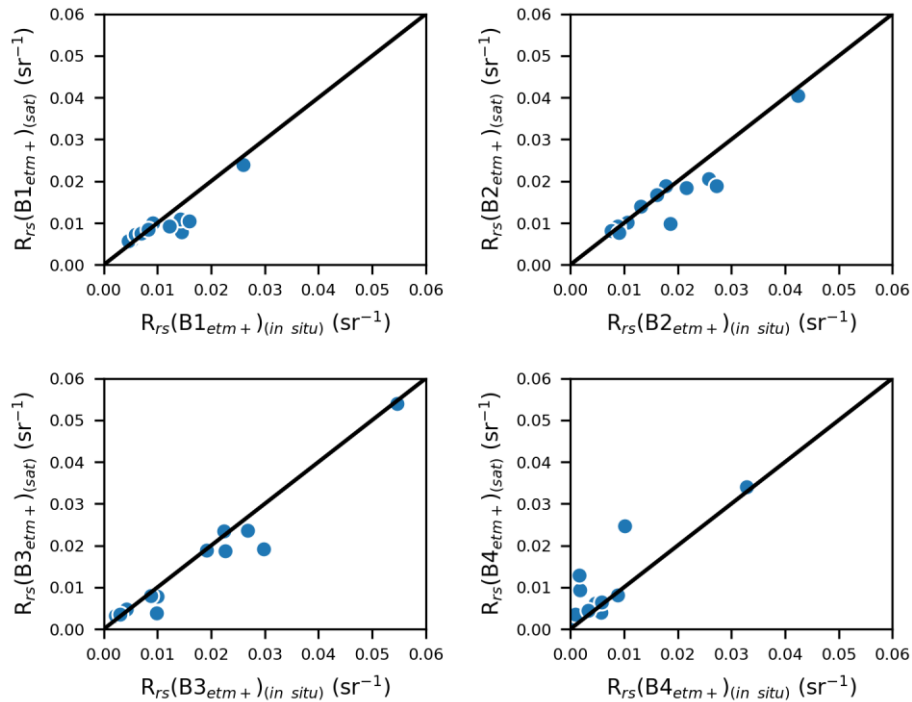
al., 2017). A possible explanation for this overestimation might be adjacency effects, which have a significant influence in inland waters (WARREN et al., 2019) and is pronounced in the NIR (VANHELLEMONT, 2019). In such cases, Martins et al. (2017) substantially improved the Sentinel-2 MSI NIR estimations using adjacency corrections in small lakes of Amazon floodplain. Furthermore, the possible influence of low signal and SNR of NIR bands, and radiometric resolution (regarding ETM+), should also be considered. Since high errors in the NIR bands have also been reported in many different atmospheric corrections such as Acolite, Sen2Cor, C2RCC, Icor, I2gen, and Polymer (MARTINS et al., 2017; WARREN et al., 2019), approaches to enhance the estimations of R_{rs} in the NIR range might involve adjacency corrections. Up until such corrections make available, the next section evaluates how those uncertainties assessed from satellite estimation of R_{rs} influence the classification.

Figure 5.4. Satellite and *in situ* match-ups of Landsat-8 OLI bands corresponding to data measured on the same day.



Source: The author.

Figure 5.5 - Satellite and *in situ* match-ups of Landsat-7 ETM+ bands, corresponding to data measured on the same day.



Source: The author.

5.3.2.2 *In situ* and satellite classification match-ups

The 24 match-ups of Sentinel-2 MSI were composed of OWTs 1, 4, 5, 8, and 9, where each algorithm classified 12 match-ups, and the remaining match-ups were considered as a novel (Table 5.6). MSI SVC and RF algorithms presented a satisfactory agreement with *in situ* OWTs and correctly classified 10 of 12 match-ups. The two misclassifications of both algorithms occurred where one OWT 9 match-up was classified as OWT 8, and one OWT 8 match-up was classified as OWT 4. Those misclassifications resulted in minimum Precision and Recall of 0.75 and 0.83, respectively.

The 75 match-ups of Landsat-8 OLI covered almost all OWTs, where the SVC algorithm classified 26 match-ups, and the RF algorithm classified 41 match-ups. OLI RF algorithm classified more match-ups than the SVC algorithm because of its higher detection rate. For example, the OWT 4 detection rate of the SVC algorithm (33%) resulted in a reduction from 26 to 4 classified match-ups. On the

other hand, the OWT 4 detection rate of the RF algorithm (89%) resulted in a reduction from 26 to 8 match-ups classified. Indeed, the detection rate in the satellite images was significantly lower than that observed *in situ*. This significant reduction shows that the uncertainty in R_{rs} estimate from satellite data is a crucial factor for classifying the OWTs.

Table 5.6 - Performance of the classifications models for each OWT in the satellite and *in situ* match-ups.

Sensor	Classification Model	OWTs								
		1	2	4	5	6	7	8	9	
MSI	SVC	Precision	-	-	0.8	1	-	-	0.8	0
		Recall	-	-	1	1	-	-	0.8	0
		Targets classified	0	-	4	2	-	-	5	1
	RF	Precision	-	-	0.75	1	-	-	0.83	0
		Recall	-	-	1	1	-	-	0.83	0
		Targets classified	0	-	3	2	-	-	6	1
n	2	-	8	6	-	-	7	1		
OLI	SVC	Precision	1	1	1	1	0.44	-	1	-
		Recall	1	1	0.75	0.43	1	-	1	-
		Targets classified	1	1	4	7	4	-	9	-
	RF	Precision	1	0.33	1	0.8	0.33	-	1	-
		Recall	0.33	0.5	0.5	0.4	1	-	1	-
		Targets classified	3	2	8	10	5	-	13	-
n	4	4	26	12	7	-	22	-		
ETM+	SVC	Precision	1	0	0	-	0	-	0.86	1
		Recall	1	0	0	-	-	-	0.86	1
		Targets classified	1	1	3	-	-	-	7	1
	RF	Precision	-	0	0	-	0	-	0.89	1
		Recall	-	0	0	-	-	-	0.89	1
		Targets classified	0	1	3	-	-	-	9	1
n	7	2	4	-	-	-	15	1		

Source: The author.

Regarding OLI match-ups classified, SVC and RF algorithms had a poor performance for classifying OWTs 4, 5, and 6. The poor performance resulted from a large number of OWTs 4 and 5 match-ups that were classified as OWT 6. The reason why OWTs 4 and 5 were estimated as OWT 6 is the $R_{rs}(B5_{oli})$ overestimation in the satellite images. The $R_{rs}(B5_{oli})$ is one of the bands with the highest difference in OWTs 4 and 5 to 6, with OWT 6 having the highest values. With $R_{rs}(B5_{oli})$ overestimated in the satellite images, the R_{rs} magnitude increased and modified the classification to OWT 6. Thus, the $R_{rs}(B5_{oli})$ overestimation has been an essential factor in the lower OWT 6 Precision and Lower OWT 4 and 5 Recall.

The 29 match-ups of Landsat-7 ETM+ comprised the OWTs 1, 2, 4, 8, and 9, where SVC and RF algorithms classified 13 and 14, respectively. Considering the number of match-ups that were classified, the ETM+ SVC and RF algorithms showed unsatisfactory performance. The SVC and RF algorithms correctly classified 8 and 9 match-ups, respectively. However, most correct classifications were biased by OWT 8, which corresponds to the majority of samples, while other OWTs showed Precision and Recall equal 0.

As mentioned in section 5.2.4, the main issue of the satellite and *in situ* match-ups is that it does not represent all OWTs. For this reason, the satellite uncertainty was simulated in the *in situ* data for all OWTs, and the classification performance in satellite images could be simulated (see section 5.2.4).

Using the *in situ* data with simulated uncertainties (Table 5.7), the MSI SVC algorithm exhibited a Recall ranging from 0.61 to 0.99 and Precision ranging from 0.44 to 0.98, while the MSI RF algorithms showed a Recall varying from 0.4 to 1 and a Precision varying from 0.41 to 1. The detection rates of MSI SVC ranged from 52% to 96%, while the detection rate of MSI RF varied from 35 to 88%. The OWT 4 exhibited the lowest Precision for SVC and RF algorithm, while the lowest Recall was in the OWT 7 and OWT 4 for SVC and RF, respectively. Similar to the *in situ* results, the lower detection rate occurred in OWTs 4, 5, and 6. Compared to the MSI RF, MSI SVC exhibited a higher detection with similar accuracy.

Table 5.7 - Performance of the classifications models for each OWT using the *in situ* simulated sensors with uncertainty estimated from the satellite and *in situ* match-ups.

Sensor	Classification Model	OWTs								
		1	2	4	5	6	7	8	9	
MSI	SVC	Precision	0.98	0.81	0.44	0.87	0.92	0.79	0.96	0.84
		Recall	0.99	0.89	0.77	0.76	0.97	0.61	0.73	0.9
		Detection rate (%)	96	83	67	52	87	80	81	84
	RF	Precision	0.99	0.86	0.42	0.93	1	0.91	0.87	0.86
		Recall	1	0.84	0.4	0.91	1	0.94	0.87	0.98
		Detection rate (%)	88	35	35	35	77	40	63	92
OLI	SVC	Precision	0.93	0.53	0.17	0.32	0.2	0.28	0.84	0.72
		Recall	0.8	0.26	0.25	0.51	0.91	0.6	0.68	0.36
		Detection rate (%)	77	65	64	70	96	74	77	74
	RF	Precision	0.89	0.54	0.17	0.28	0.28	0.32	0.78	0.87
		Recall	0.92	0.19	0.18	0.33	0.88	0.68	0.81	0.64
		Detection rate (%)	82	60	63	57	86	57	75	84
ETM+	SVC	Precision	0.65	0.39	0.08	0.06	0.04	0.13	0.79	0.15
		Recall	0.46	0.13	0.02	0.1	0.92	0.23	0.27	0.04
		Detection rate (%)	93	88	92	95	98	95	88	92
	RF	Precision	0.82	0.39	0.12	0.05	0.08	0.14	0.75	0.58
		Recall	0.75	0.13	0.04	0.02	0.93	0.15	0.6	0.25
		Detection rate (%)	36	41	56	60	83	41	65	80
n		19	23	9	4	2	5	31	4	

Source: The author.

Regarding the OLI sensor performance with simulated uncertainties, the SVC and RF algorithms presented the same misclassifications observed in the satellite classification, where OWTs 4 and 5 were classified as OWT 6. Furthermore, the poor accuracy of OWTs 2 and 4 was also related to the overlap between them in the OLI bands. Regarding the OWTs 7 and 9 that had no match-ups, the results were unsatisfactory, with Precision varying from 0.28 to 0.87, and Recall from 0.36 to 0.68.

The performance of ETM+ algorithms was generally poor, where ETM+ SVC Precision and Recall varied from 0.04 to 0.65 and from 0.02 to 0.92, respectively, while the ETM+ RF Precision and Recall varied from 0.05 to 0.82 and from 0.02 to 0.93, respectively. Analogous to the OLI algorithms, the ETM+ classifications wrongly classified the OWTs 4 and 5 as OWT 6 due to the $R_{rs}(B_{4_{etm+}})$ overestimation. The unsatisfactory performance of ETM+ classifications may be due to the high $R_{rs}(B_{4_{etm+}})$ overestimation with a mean bias of 215%.

Compared to Lobo et al., (2012) and Uudeberg et al., (2019), we also found that uncertainties of R_{rs} estimation in satellite images have a more significant role in the OWT classification accuracy. A notable example is a misclassification in classifying OWTs 4 and 5 as OWT 6 using the OLI sensor due to the $R_{rs}(B_{5_{oli}})$ overestimation. Moreover, even for MSI classification which has the lowest bias between *in situ* and satellite $R_{rs}(B_{msi})$, there is a significant drop in the classification accuracy. For example, the *in situ* MSI RF Precision of all OWTs was 1; further, with the satellite uncertainty added to *in situ* data, the Precision reduced to 0.42 in OWT 4 and 0.86 in OWT 9. This is an essential issue since OWTs has been used for detecting abrupt changes or boosting bio-optical algorithms. However, many of those studies relying on OWTs have no dealt with OWT classification accuracy in satellite images (D'ALIMONTE et al., 2003; ELEVELD et al., 2017; MOORE et al., 2014; SUN et al., 2014; VANTREPOTTE et al., 2012).

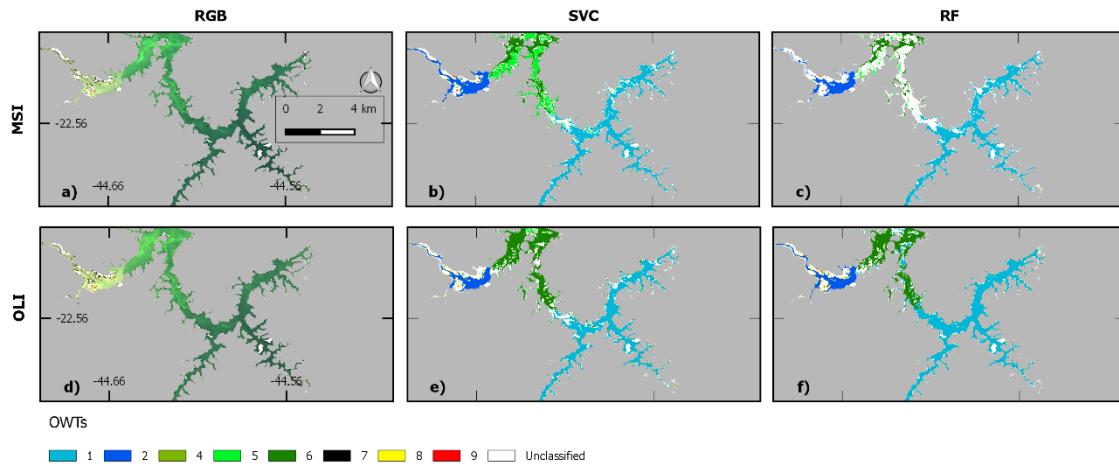
5.3.2.3 Image analyses of OWTs

The Sentinel-2 MSI and Landsat- OLI algorithms were evaluated in simultaneous images of the Funil reservoir (Figure 5.6), Curuai lake (Figure 5.7), and in the lower Amazon basin (Figure 5.8). Landsat-7 ETM+ was excluded from this analysis due to its poor performance exposed in the previous section; thus, the focus was to compare daily match-ups of Sentinel-2 MSI and Landsat-8 OLI images.

The OWTs detected in the Funil reservoir were OWT 1, 2, 5, and 6. All algorithms of MSI and OLI detected the OWT 1 in the south and east and the OWT 2 in the

farther west. Otherwise, the sensors and algorithms disagreed in estimating the OWTs 5 and 6. In the northwest region, the MSI SVC algorithms retrieved OWTs 5 and 6; the MSI RF only retrieved OWT 6 and a considerable unclassified area; the OLI SVC and RF algorithms only retrieved OWT 6.

Figure 5.6 - OWT classification in the Funil reservoir in 2019-08-10



The a), b), and c) are the RGB (B4, B3, B2) image, SVC, and RF classifications of the MSI sensor, respectively. The d), e), and f) are the RGB (Red, Green, NIR) image, SVC, and RF classifications of OLI sensor, respectively.

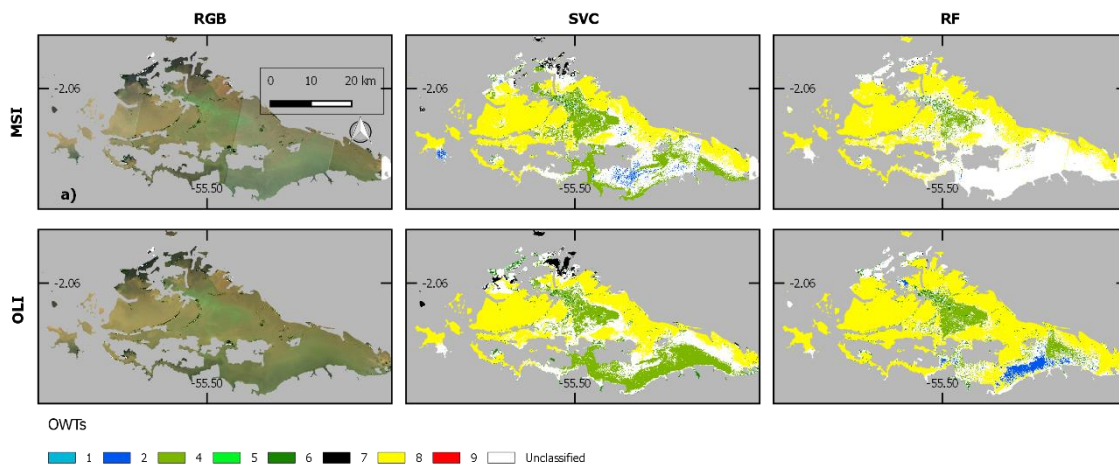
Source: The author.

As previously shown in section 5.3.2.2, the detection rate of MSI SVC is 52% for OWT 5, while that of MSI RF is 33%. Since MSI RF has lower detection than that of the MSI SVC, MSI RF misclassify most of the areas (Table 5.8), which were classified as OWT 5 by the SVC. MSI SVC retrieved 3.68 km² of OWT 5 and 6.47 km² of unclassified area, while MSI RF algorithm retrieved 0.45 km² of OWT 5 and 11.83 km² of unclassified area. The low detection rate of MSI RF would undoubtedly lead to underestimation of OWT 5 in eutrophic reservoirs, where OWT 5 is likely to occur. Therefore, due to the low detection rate, MSI RF should not be employed in those environments.

Using MSI SVC as a reference of OWT 5 and OWT 6 classification, the impact of $R_{rs}(B5_{oli})$ overestimation in the OLI classification can be demonstrated. The classifications using the OLI sensor retrieved OWT 6 the double in areas where

MSI SVC identified OWT 5. OLI classification estimated an area that approximately doubled the area of OWT6 predicted by MSI SVC classification. Furthermore, MSI SVC estimated an OWT 5 area 600% to 2500% larger than that of OLI classification. This comparison exhibits how a low Precision of an OWT causes area overestimation, while lower Recall causes area underestimation.

Figure 5.7 - OWT classification in the Curuai lake in 2018-08-01.



The a), b), and c) are the RGB (B4, B3, B2) image, SVC, and RF classifications of the MSI sensor, respectively. The d), e), and f) are the RGB (Red, Green, NIR) image, SVC, and RF classifications of OLI sensor, respectively.

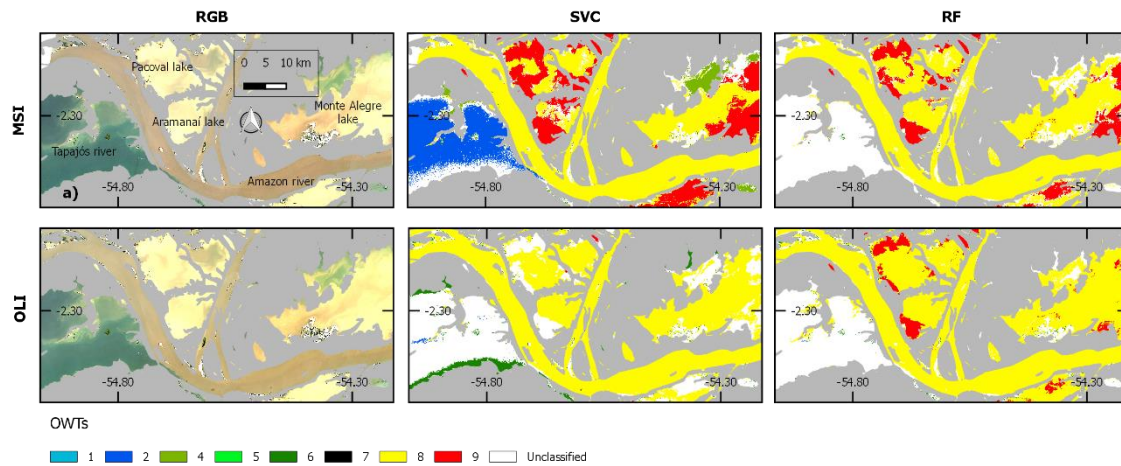
Source: The author.

The optical variability of Curuai lake contained OWTs 2, 4, 7, and 8. All algorithms classified the OWT 8 from west to east with similar coverage areas. In contrast, the remaining OWTs areas estimated were different depending on the algorithms or sensors. The MSI SVC and OLI RF algorithms classified OWT 2 in the south region, MSI RF unclassified it, and OLI SVC classified as OWT 4. MSI and OLI SVC algorithms classified OWT 7 northern area while the MSI and OLI RF algorithms assigned the pixels as unclassified.

The higher detection rate of MSI SVC than MSI RF caused SVC to estimate more areas of OWTs 2, 4, and 7 in Curuai lake. The satellite detection rates of MSI SVC are 83%, 67%, and 80% for OWTs 2, 4, and 7, respectively, while the RF

detection rates are 35%, 35%, and 40%, respectively. Consequently, for OWTs 2, 4, and 7, MSI SVC classification retrieved approximately 3100%, 500%, and 700% more areas than those of MSI RF classification, respectively. This example also shows that a low detection rate reduces the estimated area of its respective OWT.

Figure 5.8. OWT classification in the lower Amazon basin in 2017-09-04.



The a), b), and c) are the RGB (B4, B3, B2) image, SVC, and RF classifications of the MSI sensor, respectively. The d), e), and f) are the RGB (Red, Green, NIR) image, SVC, and RF classifications of OLI sensor, respectively.

Source: The author.

The overlap between OWTs 2 and 4 in OLI bands led to a confusion between those OWTs in the south region of maps in the OLI classification, where OLI SVC estimated OWT 4 in the south while OLI RF estimated OWT 2. Regarding OWT 7 estimation, its detection rate using OLI SVC was 77%, while that of the OLI RF was 57%., which resulted in the OLI SVC algorithm classifying OWT 7 in the northern area while the RF algorithm unclassified it. Consequently, OLI SVC estimated 41.11 km² of OWT 7 while OLI RF retrieved 1.01 km².

Table 5.8 - Total OWTs areas estimated in Funil Reservoir, Curuai Lake, and the Lower Amazon Basin.

Region	Sensor	Algorithm	OWTs Area (km ²)								
			1	2	4	5	6	7	8	9	Unclassified
Funil Reservoir	MSI	SVC	9.38	2.17	-	3.68	3.37	-	-	-	6.47
		RF	9.16	1.7	-	0.47	2.06	-	-	-	11.83
	OLI	SVC	8.37	1.57	-	0.51	7.41	-	-	-	7.49
		RF	10.87	1.85	-	0.14	6.29	-	-	-	5.98
Curuai Lake	MSI	SVC	-	18.79	235.07	-	33.86	24.27	489.97	-	459.68
		RF	-	0.57	41.96	-	19.93	1.44	555.04	-	642.67
	OLI	SVC	-	0.22	248.96	-	67.18	41.11	517.36	-	400.84
		RF	-	37.32	124.05	-	48.71	1.01	759.16	-	305.42
Lower Amazon Basin	MSI	SVC	-	200.82	62.1	-	13.07	0.49	887.17	305.42	236.4
		RF	-	-	0.3	-	5.74	-	1024.29	189.14	486
	OLI	SVC	-	0.37	0.71	-	72.42	0.6	1120.28	2.37	516.53
		RF	-	0.3	-	-	26.7	-	1231.25	82.74	372.26

Source: The author.

The OWTs detected in the lower Amazon basin were 2, 4, 6, 8, and 9. All algorithms of MSI and OLI detected OWT 8 in the Amazon river, Aramanaí lake, Pacoval lake, and Monte Alegre lake. The MSI SVC was the only algorithm that classified OWT 2 in the Tapajós River and OWT 4 in the northern part of Monte Alegre lake. The OWT 2 classification by MSI SVC algorithm reflected its higher detection rate (83%), while the MSI RF, OLI SVC, and OLI RF are 35%, 65%, and 60%, respectively. Regarding OWT 4, the MSI SVC detection rate is higher than that of RF, and it explains why MSI SVC classified 62.1 km² of OWT 4, while MSI RF classified as 0.3 km². Controversially, the OWT 4 detection rate of OLI algorithms was similar to the MSI SVC algorithm, but only the MSI SVC detected OWT 4.

Regarding the remaining OWTs, the MSI SVC, MSI RF, and OLI RF algorithms classified OWT 9 in the Aramanaí, Pacoval, and Monte Alegre lakes. The computed areas of OWT 9 slightly differed in each algorithm by mixing with OWT

8. Since the MSI RF algorithm exhibited the best Precision and Recall for OWTs 8 and 9, its OWT 8 and 9 areas were used as a reference. Thus, MSI SVC estimated 1.62 of OWT 9 area and 0.87 of OWT 8 area when compared to that retrieved from MSI RF classification, while OLI RF estimated 0.44 of OWT 9 area and 1.1 of OWT 8 area.

Considering all study sites analyzed, MSI SVC provided the best estimates and is the most suitable algorithm for future monitoring of the OWTs. MSI SVC has satisfactory accuracy in satellite images (Table 5.6 and Table 5.7), allowing the detection of OWTs 1, 5, and 6 in eutrophic reservoirs, and OWTs 2, 4, 7, 8, and 9 in the Amazon floodplain lakes. MSI RF fails to detect OWTs due to its low detection rate obtained in novelty detection. Moreover, the classification based on the OLI sensor failed to estimate OWTs related to algae bloom due to its $R_{rs}(B5_{oli})$ overestimation.

5.4 Conclusion

This study assessed the feasibility of detecting the Brazilian OWTs using remote sensing through Sentinel-2 MSI, Landsat-8 OLI, and Landsat-7 ETM+ sensors. The Sentinel-2 MSI is the best sensor for detecting the OWTs, while further employment of Landsat-8 OLI and Landsat-7 ETM+ for classifying the OWTs will require a better estimation of the of R_{rs} in the NIR. Thus, for further monitoring the OWTs, we recommend using the Sentinel-2 MSI sensor throughout the SVC classification algorithm.

The method of classifying OWTs using remote sensing proposed here can support a series of applications: i) provide a preliminary study of a new environment, which can further support sampling designs for survey campaigns; ii) afford an anomaly detection method for abrupt changes caused by sediment loading or algae blooms; iii) support the development of bio-optical algorithms for specific OWTs; iv) create a census of the Brazilian inland waters and provide reliable data of the water status in a macroscale level. Thus, it could fill the spatial-temporal gap of data in the Brazilian inland waters.

6 FINAL CONSIDERATIONS

Monitoring OWTs of Brazilian water can be a reliable tool for monitoring water quality with broad spatial-temporal coverage. In this study, we have provided detailed information about: (1) if Kutser's and Mobley's deglint methods can be used in the same database for computing OWTs, (2) the Brazilian OWTs, and (3) how to detect those OWTs in satellite images. Thus, we could answer the following questions:

- Can the Kutser's and Mobley's glint removal be used in the same dataset for computing OWT?

No, the spectra samples can be assigned to different OWTs depending on the glint removal method instead of being a response from the OACs. For this reason, the OWTs was only based on Mobley's method.

- What are the Brazilian OWTs?

There are eight OWTs identified in Brazilian waters used in this study: OWT 1 is transparent waters, OWT 2 is transparent waters with moderate CDOM and TSM; OWT 4, 5, and 6 are waters with algae blooms; OWT 7 is waters with the highest CDOM concentration; OWT 8 and 9 is turbid waters with high TSM concentration. However, this number of OWT may increase since the samples available are limited and do not cover the transition areas. In this case, the classification of Sentinel images using the OWTs may identify unclassified areas as key regions for sampling acquisition.

- How to detect the Brazilian OWTs in satellite images?

Our findings show that using Sentinel-2 MSI with the SVC algorithm is the best method for monitoring the OWTs in satellite images. Landsat-8 OLI and Landsat-7 ETM+ are not capable of monitoring the OWTs due to the overestimation of NIR bands.

Considering all the answers showed above, we can confirm our hypothesis that OWTs can be used for satellite monitoring of Brazilian inland waters.

REFERENCES

- AGÊNCIA NACIONAL DE ÁGUAS. **Panorama da qualidade das águas superficiais do Brasil**: 2012. Brasília: ANA, 2012.
- ALVES, T. P.; MAFRA, L. L. Diel variations in cell abundance and trophic transfer of diarrhetic toxins during a massive dinophysis bloom in Southern Brazil. **Toxins**, v. 10, n. 6, 2018.
- AMERICAN PUBLIC HEALTH ASSOCIATION; AMERICAN WATER WORKS ASSOCIATION; WATER POLLUTION CONTROL FEDERATION. **Standard methods for the examination of water and wastewater**. 16. ed. Washington, D. C.: American Public Health Association, 1985.
- ARTHUR, D.; VASSILVITSKII, S. k-means++: the advantages of careful seeding. In: ANNUAL ACM-SIAM SYMPOSIUM ON DISCRETE ALGORITHMS, 8., 2007. **Proceedings...** New Orleans, 2007.
- AUGUSTO-SILVA, P. B. **Caracterização e avaliação da dinâmica sazonal das propriedades bio-ópticas do Reservatório de Funil com apoio de sensoriamento remoto, dados in situ e modelos ópticos**. 2013. 185p. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2013.
- AUGUSTO-SILVA, P. B.; OGASHAWARA, I.; BARBOSA, C. C. F.; CARVALHO, L. A. S.; JORGE, D. S. F.; FORNARI, C. I.; STECH, J. L. Analysis of MERIS reflectance algorithms for estimating chlorophyll-a concentration in a Brazilian reservoir. **Remote Sensing**, v. 6, n. 12, p. 11689–11707, 2014.
- BARBARISI, B. F. **Integração de imagem MODIS e de dados censitários na avaliação do impacto da pecuária sobre o estado trófico de sistemas aquáticos do Baixo Amazonas**. 2010. 169p. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2010.
- BARBOSA, C. C. F. **Sensoriamento remoto da dinâmica da circulação da água do sistema planície de Curuai/Rio Amazonas**. 2005. 287p. Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2007.
- BARBOSA, C.; FERREIRA, R.; ARAUJO, C.; NOVO, E. Bio-optical characterization of two brazilian hydroelectric reservoirs as support to understand the carbon budget in hydroelectric reservoirs. **International Geoscience and Remote Sensing Symposium (IGARSS)**, p. 898–901, 2014.
- BARBOSA, C.; LEE, Z.; CARVALHO, L. S.; NOVO, E. Effect of stratified water column on chlorophyll estimate by remote sensing algorithms in a highly eutrophic hydroelectric reservoir. In: OCEAB OPTICS, 2016. **Proceedings...** 2016.

- 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, p. 1–22, 2018.
- BIDIGARE, R. R.; ONDRUSEK, M. E.; MORROW, J. H.; KIEFER, D. A. In vivo absorption properties of algal pigments. **Ocean Optics**, v. 1302, p. 290–302, 1990.
- BISHOP, C. M. Novelty detection and neural network validation. **IEEE Proceedings: Vision, Image and Signal Processing**, v. 141, n. 4, p. 217–222, 1994.
- BOURGOIN, L. M.; BONNET, M. P.; MARTINEZ, J. M.; KOSUTH, P.; COCHONNEAU, G.; MOREIRA-TURCQ, P.; GUYOT, J. L.; VAUCHEL, P.; FILIZOLA, N.; SEYLER, P. Temporal dynamics of water and sediment exchanges between the Curuaí floodplain and the Amazon River, Brazil. **Journal of Hydrology**, v. 335, n. 1–2, p. 140–156, 2007.
- BRICAUD, A.; MOREL, A.; LOUIS, P. Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. **Limnology and Oceanography**, v. 26, n. 1, p. 43–53, 1981.
- BUSHAW, K. L.; ZEPPT, R. G.; TARR, M. A.; SCHULZ-JANDER, D.; BOURBONNIERE, R. A.; HODSON, R. E.; MILLER, W. L.; BRONK, D. A.; MORAN, M. A. Photochemical release of biologically available nitrogen from aquatic dissolved organic matter. **Letters to Nature**, v. 381, p. 404–407, 1996.
- CAIRO, C.; BARBOSA, C.; LOBO, F.; NOVO, E.; CARLOS, F.; MACIEL, D.; FLORES JÚNIOR, R.; SILVA, E. F. F.; 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, 2019a.
- CAIRO, C.; BARBOSA, C.; LOBO, F.; NOVO, E.; CARLOS, F.; MACIEL, D.; FLORES JÚNIOR, R.; SILVA, E. F. F.; 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.
- CAIRO, C. T.; BARBOSA, C. C. F.; NOVO, E. M. L. M.; CARMO CALIJURI, M. Spatial and seasonal variation in diffuse attenuation coefficients of downward irradiance at Ibitinga Reservoir, São Paulo, Brazil. **Hydrobiologia**, v. 784, n. 1, p. 265–282, 2016.
- CAIRO, C. T.; MACIEL, D. A.; FLORES JÚNIOR, R.; CARLOS, F. M.; BARBOSA, C. C. F.; LOBO, F. L.; NOVO, E. M. L. M. Estimativa da concentração da clorofila-a no reservatório de Ibitinga/Sp – resultados preliminares. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 19., 2019, Santos, SP. **Anais...** São José dos Campos: INPE, 2019b.
- CARLOS, F. M.; MARTINS, V. D. S.; BARBOSA, C. C. F. Sistem semi-automático de correção atmosférica para multi-sensores orbitais. In: SIMPÓSIO

- BRASILEIRO DE SENSORIAMENTO REMOTO, 19., 2019m Santos, SP. **Anais...**São José dos Campos: INPE, 2019.
- CHEN, P.; PAN, D.; WANG, T.; MAO, Z.; ZHANG, Y. Coastal and inland water monitoring using a portable hyperspectral laser fluorometer. **Marine Pollution Bulletin**, v. 119, n. 1, p. 153–161, 2017.
- CHEN, X.; LI, Y. S.; LIU, Z.; YIN, K.; LI, Z.; WAI, O. W. H.; KING, B. Integration of multi-source data for water quality classification in the Pearl River estuary and its adjacent coastal waters of Hong Kong. **Continental Shelf Research**, v. 24, n. 16, p. 1827–1843, 2004.
- CIOTTI, A. M.; LEWIS, M. R.; CULLEN, J. Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient. **Limnology and Oceanography**, v. 47, n. 2, p. 404–417, 2002.
- CURTARELLI, V. P.; SILVA, E. F. F.; LOBO, F. L.; BARBOSA, C. C. F.; NOVO, E. M. L. M. Water transparency in a brazilian reservoir. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 19., 2019, Santos, SP. **Anais...** São José dos Campos: INPE, 2019.
- D'ALIMONTE, D.; MÉLIN, F.; ZIBORDI, G.; BERTHON, J. F. Use of the novelty detection technique to identify the range of applicability of empirical ocean color algorithms. **IEEE Transactions on Geoscience and Remote Sensing**, v. 41, n. 12, pt.1, p. 2833–2843, 2003.
- DA SILVA, E. F. F.; LUZ, L. F. G.; BARBOSA, C. C. F.; NOERNBERG, M. A. Turbidity distribution in a subtropical estuary: the estuarine complex of Paranaguá. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 19., 2019, Santos, SP. **Anais...** São José dos Campos: INPE, 2019.
- DAVIES-COLLEY, R. J.; PRIDMORE, R. D.; HEWITT, J. E. Optical properties of some freshwater phytoplanktonic algae. **Hydrobiologia**, v. 133, n. 2, p. 165–178, 1986.
- DE CARVALHO, L. A. S.; MARTINS, V. S.; LOBO, F. L.; BARBOSA, C. C. F.; NOVO, E. M. L. M. Investigating empirical models to retrieve the backscattering slope from Landsat 8 images - the case of Lago Grande Curuai. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 18., 2017. **Anais...** São José dos Campos: INPE, 2017.
- DEYONG, S.; YUNMEI, L.; QIAO, W.; GAO, J.; LE, C.; CHANGCHUN, H.; SHAOQI, G. Hyperspectral remote sensing of the pigment C-phycoyanin in turbid inland waters, based on optical classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 51, n. 7, p. 3871–3884, 2013.
- DOGLIOTTI, A. I.; RUDDICK, K.; GUERRERO, R. Seasonal and inter-annual turbidity variability in the Río de la Plata from 15 years of MODIS: El Niño dilution effect. **Estuarine, Coastal and Shelf Science**, v. 182, p. 27–39, 2016.

- DUYSENS, L. N. M. The flattening of the absorption spectrum of suspensions, as compared to that of solutions. **Biochimica et Biophysica Acta**, v. 19, p. 169–197, 1956.
- ELEVELD, M. A.; RUESCAS, A. B.; HOMMERSOM, A.; MOORE, T. S.; PETERS, S. W. M.; BROCKMANN, C. An optical classification tool for global lake waters. **Remote Sensing**, v. 9, n. 5, 2017.
- EVANS, C. D.; MONTEITH, D. T.; COOPER, D. M. Long-term increases in surface water dissolved organic carbon: observations, possible causes and environmental impacts. **Environmental Pollution**, v. 137, n. 1, p. 55–71, 2005.
- FERNANDA SOBRINHO, B.; LUZ, L.; FERNANDES, L. F.; MAFRA, L. Evidence of *noctiluca scintillans* grazing during a bloom of toxic *Dinophysis acuminata* complex, south Brazil. In: INTERNATIONAL CONFERENCE ON HARMFUL ALGAE, 18., 2018. **Proceedings...** 2018.
- FERREIRA, R. M. P. **Caracterização da ótica e do carbono orgânico dissolvido no reservatório de Três Marias/Mg**. 2014. 128p. Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, 2014.
- FRALEY, C. How many clusters? which clustering method? answers via model-based cluster analysis. **The Computer Journal**, v. 41, n. 8, p. 578–588, 1998.
- GANF, G. G.; OLIVER, R. L.; WALSBY, A. E. Optical properties of gas-vacuolate cells and colonies of microcystis in relation to light attenuation in a turbid, stratified reservoir (Mount bold reservoir, South Australia). **Marine and Freshwater Research**, v. 40, n. 6, p. 595–611, 1989.
- GARABA, S. P.; ZIELINSKI, O. Methods in reducing surface reflected glint for shipborne above-water remote sensing. **Journal of the European Optical Society Rapid Publications**, v. 8, n. 13058, p. 8, 2013.
- GITELSON, A. The peak near 700 nm on radiance spectra of algae and water: relationships of its magnitude and position with chlorophyll. **International Journal of Remote Sensing**, v. 13, n. 17, p. 3367–3373, 1992.
- GITELSON, A. A.; GRITZ, Y.; MARK N. MERZLYAK. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. **Journal of Plant Physiology**, v. 160, p. 271–282, 2003.
- HATJE, V.; PEDREIRA, R. M. A.; REZENDE, C. E.; SCHETTINI, C. A. F.; SOUZA, G. C.; MARIN, D. C.; HACKSPACHER, P. C. The environmental impacts of one of the largest tailing dam failures worldwide. **Scientific Reports**, v. 7, n. 1, p. 1–13, 2017.
- HUANG, C.; LI, Y.; YANG, H.; LI, J.; CHEN, X.; SUN, D.; LE, C.; ZOU, J.; XU, L. Assessment of water constituents in highly turbid productive water by optimization bio-optical retrieval model after optical classification. **Journal of**

Hydrology, v. 519, p. 1572–1583, 2014.

JOHNSEN, G.; SAKSHAUG, E. Biooptical characteristics of PSII and PSI in 33 species (13 pigment groups) of marine phytoplankton, and the relevance for pulseamplitude-modulated and fast-repetition-rate fluorometry. **Journal of Phycology**, v. 43, n. 6, p. 1236–1251, 2007.

JORGE, D. S. F. **Caracterização bio-óptica e implementação de algoritmos semi-analíticos para o estudo da qualidade da água na Reserva de Desenvolvimento Sustentável Mamirauá**. 2018. 163p. Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2018.

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

JØRGENSEN, S.; TUNDISI, J. G.; TUNDISI, T. M. **Handbook of inland aquatic ecosystem management**. [S.l.]: CRC Press, 2012.

KIRK, J. T. O. Monte Carlo modeling of the performance of a reflective tube absorption meter. **Applied Optics**, v. 31, n. 30, p. 6463, 1992.

KIRK, J. T. O. **Light and photosynthesis in aquatic ecosystems**. 3.ed. Cambridge: Cambridge University Press, 2020.

KOTCHENOVA, S. Y.; VERMOTE, E. F.; MATARRESE, R.; FRANK J. KLEMM, J. Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data: part I: path radiance. **Applied optics**, v. 45, n. 26, p. 6762–6774, 2006.

KOTEHENOVA, S. Y.; VERMOTE, E. F. Validation of a vector version of the 6S radiative transfer code for atmospheric correction of satellite data: part II: homogeneous Lambertian and anisotropic surfaces. **Applied Optics**, v. 46, n. 20, p. 4455–4464, 2007.

KRUSE, F. A.; LEFKOFF, A. B.; BOARDMAN, J. W.; HEIDEBRECHT, K. B.; SHAPIRO, A. T.; BARLOON, P. J.; GOETZ, A. F. H. The Spectral Image Processing System (SIPS): interactive visualization and analysis of imaging spectrometer data. **Remote Sensing of Environment**, v. 44, n. 145–163, p. 145–163, 1993.

KUTSER, T.; VAHTMÄE, 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. 133, p. 85–89, 2013.

LE, C.; LI, Y.; ZHA, Y.; SUN, D.; HUANG, C.; ZHANG, H. Remote estimation of chlorophyll a in optically complex waters based on optical classification. **Remote Sensing of Environment**, v. 115, n. 2, p. 725–737, 2011.

LEE, Z.; CARDER, K. L.; MOBLEY, C. D.; STEWARD, R. G.; PATCH, J. S. Hyperspectral remote sensing for shallow waters: a semianalytical model. **Applied Optics**, v. 37, n. 27, p. 6329, 1998.

LIBES, S. M. **Introduction to marine biogeochemistry**. 2. ed. [S.l.]: Academic Press, 2009.

LIU, J.; SUN, D.; ZHANG, Y.; LI, Y. Pre-classification improves relationships between water clarity, light attenuation, and suspended particulates in turbid inland waters. **Hydrobiologia**, v. 711, n. 1, p. 71–86, 2013.

LOBO, F. L.; NOVO, E. M. L. M.; BARBOSA, C. C. F.; GALVÃO, L. S. Reference spectra to classify Amazon water types. **International Journal of Remote Sensing**, v. 33, n. 11, p. 3422–3442, 2012.

LOBO, F. L.; COSTA, M. P. F.; NOVO, E. M. L. M. Time-series analysis of Landsat-MSS/TM/OLI images over Amazonian waters impacted by gold mining activities. **Remote Sensing of Environment**, v. 157, p. 170–184, 2015.

LONDE, L. R. **Comportamento espectral do fitoplâncton de um reservatório brasileiro eutrofizado – Ibitinga (SP)**. 2008. Tese (Doutorado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2018.

LUBAC, B.; LOISEL, H. Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea. **Remote Sensing of Environment**, v. 110, n. 1, p. 45–58, 2007.

MACIEL, D.; NOVO, E.; CARVALHO, L. S.; BARBOSA, C.; FLORES JÚNIOR, R.; LUCIA LOBO, F. Retrieving total and inorganic suspended sediments in Amazon floodplain lakes: a multisensor approach. **Remote Sensing**, v. 11, n. 15, p. 1744, 2019.

MACQUEEN. Some methods for classification and analysis of multivariate observations. In: BERKELEY SYMPOSIUM, 5., 1967. **Proceeding...** 1967.

MANN, H.B.; WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. **The Annals of Mathematical Statistics**, v. 18, n. 1, p. 50–60, 1947.

MARK, W. M. Bio-optical modeling of phytoplankton Chlorophyll-a. In: MISHRA, D. R.; OGASHAWARA, I.; GITELSON, A. A. (Ed.). **Bio-optical modeling and remote sensing of inland waters**. [S.l.]: Candice Janco, 2017. p. 157–188.

MARTINS, V. S.; BARBOSA, C. C. F.; CARVALHO, L. A. S.; JORGE, D. S. F.; LOBO, F. L.; 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.

MASEK, J. G.; VERMOTE, E. F.; SALEOUS, N. E.; WOLFE, R.; HALL, F. G.; HUENNRICH, K. F.; GAO, F.; KUTLER, J.; LIM, T. K. A landsat surface

reflectance dataset for North America, 1990-2000. **IEEE Geoscience and Remote Sensing Letters**, v. 3, n. 1, p. 68–72, 2006.

MÉLIN, F.; VANTREPOTTE, V.; CLERICI, M.; D'ALIMONTE, D.; ZIBORDI, G.; BERTHON, J. F.; CANUTI, E. Multi-sensor satellite time series of optical properties and chlorophyll-a concentration in the Adriatic Sea. **Progress in Oceanography**, v. 91, n. 3, p. 229–244, 2011.

MILLER, R. L.; BELZ, M.; DEL CASTILLO, C.; TRZASKA, R. Determining CDOM absorption spectra in diverse coastal environments using a multiple pathlength, liquid core waveguide system. **Continental Shelf Research**, v. 22, p. 1301–1310, 2002.

MILLER, R. L.; LIU, C. C.; BUONASSISSI, C. J.; WU, A. M. A multi-sensor approach to examining the distribution of total suspended matter (TSM) in the Albemarle-Pamlico Estuarine System, NC, USA. **Remote Sensing**, v. 3, n. 5, p. 962–974, 2011.

MILLER, R. L.; MCKEE, B. A. Using MODIS Terra 250 m imagery to map concentrations of total suspended matter in coastal waters. **Remote Sensing of Environment**, v. 93, p. 259–266, 2004.

MISHRA, A. K.; GARG, N. Analysis of trophic state index of Nainital Lake from Landsat -7 ETM data. **Journal of the Indian Society of Remote Sensing**, v. 39, n. 4, p. 463–471, 2011.

MISHRA, S.; MISHRA, D. R. Normalized difference chlorophyll index: a novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. **Remote Sensing of Environment**, v. 117, p. 394–406, 2012.

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

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

MOORE, T. S.; DOWELL, M. D.; BRADT, S.; RUIZ VERDU, A. An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. **Remote Sensing of Environment**, v. 143, p. 97–111, 2014.

MOREL, A.; BRICAUD, A. Theoretical results concerning light absorption in a discrete medium, and application to specific absorption of phytoplankton. **Deep-Sea Research**, v. 28A, n. 11, p. 1375–1393, 1981.

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

MOUNTRAKIS, G.; IM, J.; OGOLE, C. Support vector machines in remote sensing: a review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 66, n. 3, p. 247–259, 2011.

NOERNBERG, M. A.; LUZ, L.; SILVA, E. F. F.; MAFRA, L. L. Remote sensing reflectance variability along an intense HAB event in Southern Brazil. In: INTERNATIONAL OCEAN COLOUR SCIENCE MEETING (IOCS), 2017, Lisbon. **Proceedings...** 2017.

NUSH, E. A. Comparison of different methods for chlorophyll and phaeopigments determination. **Archives of Hydrobiology**, v. 14, n. 1, p. 14–39, 1980.

OGASHAWARA, I.; MISHRA, D. R.; GITELSON, A. A. Remote sensing of inland waters: background and current state-of-the-art. In: MISHRA, D. R.; OGASHAWARA, I.; GITELSON, A. A. (Ed.). **Bio-optical modeling and remote sensing of inland waters**. [S.l.]: Elsevier, 2017. p. 1–24.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D. Scikit-learn: machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

PENG, F.; EFFLER, S. W.; O'DONNELL, D.; WEIDEMANN, A. D.; AUER, M. T. Characterizations of minerogenic particles in support of modeling light scattering in Lake Superior through a two-component approach. **Limnology and Oceanography**, v. 54, n. 4, p. 1369–1381, 2009.

PETUS, C.; CHUST, G.; GOHIN, F.; DOXARAN, D.; FROIDEFOND, J. M.; SAGARMINAGA, Y. Estimating turbidity and total suspended matter in the Adour River plume (South Bay of Biscay) using MODIS 250-m imagery. **Continental Shelf Research**, v. 30, n. 5, p. 379–392, 2010.

PLATT, J. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. **Advances in Large Margin Classifiers**, v. 10, n. 3, p. 61–74, 1999.

RAMSAY, A. J. O.; WICKHAM, H.; GRAVES, S. **fda: Functional Data Analysis**. 2018. Disponível em: <<https://cran.r-project.org/package=fda>>.

RAMSAY, J. O. **Functional data analysis**. [S.l.]: Wiley Online Library, 2006.

REINART, A.; HERLEVI, A.; ARST, H.; SIPELGAS, L. Preliminary optical classification of lakes and coastal waters in Estonia and south Finland. **Journal of Sea Research**, v. 49, n. 4, p. 357–366, 2003.

REYNOLDS, C. **Ecology of phytoplankton**. Cambridge: Cambridge University Press, 2006. v. 136.

RIBEIRO FILHO, R.; PETRERE JUNIOR, M.; BENASSI, S.; PEREIRA, J. Itaipu reservoir limnology: eutrophication degree and the horizontal distribution of its limnological variables. **Brazilian Journal of Biology**, v. 71, n. 4, p. 889–902, 2011.

RODRIGUES, T.; ALCÂNTARA, E.; WATANABE, F.; IMAI, N. Retrieval of secchi disk depth from a reservoir using a semi-analytical scheme. **Remote Sensing of Environment**, v. 198, p. 213–228, 2017.

RODRIGUES, T.; IVÁNOVÁ, I.; ALCÂNTARA, E.; SHIMABUKURO, M.; WATANABE, F.; ROTTA, L.; IMAI, N.; BARBOSA, C. Analysis of data quality element's applicability for radiometric measurements in remote sensing of water : a case study in Nova Avanhandava Reservoir , São Paulo, Brazil. **Revista Brasileira de Cartografia**, n. 67/7, p. 1327–1339, 2015.

ROESLER, C. S.; BOSS, E. S. In situ measurement of the Inherent Optical Properties (IOPs) and potential for Harmful Algal Bloom (HAB) detection and coastal ecosystem observations. In: BABIN, M.; ROESLER, C. S.; CULLEN, J.J. (Ed.). **Real-time coastal observing systems for marine ecosystem dynamics and harmful algal blooms: theory, instrumentation and modelling**. [S.l.]: United Nations Educational, Scientific and Cultural Organization, 2008. p. 153–206.

ROTTA, L. H.; MISHRA, D. R.; ALCÂNTARA, E.; IMAI, N.; WATANABE, F.; RODRIGUES, T. Kd(PAR) and a depth based model to estimate the height of submerged aquatic vegetation in an oligotrophic reservoir: a case study at Nova Avanhandava. **Remote Sensing**, v. 11, n. 3, p. 1–21, 2019.

ROTTA, L. H. S.; MISHRA, D. R.; ALCÂNTARA, E. H.; IMAI, N. N. Analyzing the status of submerged aquatic vegetation using novel optical parameters. **International Journal of Remote Sensing**, v. 37, n. 16, p. 3786–3810, 2016.

ROUSSEUW, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. **Journal of Computational and Applied Mathematics**, v. 20, p. 53–65, 1987.

SANTOS, M. A. P. F.; MELÃO, M. G. G.; LOMBARDI, A. T. The effects of humic substances on copper toxicity to *Ceriodaphnia silvestrii* Daday (Crustacea, Cladocera). **Ecotoxicology**, v. 17, n. 6, p. 449–454, 2008.

SHI, K.; LI, Y.; LI, L.; LU, H. Absorption characteristics of optically complex inland waters: implications for water optical classification. **Journal of Geophysical Research: Biogeosciences**, v. 118, n. 2, p. 860–874, 2013a.

SHI, K.; LI, Y.; LI, L.; LU, H.; SONG, K.; LIU, Z.; XU, Y.; LI, Z. Remote chlorophyll-a estimates for inland waters based on a cluster-based classification. **Science of the Total Environment**, v. 444, p. 1–15, 2013b.

SHIGAKI, F.; SHARPLEY, A.; PROCHNOW, L. I. Animal-based agriculture, phosphorus management and water quality in Brazil : options for the future. **Scientia Agricola**, v. 63, n. 2, p. 194–209, 2006.

SILVA, M. P. **Modelos de sensoriamento remoto no estudo da variabilidade espacial e sazonal da matéria orgânica dissolvida em lagos da planície de inundação do Solimões/Amazonas a partir de suas propriedades óticas**. 2018. 97p. Dissertação (Mestrado em Sensoriamento

Remoto) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2018.

SIOLI, H. Alguns resultados e problemas da limnologia amazônica. **Boletim Técnico**, n. 3, p. 3–44, 1951.

SIOLI, H. Hydrochemistry and geology in the Brazilian Amazon region. **Amazoniana: Limnologia et Oecologia Regionalis Systematis Fluminis Amazonas**, v. 1, n. 3, p. 267–277, 1968.

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, 2018.

SUN, D.; HU, C.; QIU, Z.; CANNIZZARO, J. P.; BARNES, B. B. Influence of a red band-based water classification approach on chlorophyll algorithms for optically complex estuaries. **Remote Sensing of Environment**, v. 155, p. 289–302, 2014.

SUN, D.; LI, Y.; WANG, Q.; LV, H.; LE, C.; HUANG, C.; GONG, S. Partitioning particulate scattering and absorption into contributions of phytoplankton and non-algal particles in winter in Lake Taihu (China). **Hydrobiologia**, v. 644, n. 1, p. 337–349, 2010.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introdução ao data mining: mineração de dados**. [S.l.]: Ciência Moderna, 2009.

TIBSHIRANI, R.; WALTHER, G.; HASTIE, T. Estimating the number of clusters in a data set via the gap statistic. **Journal of the Royal Statistical Society Series B (Statistical Methodology)**, v. 63, n. 2, p. 411–423, 2001.

TUNDISI, J. G.; MATSUMURA-TUNDISI, T.; PARESCHI, D. C.; LUZIA, A. P.; HAELING, P. H. VON; FROLLINI, E. H. The Tietê/Jacaré watershed: a case study in research and management. **Estudos Avançados**, v. 22, n. 63, p. 159–172, 2008.

UUDEBERG, K.; ANSKO, I.; GETTER, P.; ANSPER, A. Using optical water types to monitor changes in optically complex inland and coastal waters. **Remote Sensing**, v.11, n.10, p. 1–20, 2019.

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.

VANTREPOTTE, V.; LOISEL, H.; DESSAILLY, D.; MÉRIAUX, X. Optical classification of contrasted coastal waters. **Remote Sensing of Environment**, v. 123, p. 306–323, 2012.

VERMOTE, E. F.; TANRÉ, D.; DEUZÉ, J. L.; HERMAN, M.; MORCRETTE, J. J. Second simulation of the satellite signal in the solar spectrum, 6s: an overview. **IEEE Transactions on Geoscience and Remote Sensing**, v. 35, n. 3, p. 675–686, 1997.

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

WET LABS. **Spectral absorption and attenuation meter: user's guide**. 2009. Available from: https://www.comm-tec.com/prods/mfgs/Wetlabs/Manuals/ac-9_manual.pdf.

WETZEL, R. G.; LIKENS, G. E. **Limnological analysis**. 2. ed. New York: Springer Verlag, 1991.

WILCOXON, F. Individual comparisons by ranking methods. **Biometrics Bulletin**, v. 1, n. 6, p. 80–83, 1945.

WILLIAMSON, C. E.; NEALE, P. J.; GRAD, G.; LANGE, H. J.; HARGREAVES, B. R. Beneficial and detrimental effects of UV radiation on aquatic organisms: implications of variation in spectral composition. **Ecological Applications**, v. 11, n. 6, p. 1843–1857, 2001.

WU, T. F.; LIN, C. J.; WENG, R. C. Probability estimates for multi-class classification by pairwise coupling. **Journal of Machine Learning Research**, v. 5, p. 975–1005, 2004.

WÜNSCH, U. J.; STEDMON, C. A.; TRANVIK, L. J.; GUILLEMETTE, F. Unraveling the size-dependent optical properties of dissolved organic matter. **Limnology and Oceanography**, v. 63, n. 2, p. 588–601, 2018.

ZHANG, F.; LI, J.; SHEN, Q.; ZHANG, B.; WU, C.; WU, Y.; WANG, G.; WANG, S.; LU, Z. Algorithms and schemes for chlorophyll a estimation by remote sensing and optical classification for turbid lake Taihu, China. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 8, n. 1, p. 350–364, 2015.

ZIBORDI, G.; RUDDICK, K.; ANSKO, I.; MOORE, G.; KRATZER, S.; ICELY, J.; REINART, A. In situ determination of the remote sensing reflectance: an inter-comparison. **Ocean Science**, v. 8, n. 4, p. 567–586, 2012.

APPENDIX A - HOW THE ACCURACY CLASSIFICATION IMPROVES USING PROBABILITY ESTIMATES AND THE PROBABILITY THRESHOLDS FOR ALL CLASSIFICATION MODELS AND OWTs

A.1 Probability thresholds

Table A.1 – SVC and RF probability thresholds for the MSI, OLI, and ETM+ sensors.

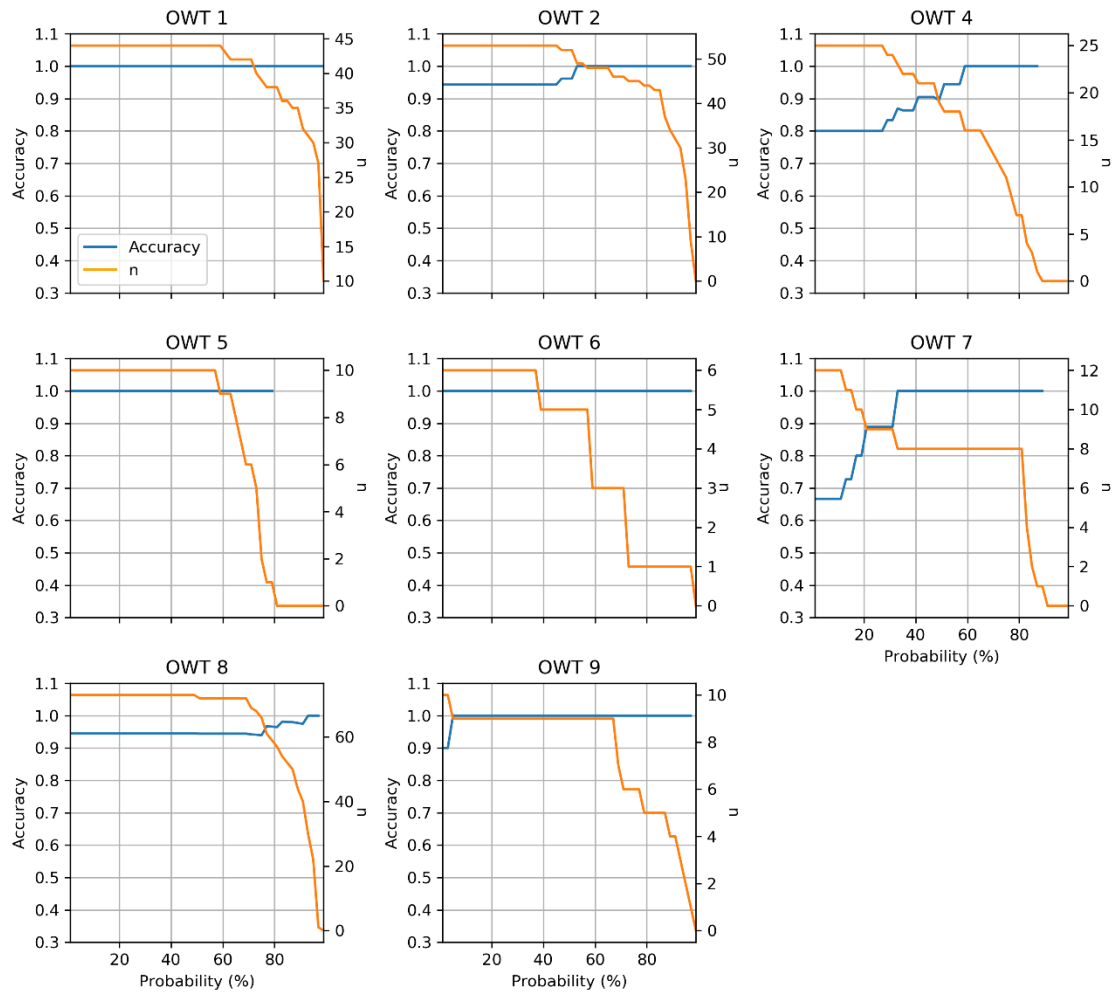
Sensor	Classification Model	Prob. Threshold (%)	OWTs							
			1	2	4	5	6	7	8	9
MSI	SVC	Prob. Threshold (%)	59	53	59	57	37	33	60	5
	RF	Prob. Threshold (%)	73	77	63	85	71	79	75	57
OLI	SVC	Prob. Threshold (%)	57	79	65	57	21	11	60	60
	RF	Prob. Threshold (%)	63	63	65	73	63	75	63	59
ETM+	SVC	Prob. Threshold (%)	43	73	67	23	13	23	79	51
	RF	Prob. Threshold (%)	75	71	67	77	73	93	73	57

Source: The author.

A.2 Increasing accuracy using probability

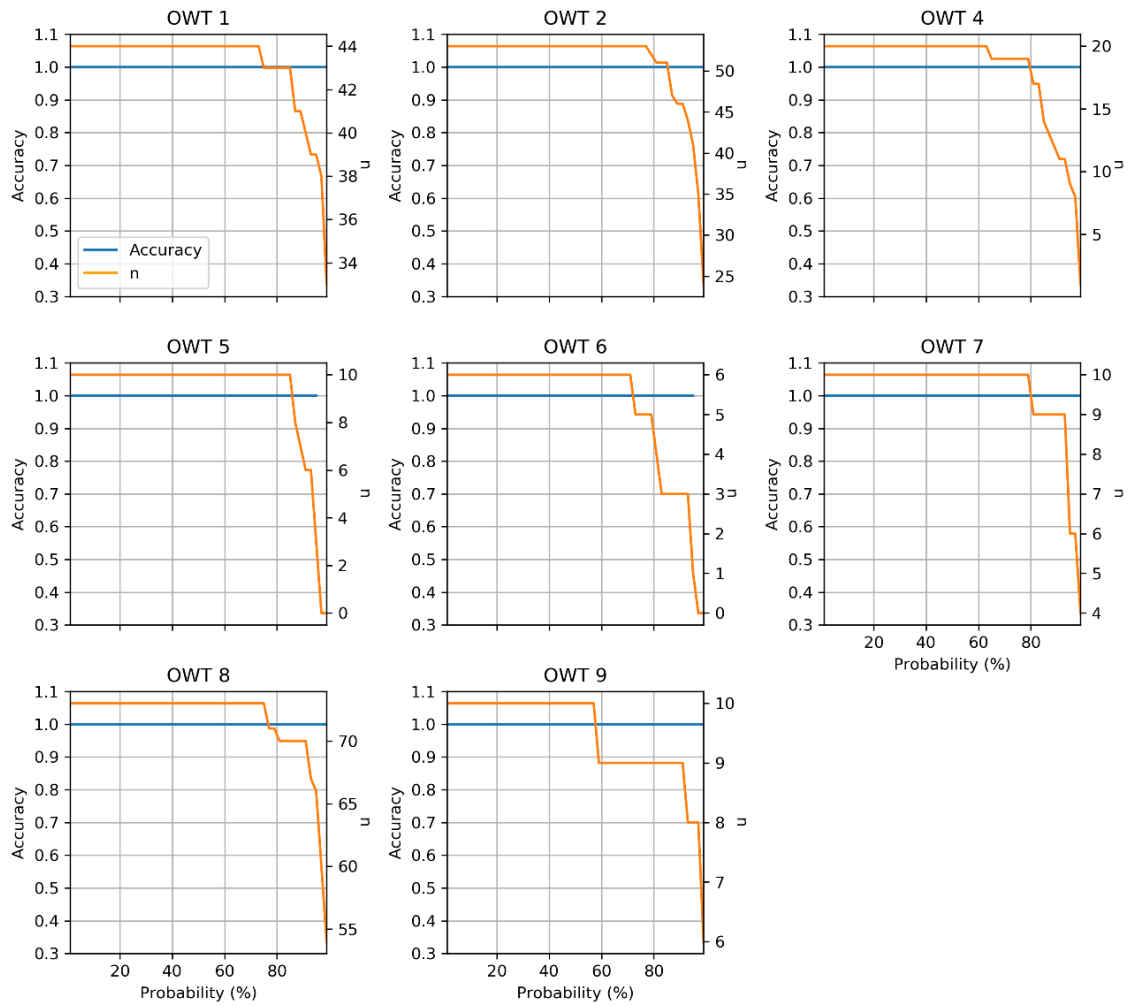
This section demonstrates how the accuracy of each class increases and the number of targets classified (n) reduces as the probability thresholds increase. Those estimates were done in the training dataset.

Figure A.1 - How accuracy and number of targets detected (n) change as the probability threshold increases, for each OWT in the MSI SVC algorithm.



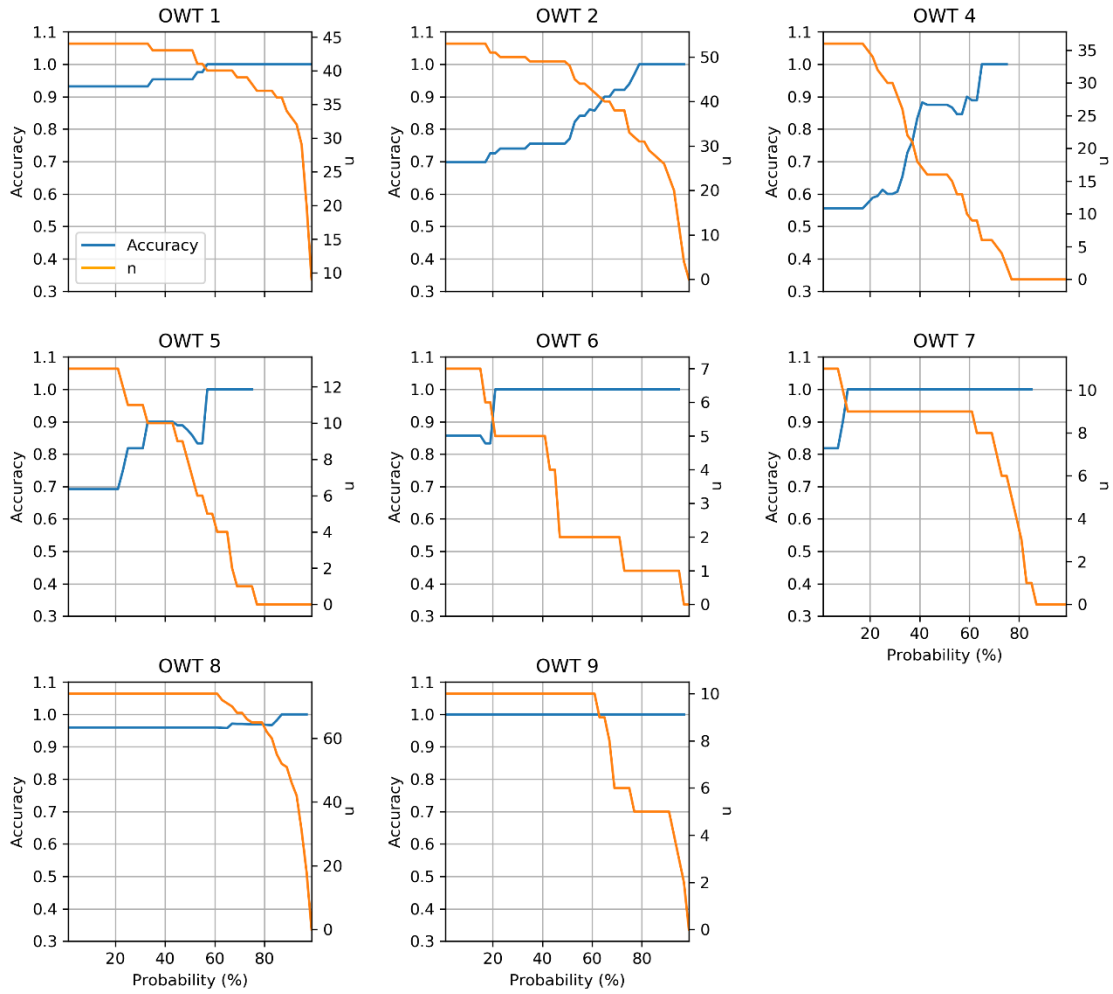
Source: The author.

Figure A.2 - How accuracy and number of targets detected (n) change as the probability threshold increases, for each OWT in the MSI RF algorithm.



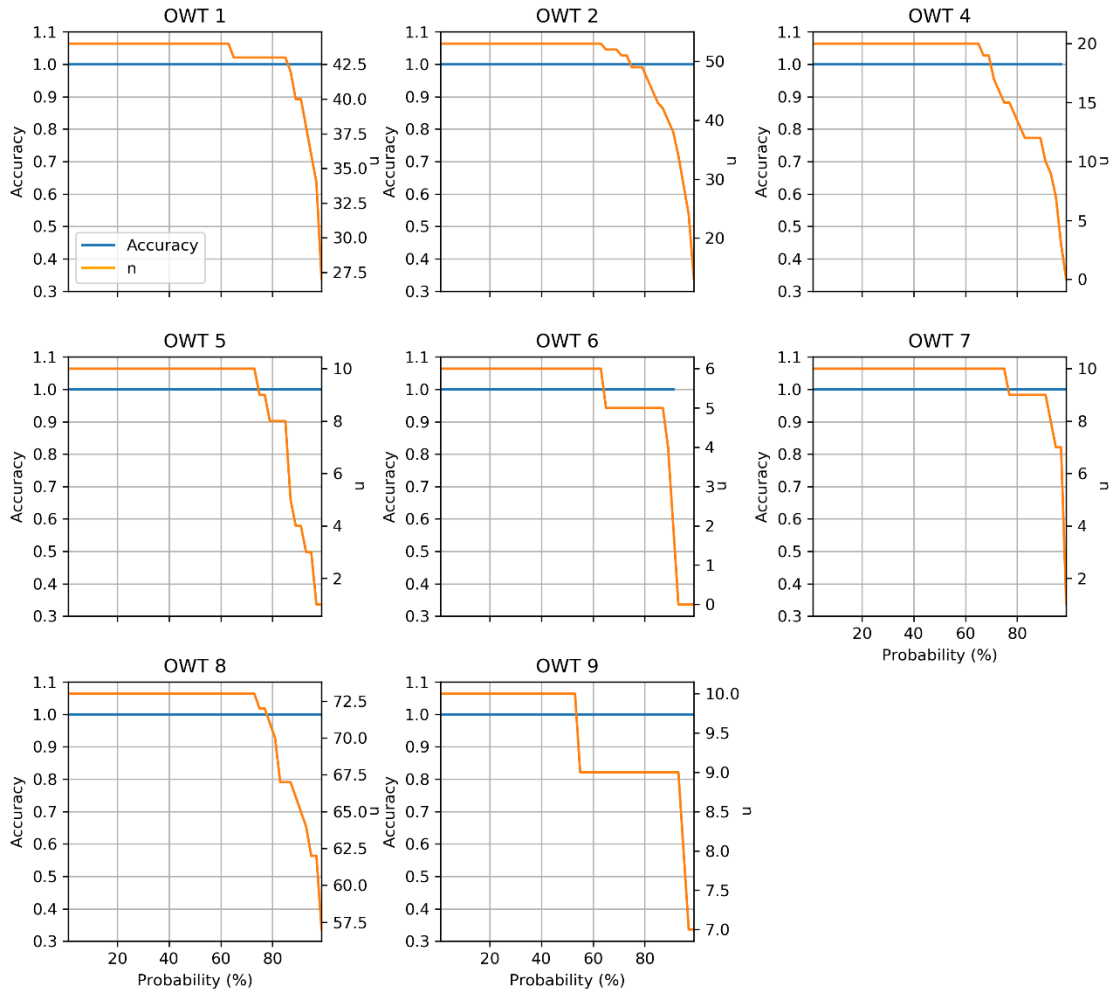
Source: The author.

Figure A.3 - How accuracy and number of targets detected (n) change as the probability threshold increases, for each OWT in the OLI SVC algorithm.



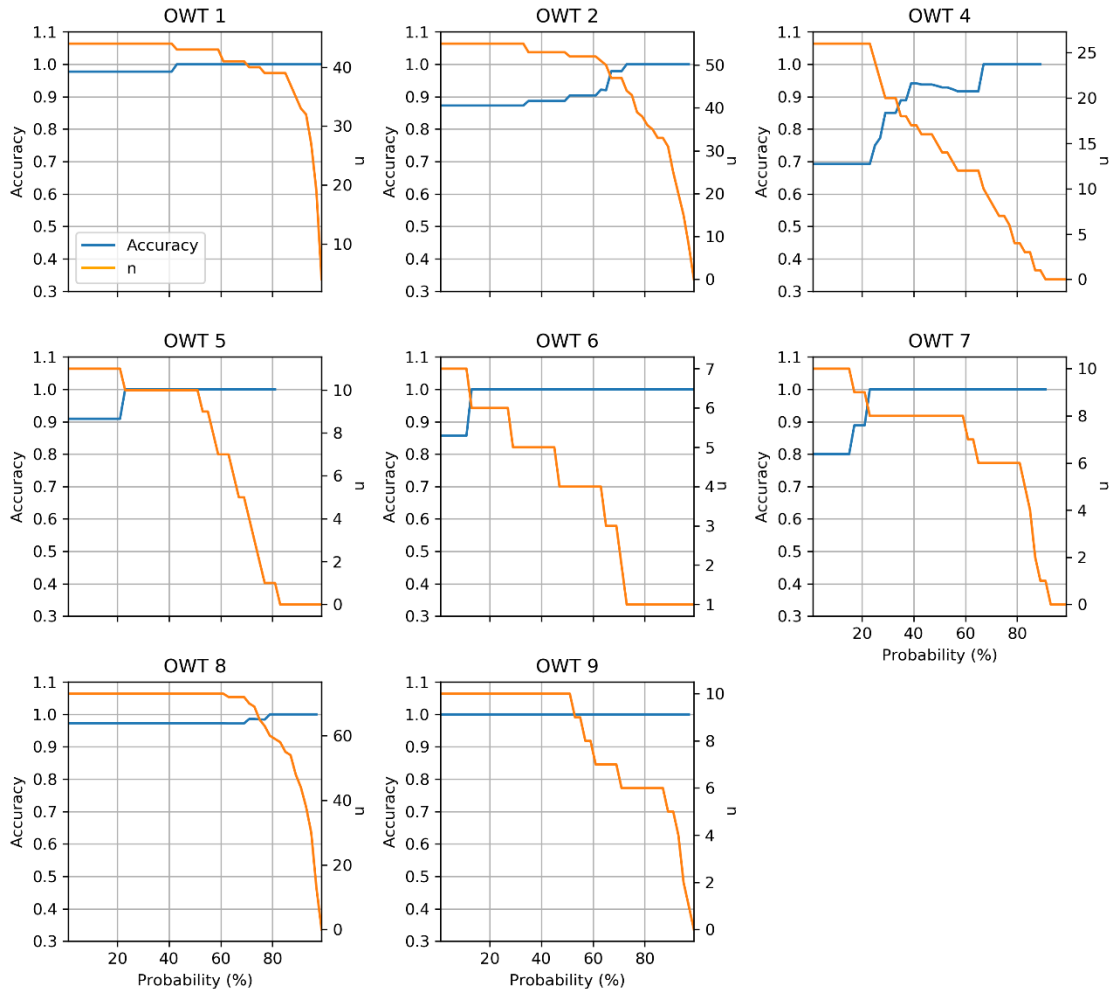
Source: The author.

Figure A.4 - How accuracy and number of targets detected (n) change as the probability threshold increases, for each OWT in the OLI RF algorithm.



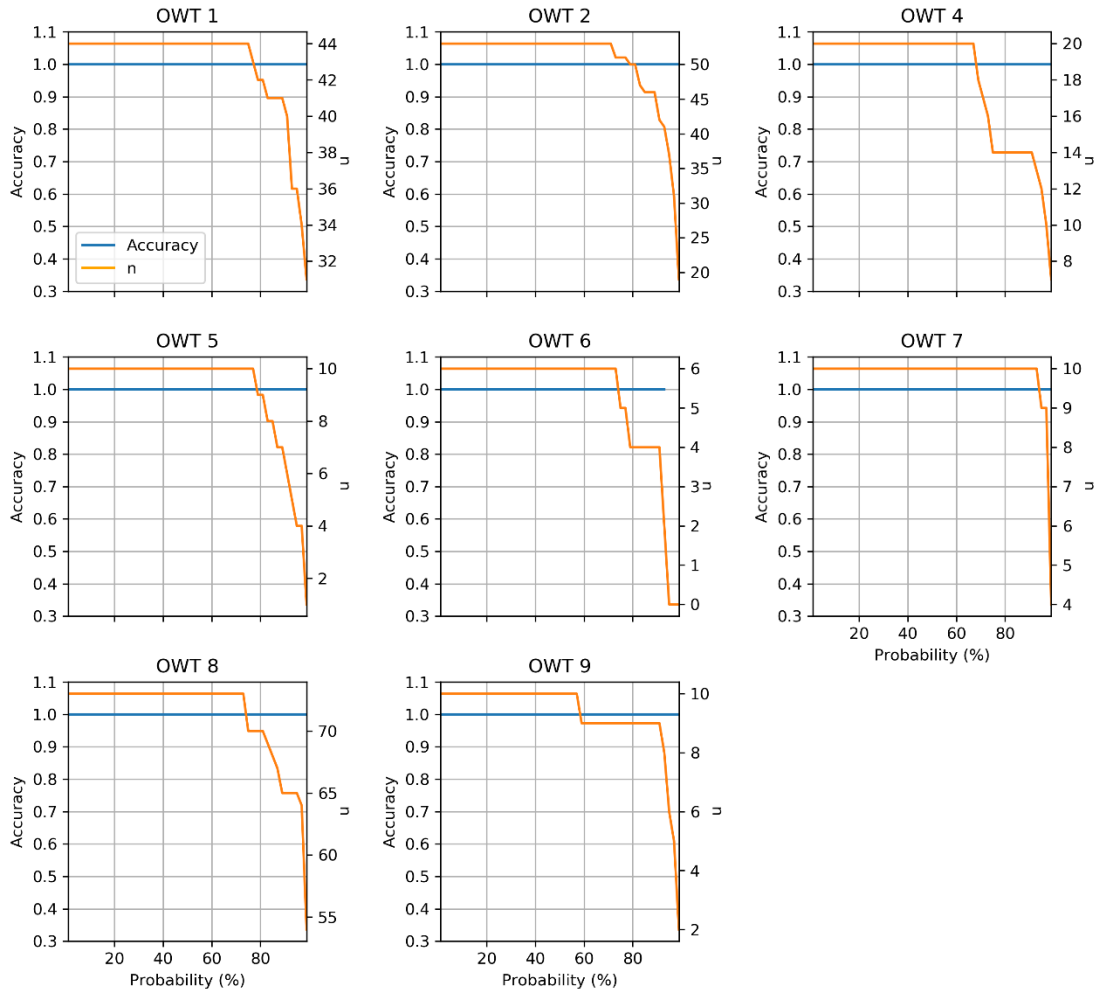
Source: The author.

Figure A.5 - How accuracy and number of targets detected (n) change as the probability threshold increases, for each OWT in the ETM+ SVC algorithm.



Source: The author.

Figure A.6 - How accuracy and number of targets detected (n) change as the probability threshold increases, for each OWT in the ETM+ RF algorithm.



Source: The author.