



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

sid.inpe.br/mtc-m21b/2016/02.05.11.51-TDI

**RE-PARAMETERISATION OF A THREE-BAND  
SEMI-EMPIRICAL ALGORITHM FOR  
C-PHYCOCYANIN-PIGMENT ESTIMATION IN AN  
OPTICALLY COMPLEX TROPICAL RESERVOIR**

Diogo de Jesus Amore

Master's Thesis for the Post-Graduate Course in Remote Sensing, advised by Dr. Jose Luiz Stech, approved in February 24, 2016.

URL of the original document:

<<http://urlib.net/8JMKD3MGP3W34P/3L598P2>>

INPE  
São José dos Campos  
2016

**PUBLISHED BY:**

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GB)

Serviço de Informação e Documentação (SID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

Fax: (012) 3208-6919

E-mail: pubtc@inpe.br

**COMMISSION OF BOARD OF PUBLISHING AND PRESERVATION  
OF INPE INTELLECTUAL PRODUCTION (DE/DIR-544):****Chairperson:**

Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação (CPG)

**Members:**

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (CST)

Dr. André de Castro Milone - Coordenação de Ciências Espaciais e Atmosféricas (CEA)

Dra. Carina de Barros Melo - Coordenação de Laboratórios Associados (CTE)

Dr. Evandro Marconi Rocco - Coordenação de Engenharia e Tecnologia Espacial (ETE)

Dr. Hermann Johann Heinrich Kux - Coordenação de Observação da Terra (OBT)

Dr. Marley Cavalcante de Lima Moscati - Centro de Previsão de Tempo e Estudos Climáticos (CPT)

Silvia Castro Marcelino - Serviço de Informação e Documentação (SID) **DIGITAL**

**LIBRARY:**

Dr. Gerald Jean Francis Banon

Clayton Martins Pereira - Serviço de Informação e Documentação (SID)

**DOCUMENT REVIEW:**

Simone Angélica Del Duca Barbedo - Serviço de Informação e Documentação (SID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SID)

**ELECTRONIC EDITING:**

Marcelo de Castro Pazos - Serviço de Informação e Documentação (SID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SID)



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

sid.inpe.br/mtc-m21b/2016/02.05.11.51-TDI

**RE-PARAMETERISATION OF A THREE-BAND  
SEMI-EMPIRICAL ALGORITHM FOR  
C-PHYCOCYANIN-PIGMENT ESTIMATION IN AN  
OPTICALLY COMPLEX TROPICAL RESERVOIR**

Diogo de Jesus Amore

Master's Thesis for the Post-Graduate Course in Remote Sensing, advised by Dr. Jose Luiz Stech, approved in February 24, 2016.

URL of the original document:

<<http://urlib.net/8JMKD3MGP3W34P/3L598P2>>

INPE  
São José dos Campos  
2016

Cataloging in Publication Data

---

Amore, Diogo de Jesus.

Am68p Re-parameterisation of a three-band semi-empirical algorithm for c-phycoyanin-pigment estimation in an optically complex tropical reservoir / Diogo de Jesus Amore. – São José dos Campos : INPE, 2016.

xxxii + 198 p. ; (sid.inpe.br/mtc-m21b/2016/02.05.11.51-TDI)

Dissertation (Master in Remote Sensing) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2016.

Guiding : Dr. José Luiz Stech.

1. Bio-optical modelling. 2. Remote Sensing. 3. Water quality. 4. C-phycoyanin. 5. Signal processing. I.Title.

CDU 528.81:556.55

---



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

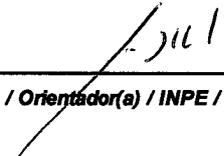
This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

Aluno (a): **Diogo de Jesus Amore**

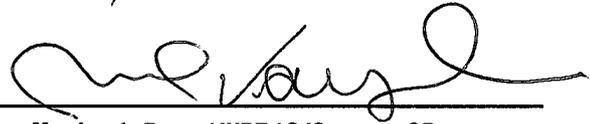
Título: " RE-PARAMETERISATION OF A THREE-BAND SEMI-EMPIRICAL ALGORITHM FOR C-PHYCOCYANIN-PIGMENT ESTIMATION IN AN OPTICALLY COMPLEX TROPICAL RESERVOIR".

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

Dr. José Luiz Stech

  
\_\_\_\_\_  
**Presidente / Orientador(a) / INPE / SJC Campos - SP**

Dr. Milton Kampel

  
\_\_\_\_\_  
**Membro da Banca / INPE / SJC Campos - SP**

Dr. Cláudio Clemente Faria Barbosa

  
\_\_\_\_\_  
**Membro da Banca / INPE / SJC Campos - SP**

Dr. Enner Herenio de Alcântara

  
\_\_\_\_\_  
**Convidado(a) / UNESP / São Paulo - SP**

**Este trabalho foi aprovado por:**

**maioria simples**

**unanimidade**

**São José dos Campos, 24 de Fevereiro de 2016**



*“Audaces fortuna iuvat”.*

*Virgil*

*“First they ignore you, then they laugh at you, then they fight you, then you win”.*

*Ghandi*

*“Better than the youngster's knowledge is the elderly's experience”.*

*Tibetan Proverb*

*“The planet does not need more successful people. But it does desperately need more peacemakers, healers, restorers, storytellers, and lovers of every kind. It needs people of moral courage willing to join the fight to make the world habitable and humane”.*

*David Orr*



*To my father, who would always silently  
smile while feeling proud.*



## **ACKNOWLEDGEMENTS**

I thank God for the ever-lasting opportunity to be better today than I was yesterday.

I am very grateful for all of those who contributed to the successful completion of this work. My most heartfelt appreciation goes to my advisor: Dr. José Luiz Stech, for believing in my potential and for the immeasurable support, liberty to do a work of my own, and utmost trust, crucial for the development of this thesis.

Acknowledgements must be given to the Brazilian National Foundation for the Undergraduates and Graduates Upgrade (CAPES), for the scholarship given through the Post-Graduation Program in Remote Sensing of INPE, and for funding my field campaign.

I should also like to acknowledge the SABESP and CETESB for allowing me clearance to enter Guarapiranga Reservoir, in São Paulo, SP.

I am thankful to INPE for allowing me to attend the Master's program in Remote Sensing and for rendering available its physical infrastructure and technical support. I must also address my gratitude to the professors of the Postgraduate Program in Remote Sensing for the knowledge and experience they shared. I am also especially grateful to the library team at INPE for their dedication and competence in helping me finding articles in several journals during the execution of this work.

My sincere and deep thanks go to the team of friends who helped me in my field campaign: Carlos Alberto Sampaio de Araújo, whose support in preparing the field work was crucial for the success of the endeavour, Joaquim Antônio Dionísio Leão, Gabriel Moiano, João Felipe Cardoso Santos, and Pétala Bianchi Augusto Silva, for the field work partnership. Also thanks to Marcelo Pedroso Curtarelli who throughout the past two years contributed immensely to

my work via many collaborative conversations, and thanks to Rafael Damiati Ferreira for help in understanding the fluorometer equipment details. A special thanks to Daniel Schaffer for uncountable theoretical and practical inputs towards to conclusion of this thesis. Lastly, a special thanks to Igor Ogashawara who contributed intellectually for my personal improvement in this field of research, and via suggesting possible topics for this thesis.

Also thanks to the Laboratório de Aerossóis, Soluções Aquosas e Tecnologias - LAQUATEC, INPE, for making its facilities available for the water quality analyses.

Special words of thanks also go to my office mates at INPE: Arley Ferreira de Souza, and Marcelo Pedroso Curtarelli for the laughs, talks and work during the last 2 years. I should also acknowledge the help, moral support and companionship of my friends at PGSER

Last but not the least, I thank Alana Toniol, my partner, who has been by my side throughout the most difficult emotional moments, and my deepest and greatest thanks go to my family who indirectly always respected my scientific endeavours, and above all else, thanks to my parents who somehow always managed to provide me with the maximum support possible to achieve my dreams, and my father who would feel proud to know of my constant attempt to get better.

## ABSTRACT

The eutrophication of aquatic systems is a worldwide environmental problem. A major aftermath is health-inflicting toxic algal bloom, which can affect humans. Therefore, aquatic systems, mostly near urban environments require environmental monitoring. The use of remote sensing for monitoring algal blooms via bio-optical modelling is based on the spectral behaviour of the optically active components (OACs) in the water to estimate their concentrations. The detection of cyanobacteria, one of the main phyla of harmful algae, takes place via the identification of a unique pigment in inland waters cyanobacteria, the phycocyanin (PC). Remote sensing techniques, such as semi-empirical algorithms - a sort of bio-optical model - have been used to estimate PC concentration in aquatic systems using *in situ* hyperspectral data and satellite multispectral data. However, there is a lack in scientific works tackling PC prediction in tropical inland waters bearing PC in low concentration such as in city-supplying Guarapiranga reservoir at the southwestern region of São Paulo city. This is mostly true because scientific studies attempt to generate models based on bloom events. However, much uncertainty is associated with models results at the low concentration ranges. Therefore the goal of this research was to evaluate the re-parameterization of a semi-empirical algorithm for a tropical oligo-mesotrophic inland water. Radiometric, fluorometric, limnological, and multi-parameter sonde data were collected in Guarapiranga Reservoir, located. This thesis presents the findings which led to the algorithm re-parameterisation. Results showed that the calibration dataset (n=15) improved PC prediction  $R^2$  by 15.3% after the re-parameterisation; and for the validation dataset (n=19), PC prediction  $R^2$  was improved by 4.79%. NRMSE for the calibration dataset was bettered by 1.76%; and it was almost equalised for the validation dataset (differed by 0.19%). The new re-parameterisation correlation coefficient developed in this study presented a better  $R^2$  (68%) than that of the original algorithm (46%). These correlations linked the band ratios used as enhancing coefficients to known PC spectral features. The bio-optical, radiometric, and water quality characterisation of Guarapiranga reservoir, and the evaluation of signal processing techniques of radiometric data yielded results that supported the generation of the new re-parameterisation coefficient. Such results were related to features in the blue-to-green spectral region capable of improving PC prediction. Uncertainties in the estimations are mainly due to the lack of *in situ* data. The re-parameterization was also considered for a synthetic dataset of the Ocean & Land Colour Imager (OLCI) sensor/Sentinel 3. The simulation of OLCI data was conducted using its spectral response function, and it was important because of its potential use in environmental monitoring. Overall results were encouraging, however, further studies are suggested to further validate this new algorithm. Nevertheless, the development of a semi-empirical algorithm for low-concentration PC prediction in tropical inland waters is an important step for the development of an ever-improving robust tool for water quality monitoring.

Keywords: Bio-optical modelling. Remote Sensing. Water Quality. Cyanobacteria. C-Phycocyanin. Signal Processing

# REPARAMETRIZAÇÃO DE UM ALGORITMO SEMI-EMPÍRICO DE TRÊS BANDAS PARA A ESTIMAÇÃO DO PIGMENTO C-FICOCIANINA DE UM RESERVATÓRIO TROPICAL OPTICAMENTE COMPLEXO

## RESUMO

A eutrofização de sistemas aquáticos é um problema ambiental global. Uma consequência de alta importância disso é a floração de algas tóxicas a qual pode afetar os seres humanos. Portanto, sistemas aquáticos, principalmente aqueles próximos a centros urbanos requerem monitoramento ambiental. A utilização de sensoriamento remoto para o monitoramento de florações de algas através da modelagem bio-óptica é baseada no comportamento espectral dos componentes opticamente ativos (OACs) da água, de modo a estimar suas concentrações. A detecção de cianobactérias, um dos filos mais importantes dentre as algas tóxicas, ocorre através da identificação de um pigmento nestes micro-organismos de águas interiores denominado, ficocianina (PC). Técnicas de sensoriamento remoto, tais como algoritmos semiempíricos – um tipo de modelo bio-óptico – tem sido utilizado para estimar concentrações de PC em sistemas aquáticos utilizando dados hiperespectrais *in situ* e dados de satélite multiespectrais. Entretanto, há uma deficiência em trabalhos científicos abordando predição de PC em águas interiores tropicais com baixas concentrações de PC tais como o reservatório de Guarapiranga, no sudoeste da capital de São Paulo. Isto é uma realidade, pois estudos científicos tendem a gerar modelos baseados em eventos de florações em que há maiores concentrações de PC. Porém, muita incerteza está associada a resultados de modelos em baixas concentrações de PC. Portanto, o objetivo deste trabalho foi avaliar a reparametrização de um algoritmo semiempírico para uma água interior oligo-mesotrófica tropical. Dados radiométricos, fluorométricos, limnológicos, e de uma sonda multiparâmetro foram coletados no reservatório de Guarapiranga. Esta dissertação apresenta os resultados que levaram à reparametrização do algoritmo. Os resultados mostraram que o  $R^2$  do conjunto de dados de calibração ( $n=15$ ) melhorou a predição de PC em 15.3% após a reparametrização; e o  $R^2$  do conjunto de dados de validação ( $n=19$ ), melhorou a predição em 4.79%. O NRMSE para o conjunto de dados de calibração foi melhorado em 1.76%; e os NRMSEs da validação praticamente estabilizaram (diferenciando-se em 0.19%). O coeficiente de correlação desenvolvido para a nova reparametrização neste estudo apresentou um  $R^2$  (68%) melhor do que o coeficiente do algoritmo original ( $R^2 = 46\%$ ) com o qual foi comparado. Estas correlações associam razões de bandas, utilizadas como coeficientes de

melhoramento, a características espectrais de PC já conhecidas. A caracterização bio-óptica, radiométrica, e de qualidade da água do reservatório, bem como a avaliação de técnicas de processamento de sinais dos dados radiométricos geraram resultados que apoiaram a geração do coeficiente de reparametrização. Tais resultados estiveram relacionados a características nas regiões espectrais do azul-verde como sendo capazes de melhorar a predição de PC. Incertezas nas estimações são essencialmente devidas à falta de dados *in situ*. A reparametrização também foi considerada para um conjunto de dados sintético do sensor *Ocean & Land Colour Imager (OLCI)/Sentinel3*. A simulação dos dados do OLCI foi realizada utilizando a sua função de resposta espectral, e foi importante por causa do seu uso potencial em monitoramento ambiental. De maneira geral os resultados foram encorajadores, entretanto, estudos futuros são sugeridos para uma validação mais robusta deste novo algoritmo. Contudo, o desenvolvimento de um algoritmo semiempírico para a predição de PC em baixas concentrações em águas interiores tropicais é um passo importante para o desenvolvimento de um monitoramento de qualidade da água passível de ser melhorado continuamente.

## LIST OF FIGURES

|  | <u>Page</u> |
|--|-------------|
| Figure 2.1 – Light as a function of depth in different trophic states temperate lakes. ....  | 9           |
| Figure 2.2 – Schematic of light entering water, where it can be reflected back, scatter off of a particle, or be absorbed in the water column .....  | 10          |
| Figure 2.3 – Typical spectral regions investigated in bio-optical models:<br><i>mia<sub>highpig</sub></i> + <i>bb<sub>spm</sub></i> , as minimum absorption by pigments; <i>a<sub>cya</sub></i> , as absorption by phycocyanin-containing cyanobacteria; <i>flu<sub>chl</sub>*</i> , as fluorescence by chlorophyll; <i>red.a<sub>chl</sub></i> , as red chlorophyll absorption trough; <i>ref<sub>n,chl+cell</sub></i> , as reflectance from high chlorophyll concentration and phytoplankton cell structure; <i>abs<sub>w</sub></i> as high absorption by water at longer wavelengths; <i>ref<sub>spm</sub></i> , as reflectance by other particulate matter. .... | 14          |
| Figure 2.4 – Flowchart of bio-optical models classification.....   | 16          |
| Figure 3.1 – Study Area Location. Sampling stations are depicted as red dots. Local bathymetry is in shaded blue. Surrounding altitude is expressed in green-brown shades. Bathymetry shown in this figure was acquired in this study. Altitude data is from Shuttle Radar Topography Mission (SRTM)..   | 29          |
| Figure 3.2 – Local measurements buoys utilised by SABESP for <i>in situ</i> limnological data collection.....  | 32          |
| Figure 4.1– Sonde YSI 6600 in operation.....   | 35          |
| Figure 4.2– Two different ways to collect PC data via fluorometric approach. The continuous flow requires the use of a pump and hoses. And the discrete analysis simply uses individual vials for each sample measuring the raw signal value displayed by the equipment after the appropriate calibration. ....  | 38          |
| Figure 4.3 – AC-S components’ illustrative scheme. ....  | 40          |
| Figure 4.4 – AC-S instrument in operation. ....  | 40          |
| Figura 4.5 – Map showing sampling stations with local depth measured <i>in situ</i> ; showing the depth points estimated from both the drainage basin and from linear estimation for the areas between the basin lines and the reservoir border; and showing the basin lines themselves.....   | 44          |
| Figure 4.6 – Main pigments percent absorption spectra. Highlighting the absorption range associated with phycocyanin.....  | 46          |

|   |    |
|---|----|
| Figure 4.7 – Fluorometer adjustment curve between fluorometer PC standard raw values and known concentration for that reading for the estimation of values collected during field campaign.....   | 47 |
| Figure 4.8 – Generalised water quality parameters plot for Coos Bay Estuary, Oregon. CTD measurement, April, 2008. Measurements depth range from 50 to 2800 m. ....   | 48 |
| Figure 4.9 – Laboratory-measured chl-a regression as a function of Sonde chl-a fluorometric sensor raw values for the first 30-cm subsurface depth.....   | 49 |
| Figure 4.10 – Radiometric data collection sensor set-up: (a) The boat was kept aligned with solar direct beam direction as to keep above-water sensors orthogonal to these beams. (b) Downwelling Irradiance ( $E_d$ ), Upwelling Radiance ( $L_u$ ), Upwelling Irradiance ( $E_u$ ). (c) Total Water Radiance ( $L_t$ ), Diffuse Sky Radiance ( $L_{sky}$ ), Global Solar Irradiance ( $E_s$ ). (d) Example of underwater measurement. ....  | 51 |
| Figure 4.11 – Multi resolution analysis subsampling characteristics: (a) Scaling function: the subspaces spanned by the scaling function at low scales are nested within those spanned at higher scales. (b) Wavelet function: fill up the gap of any two adjacent scaling subspaces. (c) How an infinite set of wavelets is replaced by one low-pass spectrum approximation scaling function <i>cork</i> in order to avoid covering the spectrum all the way to zero with detailed wavelet spectra (MALLAT, 1989)..... | 54 |
| Figure 4.12 – Remote Sensing Reflectance ( $R_{rs}$ ) for 15 sampling stations. Stations 1 and 6 were excluded from this calculation due to faulty equipment at the measurement site. ....  | 64 |
| Figure 4.13 – $R_{rs}$ spectra from Funil reservoir collected in May, 2012 (AUGUSTO-SILVA et al., 2014).....  | 68 |
| Figure 4.14 – Spectral response functions for OLCI sensor. Twenty one bands over the visible-NIR spectral range (ESA, 2016).....  | 69 |
| Figure 4.15 – Methodological flowchart for this study. Square boxes indicate datasets collected or generated for this work. Circular or ellipsoidal shapes represent the objectives-related steps within the methodology. Rhombus represents the main statistics applied to the dataset. Flags represent specific aims and main goal. ....  | 71 |
| Figure 5.1 – Topological bathymetric plot for this study’s individual sampling stations. Grey bars represent the euphotic zone depth where the radiation reaching the water surface extinguishes at the 1% level. The white bars  |    |

|   |    |
|---|----|
| represent the difference between the $Z_{eu1}$ and the actual local reservoir depth. ....   | 73 |
| Figure 5.2 – Guarapiranga north-south bathymetry cross-section. Leftmost part of figure represents the southernmost area, and the rightmost part, the northernmost. Different shades of purple represent the various regions along the cross section where the reservoir has more than one depth variation for that latitude. ....  | 76 |
| Figure 5.3 – Guarapiranga west-east bathymetry cross-section. Leftmost part of figure represents the westernmost area, and the rightmost part, the easternmost. Different shades of purple represent the various regions along the cross section where the reservoir has more than one depth variation for that longitude. ....   | 77 |
| Figure 5.4 – (a) Regression analysis between secchi depth and one-per-cent euphotic zone depth $Z_{eu1}$ . (b) Secchi depth and $Z_{eu1}$ ratio for all points with standard error bars. Trend line indicates the average ratio value of 48.47%. ....   | 78 |
| Figure 5.5 – Irradiance extinction depth per unit wavelength and integrated over the PAR region per sampling point. White line represents the reciprocal of averaged $K_d$ per unit wavelength for all irradiance-measured depths. Red line represents the irradiance per unit wavelength extinguished by one per cent of above-water ( $0^+$ ) irradiance values. Yellow line represents the integrated $Z_{eu1}$ values. .... | 80 |
| Figure 5.6 – Normalised remote sensing reflectance ( $nR_{rs}$ ) spectra at 550 nm for fifteen samples. ....  | 82 |
| Figure 5.7 – Limnological water quality parameters boxplots for all samples....   | 84 |
| Figure 5.8 – Kendall (a), Spearman (b), and Pearson (c) correlation coefficients for limnological water quality parameters. $NH_4^+$ was removed from this figure due to difference in sample numbers. ....   | 88 |
| Figure 5.8 – Kendall (a), Spearman (b), and Pearson (c) correlation coefficients for limnological water quality parameters. $NH_4^+$ was removed from this figure due to difference in sample numbers. – Continuation .....   | 89 |
| Figure 5.9 – Kendall, Spearman, and Pearson correlation coefficients for $NH_4^+$ against all other parameters. $NH_4^+$ had three inappropriate samples; therefore, the whole dataset was adjusted to the $NH_4^+$ sample range (fourteen rather than seventeen samples). ....   | 89 |
| Figure 5.10 – Laboratory-calculated CDOM (a), total (b), phytoplankton (c), and non-algal particles (tripton) (d) absorption coefficients.....  | 92 |

- Figure 5.11 – Ternary plot representing the relative absorption coefficients for phytoplankton ( $a_{phy}$ ), tripton ( $a_{nap}$ ), and CDOM ( $a_{CDOM}$ ). Wavelengths (nm) represent the ones most used in bio-optical algorithms, imaging satellite sensors bands, and OAC-related minimum/maximum absorption bands. 93
- Figure 5.12 – ACS-measured total absorption coefficient per unit depth for fifteen sampling stations. Individual figures for each station are standardised by the point with maximum depth (y-axis) and absorption magnitude (x-axis). Stations with fewer measurements represent points with shorter maximum depth. Data colours are linked to their respective spectrum wavelength colour, black being the NIR range. Two sampling stations failed to appropriately collect data and were discarded. ....99
- Figure 5.13 – ACS-measured total attenuation coefficient per unit depth for fifteen sampling stations. Individual figures for each station are standardised by point with maximum depth (y-axis). P9, P15, and P16 absorption magnitudes have their own plot ranges, and all other points' absorption magnitudes are standardised by P14's magnitude range. Stations with fewer measurements represent points with shorter maximum depth. Data colours are linked to their respective spectrum wavelength colour, black being the NIR range. Two sampling stations failed to appropriately collect data and were discarded. ....100
- Figure 5.14 – ACS-measured total absorption coefficient per unit wavelength for fifteen sampling stations. Individual figures for each station are standardised by the wavelength (x-axis) range (400-750 nm), and by  $10 \text{ m}^{-1}$  absorption magnitude. Data colours are linked to the different correction methods applied to the dataset. Two sampling stations (P1 and P12) failed to appropriately collect data and were discarded. ....101
- Figure 5.15 – Dissolved Oxygen (DO) versus chlorophyll-a plot over depth for fourteen sampling stations. Red lines represent DO data, and green lines represent chlorophyll data. Black straight line across depth axis represents the one-per-cent euphotic zone depth for that point. Missing sampling points were excluded from data plotting due to faulty *in situ* measurements. ....107
- Figure 5.16 – Temperature versus conductivity plot over depth for fifteen sampling stations. Maroon lines represent temperature data, and blue lines represent conductivity data. Black straight line across depth axis represents the one-per-cent euphotic zone depth for that point. Missing sampling points were excluded from data plotting due to faulty *in situ* measurements. ....108

|  |     |
|--|-----|
| Figure 5.17 – Regression analysis plots for $a_{CDOM}(440)$ as a function of dissolved organic matter (DOC) concentrations.....  | 111 |
| Figure 5.18 – Regression analysis plots for (a) $b_{bp}(555)$ and (b) $a_{phy}(620)$ as a function of phycocyanin (PC) concentrations. ....  | 112 |
| Figure 5.19 – Regression analysis plots for (a) $a_{phy}(620)$ as a function of chlorophyll-a (chl-a) plus phycocyanin (PC) concentrations, and (b) $b_{bp}(555)$ as a function of only chl-a. ....  | 113 |
| Figure 5.20 – Regression analysis plots for (a) $a_{CDOM}(440)$ and (b) $a_{phy}(440)$ as a function of chlorophyll-a (chl-a) concentrations.....  | 114 |
| Figure 5.21 – Regression analysis plots for (a) $a_{phy}(675)$ as a function of chlorophyll-a concentrations, and (b) $a_{nap}(440)$ as a function of total suspended inorganics (TSI) concentrations. ....  | 115 |
| Figure 5.22 – Regression analysis plots for (a) $b_{bp}(555)$ and (b) $a_{CDOM}(440)$ as a function of total suspended matter (TSM) concentrations.....  | 116 |
| Figure 5.23 – Haar wavelet clustering analysis dendrogram for fifteen sampling points.....   | 122 |
| Figure 5.24 – Discrete Meyer wavelet clustering analysis dendrogram for fifteen sampling points. ....  | 123 |
| Figure 5.25 – Reference spectra generated from WASI software used for spectral angle mapping (SAM) classification and spectral angle errors for fifteen sampling stations correlated with each reference class. The redder (smaller) the error value the more correlated the station spectrum is to the reference class. Reddest values represent the reference class to which the station spectrum was allocated during the SAM classification..... | 127 |
| Figure 5.26 – Chlorophyll-a (chl-a), phycocyanin (PC), and chl-a to total suspended matter (TSM) ratio bio-optical algorithms evaluation. Statistics for the correlations are given by Pearson-coefficient (R) for predicted against measured values. ....   | 130 |
| Figure 5.27 – PC to chl-a ratio (PC:chl-a) monotonically increasing for all sampling stations from this study’s dataset and from Ogashawara et al. (2013). Horizontal line at the ratio value of one indicates how many samples have higher concentrations of PC (above line) or chl-a (beneath line).....   | 134 |
| Figure 5.28 – Scatterplot between a reflectance band ratio [ $R_{rs}(620):R_{rs}(709)$ ] and measured PC concentration.....  | 135 |

|  |     |
|--|-----|
| Figure 5.29 – Remote sensing reflectance ( $R_{rs}$ ) spectrum for sampling station three, slope one (S1), and phycocyanin line height (PLH). The coefficient of proportionality represents the correlation between S1 and PLH for all sampling points. The slope between $PLH$ and $S1$ (Figure 30a) represents the re-parameterisation coefficient. .... | 135 |
| Figure 5.30 – (a) Regression analysis between phycocyanin (PC) line height (PLH) and slope one (S1). (b) Sensitivity plot between PLH, S1 and PC concentration. ....   | 136 |
| Figure 5.31 – Calibration (this study’s dataset) and validation (dataset from Augusto-Silva, 2014) of the adjusted MI14 algorithm, and comparison with the original MI14 algorithm. Brackets indicate 95% confidence level. ....   | 139 |
| Figure 5.32 – Remote sensing reflectance ( $R_{rs}$ ) (red line) for each sampling station calculated from <i>in situ</i> -collected radiance and irradiance data via Ramses TRioS; and resampled $R_{rs}$ data to the first sixteen OLCI sensor bands (400- 779 nm) (blue line). ....   | 142 |
| Figure A.1– List of some bio-optical models found in literature covering from empirical to QAA algorithms.....   | 168 |
| Figure B.1– Haar and Discrete Meyer wavelets scaling and wavelet functions .....   | 170 |
| Figure C.1– Water Quality Parameters Cross Correlation Scatterplots and histogram.....   | 172 |
| Figure E.1 – TLS scatterplots for laboratory-calculated, ACS-measured, and QAA-estimated absorption coefficients .....   | 178 |
| Figure E.2 – TLS scatterplots for laboratory-calculated, ACS-measured, and QAA-estimated absorption coefficients .....   | 179 |
| Figure E.3 – TLS scatterplots for laboratory-calculated, ACS-measured, and QAA-estimated absorption coefficients .....   | 180 |
| Figure E.4 – TLS scatterplots for laboratory-calculated, ACS-measured, and QAA-estimated absorption coefficients .....   | 181 |
| Figure E.5 – TLS scatterplots for laboratory-calculated, ACS-measured, and QAA-estimated absorption coefficients .....   | 182 |
| Figure F.1 – DWTCA-Clustered $R_{rs}$ Spectra for signal decomposition level (Haar wavelet). ....  | 184 |
| Figure F.2 – DWTCA-Clustered $R_{rs}$ Spectra for decomposition level one (Haar wavelet). ....   | 185 |

|   |     |
|---|-----|
| Figure F.3 – DWTCA-Clustered $R_{rs}$ Spectra for decomposition level two (Haar wavelet).....   | 186 |
| Figure F.4 – DWTCA-Clustered $R_{rs}$ Spectra for decomposition level three (Haar wavelet)..... | 187 |
| Figure F.5 – DWTCA-Clustered $R_{rs}$ Spectra for decomposition level four (Haar wavelet).....  | 188 |
| Figure F.6 – DWTCA-Clustered $R_{rs}$ Spectra for decomposition level five (Haar wavelet).....  | 189 |
| Figure F.7 – DWTCA-Clustered $R_{rs}$ Spectra for decomposition level six (Haar wavelet).....   | 190 |
| Figure F.8 – DWTCA-Clustered $R_{rs}$ Spectra for decomposition level seven (Haar wavelet)..... | 191 |
| Figure F.9 – DWTCA-Clustered $R_{rs}$ Spectra for decomposition level eight (Haar wavelet)..... | 192 |



## LIST OF TABLES

|  | <u>Page</u> |
|--|-------------|
| Table 2.1 - IOPs, AOPs and Radiometric Quantities commonly used in optical hydrology .....   | 11          |
| Table 2.2 – Specifications of the OLCI on the Sentinel-3 satellite system; shaded areas are the ones that were included from MERIS specifications .....  | 28          |
| Table 3.2 – Location of sampling stations within Guarapiranga reservoir .....  | 31          |
| Table 4.1 – Available wavelets within Matlab <sup>®</sup> . Highlighted are the only wavelets suitable for this study’s discrete analysis. ....  | 53          |
| Table 4.2 – OACs concentration magnitude within each spectral class generated by WASI. ....  | 57          |
| Table 4.3 – Summary of $R_{rs}$ -based bio-optical algorithms used in this study. ....   | 65          |
| Table 4.4 –Summary of error estimator, correlation coefficients, and estimation curve robustness used in this study. ....  | 66          |
| Table 5.1 – Principal-component (PCt) fitting total least squares regression sum of squared errors and root mean squared error (RMSE) for PCts one, two, and three for each wavelength and absorption correction method. PCts explanatory variances for each wavelength and correction method are also presented. .... | 104         |
| Table 5.2 – Groups allocated by the discrete wavelet transform clustering analysis for each sampling station for each decomposition level for both Haar and Discrete Meyer wavelets. ....  | 118         |
| Table 5.3 – Haar wavelet cophenetic and inconsistency coefficients.....  | 122         |
| Table 5.4 – Discrete Meyer cophenetic and inconsistency coefficients.....  | 123         |
| Table 5.5 – Spectral angle error from spectral angle mapping (SAM) analysis for each sampling point (column-title) against reference classes (row-title) generated in WASI. Bold numbers represent to which class each sampling point has been allocated. ....   | 126         |
| Table D.1 - Kendall correlation coefficients for the limnological dataset .....  | 174         |
| Table D.2 - Kendall p-value for the limnological dataset .....   | 174         |
| Table D.2 - Kendall p-value for the limnological dataset - Continuation .....  | 175         |
| Table D.3 – Spearman correlation coefficient for the limnological dataset.....   | 175         |

|  |     |
|--|-----|
| Table D.4 - Spearman p-value for the limnological dataset.....                                   | 176 |
| Table D.5 – Pearson correlation coefficient for the limnological dataset.....                    | 176 |
| Table D.5 - Pearson correlation coefficient for the limnological dataset –<br>Continuation ..... | 177 |
| Table D.6 - Pearson p-value for the limnological dataset.....                                    | 177 |

## LIST OF ACRONYMS AND ABBREVIATIONS

|               |   |
|---------------|---|
| <i>a</i>      | Absorption coefficient                                      |
| ADEOS         | Advanced Earth Observing Satellite                          |
| AISA Eagle    | Airborne Imaging Spectrometer for Applications Eagle        |
| AOPs          | Apparent Optical Properties                                 |
| ARIES         | Australian Resource Information and Environmental Satellite |
| ASI           | Italian Space Agency  |
| <i>b</i>      | Scattering coefficient                                      |
| BUS           | Bottom-Up Strategy  |
| <i>c</i>      | Attenuation coefficient                                     |
| <i>ca</i>     | Coefficient of Approximation                                |
| <i>cd</i>     | Coefficient of Detail                                       |
| CASI-2        | Compact Airborne Spectrographic Imager-2                    |
| CDM           | Coloured Detrital Matter                                    |
| CDOM          | Coloured Dissolved Organic Matter                           |
| CETESB        | <i>Companhia de Tecnologia de Saneamento Ambiental</i>      |
| CHABs         | Cyanobacterial Harmful Algal Blooms                         |
| Chl- <i>a</i> | Chlorophyll- <i>a</i>                                       |
| CHRIS         | Compact High Resolution Imaging Spectrometer                |
| CI            | Cyanobacteria Index   |
| COD           | Chemical Oxygen Demand                                      |
| CZCS          | Coastal Zone Color Scanner                                  |
| DIC           | Dissolved Inorganic Carbon                                  |
| DO            | Dissolved Oxygen  |
| DOC           | Dissolved Organic Carbon                                    |
| DWTCA         | Discrete Wavelet Transform Clustering Analysis              |
| EMR           | Electromagnetic Radiation                                   |
| EnMAP         | Environmental Mapping and Analysis Program                  |
| ENVISAT       | Earth Observing Satellite                                   |
| ERTS-1        | Earth Resources Technology Satellite                        |
| ESA           | European Space Agency                                       |
| ETM+          | Enhanced Thematic Mapper Plus                               |

|                              |   |
|------------------------------|---|
| FOV                          | Field Of View                                       |
| GA-PLS                       | Genetic Algorithm and Partial Least Squares         |
| HABs                         | Harmful Algal Blooms                                |
| HICO                         | Hyperspectral Imager for the Coastal Ocean          |
| HyspIRI                      | Hyperspectral Infra-red Imager                      |
| INPE                         | National Institute for Space Research               |
| IOPs                         | Inherent Optical Properties                         |
| LED                          | Light-Emitting Diode                                |
| LMI                          | Linear Matrix Inversion                             |
| MERIS                        | Medium Resolution Imaging Spectrometer              |
| MODIS                        | Moderate Resolution Imaging Spectroradiometer       |
| NaClO                        | Sodium Hypochlorite                                 |
| NAP                          | Non-Algal Particles                                 |
| NASA                         | National Aeronautics and Space Administration       |
| NDPCI                        | Normalized Difference Phycocyanin Index             |
| NH <sub>4</sub> <sup>+</sup> | Ammonium cation                                     |
| NIR                          | Near Infrared                                       |
| NRMSE                        | Normalized Root Mean Square Error                   |
| OACs                         | Optically Active Components                         |
| OCW                          | Optically Complex Waters                            |
| OBM                          | Optimal Band ratio Modelling                        |
| OCM                          | Ocean Colour Monitor                                |
| OLCI                         | Ocean & Land Color Imager                           |
| PAR                          | Photosynthetic Active Radiation                     |
| PCA                          | Principal Component Analysis                        |
| PIXEL                        | Picture Element                                     |
| PC                           | C-Phycocyanin pigment                               |
| Pct                          | Principal Component                                 |
| PM                           | Particulate Matter                                  |
| PML                          | Plymouth Marine Laboratory                          |
| PPMCC                        | Pearson Product Moment Correlation Coefficient      |
| PRISMA                       | PREcursore IperSpettrale della Missione Applicativa |

|           |   |
|-----------|---|
| QAA       | Quasi-Analytical Algorithm  |
| $R$       | Irradiance Reflectance Ratio  |
| RMSE      | Root Mean Square Error  |
| $R_{rs}$  | Remote Sensing Reflectance from above surface                           |
| $r_{rs}$  | Remote Sensing Reflectance from subsurface                              |
| SABESP    | <i>Companhia de Saneamento Básico do Estado de São Paulo</i>            |
| SABIA-MAR | Argentinean-Brazilian Satellite of Environmental Information of the Sea |
| SAM       | Spectral Angle Mapper   |
| SeaWiFS   | Sea-viewing Wide Field-of-view Sensor                                   |
| SNR       | Signal-to-Noise Ratio   |
| SS        | Spectral Shape  |
| SSE       | Sum of Squared Errors   |
| SOA       | Spectral Optimisation Algorithm   |
| SRTM      | Shuttle Radar Topography Mission  |
| SWIR      | Short Wave Infrared   |
| TDS       | Top-Down Strategy   |
| TIR       | Thermal Infrared  |
| TLS       | Total Least Squares   |
| TM        | Thematic Mapper   |
| TN        | Total Nitrogen  |
| TSI       | Total Suspended Inorganics  |
| TSIx      | Trophic State Index   |
| TSM       | Total Suspended Matter  |
| TSO       | Total Suspended Organics  |
| ULF       | Underwater Light Field  |
| UV        | Ultra-Violet  |
| UV-A      | Ultra-Violet A  |
| VNIR      | Visible and Near Infrared   |
| VSWIR     | Visible and Shortwave Infrared  |
| WT        | Wavelet Transform   |
| WD        | Wavelet Domain  |
| WQP       | Water Quality Parameter   |



## TABLE OF CONTENTS

|  | <u>Page</u> |
|--|-------------|
| 1 INTRODUCTION.....  | 1           |
| 1.1. Hypothesis .....  | 6           |
| 1.2. Objectives .....  | 6           |
| 2 THEORETICAL BACKGROUND .....                                   | 7           |
| 2.1. Remote Sensing of Aquatic Environments.....                 | 7           |
| 2.2. Optical Properties of Water and Underwater Light Field..... | 8           |
| 2.3. Bio-optical modelling.....                                  | 12          |
| 2.3.1. Types of bio-optical algorithms .....                     | 13          |
| 2.3.2. PC-predictive bio-optical models .....                    | 17          |
| 2.4. Signal processing.....                                      | 21          |
| 2.5. Sensor applicability .....                                  | 24          |
| 3 STUDY AREA .....   | 29          |
| 4 MATERIALS AND METHODS .....                                    | 35          |
| 4.1. Equipment.....  | 35          |
| 4.1.1. YSI Sonde.....  | 35          |
| 4.1.2. Fluorometer Turner 10-AU.....                             | 37          |
| 4.1.3. RAMSES hyperspectral radiometers .....                    | 38          |
| 4.1.4. Spectrophotometer .....                                   | 39          |
| 4.1.5. AC-S In-Situ Spectrophotometer .....                      | 39          |
| 4.2. Bathymetry .....  | 43          |
| 4.3. Limnological Analysis.....                                  | 44          |
| 4.3.1. Chlorophyll-a Analysis .....                              | 45          |
| 4.3.2. Total Suspended Matter (TSM) Analysis .....               | 45          |
| 4.3.3. PC Analysis .....   | 46          |
| 4.3.4. Ion chromatography and Total Carbon Analyser .....        | 48          |
| 4.3.5. Sonde Analysis .....                                      | 48          |
| 4.4. Radiometric Analysis.....                                   | 50          |
| 4.4.1. $R_{rs}$ .....  | 50          |
| 4.4.2. Underwater Light Field.....                               | 52          |

|         |   |      |
|---------|---|------|
| 4.4.3.  | Discrete Wavelet Transform Clustering Analysis (DWTCA) .....                                      | 53   |
| 4.4.4.  | Spectral Angle Mapper (SAM) .....   | 55   |
| 4.4.5.  | Laboratory-calculated, <i>In-Situ</i> -measured, and QAA-estimated IOPs .....                     | 57   |
| 4.4.5.1 | Ternary Plot, IOPs-WQPs Least Squares Regression, and $a_t$ Total Least Squares Regressions ..... | 61   |
| 4.4.6.  | Semi-empirical bio-optical models and re-parameterisation .....                                   | 63   |
| 4.5.    | OLCI/Sentinel 3 simulation.....   | 68   |
| 4.6.    | Methodology Flowchart.....  | 70   |
| 5       | RESULTS AND DISCUSSION.....   | 73   |
| 5.1.    | Bathymetry .....  | 73   |
| 5.1.1.  | Underwater Light Field.....   | 78   |
| 5.2.    | AOP dataset.....  | 82   |
| 5.3     | Limnological and Fluorometric Dataset.....  | 83   |
| 5.3.1   | Correlation Matrices.....   | 86   |
| 5.4     | Laboratory spectrophotometric dataset.....  | 92   |
| 5.4.1   | Ternary Plot .....  | 93   |
| 5.5     | AC-S plots .....  | 96   |
| 5.5.1.  | Total Least Squares Regression.....   | 102  |
| 5.6.    | Sonde measurements .....  | 1025 |
| 5.7.    | Least Squares Regression.....   | 109  |
| 5.8.    | Signal processing techniques .....  | 117  |
| 5.8.1.  | DWTCA.....  | 117  |
| 5.8.2.  | Spectral Angle Mapping (SAM) .....  | 124  |
| 5.9.    | Semi-empirical algorithms.....  | 128  |
| 5.10.   | Semi-empirical three-band PC-retrieval algorithm re-parameterisation .....                        | 131  |
| 5.11.   | OLCI/Sentinel 3 simulation.....   | 140  |
| 6       | FINAL CONSIDERATIONS AND CONCLUDING REMARKS .....   | 145  |
|         | BIBLIOGRAPHIC REFERENCES .....  | 149  |
|         | APPENDIX A .....  | 168  |
|         | APPENDIX B .....  | 170  |
|         | APPENDIX C .....  | 172  |
|         | APPENDIX D .....  | 174  |

|  |     |
|--|-----|
| APPENDIX E .....   | 178 |
| APPENDIX F.....  | 184 |
| ANNEX A. UTM Coordinate Table Location of All Water Quality Sampling<br>Stations Monitored by <i>SABESP</i> and <i>CETESB</i> at Guarapiranga Reservoir<br>..... | 194 |
| ANNEX B. Input Feeding Streams by Socio-Economic-Environmental Areas<br>and Water Quality Sampling Stations by <i>SABESP</i> and <i>CETESB</i> .....             | 196 |
| ANNEX C. Sentinel-3 OLCI instrument technical characteristics .....  | 198 |



## 1 INTRODUCTION

The eutrophication process in aquatic systems is becoming a significant water quality problem, affecting inland water bodies all over the world (UNEP, 2000). It has been a concern for environmental and public health managers since one of its main consequences is the onset of Harmful Algal Blooms (HABs) (SIVONEN; JONES, 1999). One of the most common phylum of HABs is the cyanophyta, which has some species of cyanobacteria (also known as blue-green algae) capable of producing toxins. These species have been occurring in aquatic systems worldwide over the past twenty years (FALCONER; HUMPAGE, 2006). Thus, there is an increasing need for water governance systems, mainly in places with lack of potable water and environmental management. Seven key challenges to practice a good management were described by the National Research Council (NRC) in their document "The Drama of the Commons" (NRC, 2002). The first one is the monitoring of resources, and the second, a low cost enforcement of rules. To face these challenges, Dietz et al. (2003) believes the use of current research in collaboration with management strategies could solve particular challenges.

The monitoring of Cyanobacterial Harmful Algal Bloom (CHAB) is an important task for aquatic systems, mainly in water bodies used for water supply. Its importance is justified because cyanobacteria have been considered the largest and most diverse group of prokaryotes with very fast growth rates, especially in warm summer, when temperature, light and nutrients from agriculture fertilizers and other sources increase (MISHRA et al., 2009). They usually dominate the phytoplankton in inland and coastal areas because of their capacity of buoyancy regulation, elementary nitrogen fixing capability, and efficient use of yellowish-orange light for photosynthesis (REYNOLDS, 2006). All these capabilities make cyanobacteria one of the main phyla present in inland eutrophic waters. Furthermore, aquatic environments with CHABs develop thick surface scums, and they also have a distinct taste and odour (WORLD HEALTH ORGANIZATION, 1999; RANDOLPH et al., 2008; MISHRA et al., 2009).

However, the main problem of CHABs is their capacity to produce toxins also known as "cyanotoxins", which are a major concern for human health. Some species of cyanobacteria produce toxins which cause hepatotoxic, neurotoxic and dermatotoxic effects and general inhibition of protein synthesis in animals and humans (WORLD HEALTH ORGANIZATION, 1999).

Traditional CHAB's monitoring methods consist of collection of field samples, laboratory analysis, and manual cell counts. These methods are time consuming, labour-intensive, and costly (LE et al., 2011a). Additionally, spatial and temporal heterogeneity of water bodies often result in inadequate monitoring and characterization of CHABs, since they must rely on sampling methods, interpolation and extrapolation between sample points (KHORRAM et al., 1991). Another problem of the traditional methods of monitoring CHABs is related to their regulation of buoyancy which allows them to move upwards and downwards in the water column. This characteristic affects the collection of water since a floating ship could disturb the natural spatial distribution of a bloom (KUTSER, 2004). Besides, planning field trips to monitor CHABs is extremely difficult, because algal blooms may be ephemeral and vary according to weather conditions, capable of changing the dynamics of the water column in few days through mixing and stratification processes (TUNDISI et al., 2004; OGASHAWARA et al., 2014).

Efficient alternative methods should be developed to improve the monitoring of CHABs, combining spatial and temporal approaches with low cost analysis. These characteristics enhance remote sensing as a valuable tool for a potentially effective solution to monitoring inland water quality. This idea is supported by Kutser (2004), who emphasized that the use of remote sensing to provide information about the extent of CHABs is more reliable if compared to traditional monitoring methods, because it does not break (or interfere with) the CHABs to collect data. Metsamaa et al. (2006) also highlighted the use of remote sensing by describing it as the only efficient technique to map the spatial distribution of CHABs over time. The authors also stressed that remote sensing

is an alternative to estimate the amount of cyanobacteria just below the surface, however it could not be used to estimate it down the water column. In order to explore the water column characteristics, the underwater light field should be studied. Harrison and Platt (1986), Falkowski and Laroche (1991) studied how changes in light intensity and its spectral distribution in the water column govern the physiological acclimation of phytoplankton cells. A further study from Stoń-Egiert et al. (2012) determined the statistical relationships between the concentrations of the major groups of pigments and the various optical characteristics of the light fields in the waters of the southern Baltic Sea.

The nomenclature Case 1 and Case 2 waters have been used as a way to classify aquatic environments (MOREL; PRIEUR, 1977). The ratio of chlorophyll-*a* (chl-*a*) concentration (in  $\text{mg m}^{-3}$ ) to the scattering coefficient at 550 nm (in  $\text{m}^{-1}$ ) was the main factor to develop such a nomenclature. Case 1 waters should have a ratio larger than one, whereas in Case 2 waters, this ratio should be less than one. However, newer definitions for these water environment classifications were proposed by Gordon and Morel (1983). For the authors, phytoplankton is the constituent which controlled Case 1 waters optical properties those, and Case 2 waters are those whose optical properties are controlled by other constituents (e.g., mineral particles and CDOM), and these constituents concentrations do not covary with phytoplankton concentration. Mobley et al. 2004 assess that there are issues arising from such a simplistic classification, such as the misinterpretation over whether or not inland waters belong to Case 2. Such assumption is true because it is possible for a continental aquatic system to be dominated only by the phytoplankton (Case 1). Nevertheless, this classification is still used because it provides a synoptic view of study sites.

Remote sensing of water quality monitoring has not shown as much efficacy in Case 2 (hence forth, optically complex waters - OCW) waters studies if compared to Case 1 waters. Such evidence can be attributed to the major optically active components (OACs) in an OCW: phytoplankton, non-algal

particles (NAP), CDOM and pure water. Anthropogenic activity has increased the interaction among the constituents adding to the heterogeneity of the water systems. Uncertainties can arise from such heterogeneous environments. One example is bottom effect. Giardino et al., (2007) considered during data analysis how the semi-analytic relation between  $R_{rs}$  and IOPs are affected by bottom effect. Another example is the spatial and temporal resolutions of remote sensors. These resolutions are crucial for efficient water monitoring. Hyperspectral orbital sensors have the potential to map CHABs (Kutser (2004), however, environmental processes linked to water quality monitoring would require better temporal resolution (preferably daily monitoring) than those provided by such sensors. Although there are not many orbital optical sensors focused on aquatic studies, remote sensing techniques are potentially applicable tools for monitoring water quality.

Chl-*a* concentration was the initial variables used by remote sensing of aquatic systems for the identification of CHABs. The main reason for such attempt was the fact that chl-*a* is the primary photosynthetic pigment in cyanobacteria (REINART; KUTSER, 2006). Nevertheless, due to the fact chl-*a* is the most dominant photosynthetic pigment among all phytoplankton groups (HUNTER et al., 2009), recent studies considered using the spectral features of c-phyococyanin (PC) for the identification of cyanobacterial biomass (REYNOLDS, 2006; LI et al., 2015) in inland waters.

In hyperspectral remote sensing, spectra are increasingly analysed using methods developed for laboratory studies, such as derivative analysis, and spectral angle mapping techniques (ZHANG; LI, 2014). These techniques require smooth reflectance spectra. Therefore, there is a need for smoothing algorithms that fulfil the requirement of preserving local spectral features (SCHMIDT; SKIDMORE, 2004). Following up on this premise, PC (as well as other OACs) spectral characteristics can be appraised via signal processing techniques. Local spectral features are identifiable and evaluable in different levels of signal decomposition, and such achievement is capable via wavelet

transform clustering techniques (AMPE et al., 2014; 2015) and via spectral angle mapping algorithms (ZHANG; LI, 2014).

PC has been used as a cyanobacteria proxy because of its two distinct spectral characteristics: the absorption feature around 620 nm and the fluorescence around 650 nm (DEKKER; HESTIR, 1993; SIMIS et al., 2005; RUIZ-VERDÚ et al., 2008; ZHOU et al., 2014), the latter being variable for different species (MACINTYRE et al., 2010). Thus, the development of remote-sensing-based bio-optical algorithms that include PC absorption feature spectral bands have been used to estimate PC concentrations (VINCENT et al., 2004) or PC's Inherent Optical Properties (IOPs) (MISHRA et al., 2013).

The red part of the spectrum at around 665 nm is closely associated with absorption of chl-*a*. Due to the fact this spectral region is highly correlated with chl-*a* concentration, studies such as Mishra and Mishra (2014) used this spectral region to appropriately correct PC prediction for chl-*a* influence. However, there is a lack of studies which consider shorter wavelengths for the correction of such confusion. Also, bio-optical studies on PC over inland waters tend to focus their efforts on collecting data where PC is found in higher concentrations. Therefore, the scientific literature lacks on studies covering the shorter wavelength spectral region as well as waters bearing PC in low concentrations.

## **1.1. Hypothesis**

Taking into account the aforementioned, this thesis is based on the following hypothesis:

- ✓ A parameterisation coefficient using the blue-to-green spectral region applied to a semi-empirical bio-optical algorithm for PC-chl-a confusion improvement increases the low-concentration PC-pigment estimation efficacy in a tropical reservoir.

## **1.2. Objectives**

To test that hypothesis, a main goal was elaborated. The main goal was to enhance the performance of a phycocyanin-predictive semi-empirical bio-optical algorithm for a city-supplier water reservoir. In order to achieve such goal the following specific aims were elaborated:

1. Bio-optical, radiometric and water quality characterization of this work's study site, Guarapiranga reservoir;
2. Evaluation of feasibility for comparison and contrast of two signal processing techniques (i.e., discrete wavelet transform and spectral angle mapper) over the radiometric dataset clustering;
3. Calibration and validation of a novel semi-empirical bio-optical algorithm.

## **2 THEORETICAL BACKGROUND**

### **2.1. Remote Sensing of Aquatic Environments**

In the early seventies, Earth Observations from space started after Earth Resources Technology Satellite (ERTS-1) was sent to space by the National Aeronautics and Space Administration (NASA) (JENSEN, 2000). Some environmental variables such as chl-a concentration and temperature were noticed by oceanographers as parameters capable of being measured remotely (BUKATA, 2013). Limnologists, though, only started over the past two decades measuring water quality parameters remotely (SIMIS et al., 2005; PALMER et al., 2015; MOUW et al., 2015).

Optical properties of an aquatic system can be monitored by remote sensing techniques. This way, spatiotemporal assessment of such properties can be quantified. Attenuation properties are acquired from both orbital or proximal measurements datasets regarding the visible part of solar electromagnetic spectrum. Therefore, environmental parameters such as primary production and natural events such as algal blooms can be estimated (PLATT et al., 2008)

Oceanographic researchers advanced the theoretical and practical basis for marine optics. The interaction between the incoming radiation from sun and diffuse-sky with the water medium could be modelled by the marine optics accomplishments. Many authors among which one can include Cox and Munk (1954), Jerlov (1968), Petzold (1972), Preisendorfer (1976) created the basis for marine optics main theory by the time of ERTS-1 launch.

Gordon et al., (1975a) developed the early practical oceanic bio-optical model. The radiative transfer equation related apparent optical properties (AOPs) to the IOPs via an stochastic simulation approach in OACs-bearing oceanic waters. Bukata et al., (1979) modelled the first bio-optical algorithm associating AOPs to IOPs. Other authors made use of earlier sensors to investigate the connections

between optical properties and OACs that allowed for these constituents' retrieval from remote sensing data (BUKATA et al., 1981a; 1981b).

## 2.2. Optical Properties of Water and Underwater Light Field

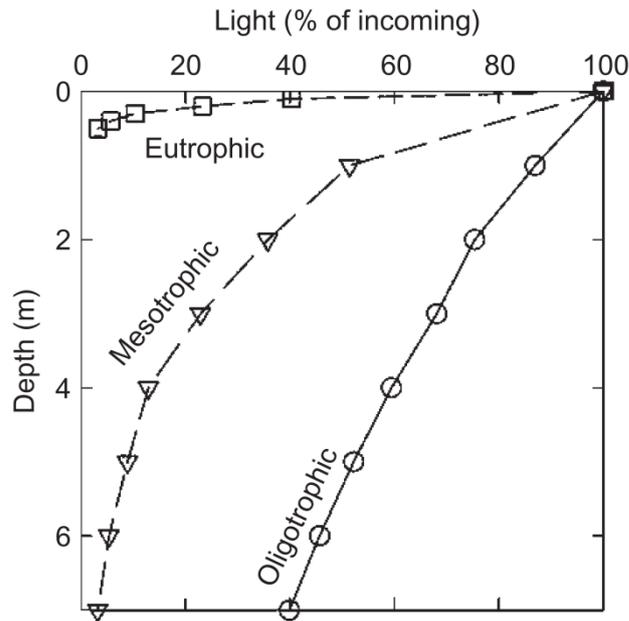
IOPs and AOPs primarily constitute the two groups regarding optical properties of an aquatic system. IOPs depend only upon the OACs concentration within the water body, whereas the AOPs depend not only on that but also on the underwater light field directionality. Examples of IOPs are the total absorption coefficient ( $a$ ) and the total scattering coefficient ( $b$ ), and examples of AOPs are the irradiance reflectance ( $R$ ), the remote-sensing reflectance ( $R_{rs}$ ), and various diffuse attenuation functions (MOBLEY, 2001).

The coefficient of attenuation ( $c$ ) influence over the incoming electromagnetic radiation (EMR) is depicted in of Figure 2.1, in which it is possible to perceive how the incident EMR magnitude is attenuated within the water body as a function of different constituent concentrations. Firstly, the EMR touches the water body surface layer. Then, part of it is absorbed and/or scattered back toward the radiation source direction and some is transmitted through the medium. The portion that is transmitted is again absorbed and/or scattered obeying Beer's Law (Equation 2.1) until all EMR is extinguished or reflected from the water column (figure 2.2) or surface bottom (BATIUK et al., 2015), and once more be under the influence of beer's law in the opposite direction.

$$E = E_0 \times e^{(-Kz)} \quad (2.1)$$

Where,  $E$  is the irradiance at a given depth moving down the water column,  $E_0$  is the irradiance at the surface,  $K$  is the attenuation coefficient, and  $z$  is depth.

Figure 2.1 – Light as a function of depth in different trophic states temperate lakes.



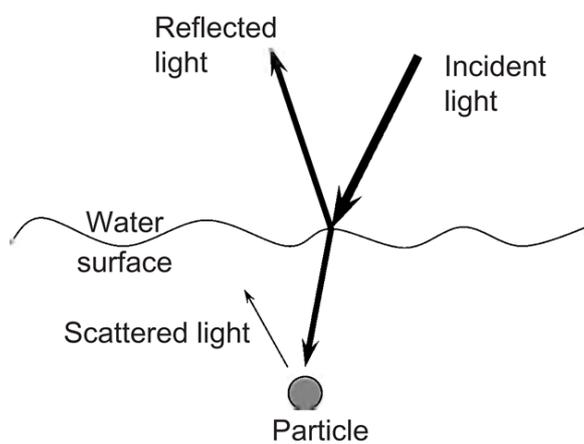
Source: Dodds and Whiles (2010).

From the above discussion it can be stated that the attenuation ( $c$ ) is the sum of total absorption ( $a$ ) plus total scattering ( $b$ ), as in equation 2.2.

$$c = a + b \quad (2.2)$$

AOPs, in the form of  $R$  or  $R_{rs}$ , are the ratio of two radiometric quantities (irradiance ratio (for  $R$  calculations) or radiance over irradiance (for  $R_{rs}$  calculations)) which are the basic properties of light measured by optical sensors. This rationing removes effects of the magnitude of the incident sky radiance onto the aquatic surface. One example occurs when the sun goes behind a cloud, the downwelling and upwelling irradiances within water may change by an order of magnitude within a few seconds, but their ratio will be nearly constant. This is why AOPs cannot be measured in laboratory or in water sample; they must be measured *in situ*. A list of the main IOPs, AOPs and radiometric quantities are shown on Table 2.1.

Figure 2.2 – Schematic of light entering water, where it can be reflected back, scatter off of a particle, or be absorbed in the water column



Source: Dodds and Whiles (2010).

Table 2.1 - IOPs, AOPs and Radiometric Quantities commonly used in optical hydrology

| Quantity                            | Units (SI)                | Simbology            |
|-------------------------------------|---------------------------|----------------------|
| <b>Inherent optical properties</b>  |                           |                      |
| Absorption coefficient              | $m^{-1}$                  | $a$                  |
| Volume scattering function          | $m^{-1} sr^{-1}$          | $\beta$              |
| Scattering coefficient              | $m^{-1}$                  | $b$                  |
| Backscatter coefficient             | $m^{-1}$                  | $b_b$                |
| Beam attenuation coefficient        | $m^{-1}$                  | $c$                  |
| <b>Apparent optical properties</b>  |                           |                      |
| Irradiance reflectance (ratio)      | -                         | $R$                  |
| Remote sensing reflectance          | $sr^{-1}$                 | $R_{rs}$             |
| Remote sensing reflectance (sub)    | $sr^{-1}$                 | $r_{rs}$             |
| Attenuation coefficients:           |                           |                      |
| of radiance $L(z, \theta, \varphi)$ | $m^{-1}$                  | $K(\theta, \varphi)$ |
| of downwelling irradiance $E_d(z)$  | $m^{-1}$                  | $K_d$                |
| of PAR                              | $m^{-1}$                  | $KPAR$               |
| <b>Radiometric Quantities</b>       |                           |                      |
| Quantity of radiant energy          | $J nm^{-1}$               | $Q$                  |
| Power                               | $W nm^{-1}$               | $\Phi$               |
| Intensity                           | $W sr^{-1} nm^{-1}$       | $I$                  |
| Radiance                            | $Wm^{-2} sr^{-1} nm^{-1}$ | $L$                  |
| Upwelling radiance                  | $Wm^{-2} sr^{-1} nm^{-1}$ | $L_u$                |
| Sky radiance                        | $Wm^{-2} sr^{-1} nm^{-1}$ | $L_s$                |
| Water leaving radiance              | $Wm^{-2} sr^{-1} nm^{-1}$ | $L_w$                |
| Downwelling plane irradiance        | $Wm^{-2} nm^{-1}$         | $E_d$                |
| Upwelling plane irradiance          | $Wm^{-2} nm^{-1}$         | $E_u$                |
| Net irradiance                      | $Wm^{-2} nm^{-1}$         | $E$                  |
| Scalar irradiance                   | $Wm^{-2} nm^{-1}$         | $E_0$                |
| Downwelling scalar irradiance       | $Wm^{-2} nm^{-1}$         | $E_{0d}$             |
| Upwelling scalar irradiance         | $Wm^{-2} nm^{-1}$         | $E_{0u}$             |
| Incident spectral irradiance        | $Wm^{-2} nm^{-1}$         | $E_s$                |
| Photosynthetic available radiation  | Photons $s^{-1}m^{-2}$    | $PAR$                |

Source: adapted from Mobley (2001).

### **2.3. Bio-optical modelling**

The expression 'bio-optical' was first used in a report of the Scripps Institute of Oceanography by Smith and Baker (1978) to describe a 'state of ocean waters'. The 'state' that the authors referred to is, in their definition, the fact that in many oceanic waters, optical properties of water are essentially subordinated to the optical properties of biological materials, mainly phytoplankton and its derivatives (MOREL, 2001). However, currently, the term 'bio-optical' has been followed by nouns like model or algorithms (OGASHAWARA, 2015).

Bio-optical models can be classified in mainly two different groups. The first group is that one which attempts to establish an empirical relationship between biological activity and radiometric quantities. This affirmation directly links radiometric quantities as a function of biological activities (MOREL, 2001; OGASHAWARA, 2015). The second group uses the radiative transfer theory in order to estimate optical properties from biological constituents. After such derivation of these optical properties, the biological activity would be retrieved from further correlation between the optical properties and the specific optical properties (GONS, 1999).

### 2.3.1. Types of bio-optical algorithms

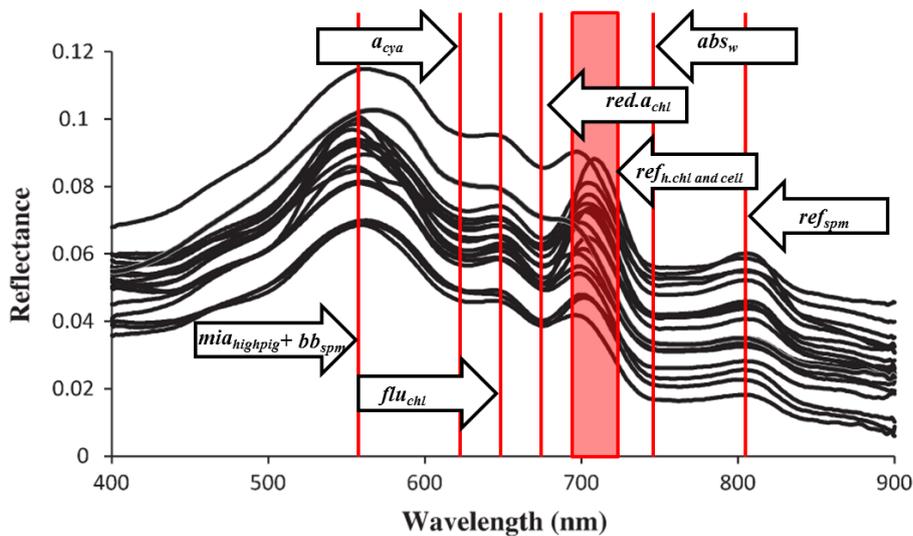
Different terms have been used in the literature to classify bio-optical algorithms according to their characteristics, however, there is no clearly known definition of bio-optical modelling terms, and their use is still very subjective. An example of the subjectivity in bio-optical modelling terminology is to notice that Simis et al. (2005) classify their algorithm as 'semi-empirical' while Hunter et al. (2010) classified Simis' algorithm as 'semi-analytical'. The most common terms for bio-optical algorithms are empirical, semi-empirical, semi-analytical (i.e., bottom-up-strategy (BUS), and top-down-strategy (TDS)), and analytical. Regarding BUS, it can be cited the spectral optimisation algorithm (SOA) (MARITORENA et al., 2002), and the linear matrix inversion (LMI) (HOGE; LYON, 1996); as for TDS, it contemplates algorithms such as the quasi-analytical algorithm (QAA) (LEE et al., 2002) and the Plymouth Marine Laboratory (PML) (SMYTH et al., 2006).

In the previous section, two classes of bio-optical algorithms were presented: the ones that derive the biological activity (statistical) and the ones that derive the optical properties (physical). Empirical and semi-empirical algorithms usually used statistical regressions between *in situ* measurements of water quality parameters and radiometric data from satellite or proximal remote sensing devices (e.g. spectroradiometer). The difference between empirical and semi-empirical algorithms relies on the physical background for their development (OGASHAWARA, 2015).

In an empirical algorithm, the selection of spectral bands is based on a combination of different spectral bands which will provide the best correlation between radiometric data and biological activity. In this type of algorithm, statistical tools such as genetic algorithms, neural networks and stepwise regressions are used to compute the best combination of spectral bands. In empirical algorithms, the combination of spectral bands does not follow any physical background; therefore, there is not any relation among IOPs to explain their selection.

Semi-empirical algorithms, on contrary, usually use spectral bands selected according to physical assumptions based on the spectral behaviour of the target (MOREL; GORDON, 1980). An example of a physical assumption is the 'Red-Edge', commonly used in vegetation indices, which uses the concepts of absorption and scattering in the red and near-infrared (NIR) spectral bands, respectively. The results of semi-empirical models are then correlated to the biological activity through a statistical analysis. Typical spectral regions investigated in such models are depicted in Figure 2.3.

Figure 2.3 – Typical spectral regions investigated in bio-optical models:  $mia_{highpig} + bb_{spm}$ , as minimum absorption by pigments;  $a_{cya}$ , as absorption by phycocyanin-containing cyanobacteria;  $flu_{chl}$ , as fluorescence by chlorophyll;  $red.a_{chl}$ , as red chlorophyll absorption trough;  $ref_{n,chl+cell}$ , as reflectance from high chlorophyll concentration and phytoplankton cell structure;  $abs_w$  as high absorption by water at longer wavelengths;  $ref_{spm}$ , as reflectance by other particulate matter.



Source: Adapted from Zhou et al. (2014). \*Variable for different species (MACINTYRE et al., 2010).

BUS and TDS algorithms are usually used to estimate optical properties in the water column, and both can be considered semi-analytic algorithms. They rely on the inversion of the radiative transfer theory to establish relationships among AOPs and IOPs which are computed through several analytical and empirical

steps. The derivation of IOPs commonly uses reflectance and this relationship is described by Equation (3.1) (GORDON et al., 1975a; GORDON et al., 1988).

$$r_{rs}(\lambda) = \frac{L_u}{E_d} = \frac{f}{Q} \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)} \quad (3.1)$$

Where  $r_{rs}(\lambda)$  is the remote sensing reflectance just below the water surface,  $L_u(0^-, \lambda)$  and  $E_d(0^-, \lambda)$  are upwelling radiance and downwelling irradiance, respectively,  $f$  is the light field factor, and  $Q$  is the light distribution factor (Li et al., 2015),  $a(\lambda)$  is the total spectral absorption coefficient,  $b_b(\lambda)$  is the total spectral backscattering coefficient.

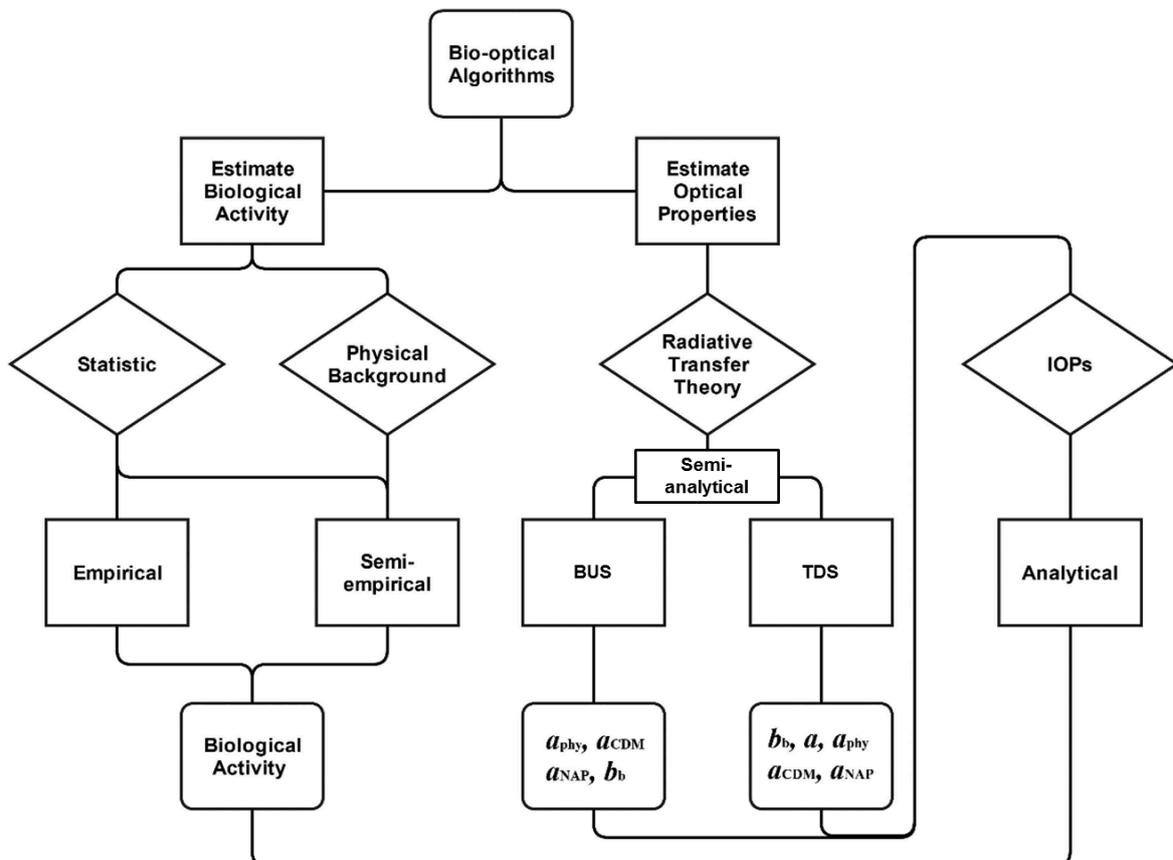
The main difference between BUS and TDS algorithms is based on the process to estimate  $a(\lambda)$  and  $b_b(\lambda)$ . In BUS algorithms, the estimation of the  $a(\lambda)$  is computed by the sum of absorption coefficients of phytoplankton, non-algal particles and Coloured Dissolved Organic Matter (CDOM). TDS algorithms do not depend on the estimation of constituent-specific IOPs (SIOPs), they estimate  $a(\lambda)$  directly from  $R_{rs}$  and constituent-specific IOPs are computed from the spectral decomposition of the estimated  $a(\lambda)$  (LEE et al., 2002). For  $b_b(\lambda)$  estimation, BUS algorithms are usually computed as the sum of the backscattering coefficients for each constituent in water except for CDOM (it is generally assumed that CDOM is nonscattering).

The outputs from BUS and TDS algorithms are then used to compute the biological activity through an analytical approach using IOPs. An analytical algorithm is the one based only in physical properties, such as the one proposed by Gons (1999) which relates the concentration of chl-*a* to the ratio between the absorption coefficient of phytoplankton ( $a_{phy}(\lambda)$ ) - based on the irradiance reflectance at 675 nm over that at 705 nm - and the specific

absorption coefficient of phytoplankton ( $a_{phy}^*$  - the absorption coefficient normalised by the constituent of interest concentration).

To summarize the classification of bio-optical algorithms, Figure 2.4 presents a flowchart which shows the division of bio-optical algorithms into two major goals: (1) estimation of biological activity and (2) estimation of optical properties. The left part of the flowchart was designated for the empirical and semi-empirical algorithms, the central part was designated for the semi-analytical algorithms and the right part was designated for analytical algorithms. This flowchart also tried to establish the procedures of bio-optical modelling used by each type of algorithm: statistical correlations (with or without physical background assumptions); the radiative transfer theory assumption; and IOP-seizing analytical procedure.

Figure 2.4 – Flowchart of bio-optical models classification



Source: adapted from Ogashawara (2015).

### **2.3.2. PC-predictive bio-optical models**

Literature has shown different methods for modelling PC estimation. A regression procedure was used by Vincent et al. (2004) used Thematic Mapper (TM) and Enhanced Thematic Mapper Plus (ETM+) sensors from Landsat 5 and 7, respectively, to collect radiometric data. The authors placed all TM and ETM+ spectral bands into the regression analysis and let the algorithm sort the best ones out. Thence, two approaches were further investigated in that study, one combining single bands, and the other combining spectral band ratios (VINCENT et al., 2004).

Genetic algorithms and Partial Least Squares method (GA-PLS) serve as yet other empirical methods for selecting spectral bands. Spectroscopic analysis from hyperspectral data were evaluated via the GA-PLS. GA selected spectral variables and PLS evaluated the relationship between spectral variables and biological activity. An optimal band ratio was used by Song et al., (2013b) to retrieve phycocyanin concentration. The use of optimal band ratio causes dependence between spectral bands and correlogram. Sun et al. (2013a) also used a correlogram analysis to estimate the best correlations among band ratios for PC detection. They classified waters in three different types and then related each water type to several band-ratios.

Characteristics absorption and fluorescence features from PC at 620 and 650 nm, respectively, have been used in several PC-retrieving empirical algorithms. The first semi-empirical bio-optical model was developed by Dekker and Hestir (1993), and relates PC concentration to the distance from the centre point of a baseline to the reflectance at 624 nm which is related to the PC absorption feature. Another semi-empirical algorithm was proposed by Schalles and Yacobi (2000) known as the single reflectance ratio, which uses PC fluorescence peak reflectance at 650 nm as reference, and then targets PC absorption at 625 nm.

A greatly literature-cited algorithms is the empirical algorithm developed for spaceborne Medium Resolution Imaging Spectrometer (MERIS) sensor. This algorithm relates PC to its absorption feature at 620nm which is also the centre of the band 6 in MERIS. PC is retrieved via the ratio between the 620 nm band over a NIR band, centred at 709 nm, as reference. A second band ratio (709/665nm) is used to retrieve a correction factor for the absorption of chl-a at the 620 nm spectral band. A band (centred at 779 nm) is used to acquire the  $b_b$  coefficient which is introduced in the two band ratios to retrieve the  $a$  coefficients by inverting a reflectance model (SIMIS et al., 2005).

Wynne et al. (2008) developed a spectral shape (SS) algorithm using the water leaving radiance ( $L_w$ ) from MERIS at 681, 709 and 665 nm. This algorithm allowed the authors to distinguish bloom and non-bloom conditions in Bear Lake, Michigan, USA. However, this algorithm does not retrieve PC concentration per se but, instead, a nominal positive- and negative-like classification. Wynne et al. (2010) proposed a cyanobacteria index (CI) using the SS algorithm in order to classify the occurrence or not of CHABs. A positive CI is indicative of elevated densities of cyanobacteria while the negative CI is indicative of no CHABs.

Hunter et al. (2008) adapted the three-band model proposed by Gitelson et al. (2008) used for the retrieval of chl-a from MERIS data to propose a PC three band algorithm. They applied their algorithm to experimental cultures of cyanobacteria. Hunter et al. (2010) proposed an update on their previous three band algorithm changing the spectral bands to 600, 615 and 725 nm reflecting the fact that measurements in the foregoing study were made from laboratory cultures rather than from natural waters.

The new single reflectance ratio proposed by Mishra et al. (2009) used the reflectance at 700 nm as reference and aimed to PC absorption at 600 nm as to minimize the chl-a interference. Le et al. (2011) based on a four-band algorithm for chl-a estimation (LE et al., 2009) developed a four-band algorithm for PC

concentration estimation. The spectral bands for this algorithm were chosen according to the same characteristics used in the chl-*a* four-band model. Thus the first band should be the most sensitive for PC, which the authors chose 630 nm. The second spectral band should have less interference of PC absorption and should have a difference on the absorption coefficients of NAP and CDOM, so it was chosen the 645 nm band. Third and fourth bands were chosen to minimize the effect of backscattering from TSS, so 730 and 695 nm were chosen as the wavelengths for the third and fourth bands, respectively.

Dash et al. (2011) used the Ocean Colour Monitor (OCM) sensor to develop a spectral reflectance slope algorithm in order to map cyanobacteria in a small freshwater lake. The slope algorithm proposed by the authors uses OCM bands 4 (510.6 nm) and 5 (556.4 nm). This spectral range was used to identify PC even without the absorption feature at 620 nm. Thus, due to OCM spectral resolution (404-885 nm) the slope between bands 4 and 5 was used to identify the low PC absorption coefficient values. Gómez et al. (2011) proposed a Normalized Difference Phycocyanin Index (NDPCI) for MERIS and Compact High Resolution Imaging Spectrometer (CHRIS) spectral bands. The authors used indices bands around 705 nm, which are located proximal to the chl-*a* fluorescence (MERIS band 9 and CHRIS band 14) and the PC absorption feature, centred at 620 nm (MERIS band 6 and CHRIS band 9). Wheeler et al. (2012) also applied the concept of Red/NIR algorithms, from chl-*a* bio-optical models for OCW, to estimate PC using Quickbird and MERIS images from Missisquoi Bay, USA.

Mishra and Mishra (2014) improved the three-band algorithm proposed by Gitelson et al. (2008) and modified by Hunter et al., (2010) by introducing a confusion-improver coefficient for the prediction of PC. This coefficient was based on the band ratio for the green-red spectral region. This ratio correlated well with chl-*a* laboratory data and therefore was tested to evaluate whether it could remove the confusion between PC and chl-*a*. However, the authors emphasize the algorithm low efficacy for inland water bearing low

concentrations of PC. LI et al. (2015) suggested that for their semi-analytical PC-predicting algorithm the blue range of the spectrum increases uncertainty within the algorithm. This uncertainty is due to higher variability in the coefficient of absorption values increasing uncertainty over the the subsequent steps estimation process. Therefore, it is noted here that further investigations are needed to evaluate whether these low-concentration PC-prediction studies can undergo improvement, and whether this improvement can be assessed for the higher-frequency wavelengths. A list of all models referenced so far is shown on Appendix A.

## 2.4. Signal processing

Wavelet transform (WT) is an analysis tool well suited for the study of multi-scale, non-stationary processes occurring over finite spatial and temporal domains (LAU; WENG, 1995). Natural events (e.g., atmospheric turbulence; and hydrological systems interaction with EMR characterised by a set of global and local parameters such as frequency, intensity, time position, and duration can be considered non-stationary processes (BOLZAN, 2004). WT can evaluate these processes due to its multi-scale capability to sieve through the many frequency levels that compose a signal (AMPE et al., 2014; 2015).

Non-stationary processes have their statistical data such as mean and variance changing over any segment of a time series (BOLZAN, 2004). On contrary, stationary processes do not have this characteristic. They are well understood from the Fourier-transform perspective, since it aims to select the period of each oscillatory component present in a time series (BOLZAN, 2004). However, non-stationary processes have more complex signal information, requiring a more flexible and robust tool.

Fourier transforms or time filters (CHAPARRO, 2015) do not provide much flexibility in the representation of both local and global signals. Instead, WT is a tool that is suited for such a purpose. In this regard, WT can be considered a flexible mathematical framework whose potential to adapt to the signal data being analysed is comparable to that of a covariance matrix used in a spatial Kriging interpolation (DEUTSCH; JOURNEL, 1992). Therefore, in the same way that Kriging is a powerful and adaptative spatial interpolation method, WT is an efficient signal clustering technique for its flexible adaptation to the many decomposition levels of a signal.

WT uses generalized local base functions (wavelets) that can be stretched and translated with a flexible resolution in both frequency and time. The flexible windows are adaptive to the entire time frequency domain, known as the

wavelet domain (WD), narrowing while focusing on high-frequency signals and widening while searching the low-frequency background (WENG; LAU, 1994).

Mathematically, a WT decomposes a signal  $s(t)$  in terms of some elementary functions  $\psi_{a,b}(t)$  derived from a "mother wavelet" or "analysing wavelet"  $\psi(t)$  by dilation and translation (Equation 3):

$$\psi_{a,b}(t) = \frac{1}{\sqrt[2]{a}} \psi\left(\frac{t-b}{a}\right) \quad (3)$$

where  $b$  denotes the position (translation) and  $a (>0)$  the scale (dilation) of the wavelet;  $\psi_{a,b}(t)$  are called "daughter wavelets" or, simply, "wavelets." An energy normalization factor  $\sqrt[2]{a}$  keeps the energy of daughter wavelets the same as the energy of the mother wavelet. The wavelet transform of a real signal ( $t$ ) with respect to the analyzing wavelet  $\psi(t)$  may be defined as a convolution integral (Equation 4):

$$W(b, a) = \frac{1}{\sqrt[2]{a}} \int \psi^*\left(\frac{t-b}{a}\right) s(t) dt \quad (4)$$

where  $\psi^*$  is the complex conjugate of  $\psi$  defined on the open "time and scale" real ( $b, a$ ) half plane. The function  $s(t)$  can be formally reconstructed from the wavelet coefficients by the inversion formula (Equations 5 and 6):

$$s(t) = \frac{1}{C_\psi} \int \frac{da}{a^2} \int db \frac{1}{\sqrt[2]{a}} \psi\left(\frac{t-b}{a}\right) W(b, a) \quad (5)$$

where

$$C_\psi = \int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega < +\infty \quad (6)$$

and  $\hat{\psi}$  is the Fourier transform of  $\psi$ .

An intuitive way to represent WT coefficients is similar to the way in which the musical tones are represented in a musical score. Pioneers in the WT field have borrowed the concept of "octave" which is logarithmic with the base of 2 for frequency or timescale, as a unit to divide the frequency domain. This unit allows us to include a broad range of scales, from very small to very large, in an efficient way in a coordinate system with linear interval in octave while logarithmic in frequency scale (WENG; LAU, 1994). Hence, unlike the FT which maps a 1D time series to a 1D spectrum, the WT maps a 1D time series to a 2D image that portrays the evolution of scales and frequencies with time or some other abscissa range (e.g., EMR wavelength).

The scale range is chosen depending on the frequency content of the timescales. For low-frequency (or high frequency) variability a scale range containing only lower (or higher) octaves is needed, similar to a bass (treble) clef in music. In general, a continuous scale covering both ends of the spectrum is used. As a result, WT decomposes a signal into "localized" or "instantaneous" frequencies complete with the measure of intensity and duration for each frequency, analogous to the bass/treble clef, the crescendo/decrescendo, and the tempo in a piece of music (LAU; WENG, 1995).

Wavelets can be mainly divided into two categories: continuous or orthogonal. When it comes to choosing which WT, orthogonal (which may not be a discrete one) or continuous, to be used in a specific application, one needs to have it clear what their differences are. Orthogonal wavelets are desirable for use in decomposition and in reconstruction of time series. Continuous wavelets yield enhanced information on the timescale localization. Therefore, orthogonal wavelets are ideal for clustering and synthesis, whereas continuous wavelets are more suitable for easier graphical interpretation due to its redundancy reinforcing certain traits within a signal.

Freitas and Shimabukuro (2008) applied a discrete wavelet transform for filtering MODIS fraction images time-series. The filtered signal was

reconstructed excluding high frequencies for each pixel in the fraction images time series. This procedure allowed a viewing of original signal without clouds and other noises. This work is a real application of wavelet transform in remote sensing. It is clear how the signal of interest was able to be sorted and clean from unwanted frequencies. Regarding water studies, Ampe et al. (2014) attempted to isolate the most informative wavelet regions via thresholding, and then, relate all five regions to known inherent optical properties. And also, Ampe et al. (2015) proposed a wavelet-enhanced inversion method, specifically designed for complex waters in which it integrated wavelet-transformed high-spectral resolution reflectance spectra in a multi-scale analysis tool.

Signal clustering can be of great support for the analysis of radiometric data related to water quality studies. It can support the development of a constituent-identification procedure for a water body management strategy. Therefore, it seems more appropriate to consider for this kind of study the application of an orthogonal wavelet, in this case, a discrete one. Moreover, comparison between different signal processing techniques in order to evaluate methodological efficacy is crucial for an ever-enhancing application of such techniques. Thereby, assessment of a discrete wavelet transform clustering against another widely used signal processing technique such as the spectral angle mapper (SAM) (ZHANG; LI, 2014; SANDER DE CARVALHO et al., 2015) seems appropriate to be performed.

## **2.5. Sensor applicability**

Coastal Zone Colour Scanner (CZCS) aboard Nimbus-7 launched in 1978 was the first sensor dedicated to aquatic systems. Therefore, marine optics improved, enabling remote multispectral sensors to monitor water colour and, thence, water quality. Other satellite sensors for water monitoring were launched after the CZCS such as the Landsat family (2 to 8), Sea-viewing Wide Field-of-view Sensor (SeaWiFS), Advanced Earth Observing Satellite (ADEOS), Hyperion, Australian Resource Information and Environmental Satellite

(ARIES), Moderate Resolution Imaging Spectroradiometer (MODIS), and MERIS.

Water quality monitoring still has limitations regarding Earth-observing satellite sensors. Examples of such limitations are the ones associated with the sensor capabilities and range of data acquisition. The four domains linked to remote sensors (spatial, temporal, spectral, and radiometric) impose several degrees of limitation. Depending on for what purpose the sensor was designed for, it could have better or worse resolution for each of the domains. And as such, the sensor ability to acquire certain types of data is compromised. For example, better temporal resolution contributes for the acquisition of more frequent data on cyanobacteria dynamics over an aquatic system. However, better spectral resolution highly contributes to acquisition of precise mineral absorption bands over geological features. Therefore, each Earth observation system must have its priorities in order to develop the adequate type of sensor system.

Environmental limitations are those not caused by a sensor, but by environmental factors. Weather dependency is one of the most important factors, since the image acquisition may be compromised by adverse atmospheric conditions, which are dependent on season and geographical location. Other atmospheric compounds such as cloud, haze, fog, smoke or dust can also compromise the frequency of earth observations over a target area. It is important to enhance that the detection of water quality variables is restricted to variables that have a direct influence on water optical conditions. It is also important to enhance that the variables influencing bio-optical models outputs are located at water bodies' near-surface regions, since most of the optical remote sensing signal cannot be derived from deeper zones. In marine optics the sea state is also an important environmental factor for the success of bio-optical modelling.

Not all remote sensors have the appropriate resolution for the identification of PC. This is true because the absorption band depth at 620 nm is shorter

compared to that of chl-a pigment. Chl-a is the major pigment of photosynthetic process whereas phycocyanin (the pigment associated with the absorption band depth at 620 nm) is an accessory pigment interacting less intensely with the incoming radiation. Sensors with higher spectral resolution such as CHRIS, Hyperion and Hyperspectral Imager for the Coastal Ocean (HICO) have the potential for mapping PC, however, worse temporal resolution, hinder the full potential of such sensors. Also, low signal-to-noise ratio (SNR) associated with hyperspectral sensors can be a great source of low-quality data. The use of multi-spectral orbital sensors is reduced due to the lack of spectral bands over the PC absorption feature. MERIS was the only multispectral sensor with a spectral band around 620 nm, however as ENVISAT stopped communicating with ground stations in middle 2012, the MERIS use for aquatic studies was compromised.

New hyperspectral sensors such as Environmental Mapping and Analysis Program (EnMAP), Hyperspectral Infra-red Imager (HyspIRI) and PRRecursore IperSpettrale della Missione Applicativa (PRISMA) have been developed to improve hyperspectral measures from the space. EnMAP is a German hyperspectral satellite mission providing high quality hyperspectral image data on a timely and frequent basis. It will provide over 240 continuous spectral bands in the range between 420 and 2450 nm with a spatial resolution of 30 m. The sensor will work in a push broom configuration, and will provide global coverage in quasi-nadir mode ( $\pm 5^\circ$ ) from a sun-synchronous orbit (STUFFLER et al., 2009).

HyspIRI is a global mapping mission that was recommended by the National Research Council's Earth Sciences Decadal Survey (NRC, 2007). Two sensors will be available: a visible-near-infrared (VNIR) and shortwave infrared (VSWIR) imaging spectrometer; and a multispectral thermal infrared (TIR) instrument. The VSWIR instrument will have 213 spectral channels between 380 and 2500 nm centred every 10 nm. Both instruments will have a 60-m spatial resolution at nadir (ROBERTS et al., 2012). PRISMA is a remote

sensing space mission under development by the Italian Space Agency (ASI). It will cover the Earth with a spatial resolution of 30 m, and 250 spectral bands with a spectral resolution better than 10 nm. The sensor spectral range covered is visible and near infrared (VNIR) and short wave infrared (SWIR) bands. It will also have a panchromatic band with 5 m of spatial resolution which will be co-registered to the hyperspectral bands in order to allow fusion techniques on PRISMA's images (LABATE et al., 2009). Despite these improvements, these sensors' equatorial revisit frequency will occur on demand, 19 days and 25 days for EnMAP, HypSIRI and PRISMA respectively (DEKKER; HESTIR, 2012).

Sentinel constellation is meant to be an ENVISAT heritage. The Ocean and Land Colour Instrument (OLCI) is on-board Sentinel-3 satellite. It is a MERIS-like sensor (DONLON et al., 2012; MALENOVSKÝ et al., 2012). OLCI will have 21 spectral bands, and the extra bands are for better atmospheric correction. Table 2.2 shows all OLCI bands and shaded areas are the new bands. The sensor will have a swath width of 1270 km (FOV of 68°, but slightly tilted) and a 300-m spatial resolution. It is tilted 12.6° westwards to avoid sun-glint over water, which may have some effects over land too (CLEVERS; GITELSON, 2013). Due to the satellites constellation strategy, two Sentinel-3 satellites have been scheduled.

OLCI's equatorial revisit time using one satellite is 3.8 days, however, if using the two-satellite strategy, revisit time will reduce to 1.9 days (ESA, 2012). For more details on OLCI sensor, a table with key characteristics is shown on Annex C. Another multispectral mission that will be developed is the SABIA-MAR mission which will have two sensors imaging earth in a global and regional scale. The mission was conceived to provide information and products to studies of ocean ecosystems, carbon cycling, marine habitats mapping, coastal hazards, and coastal land cover/land use (CHAMON, 2013).

Table 2.2 – Specifications of the OLCI on the Sentinel-3 satellite system; shaded areas are the ones that were included from MERIS specifications

| Spectral band | Centre wavelength (nm) | Bandwidth (nm) |
|---------------|------------------------|----------------|
| O1            | 400                    | 15             |
| O2            | 412.5                  | 10             |
| O3            | 443                    | 10             |
| O4            | 490                    | 10             |
| O5            | 510                    | 10             |
| O6            | 560                    | 10             |
| O7            | 620                    | 10             |
| O8            | 665                    | 10             |
| O9            | 673.75                 | 7.5            |
| O10           | 681                    | 7.5            |
| O11           | 709                    | 10             |
| O12           | 754                    | 7.5            |
| O13           | 761                    | 2.5            |
| O14           | 764.375                | 3.75           |
| O15           | 767.5                  | 2.5            |
| O16           | 779                    | 15             |
| O17           | 865                    | 20             |
| O18           | 885                    | 10             |
| O19           | 900                    | 10             |
| O20           | 940                    | 20             |
| O21           | 1020                   | 40             |

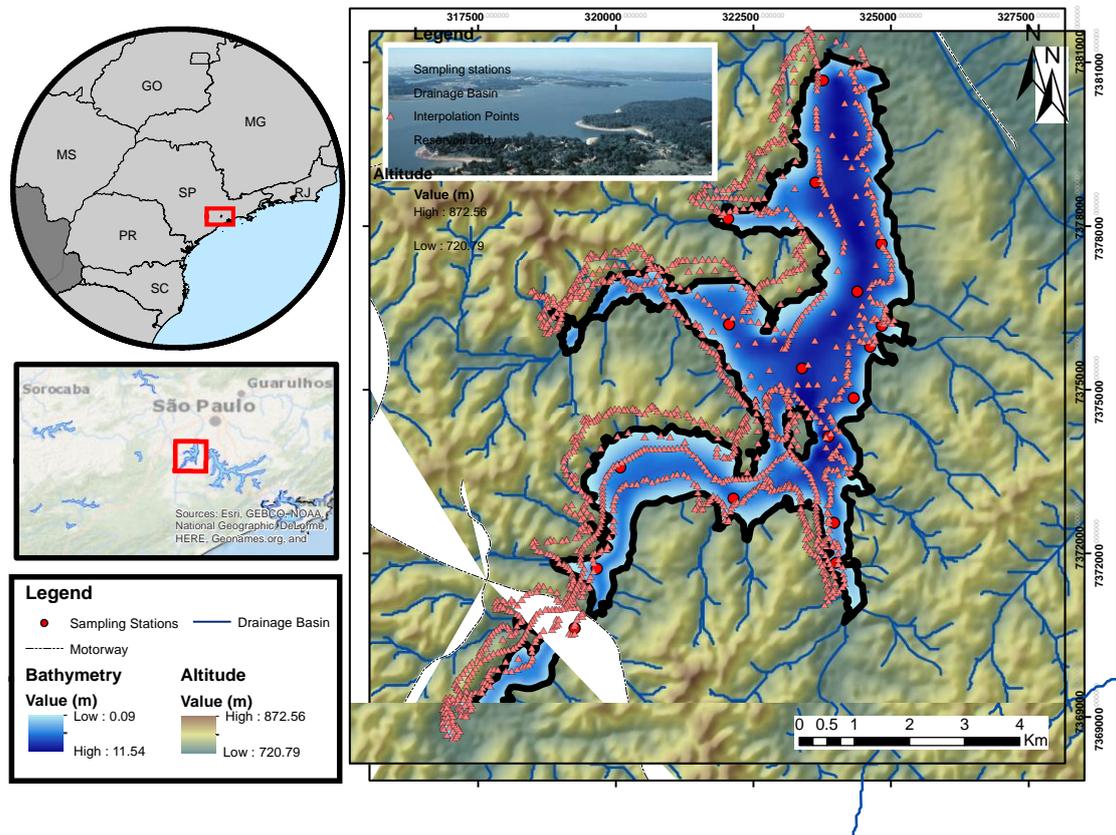
Source: Ogashawara (2014)

Considering the above-mentioned OLCI characteristics, one notes how this sensor is a potential payload for CHABs monitoring because of the spectral bands located in areas of PC interaction with incoming radiation. Dekker and Hestir (2012) attributed MERIS as the most suitable satellite sensor system with sufficient temporal frequency for inland and near-coastal water quality. As a MERIS-heritage programme sensor, OLCI has been considered a potential tool for remote sensing of water quality studies (DEKKER; HESTIR, 2012; SONG et al., 2013c).

### 3 STUDY AREA

The Guarapiranga reservoir (Figure 3.1) is located at 23°43' S e 46°32' W, and has a flood area of approximately 33 km<sup>2</sup>, maximum depth of 13 meters and retention period of 185 days; providing 11 thousand liters of water per second for over 4 million people (GEMELGO et al., 2008). Its morphology is dendritic, narrow and elongated, increasing the land use influence in its watershed (CPLEA; COBRAPE, 2006).

Figure 3.1 – Study Area Location. Sampling stations are depicted as red dots. Local bathymetry is in shaded blue. Surrounding altitude is expressed in green-brown shades. Bathymetry shown in this figure was acquired in this study. Altitude data is from Shuttle Radar Topography Mission (SRTM).



The reservoir was built between 1906 and 1908, for the purpose of electric energy generation and Tietê river flow regulation; it is located in the southwestern portion of Tietê river watershed (GEMELGO et. al., 2008). Currently, the reservoir has multiple uses, and they include leisure, bathing, and water sports. In both, Billings and Guarapiranga reservoirs, within the

metropolitan region of São Paulo, different studies on phytoplankton communities were developed. They beared an ecological- and health-related focus since the importance of these reservoirs for the region dwellers supply is highly important and requires constant monitoring. Therefore, due to its importance, Guarapiranga reservoir was chosen as this thesis' study site. Limnological characteristics showed the environment is a site where cyanobacteria are likely to be encountered (LORENZI, 2004), and this fact has also been considered for the study site choice.

Seventeen sampling stations were chosen for the field campaign (Figure 3.1) from April 14<sup>th</sup> till April 17<sup>th</sup>, 2015. The sampling stations visited were part of the stations set by *Tecnological Company of Environmental Sanitation* (CETESB) and *Basic Sanitation Company of São Paulo State* (SABESP) as water samples collection stations (Annex A; Annex B; Table 3.2). Each station has a strategic meaning for both SABESP and CETESB and this meaning was seized for the stations choice in this study. Each point allows the assessment of local input streams and therefore the evaluation of how these streams influence the reservoir and whether they have a positive or negative contribution toward the overall reservoir water quality.

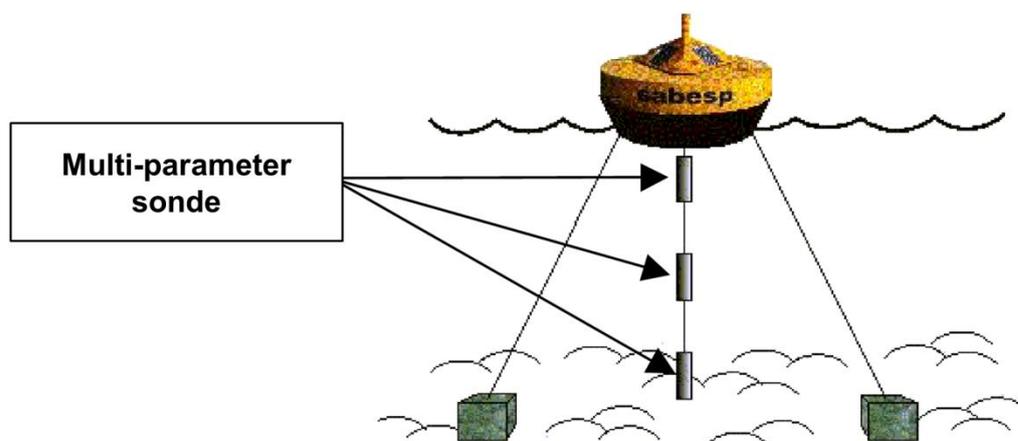
Table 3.2 – Location of sampling stations within Guarapiranga reservoir

| <b>Station ID for this study</b> | <b>Station Name (SABESP/CETESB)</b> | <b>Latitude (UTM-23S)</b> | <b>Longitude (UTM-23S)</b> |
|----------------------------------|-------------------------------------|---------------------------|----------------------------|
| P1                               | GU-103                              | 7378127                   | 322039                     |
| P2                               | GU-102                              | 7378791                   | 323625                     |
| P3                               | GUAR00900                           | 7380664                   | 323770                     |
| P4                               | GUAR00602                           | 7377672                   | 324823                     |
| P5                               | GU-104                              | 7376799                   | 324390                     |
| P6                               | GU-210                              | 7376174                   | 324831                     |
| P7                               | GUAR00452                           | 7375780                   | 324621                     |
| P8                               | GUAR00202                           | 7374841                   | 324321                     |
| P9                               | GUAR00100                           | 7371820                   | 323999                     |
| P10                              | GU-107                              | 7372559                   | 323967                     |
| P11                              | GU-108                              | 7374134                   | 323864                     |
| P12                              | Main Body*                          | 7375387                   | 323387                     |
| P13                              | GU-105                              | 7376197                   | 322059                     |
| P14                              | GU-106                              | 7373567                   | 320082                     |
| P15                              | X1*                                 | 7371717                   | 319654                     |
| P16                              | X2*                                 | 7370628                   | 319254                     |
| P17                              | GU-106_107*                         | 7373001                   | 322140                     |

\*These stations were chosen as extra stations. They had particular characteristics related to the OACs concentration and were considered relevant for this study.

The Figure 3.2 shows an example of a water quality parameter automatic collection station from SABESP. This location was also used by this study in order to acquired field data. These stations are strategically placed within the reservoir according to the nearest river inflow point junction. These junction points allow for water quality parameter (WQP) data collection. Local rivers are the sub-basins main feed into the reservoir.

Figure 3.2 – Local measurements buoys utilised by SABESP for *in situ* limnological data collection.



Adapted from Flores et al. (2001). Multi-parameter buoys installed in Guarapiranga reservoir.

CPLEA and COBRAPE (2006) grouped these local tributaries into five main groups. The groups allow for the main pollution sources identification that flow into the reservoir. Therefore, this study based itself in this grouping for the collected data assessment. The groups are described below and their numbering location is depicted in Annex B. All figures mentioned in the grouping list below represent average values from the 1995 until 2005 (CPLEA; COBRAPE, 2006).

Groups:

1. *Bonito/Pedras* (GU-210), *São José* (GU-211), *Tanquinho* (GU-212), *Itupu* (GU-219) e *Guavirutuba/Talamanca* (GU-220) are highly polluted

streams, located in highly dense urban areas, predominantly low-income areas which receive high levels of domestic sewage drain-off; Guavirutuba stream has the highest observed concentrations; despite the fact that all these sub-basins have sewage systems set-up around all public roadways, it is exactly inside them that local slums are located (ca. 50,909 inhabitants, 64% of the total basin area, according to (CENSO, 2000)(census status), and inside the sub-basins, *Itupu e Guavirutuba* streams have lower indices of homes connected to a sewage system;

2. *Ribeirões Itaim* (GU-214), *Parelheiros* (GU-213 and GU-213A), *Crispim* (GU-223) e *rio Embu Mirim* (GU-218) represent mix-use sub-basins, encompassing low-income highly-dense urban expansion areas, farming, and remaining natural vegetation; *Embu Mirim* sub-basin, also receives, industrial and domestic income loadings originated from *Itapeçerica da Serra, Embu* and part of *São Paulo*; *Embu Mirim* sub-basin possess, however, extensive floodplains which contribute to significant organic loadings; higher concentrations are detected in *Bairro Crispim* stream waters;
3. *Rio Embu Guaçu* (GU-216) e *ribeirão Santa Rita* (GU-215) have 141 and 101 km<sup>2</sup> respectively, encompassing predominantly areas of non-urban use; they receive industrial and domestic contributions from *Embu Guaçu* and *Cipó* district; total phosphorous mean concentrations are on average lower than the previous group; chemical oxygen demand (COD), however, are within the same order of magnitude, clarifying that the groups water quality main difference is due to incorrect domestic wastewater disposal, which is higher in group 2;
4. *Córrego 2* (GU-209) is a smaller sub-basin (1.14 km<sup>2</sup>), located at the reservoir right-hand side margin, in a settled higher-income area, supplied with an efficient wastewater treatment network, reaching the 98% figure of dwelling homes under the network range; total phosphorus

mean concentrations and COD are respectively 0,13 mg/L and 19 mg/L; these figures should be encountered in all previous groups, and they would represent a good proxy of a higher-quality tributary; total phosphorus concentration is 25 times lower than those observed within the first group streams;

5. Golf Club stream (GU-222) is a 12 km<sup>2</sup>, located at the reservoir right-hand side margin, encompassing strictly rural areas with farming, leisure farms, predominantly little-altered most-natural areas; total phosphorus mean concentration is 0.03 mg/L and COD, 12 mg/L.

## 4 MATERIALS AND METHODS

### 4.1. Equipment

Fluorometric, radiometric and spectrophotometric equipments were used to collect data on the water column and above the surface of Guarapiranga reservoir. In this section the main equipments are described.

#### 4.1.1. YSI Sonde

The multi-parameter sonde (YSI, 2011) used in the field measurements (Figure 4.1) was the YSI/6600. This equipment measured the following parameters: Dissolved Oxygen (DO) (mg/L), temperature (°C), conductivity ( $\mu\text{S}/\text{cm}$ ), chl-*a* (raw units).

Figure 4.1– Sonde YSI 6600 in operation.



The sonde utilises a thermistor of sintered metallic oxide that changes predictably in resistance with temperature variation. The algorithm for conversion of resistance to temperature is built into the sonde's software, and accurate temperature readings in degrees Celsius are provided automatically. No calibration or maintenance of the temperature sensor is required (YSI, 2011).

For conductivity measurements, the sonde utilizes a cell with four pure nickel electrodes for measurement of solution conductance. Two of the electrodes are

current driven, and two are used to measure the voltage drop. The measured voltage drop is then converted into a conductance value in milli-Siemens (milli-ohms). To convert this value to a conductivity value in micro-Siemens per cm ( $\mu\text{S}/\text{cm}$ ), the conductance is multiplied by the cell constant that has units of reciprocal cm ( $1/\text{cm}$ ) (YSI, 2011).

The DO sensor employs the patented YSI Rapid Pulse system for the measurement of dissolved oxygen. The Rapid Pulse system utilizes a Clark-type sensor that is similar to other membrane-covered steady-state dissolved oxygen probes. This sensor measurement range varies from 0 to 50 mg/L of DO on a precision range from  $\pm 2\%$  to  $\pm 6\%$  depending on the collected measurement; and it has a resolution of 0.01 mg/L (YSI, 2011).

On chlorophyll measurements, a key characteristic of chlorophyll is that it fluoresces, that is, when irradiated with light of a particular wavelength, it emits light of a longer wavelength (lower energy). The ability of chlorophyll to fluoresce is the basis for all commercial fluorometers capable of measuring the analyte *in vivo*. Fluorometers of this type have been in use for some time. These instruments induce chlorophyll to fluoresce by shining a beam of light of the proper wavelength into the sample, and then measuring the longer wavelength light which is emitted as a result of the fluorescence process.

Most chlorophyll systems use a light emitting diode (LED) as the source of the irradiating light that has a peak wavelength of approximately 470 nm. LEDs with this specification produce radiation in the visible region of the spectrum with the light appearing blue to the eye. On irradiation with this blue light, chlorophyll resident in whole cells emits light in the 650-700 nm region of the spectrum. To quantify the fluorescence the system detector is usually a photodiode of high sensitivity that is screened by an optical filter that restricts the detected light. The filter prevents the 470 nm exciting light from being detected when it is backscattered off of particles in the water. Without the filter, turbid (cloudy)

water would appear to contain fluorescent phytoplankton, even though none were present (YSI, 2011).

Fluorescence chlorophyll measurements might be aimed to different purposes. Murchie and Lwason (2013) suggested that differences between dark-adapted measurements or daylight measurements should be considered depending on what is the primary objective of the study. Dark-adapted measurements are of great importance to plant stress assessment. Also, studies have considered making restricted measurements during periods of the day when PAR is known to be saturating for photosynthesis and electron transport (MURCHIE et al, 2002). This thesis acquired chlorophyll fluorescence data during daylight from 10am to 15pm since it is the period of day where PAR is most strong. This allowed for a better estimation of chlorophyll activity yield since the water body was not loaded with high concentration of such photosynthetic pigment.

#### **4.1.2. Fluorometer Turner 10-AU**

The 10-AU-005-CE Field Fluorometer (Turner Designs, Sunnyvale, CA, USA) is field-portable instrument that can be used for both continuous-flow or discrete sample analysis. The instrument has a watertight case, internal data logging, automatic range changing, and water-tight quick-change filter paddles (TURNER DESIGNS, 2008; 2009). The 10-AU fluorometer (Figure 4.2) was equipped with a PC Optical Kit (P/N: 10-305) including a cool-white mercury vapour lamp, a 580-620-nm excitation filter, a 640-nm emission filter, and a reference filter at >665 nm. This optical kit permits the measurement of in vivo PC. This study used the discrete analysis procedure.

Figure 4.2– Two different ways to collect PC data via fluorometric approach. The continuous flow requires the use of a pump and hoses. And the discrete analysis simply uses individual vials for each sample measuring the raw signal value displayed by the equipment after the appropriate calibration.



#### 4.1.3. RAMSES hyperspectral radiometers

Radiance and irradiance measurements were performed using the RAMSES hyperspectral radiometers (TriOS GmbH, Oldenburg, Germany). The radiometers measure in the VNIR range of the spectrum (320–950 nm) with 3.3 nm spectral resolution (0.3-nm accuracy). RAMSES radiometers can be classified in ACC-UV, ACC-VIS, ASC-VIS and ARC (TRIOS, 2009). The RAMSES-ACC-UV is an integrated Ultra-Violet (UV) hyperspectral radiometer; and the RAMSES-ACC-VIS is a Ultra-Violet A (UV-A) and visible spectral region hyperspectral radiometer, both equipped with a cosine collector.

RAMSES-ASC-VIS is equipped with a spherical collector shielded to measure radiation from one hemisphere. It can measure scalar irradiance through the use of two of these sensors pointed in opposite directions. RAMSES-ARC is a highly integrated hyperspectral radiometer for the UV and visible spectral range. All these radiometers are calibrated for underwater and air measurements using

two different calibrations. They collect the signal through a portable terminal (e.g., laptop) via the MSDA\_XE software which is used to record irradiance and radiances signal and export the data from the database for further processing. Ohde and Siegel (2003) furthered the use of such sensors applying the sensors to calculate the immersion factors of sampled radiance and irradiance.

#### **4.1.4 Spectrophotometer**

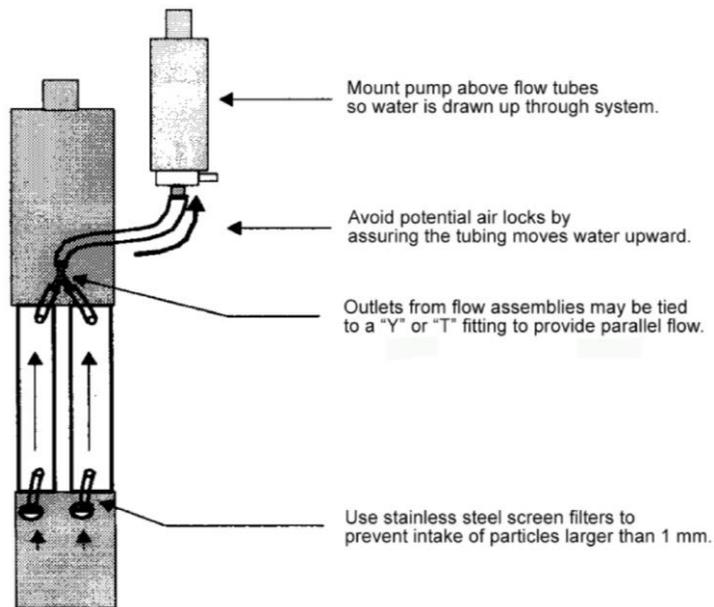
A Shimadzu UV/Vis spectrophotometer (SHIMADZU CORPORATION, Kyoto, Japan) with an integrating sphere was used to measure absorbance of the samples with 10 cm quartz cells and a spectral range from 190 - 1100 nm for particulate matter (PM) analyses. Whereas for CDOM analysis, the integrating sphere was not used, using only a 100 mm cuvette.

#### **4.1.5 AC-S In-Situ Spectrophotometer**

The OACs total absorption and attenuation were acquired by WETLabs AC-S sensor (Figures 4.3 and 4.4). The equipment operation range goes from visible to NIR (400 a 750 nm), and generated water column optical properties spectral profiles.

AC-S employs two 25 cm optical paths in order to operate in natural waters: an absorption tube, and an attenuation tube. The light source uses a collimated light beam from a tungsten lamp on a variable linear filter. The absorption tube has an internal reflective surface element and a large detecting area, whereas the attenuation tube has an opaque internal surface. The instrument has an output with a 4 nm resolution (WETLABS, 2013).

Figure 4.3 – AC-S components' illustrative scheme.



Source: WETLABS (2013).

Figure 4.4 – AC-S instrument in operation.



AC-S' output data format is binary. This is converted to a more applicable format via post-processing. The primary transfer equation for the acquisition of the attenuation coefficient ( $c$ ) or the absorption coefficient ( $a$ ) for any given wavelength is Equation 4.1.

$$T(\lambda) = e^{-c(\lambda)x} \quad (4.1)$$

where:  $T$  is transmittance;  $c$  is the attenuation coefficient; and  $x$  is the water volume optical path (WETLABS, 2013).

Transmittance can also be calculated via the ratio between the measured signal and the reference value (Equation 4.2).

$$T(\lambda) = \frac{\left( \frac{C_{sig}(\lambda)}{C_{ref}(\lambda)} \right)}{N} \quad (4.2)$$

where  $T$  is transmittance,  $C_{sig}$  is the measured signal;  $C_{ref}$  is the reference value; and  $N$  is an instrument-specific calibration constant, acquired in laboratory, using clean water (WETLABS, 2013).

Equalising Equations 4.1 and 4.2, one gets:

$$\frac{\left( \frac{C_{sig}(\lambda)}{C_{ref}(\lambda)} \right)}{N} = e^{-c(\lambda)x} \quad (4.3)$$

One can, therefore, acquire the attenuation coefficient (and similarly acquire the absorption coefficient) from Equation 4.4.

$$c(\lambda) = \frac{1}{x} \cdot \left[ \ln \left( \frac{C_{sig}(\lambda)}{C_{ref}(\lambda)} \right) - \ln(N) \right] \quad (4.4)$$

Determining  $\ln(N)$  is essential for the precise measurement of  $c$  and  $a$ . This value is determined for each channel and it is supplied by the instrument calibration data sheet.

AC-S can operate in different ways: via profiling, via hauling, and moored. In the present study only the vertical profiling was used. The profiling system allows for AC-S to be mounted in conjunction with a pump (Sea Bird Electronics, model SBE 5T), a pressure sensor (Sea Bird Electronics, model SBE 50), a battery, and a data integrator (DH4, from WETLabs). These elements were mounted into an aluminium structure (see Figure 4.4).

Pump is responsible for the water passage through the equipment pipes, and data collected by the pressure sensor can be converted to depth data. The battery supplies energy for the system and the DH4 controls data acquisition and storage. In the profiling mode, it is necessary to inform the waiting time, which accounts for the time necessary to place the equipment into the water; accounts for the system warm up; accounts for sampling frequency; and for total sampling per point.

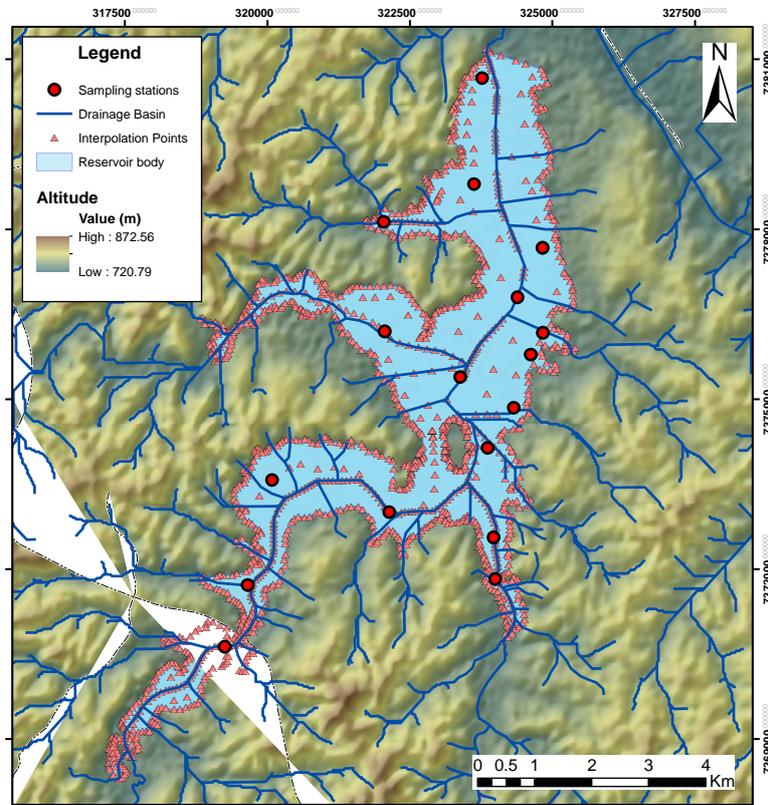
In this study the following setup times were established: 2-minute waiting time; 5-minute warm-up; 4-Hz sampling frequency; variable sampling time depending on depth of each sampling station. The profiling in each station took place as follows: the instrument was lowered to the safest deepest point (this depth was chosen to be one or two meters shorter than maximum depth); at that depth the instrument was left to warm up; then, sampling occurred bringing it upwards, then, downwards, then finally, upwards again.

## 4.2. Bathymetry

The bathymetric map was developed using: the sampling stations depth values collected during the fieldwork; the reservoir drainage basin; and depth points generated from both the combination of the previous two variables and zero-value depths (orange triangles in Figure 4.5) along the reservoir border (Figure 4.5). This method for the generation of a bathymetric map was developed after the need encountered by this study to yield the bathymetry for Guarapiranga reservoir in order to evaluate the reservoir's overall hydraulic flux. This method was the only option available, as the second fieldwork could not occur due to unforeseen field-related shortcomings. And, therefore, as the second fieldwork had been planned to also include the echobathymeter, this data could not be collected.

Ordinary kriging (OK) interpolation was applied to the reservoir measured depths and the drainage-basin-based depth point. The interpolation model was the stable model, automatically adjusted within ArcGIS<sup>®</sup> geostatistics package. This model adjusted all interpolation parameters in order to minimise the error statistics. Ordinary kriging interpolation is a spatial estimator geostatistical technique that requires the data being analysed to be normally distributed (FISCHER; GETIS, 2010). This permits the results to fall within a statistically significant framework which makes the technique relevant (CURTARELLI et al., 2015). As much as data to be interpolated for this work were not normally distributed, if log-normal transformations were applied to highly skewed datasets, such as that of this work, ordinary kriging interpolation would generate favourable results (SAITO; GOOVAERTS, 2000; WU et al., 2006). Even beyond the above-mentioned premises, the geomorphological and hydrographical features from Guarapiranga reservoir used to construct the dataset to be interpolated - explained in section 4.2 - favoured the interpolation significance. Mostly because accounting for the geomorphological and hydrographical local characteristics minimised the statistical pitfalls from the constructed dataset by allowing data interpolation within a fairly understood set of sample values.

Figura 4.5 – Map showing sampling stations with local depth measured *in situ*; showing the depth points estimated from both the drainage basin and from linear estimation for the areas between the basin lines and the reservoir border; and showing the basin lines themselves.



### 4.3. Water Quality Parameters

All limnological parameters value distribution range analysed in this study were concatenated in a boxplot. This plot was distributed in three parts (low, medium, high values). This separation allowed the assessment of which parameters had higher concentrations in the studied water body. Moreover, the trophic state index (TSI<sub>x</sub>) was calculated according to Lamparelli (2004). Furthermore, correlation among all WQP was carried out via three correlation coefficients (Table 4.4). These coefficients were the Pearson Correlation Coefficient ( $R$ ), Spearman Rank Correlation ( $\rho$ ), Kendall Rank Correlation ( $\tau$ ).

There are numerous guidelines on when to use each of these correlation coefficients. One guideline is based on the type of the data being analysed. This guideline indicates that the Pearson product moment correlation coefficient is

appropriate only for interval data, while Spearman's and Kendall's correlation coefficients could be used for either ordinal or interval data. Some guidelines also exist suggesting which correlation might be more appropriate for data that involves several types of variables (SPRENT; SMEETON, 2000). According to Khamis (2008), for data that have at least one ordinal variable, Kendall's tau is more appropriate. Other investigators suggested Spearman's correlation coefficients for the same scenarios (SIEGEL, 1957; BLOCH, 2007). However, all of these correlation coefficients could be computed for interval data (e.g. continuous) (LIEBETRAU, 1983). All three coefficients were applied to this work's dataset as to evaluate their response to the dataset. A detailed account for each WQP and their correlations was carried out in the sections 5.3 and 5.3.1.

#### **4.3.1. Chlorophyll-a Analysis**

Chl-*a* analysis were carried out after collecting water samples from the subsurface, at approximately 10-30 cm below the water surface. Samples were kept at cool temperatures until filtering procedure took place. This procedure consisted of filtering the collected samples using GF/F filters (Whatman, 0.7 µm pore size) and then extracting the samples from the filters using 90% acetone and measuring the absorbance in a Varian Cary 50 Conc UV-VIS spectrophotometer (Agilent Technologies, Santa Clara, CA, USA) (NUSH, 1980). Concentration of chl-*a* from the spectrophotometric absorbance data was calculated using the equation from LORENZEN (1967).

#### **4.3.2. Total Suspended Matter (TSM) Analysis**

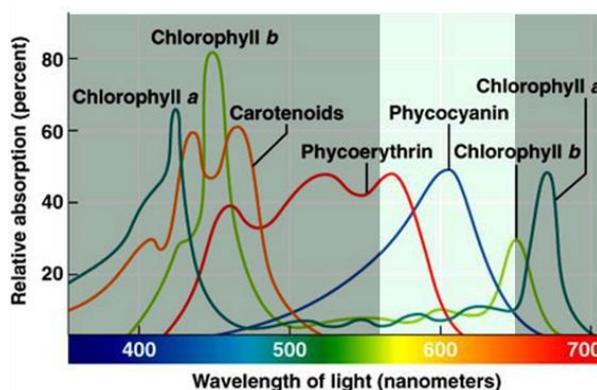
TSM concentrations (including, inorganic (TSI) and organic (TSO) components) were determined based on Wetzel and Likens (2000) from water samples filtered through pre-washed Whatman GF/F glass-fibre filters, dried at 105°C, and weighed to determine the TSM. Filters were dried and weighed and TSO calculated by the difference in weight between the pre- and post-filtered filters (TSI).

### 4.3.3. PC Analysis

PC concentration was estimated through the use of an *in situ* fluorometer (Turner Designs, model 10-AU-005) with an *in vivo* PC optical kit which uses the excitation and emission wavelengths at 580-620 nm and 640 nm, respectively. Water samples were collected and brought back to base camp, and fluorometric measurements took place as soon as the in-water data collection finished. Water samples were poured into clean transparent vials and placed inside the equipment. Once inside the measurement holder, the water sample was left still until the equipment reader settled in and displayed a stable raw value, which was then recorded for further processing.

The fluorometer emits a cool white light (TURNER DESIGNS, 2008; 2009). This light goes through the excitation filter which only allows a certain wavelength range to pass through onto the sample. The excitation range of interest for the phycocyanin study is from 580 to 620 nm, which is the EMR spectral range phycocyanin pigment interacts most with the radiation via absorption processes. Afterwards, this wavelength range reaches the sample and if there is sufficient material to interact with the radiation, emission in the 640 nm takes place activating the light detector. This wavelength range for phycocyanin detection is highly specific (Figure 4.6) and therefore, no further work in laboratory was considered as to extract only this pigment.

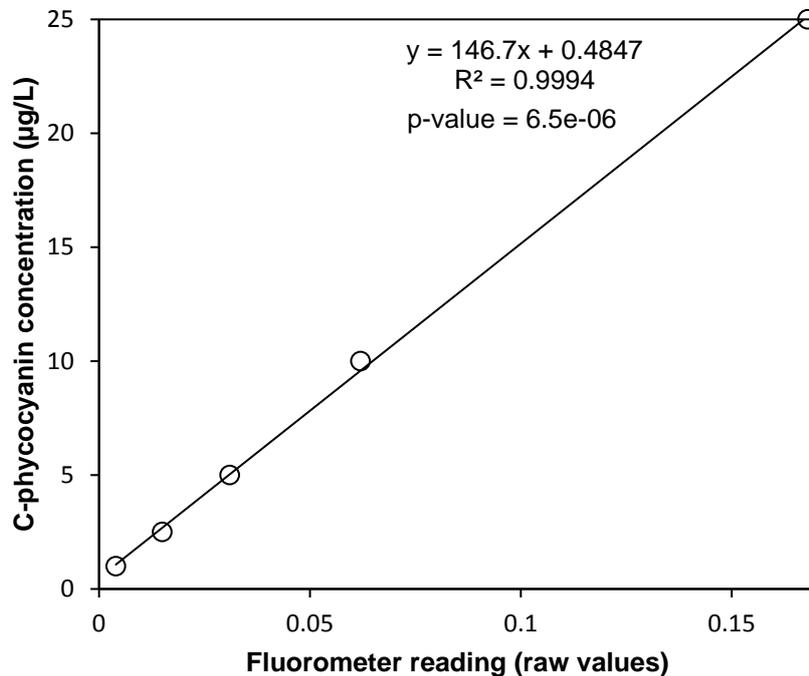
Figure 4.6 – Main pigments percent absorption spectra. Highlighting the absorption range associated with phycocyanin.



Source: <http://www.citruscollege.edu/lc/archive/biology/Pages/Chapter06-Rabbitoy.aspx>

After the field data were recorded, a calibration curve (Figure 4.7) was adjusted between the raw values acquired by the equipment and the concentration ( $\mu\text{g/L}$ ) of that respective value. This concentration was determined using a standard product (Sigma-Aldrich #P52468 - 1MG) of known concentration diluted in deionised water. This master's thesis has not used the phosphate buffer for the standard dissolution as used by Kasinak et al (2015) and suggested by the literature. The main reason for that was that the standards used by Kasinak et al (2015) and this study were different. This study used a pure solution of c-phycoocyanin pigment which only required it to be dissolved in dionised water whereas Kasinak et al (2015) used a lyophilized powder standard. The authors needed to extract the pigment from the organisms cells within the standard by means of phosphate dissolution.

Figure 4.7 – Fluorometer adjustment curve between fluorometer PC standard raw values and known concentration for that reading for the estimation of values collected during field campaign.



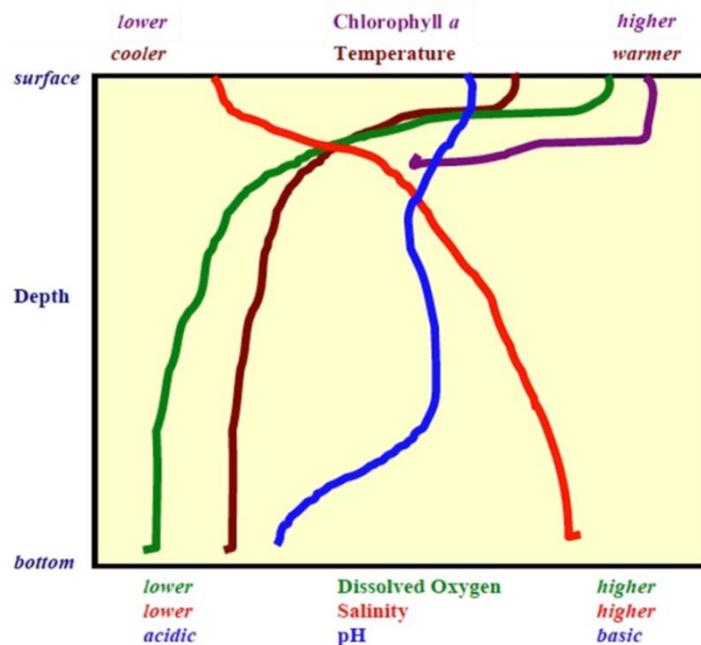
#### 4.3.4. Ion chromatography and Total Carbon Analyser

All anions and cations presented in this work were measured by the ion chromatography according to the procedures manual (FORTI et al., 2012). Also, total nitrogen (TN), dissolved organic carbon (DOC), dissolved inorganic carbon (DIC) data presented in this work were measured by the total carbon analyser (TOC-Vcpn) according to the procedures manual (ALCAIDE; FORTI, 2012).

#### 4.3.5. Sonde Analysis

Sonde WQP measurements in this study were calibrated according to the appropriate equipment manual procedure (YSI, 2011). Each sampling station had a different depth range due to depth variations within the water body. Figure 4.8 depicted a generalised YSI-measurable parameter plot over depth for Coos Bay Estuary.

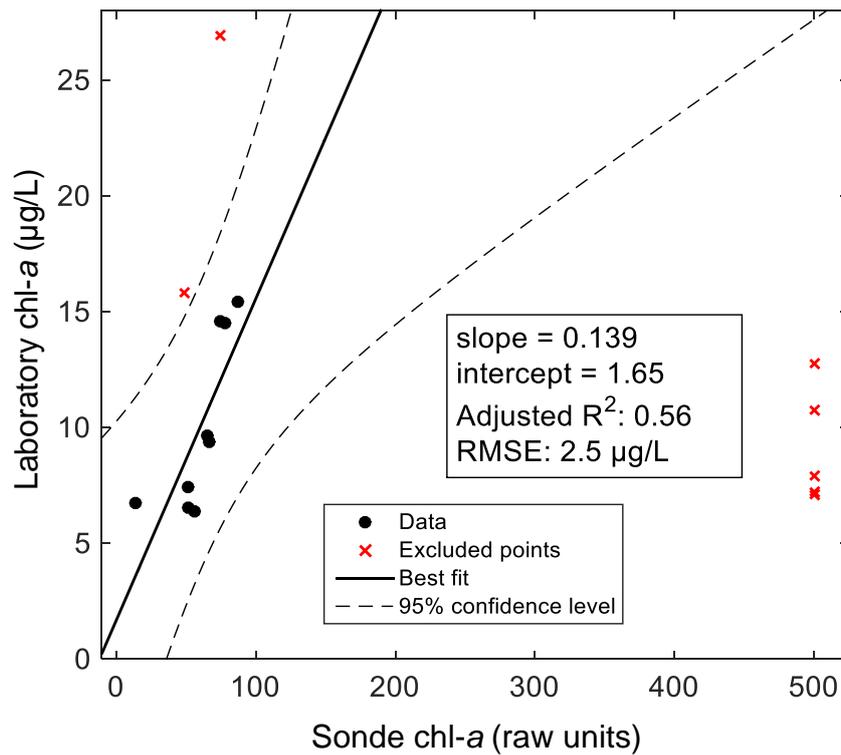
Figure 4.8 – Generalised water quality parameters plot for Coos Bay Estuary, Oregon. CTD measurement, April, 2008. Measurements depth range from 50 to 2800 m.



Source: <https://teacheratsea.wordpress.com/tag/noaa-ship-mcarthur-ii/>

In this study all parameters shown in Figure 4.8 (except pH, due to sensor fault) were measured. Chl-a was the only parameters that need further calibration after measurement as it did not yield values in the appropriate units after measurements. Therefore, a regression analysis (Figure 4.9) was applied between the water-sample laboratory-measured chl-a and sonde-measured chl-a for the first 30 cm subsurface (0<sup>1</sup>) depth. However, few points were excluded from the analysis due to incongruent extreme values from sonde measurements. These values were considered outliers because they were related to a failure within the sonde data acquisition during fieldwork, and they were removed from the adjustment curve.

Figure 4.9 – Laboratory-measured chl-a regression as a function of Sonde chl-a fluorometric sensor raw values for the first 30-cm subsurface depth.



Note: far right points at around 500 sonde raw units values were the points excluded from the regression. Higher values from laboratory-measurements were also excluded from regression in order to improve R<sup>2</sup> and Root Mean Square Error (RMSE)

## 4.4. Radiometric Analysis

### 4.4.1. $R_{rs}$

Six RAMSES sensors were used, two ACC and four ARC sensors, were acquiring data simultaneously (Figure 4.10b, and c). Two irradiance sensors centred at nominal (excluding effects of wave motion) viewing zenith angle ( $\theta_v$ ) of  $90^\circ$  were fix-pointed upward ( $E_d$ ) and downward ( $E_u$ ) directions. A radiance sensor was also pointed downward to measure the upwelling radiance ( $L_u$ ). These three measures collected above- and below-surface water measurements. An optical-fiber cosine-diffuser radiometer (yielding a hemispherical field of view, FOV), pointed upward to acquire the  $E_s$ . The other two radiometers with a  $7^\circ$  FOV were centred at nominal  $\theta_v$  of  $40^\circ$  from the water-surface horizontal plane: one upwards for the diffuse sky radiance  $L_{sky}$  and one downwards for the total water radiance  $L_t$ .

$R_{rs}$  was calculated according equation 4.5 (MOBLEY, 1999)

$$R_{rs}(\lambda, 0+) = \frac{(1 - r_F)L_u(w, \lambda)}{n^2 E_d(a, \lambda)} F_i(\lambda) \quad (4.5)$$

Where,  $r_F$  is the Fresnel reflectance of surface (0.03);  $n$  is the refractive index of water (1.33); and  $F_i(\lambda)$  is the spectral immersion factor (OHDE; SIEGEL, 2003) which is already accounted for in the Ramses TriOS calibration files; and the terms  $w$  (water) and  $a$  (air) were included in the quantities arguments in order to remind the reader that those quantities represent the underwater ( $w$ ) and above-surface ( $a$ ) measurements, respectively. The spectral immersion coefficient for each sensor was derived from Equation 4.6:

$$F_i(\lambda) = \frac{n_w(n_w(\lambda) + n_g(\lambda))^2}{(1 - n_g(\lambda))^2} \quad (4.6)$$

where,  $n_w$  is the wavelength-dependent refractive index of freshwater and can be estimated using Equation 4.7 (AUSTIN; HALIKAS, 1976).  $n_g$  is the corresponding index of the glass window of the radiance sensor.

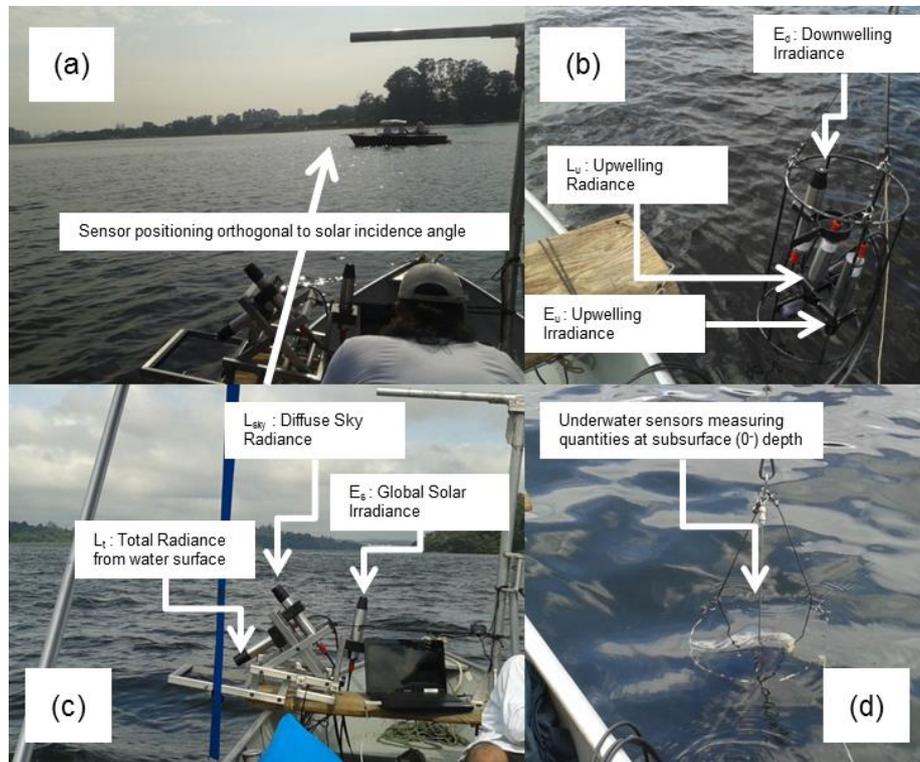
$$n_w(\lambda) = 1.325147 + \frac{6.6096}{(\lambda - 137.1924)} \quad (4.7)$$

The below surface remote sensing reflectance,  $r_{rs}$ , was calculated according to Equation 4.8 (MOBLEY, 1999; KIRK, 2011).

$$r_{rs}(\lambda) = \frac{L_u(\theta, \phi, \lambda)}{E_d(\lambda)} \quad (4.8)$$

where  $\theta$  and  $\phi$  specify the polar and azimuthal directions, respectively.

Figure 4.10 – Radiometric data collection sensor set-up: (a) The boat was kept aligned with solar direct beam direction as to keep above-water sensors orthogonal to these beams. (b) Downwelling Irradiance ( $E_d$ ), Upwelling Radiance ( $L_u$ ), Upwelling Irradiance ( $E_u$ ). (c) Total Water Radiance ( $L_t$ ), Diffuse Sky Radiance ( $L_{sky}$ ), Global Solar Irradiance ( $E_s$ ). (d) Example of underwater measurement.



Notes: boat orthogonal position varied between 90 – 135 degrees.

As described by Mishra et al. (2005), changes in sun illumination condition may cause variations in  $E_s$ . Thus, in order to normalise the radiometric measurements below water surface, Mueller (2000)'s method was adapted to quantify the variation in the obtained  $E_d$  spectra using simultaneously acquired  $E_s$  measures as showed on Figure 4.10b, and c. The methodology calculates a normalization factor ( $NF$ ) which can be estimated via equation 4.9.

$$NF = \frac{E_s(t(z_1), \lambda)}{E_s(t(z_m), \lambda)} \quad (4.9)$$

where,  $E_s(t(z_m), \lambda)$  is the  $E_s$  measured at time  $t(z_1)$  on the top of the boat at the first scan, and  $t(z_m)$  the  $E_s$  measured on top of the boat at the  $m$  scan.

#### 4.4.2. Underwater Light Field

The underwater light field was estimated via the processing of  $E'_d$  (normalised downwelling irradiance) at the varying local depths. Calculation of the vertical diffuse attenuation coefficient for downwelling irradiance,  $K_d(\lambda)$ , integrated for the Photosynthetic Active Radiation (PAR), was carried out using Equation 4.10 (MISHRA et al., 2005; KIRK, 2011) via an automated MatLab<sup>®</sup>:

$$-K_d(\lambda) = \frac{\ln\left(\frac{E'_d(z)(\lambda)}{E'_d(z-1)(\lambda)}\right)}{\Delta z} \quad (4.10)$$

where,  $z$  represents measurement-related depths (m).  $K_d(PAR)$  was acquired via the trapezoidal integration method between 400 and 700 nm.

After  $K_d(PAR)$  calculation, a mean value over all depths was taken and applied into Beer's Law (Equation 2.1) along with the over-surface ( $0^+$ ) irradiance. Moreover, a topological stacked barplot was plotted for local maximum depth comparison against euphotic zone depth for all samplings points. The underwater light field was generated as a means to understand the EMR 3-D spatial variability throughout the reservoir. Moreover, it would allow the

identification of tendencies which could provide clues on how the WQP and OAC distribution occurs not only on 2-D but also on 3-D.

#### 4.4.3. Discrete Wavelet Transform Clustering Analysis (DWTCA)

The DWTCA is a Matlab® routine already implemented in the software multi signal analysis package. This routine applies the concepts and mathematical framework from the discrete wavelet transform technique onto a set of signals. The user must set into the function arguments the following: a matrix containing the spectra to be evaluated; the number of groups to cluster the dataset (five groups were chosen for this study in an attempt to encompass the four major OACs classes plus an extra mixed-OAC class); a list indicating whether the wavelet coefficients to be used in the decomposition-reconstruction process are the coefficient of approximation (*ca*) or coefficient of detail (*cd*) (DAUBECHIES, 1992); the wavelet families available from the software list (table 4.1) to analyse the dataset (only Haar and the discrete approximation of Meyer wavelet were feasible for this study (Appendix B) as the other wavelets were not suitable for this discrete analysis); the distance computation method between two objects in the data matrix (i.e., Euclidean was used in this study); and the hierarchical cluster tree computation method (i.e., ward method was used in this study – this method was appropriate for the Euclidean distance).

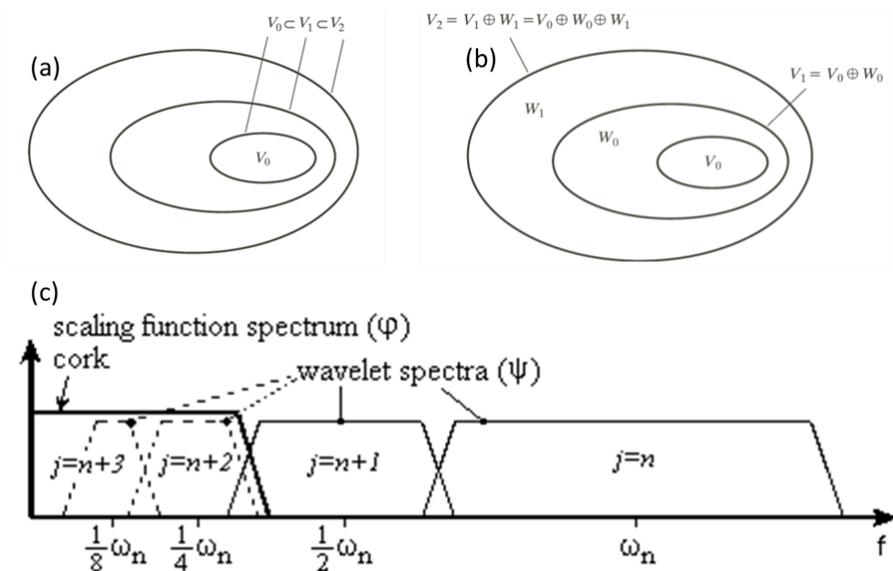
Table 4.1 – Available wavelets within Matlab®. Highlighted are the only wavelets suitable for this study’s discrete analysis.

| <b>Wavelet Family<br/>Short Name</b> | <b>Wavelet Family Name</b>    | <b>Wavelet Family<br/>Short Name</b> | <b>Wavelet Family Name</b>              |
|--------------------------------------|-------------------------------|--------------------------------------|---|
| 'haar'                               | Haar wavelet                  | 'dmey'                               | Discrete approximation of Meyer wavelet |
| 'db'                                 | Daubechies wavelets           | 'gaus'                               | Gaussian wavelets                       |
| 'sym'                                | Symlets                       | 'mexh'                               | Mexican hat wavelet                     |
| 'coif'                               | Coiflets                      | 'morl'                               | Morlet wavelet                          |
| 'bior'                               | Biorthogonal wavelets         | 'cgau'                               | Complex Gaussian wavelets               |
| 'fk'                                 | Fejer-Korovkin filters        | 'shan'                               | Shannon wavelets                        |
| 'rbio'                               | Reverse biorthogonal wavelets | 'fbsp'                               | Frequency B-Spline wavelets             |
| 'meyr'                               | Meyer wavelet                 | 'cmor'                               | Complex Morlet wavelets                 |

Once the appropriate parameters are set, the function clusters the original spectra into the five groups of interest for each level of decomposition, following the attributes space analysis metrics. *cds* rather than *cas* were chosen for this study because they are associated with wavelet functions (SHENG, 2000). Since wavelet functions fill up the gap of any two adjacent scaling subspaces (Figure 4.11), they preserve the small scales details. Nine figures using Haar wavelet (one for each decomposition level, being the first level the original  $R_{rs}$  signal level) were generated in order to depict these clusters graphically (Appendix F).

The  $R_{rs}(400-850 \text{ nm})$  spectra acquired from Trios sensors were the signal data used for clustering. This dataset was normalized over the 550 nm wavelength in order to eliminate a classification based on the signal magnitude values. Therefore, as the spectra became entangled around the 550 nm position, only the spectral shape was evaluated by the wavelet function.

Figure 4.11 – Multi resolution analysis subspace characteristics: (a) Scaling function: the subspaces spanned by the scaling function at low scales are nested within those spanned at higher scales. (b) Wavelet function: fill up the gap of any two adjacent scaling subspaces. (c) How an infinite set of wavelets is replaced by one low-pass spectrum approximation scaling function *cork* in order to avoid covering the spectrum all the way to zero with detailed wavelet spectra (MALLAT, 1989).



In order to evaluate the clusters consistency and correlation from each decomposition level cluster tree, the inconsistency coefficients (Equation 4.11) and the cophenetic correlation coefficients (Equation 4.12) were applied to the clustering results. The inconsistency coefficient (JAIN; DUBES, 1988) represents how each non-leaf cluster link height compares to average height of other links at the same level of hierarchy; the higher the values, the less similar the objects connected by the link. The cophenetic correlation (ROHLF; FISHER, 1968) for a cluster tree is defined as the linear correlation coefficient between the cophenetic distances obtained from the tree, and original distances (or dissimilarities) used to construct the tree. Thus, it is a measure of how faithfully the tree represents the dissimilarities among observations.

$$Y(k, 4) = (z(k, 3) - Y(k, 1))/Y(k, 2) \quad (4.11)$$

Where:  $Y$  is inconsistency coefficient for each link of the hierarchical cluster tree  
 $z$ ;  $k$  is the link from each tree.

$$c = \frac{\sum_{i<j} (Y_{ij} - y)(Z_{ij} - z)}{\sqrt{\sum_{i<j} (Y_{ij} - y)^2 \sum_{i<j} (Z_{ij} - z)^2}} \quad (4.12)$$

Where:  $Y_{ij}$  is the distance between objects  $i$  and  $j$  in  $Y$ ;  $Z_{ij}$  is the cophenetic distance between objects  $i$  and  $j$ , from the tree of hierarchical clusters;  $y$  and  $z$  are the average of  $Y$  and  $Z$ , respectively.

Furthermore, a dendrogram was plotted for the matrix containing the cluster groups each spectrum was associated to at each level of decomposition after the DWTCA was run. This dendrogram depicted how those spectra were classified in the same cluster over the widest  $cd$  range possible.

#### 4.4.4. Spectral Angle Mapper (SAM)

The goal of the SAM technique is to determine the degree of similarity between a test spectral dataset and a reference spectral library (ZHANG; LI, 2014). In

the present work, *in situ* hyperspectral  $R_{rs}$  data were the test data. Reference data used in this work was generated by the software water colour simulator (WASI) (GEGE, 2014). The spectral angle mapping was performed via an algorithm which determined the spectral similarity between two spectra (Equation 4.13), i.e., an automated method for comparing acquired spectra to a spectral library (KRUSE et al., 1993).

$$\alpha = \cos^{-1} \left( \frac{\vec{t} \cdot \vec{r}}{\|\vec{t}\| \cdot \|\vec{r}\|} \right) \quad (4.13)$$

where  $t$  is the test spectrum and  $r$  is the reference spectrum.

The reference spectra generated by WASI software aimed to cover five different spectral classes. These classes have different OAC concentrations. Varying absorption coefficients values for each OAC characterised the different concentrations. Table 4.2 shows the different OAC concentration magnitude used to generate the spectra for a given class. Also, a non-default AOP-related parameter (fraction of sky radiance due to direct solar radiation ( $g\_dd$ )) was used for classes four and five. This parameter represents the amount of direct radiance that is backscattered towards sky before reaching the water surface, and undergoes a diffuse reflectance back towards the water surface. Therefore, the higher this parameter value set into the  $R_{rs}$  model, the more higher-frequency radiation misses the target (i.e., water body). Thus, the slope between the blue and green spectral ranges increases due to a decrease in magnitude at the blue range. This feature associated with  $g\_dd$  was seized as to yield a different spectral class with a higher blue-green slope.

After the algorithm was applied to the dataset, a matrix was generated. This matrix contained the different spectral angle errors for each pair of measured and reference spectrum. This matrix was generated in order to evaluate how the difference reference classes were associated with the measured spectra as a function of the algorithm error values.

Table 4.2 – OACs concentration magnitude within each spectral class generated by WASI.

| Class | Phyto 1 | Phyto 2 | Phyto 5 | Particulate Matter | CDOM          | Detritus | g_dd |
|-------|---------|---------|---------|--------------------|---------------|----------|------|
| 1     | Medium  | None    | None    | Low                | High          | None     | 0.02 |
| 2     | None    | None    | Medium  | High               | Low           | None     | 0.02 |
| 3     | Low     | None    | None    | Low                | Medium        | Low      | 0.02 |
| 4     | None    | None    | None    | Medium<br>High     | Medium<br>Low | Low      | 0.3  |
| 5     | High    | High    | High    | Medium             | High          | Low      | 0.18 |

Low represents 0-25%; Medium Low represents 26-38%; Medium represents 39-62%; Medium High represents 63-75%; High represents 76-100%. Each phytoplankton type is associated with a specific phylum. Phyto 1 represents cryptophyta type L; phyto 2 represents cryptophyta type H; Phyto 5 represents green algae. The default  $g\_dd$  value set by the software is 0.02.

#### 4.4.5. Laboratory-calculated, *In-Situ*-measured, and QAA-estimated IOPs

Absorption coefficients such as  $a_t$ ,  $a_{phy}$ ,  $a_{nap}$  and  $a_{CDOM}$  – total particulate, phytoplankton, non-algal particles (tripton), and coloured dissolved organic matter, respectively - were measured at LAQUATEC Laboratory - INPE. For the  $a_{CDOM}$ , the water samples from the field campaign were temporarily stored in a cooled chamber and filtered through 0.2- $\mu$ m-large nylon filters. In order to retrieve the water samples CDOM absorption coefficient, the beam attenuation of the filtered water was measured on a Shimadzu UV/Vis spectrophotometer light detector after the beam passed through a transparent 100-mm quartz cuvette. As data from the equipment are noisy, an exponential shape of the  $a_{CDOM}$  was fitted based on the 380–800 nm wavelength range.

The absorption spectra of non-algal particles and phytoplankton were measured using the filter pad method using a Perkin Elmer integrating sphere attached to the Shimadzu UV/Vis spectrophotometer following the methods described by Tassan and Ferrari (2002). The method consisted on the determining the water samples  $a_t$ . The water samples were filtered under low pressure through a 25-mm GF/F Whatman filter.  $a_t$  in the range 400 - 800 nm was determined in the spectrophotometer. Samples were then de-pigmented by soaking the filters in a sodium hypochlorite (NaClO) 10% solution. The values of  $a_{nap}$  were then

measured as described above and  $a_{phy}$  values were calculated from Equation 4.14.

$$a_{phy} = a_t - a_{nap} \quad (4.14)$$

Measurements from AC-S allowed the development of attenuation and absorption plots. Pure water offsets supplied from the instrument's device file were applied to all vertical profiles to compute lake water total (particulate + dissolved matter) attenuation, and absorption (WETLABS, 2013). Before correcting for salinity and temperature (WETLABS, 2013), profiles were binned using a 10-cm median window. Once salinity and temperature corrections were carried out, depth per absorption and per attenuation plots were generated for five wavelengths (443, 500, 549, 670, 716 nm) which represent best the spectral regions in which the most common OACs interact with the EMR.

The absorption measurements were, then, submitted to the scattering corrections (WETLABS, 2013; SANDER DE CARVALHO et al., 2015) as follows: the Flat method subtracts the absorption measurement at a reference wavelength (Equation 4.15):

$$a_{Flat} = a_{corr} - a_{corr}(\lambda_{ref}) \quad (4.15)$$

where  $a_{Flat}$  stands for the AC-S absorption corrected by the Flat method,  $a_{corr}$  is the absorption coefficient corrected for temperature/salinity and  $\lambda_{ref}$  is the reference wavelength. Based on Leymarie et al. (2010) instead of 715 nm as reference wavelength, it was used the 743 nm wavelength in an attempt to improve this correction, since the absorption wavelength coefficient is expected to be lower at 743 nm than at 715 nm. The Proportional Correction is described by Equation 4.16

$$a_{Prop} = a_{corr} - a_{corr}(\lambda_{ref}) \cdot \frac{b_{corr}(\lambda)}{b_{corr}(\lambda_{ref})} \quad (4.16)$$

Leymarie et al. (2010), assessed two reference wavelengths, 715 and 870, for 10-cm AC-9, based on computer simulations. In this study, the wavelengths tested were 715 ( $a_{P715}$ ) and 743 ( $a_{P743}$ ) due to the spectral range limit imposed by AC-S. The temperature/salinity corrected scattering ( $b_{corr}$ ) was estimated as a difference between  $c_{corr}$  and  $a_{corr}$ . The Kirk method is described as (Eq. 4.17):

$$a_{Kirk} = a_{corr} - CFS \cdot b_{corr} \quad (4.17)$$

in which the constant fraction of scattering (CFS) proposed by Kirk (1992) for coastal water ranges from 0.121 to 0.162. Wet Labs manual (WETLABS, 2013), however, recommends 0.18 for turbid waters. Four scattering coefficients were calculated from  $c_{corr}$  and four scattering corrected absorptions. They were named as:  $b_{Flat}$  ( $c_{corr} - a_{Flat}$ ),  $b_{P715}$  ( $c_{corr} - a_{P715}$ ),  $b_{P743}$  ( $c_{corr} - a_{P743}$ ), and  $b_{Kirk}$  ( $c_{corr} - a_{Kirk}$ ). Thence, a figure was generated depicting the absorption per unit wavelength from 400 towards 750 nm.

The Quasi-Analytical Algorithm (QAA) was originally developed by Lee et al. (2002) to derive the absorption and backscattering coefficients by analytically inverting the spectral remote-sensing reflectance ( $R_{rs}(\lambda)$ ). The version used in this master's thesis was the QAAv6. QAA starts with the calculation of the total absorption coefficient ( $a$ ) at a reference wavelength ( $\lambda_0$ ), and then propagate the calculation to other wavelengths. Component absorption coefficients (contributions by non-algal particles, CDOM and phytoplankton pigments) are further algebraically decomposed from the total absorption spectrum. To summarize, briefly, QAA is consisted of the following elements:

1. The ratio of backscattering coefficient ( $b_b$ ) to the sum of backscattering and absorption coefficients ( $b_b/(a+b_b)$ ) at  $\lambda$  is calculated algebraically based on the models of Gordon et al. (1988) and Lee et al. (1999),

$$\frac{b_b(\lambda)}{a_t(\lambda) + b_b(\lambda)} = \frac{-0.0895 + \sqrt{0.008 + 0.499r_{rs}(\lambda)}}{0.249} \quad (4.18) \blacksquare$$

Here  $r_{rs}(\lambda)$  is the nadir-viewing spectral remote-sensing reflectance just below the surface and is calculated from nadir-viewing  $R_{rs}(\lambda)$  through,

$$r_{rs}(\lambda) = R_{rs}(\lambda)/0.52 + 1.7R_{rs}(\lambda) \quad (4.19)$$

2. The spectral  $b_b(\lambda)$  is modelled with the widely used expression from Gordon and Morel (1983) and Smith and Baker (1981),

$$b_b(\lambda) = b_{bw}(\lambda) + b_{bp}(\lambda_0) \left( \frac{\lambda_0}{\lambda} \right)^\eta \quad (4.20)$$

where  $b_{bw}$  and  $b_{bp}$  are the backscattering coefficients of pure seawater and suspended particles, respectively. Values of  $b_{bw}(\lambda)$  are provided in Morel (1974).

3. When  $a_t(\lambda_0)$ , the ratio of  $b_b/(a_t+b_b)$  at  $\lambda_0$ , and  $b_{bw}(\lambda_0)$  are known,  $b_{bp}(\lambda_0)$  in Equation 4.16 can be easily derived with the combination of Equations 4.14 and 4.16. The values of  $b_b(\lambda)$  at other wavelengths are then calculated after the power parameter ( $\eta$ ) is estimated (LEE et al., 2002).
4. Applying  $b_b(\lambda)$  to the ratio of  $b_b/(a_t+b_b)$  at  $\lambda$  (Equation 4.18), the total absorption coefficient at  $\lambda$ ,  $a_t(\lambda)$ , is then calculated algebraically.
5. After  $a_t(\lambda)$  is known,  $a_{CDM}(\lambda)$  (absorption coefficient for CDOM and NAP) and  $a_{phy}(\lambda)$  is calculated through

$$\left\{ \begin{array}{l} a_{CDM}(443) = \frac{(a(411) - \zeta a(443)) - (a_w(411) - \zeta a_w(443))}{\xi - \zeta} \\ a_{phy}(\lambda) = a(\lambda) - a_w(\lambda) - a_{CDM}(443)e^{-S(\lambda-443)} \end{array} \right. \quad (4.21)$$

Here  $\zeta$  equals  $a_{phy}(411)/a_{phy}(443)$  and  $\xi$  equals  $a_{CDM}(411)/a_{CDM}(443)$ .

#### 4.4.5.1 Ternary Plot, IOPs-WQPs Least Squares Regression, and $a_t$ Total Least Squares Regressions

A ternary plot containing the relationship among the relative absorption coefficients for CDOM, NAP, and phytoplankton was generated in order to evaluate which constituent dominated the water – excluding pure water absorption coefficient - in a given wavelength. These absorption coefficients used in the ternary plot were the laboratory-calculated coefficients. Seven wavelengths were chosen for this procedure: 412, 443, 560, 620, 665, 675, 700 nm (BABIN et al., 2003; RIDDICK et al., 2015).

Once the IOPs were acquired from the three different methodologies described above (laboratory-calculated, *in-situ*-measured (AC-S), and QAA-estimated), several regression analyses were carried out to evaluate the relationships among the IOPs (i.e.,  $b_{bp}(555)$ ,  $a_{CDOM}(440)$ ,  $a_{phy}(440)$ ,  $a_{phy}(620)$ ,  $a_{phy}(675)$ ,  $a_{nap}(440)$ ,) and five WQPs (i.e., chl-*a*, PC, TSM, TSI, and DOC). In order to achieve that  $b_{bp}(555)$  was correlated against PC, TSM, and chl-*a* (HUOT et al., 2007);  $a_{phy}(620)$  correlated against PC, and PC+chl-*a*;  $a_{phy}(440)$  against chl-*a*;  $a_{phy}(675)$  against chl-*a*;  $a_{CDOM}(440)$  against chl-*a* and TSM; and  $a_{nap}(440)$  against TSI (BABIN et al., 2003; WHITMIRE et al., 2010; RIDDICK et al., 2015). Statistics and error analysis carried out for these regressions are listed in Table 4.4.

The IOPs used in the correlations were the ones obtained from the laboratory analyses. These correlations intended to further evaluate the relationships between IOPs and AOPs, and assess how these correlations could lead up to the identification of a re-parameterisation approach towards the semi-empirical bio-optical algorithm. It is important to point out that the particulate backscattering at 555 nm,  $b_{bp}(555)$ , was calculated via Gordon et al. (1975a) from AC-S data. The absorption data used from AC-S instrument was the one corrected only for temperature. The choice for this method of correction was that it presented the most robust statistics from the Total least Squares

regression (TLS) analysis. It seemed a more accurate data representation of total absorption coefficient than laboratory data since it was being measured *in situ* without further error-inducing steps possibly linked with laboratory measurements which can further propagate errors after splitting the water into three components (phytoplankton, CDOM and non-algal particles absorptions) and then, summing them up again to acquire the total absorption coefficient. Nevertheless, two distinct sources of measurement errors must be considered separately. The first source mainly results from pure water calibration of the instrument, while the second source is inherent to the instrument itself (Leymarie et al, 2010). And both errors are sources of uncertainties within the data used in this work.

After the least squares regression analyses, TLS was carried out over the three differently-acquired absorption coefficients. TLS primarily uses Principal Components Analysis (PCA) within its mathematical framework. PCA is a highly sophisticated mathematical framework for dimensionality reduction of large datasets (NEUMANN et al., 1993; LELONG et al., 1998; PAL et al., 2007; ZABALZA et al., 2014). PCA minimizes the perpendicular distances from the data to the fitted model. This is the linear case of what is known as TLS, and is appropriate when there is no natural distinction between predictor and response variables, or when all variables are measured with error (NIEVERGELT, 1994; GROEN, 1998; MARKOVSKY; HUFFEL, VAN, 2007). This is in contrast to the usual regression assumption that predictor variables are measured exactly, and only the response variable has an error component (HELSEL; HIRSCH, 1992; LEGENDRE; LEGENDRE, 2012). PCA was, in this study, applied onto an orthogonal regression best fitting within a 3-dimensional scatterplot for the  $a_t$  variable which has been acquired via three different methodologies.

One TLS plot was generated per wavelength for the laboratory-measured  $a_t$ , AC-S calculated  $a_t$ , and QAA output  $a_t$  (for the five QAA output wavelengths: 412, 443, 489, 530, 555). These wavelength plots sub-plotted in themselves five different TLS regressions for the five corrections applied to ACS data (i.e., Flat

750, Prop. 715, Prop. 750, Kirk, Temperature-only corrections). This PCA application intended to evaluate the relationships among the IOP results from different methodologies; and investigate whether these correlations could yield results corroborating the identification of a possible re-parameterisation approach towards the semi-empirical bio-optical algorithm.

#### **4.4.6. Semi-empirical bio-optical models and re-parameterisation**

The semi-empirical bio-optical (Table 4.3) models were tested on the  $R_{rs}$  dataset (Figure 4.12). This dataset presented varied  $R_{rs}$  results both in magnitude and spectral behaviour. Therefore it was presented in individual plots, one per sampling station, rather than all overlapping in one plot. A linear calibration curve was set between the each model output value and measured PC concentration. Then, the bio-optical models were plotted in a 1:1 scatterplot and statistically analysed via the Pearson Product Moment Correlation Coefficient (PPMCC) - R value. After an evaluation of the best R values, the best-performance PC algorithm was selected for the re-parameterisation. Some common Chl-a and chl-a-to-TSM (chl-a:TSM) ratio algorithms were tested to evaluate which generated the best results for this study and whether much variability would be encountered among them. Chl-a:TSM algorithm was evaluated for its ability to demonstrate the level of proportionality existing between chl-a and TSM concentrations for Guarapiranga.

Figure 4.12 – Remote Sensing Reflectance ( $R_{rs}$ ) for 15 sampling stations. Stations 1 and 6 were excluded from this calculation due to faulty equipment at the measurement site.

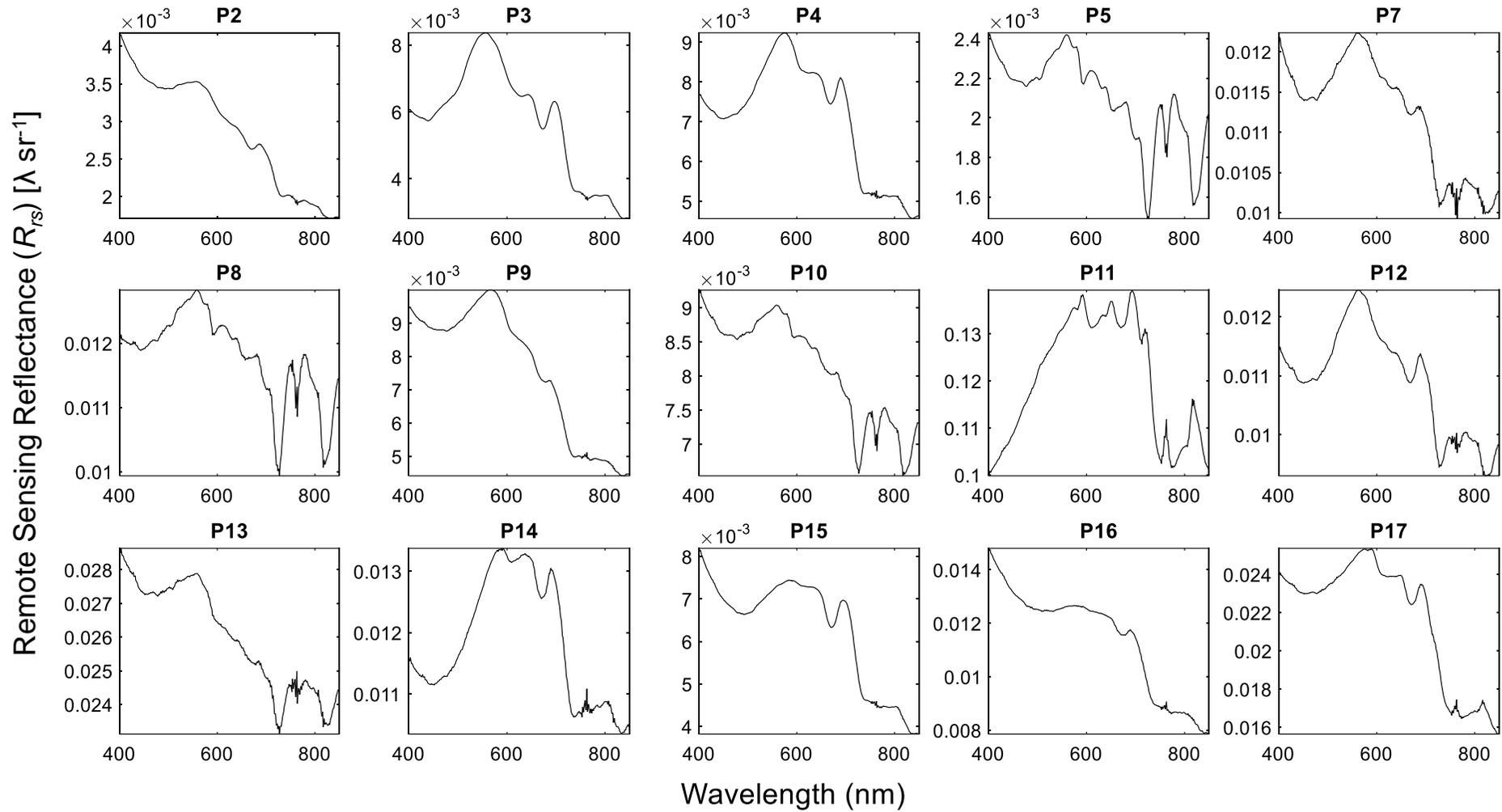


Table 4.3 – Summary of  $R_{rs}$ -based bio-optical algorithms used in this study.

| Model        | Algorithm  | Reference                        |
|--------------|--|----------------------------------|
| <b>SY00</b>  | $PC \propto \left( \left[ \frac{R_{rs}(650)}{R_{rs}(625)} \right] - 0,97 \right) \cdot 1096,5$   | Schalles and Yacobi (2000)       |
| <b>SI07*</b> | $PC \propto 170 * \left\{ \left[ \frac{R_{rs}(709)}{R_{rs}(620)} \times (a_w(709) + b_b) - b_b - a_w(620) \right] - 0,24 \times a_{phy}(665) \right\}$ | Simis et al., (2005) and (2007a) |
| <b>HU10</b>  | $PC \propto [R_{rs}^{-1}(615) - R_{rs}^{-1}(600)] \cdot R_{rs}(725)$   | Hunter et al., (2008) and (2010) |
| <b>LE11</b>  | $PC \propto [R_{rs}^{-1}(625) - R_{rs}^{-1}(650)] \cdot [R_{rs}^{-1}(730) - R_{rs}^{-1}(695)]^{-1}$  | Le et al. (2011)                 |
| <b>MI14</b>  | $PC \propto [R_{rs}^{-1}(620) - (\psi R_{rs}(665))^{-1}] \cdot R_{rs}(778)$  | Mishra and Mishra (2014)         |
| <b>OC3</b>   | $OC3M - 547,551$   | O'Reilly et al. (1998)           |
| <b>OC4</b>   | $OC4E$   | O'Reilly et al. (2000)           |
| <b>2HU14</b> | $chl - a \propto [R_{rs}^{-1}(706)/R_{rs}^{-1}(685)]$  | Huang et al. (2014)              |
| <b>3HU14</b> | $chl - a \propto [R_{rs}^{-1}(685) - R_{rs}^{-1}(707)] \cdot [R_{rs}^{-1}(722)]$   | Huang et al. (2014)              |
| <b>SU13</b>  | $\frac{Chla}{TSM} \propto 0.07 \left[ \frac{R_{rs}(696) - R_{rs}(673)}{R_{rs}(673)R_{rs}(696)} \right] + 0.349$  | Sun et al. (2013b)               |

Notes: \* $b_b$  was estimated via a band ratio (GONS, 1999);  $a_{phy}$  was estimated adapted from a band ratio (GORDON et al., 1975a; GORDON; MCCLUNEY, 1975b).

Error analysis was performed by comparing measured and predicted OAC concentrations. PPMCC, Adjusted R<sup>2</sup>, and Normalized Root Mean Square Error (NRMSE) were used to evaluate models performance and were calculated according to Table 4.4.

Table 4.4 –Summary of error estimator, correlation coefficients, and estimation curve robustness used in this study.

| Estimator               | Formulae   |
|-------------------------|--|
| PPMCC                   | $R = \frac{n(\sum_{i=1}^n x_i y_i) - (\sum_{i=1}^n x_i)(\sum_{i=1}^n y_i)}{\sqrt{n(\sum_{i=1}^n x^2) - (\sum_{i=1}^n x)^2} \sqrt{n(\sum_{i=1}^n y^2) - (\sum_{i=1}^n y)^2}}$ |
| SRC                     | $\rho = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$   |
| KRC                     | $\tau = \frac{n_c - n_d}{\frac{1}{2}n(n - 1)}$   |
| Adjusted R <sup>2</sup> | $Adj. R^2 = 1 - \frac{SS_{res}/(n - K)}{SS_t/(n - 1)}$   |
| NRMSE                   | $NRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - x_i)^2}}{x_{max} - x_{min}}$  |

**PPMCC** stands for Pearson Product Moment Correlation Coefficient where  $x$  and  $y$  are the measured and predicted values; **SRC** stands for the Spearman Rank Correlation Coefficient where  $d_i$  is the difference between ranks of corresponding values  $x$  and  $y$ ; **KRC** stands for Kendal Correlation Coefficient, where  $n_c$  is the number of concordant and  $n_d$  is the number of discordant; unlike the R<sup>2</sup>, the **adjusted R<sup>2</sup>** gives the percentage of variation explained by only those independent variables that in reality affect the dependent variable, where  $SS_{res}$  is the sum of squares error, and  $SS_t$  is the total sum of squares,  $n$  is the total number of samples, and  $k$  is the number of populations; the **NRMSE** has a percentage error output which allows easier comparison among different model results.

The re-parameterisation proposed in this study investigated the change of coefficient that aimed to reduce the chl-*a* interference in PC prediction for very low PC concentration (0-15 µg/L), which was a step further from what Yacobi et al. (2015) who explored a broader range (0.5-126.4 µg/L). A sensitivity analysis was carried out for the re-parameterising coefficient. This analysis was performed to study how the re-parameterisation coefficient was behaving in correlation with the PC concentration. The sensitivity analysis was conducted plotting a three-dimensional surface scatterplot containing the PC concentration

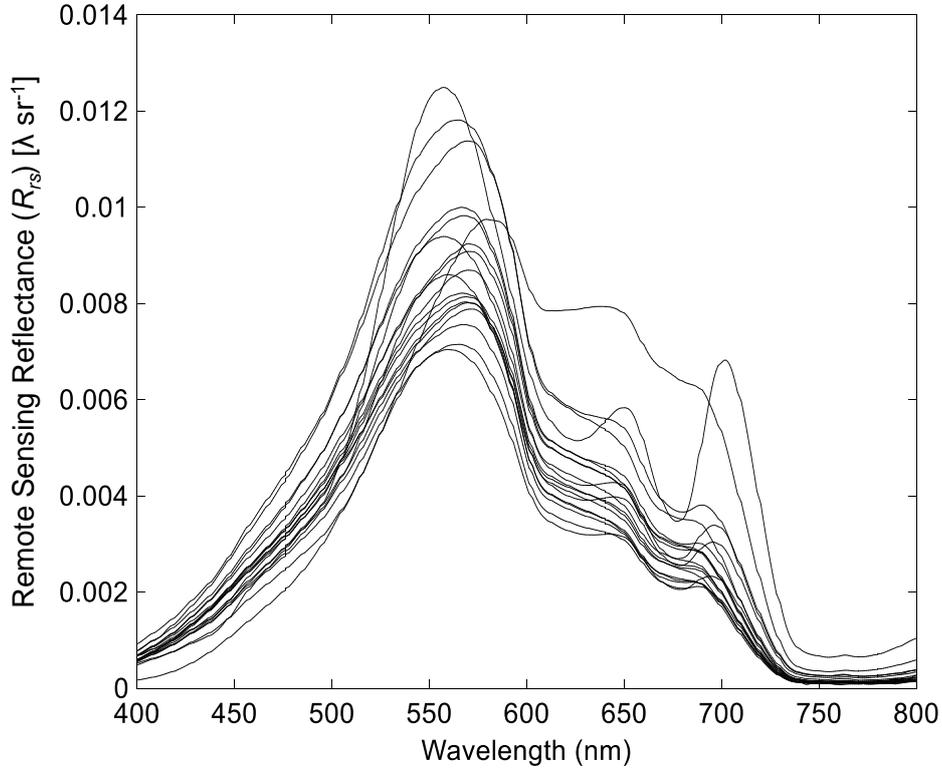
measured *in situ*, and a two-dimensional correlation between the new proposed blue-green band ratio (re-parameterising coefficient, *PLH*) and the 620-nm PC absorption band ratio.

The new coefficient, *PLH*, was then applied to the three-band semi-empirical algorithm. Thence, the original algorithm (MI14; Table 4.3) was evaluated for its  $R^2$  and NRMSE results against the re-parameterised algorithm statistics and error analysis. Furthermore, in an attempt to validate the re-parameterisation coefficient, a new set of comparison was carried out for an  $R_{rs}$  dataset from Funil reservoir collected by Augusto-Silva et al., (2014).

However, an important point must be considered. This new  $R_{rs}$  dataset did not have an *in situ* set of PC concentration measurements. Therefore, a randomly-generated set of PC concentration values within the same range of those from this study's was created by equating the sample number between this study's sample number and Augusto-Silva et al., (2014) sample number. This was achieved by adding four extra sample values to this study's sample number from the same range of values (0-15  $\mu\text{g/L}$ ) this study encountered. Then, a randomisation procedure was applied to this new dataset in order to allocate each sample value for PC concentration to a  $R_{rs}$  value from Augusto-Silva et al., (2014). At last, this new set of synthetic PC concentration samples had 19 values all within the this study's value range.

Thus, Funil dataset was not evaluated for the actual NRMSE and adjusted  $R^2$  magnitudes but only, for how these two statistical metrics varied between the algorithms, i.e., whether the new-coefficient algorithm had a higher adjusted  $R^2$  and a better NRMSE than the original algorithm. Figure 4.13 depicted Funil reservoir  $R_{rs}$  dataset.

Figure 4.13 –  $R_{rs}$  spectra from Funil reservoir collected in May, 2012 (AUGUSTO-SILVA et al., 2014).



#### 4.5. OLCI/Sentinel 3 simulation

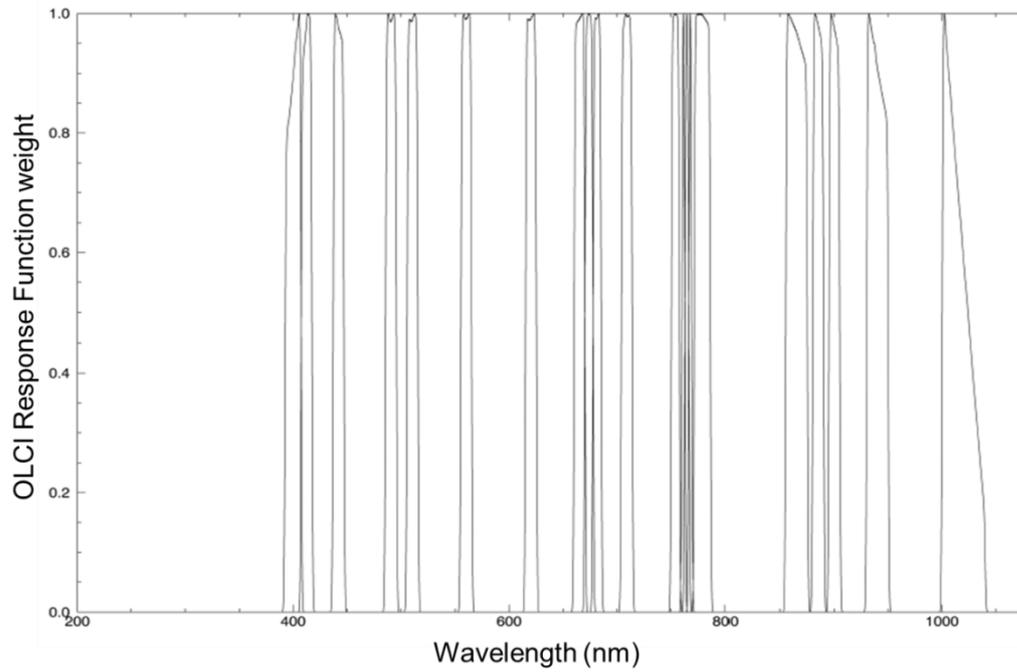
Sensors such as OLCI intend to monitor water quality, and therefore, they need to identify the major interesting spectral features for water quality monitoring. OLCI-sensor resampling from  $R_{rs}$  data was carried out to evaluate whether OLCI bands would represent the spectra peaks and troughs for the OACs absorption and reflectance maxima. To simulate the  $R_{rs}(\lambda_i)$  signals that would be recorded by satellite sensor channels centered at wavelength  $\lambda_i$ , weighted averages of each  $R_{rs}$  spectrum were calculated by using as weights the spectral band response functions of OLCI (ESA, 2016), as given below

$$R_{rs_s}(\lambda) = \frac{\sum_{\lambda} R_{rs}(\lambda) \times S(\lambda)}{\sum_{\lambda} S(\lambda)} \quad (4.22)$$

where  $S(\lambda)$  is the OLCI spectral response function, and  $R_{rs_s}(\lambda)$  is the simulated  $R_{rs}$ . Also, the re-parameterised three-band semi-empirical algorithm reference wavelengths were investigated to whether they are represented by the available OLCI bands. Finally, specific spots within the resampled spectra were evaluated

to how they might misrepresent or imbue errors in this study's calculated  $R_{rs}$  spectra. Figure 4.14 depicted the response functions for OLCI sensor bands.

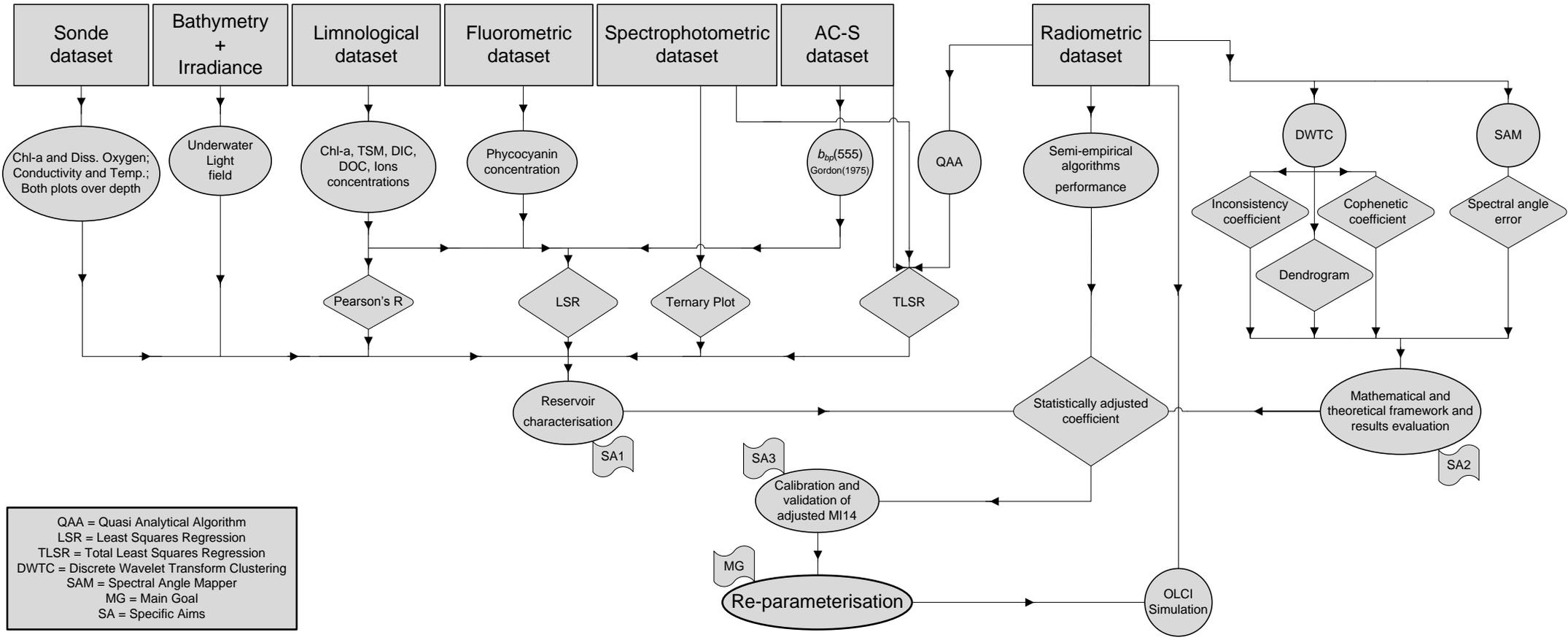
Figure 4.14 – Spectral response functions for OLCI sensor. Twenty one bands over the visible-NIR spectral range (ESA, 2016).



#### 4.6. Methodology Flowchart

Figure 4.15 depicts the methodological flowchart used in this study. On the lefthand side of the chart one observes the series of methodological steps associated with specific aim number one. On the far right side of the chart one notes the steps linked to specific aim number two. And right down the centre, there are the third specific aim steps. All these steps converge towards the main goal and a further extrapolation of the *in situ* radiometric data to OLCI's bands spectral response function.

Figure 4.15 – Methodological flowchart for this study. Square boxes indicate datasets collected or generated for this work. Circular or ellipsoidal shapes represent the objectives-related steps within the methodology. Rhombus represents the main statistics applied to the dataset. Flags represent specific aims and main goal.



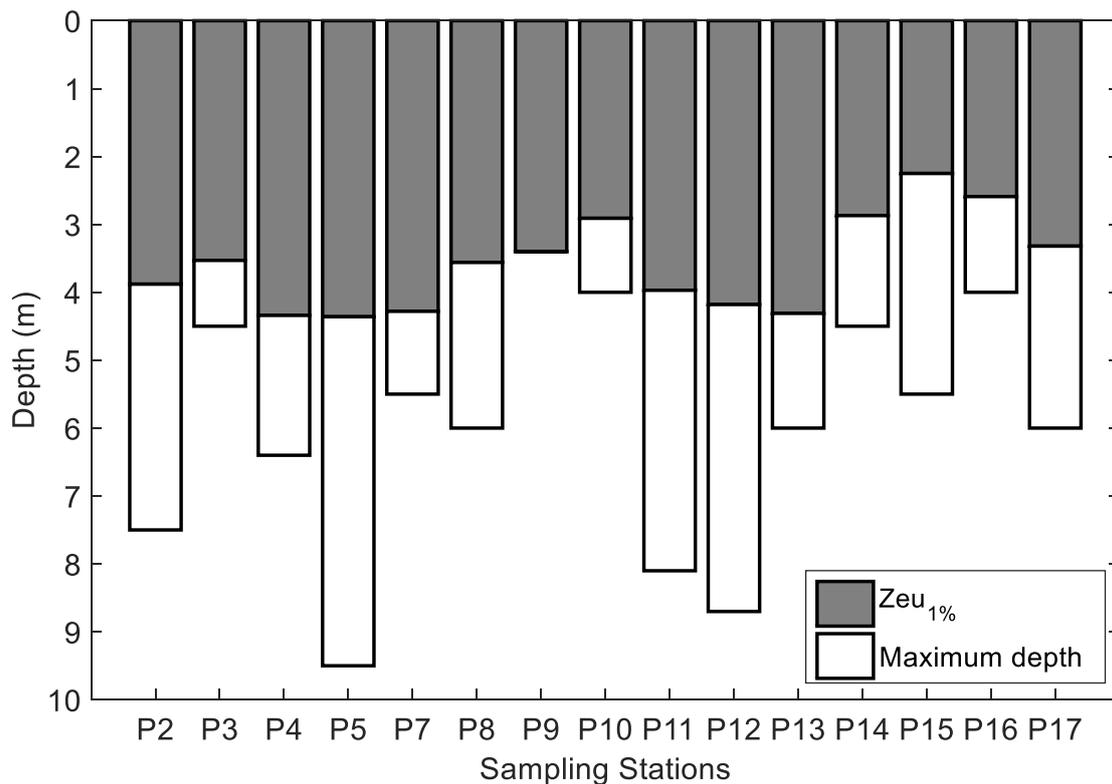


## 5 RESULTS AND DISCUSSION

### 5.1. Bathymetry

Bathymetric results for Guarapiranga reservoir are represented in Figures 5.1, 5.2, and 5.3. Figure 5.1 was a topological representation of each sampling point maximum depth as well as the one-per-cent euphotic zone depth. The solar irradiance that reaches the euphotic zone one-per-cent limit ( $Z_{eu1}$ ) has an average depth just less than four meters. The average and maximum local depths for all points assessed in this study were approximately 5.8 and 9.5 meters, respectively.

Figure 5.1 – Topological bathymetric plot for this study’s individual sampling stations. Grey bars represent the euphotic zone depth where the radiation reaching the water surface extinguishes at the 1% level. The white bars represent the difference between the  $Z_{eu1}$  and the actual local reservoir depth.



The maximum depth estimated by the interpolation was 11.54 meters. CPLEA and COBRAPE (2006) stated that the average and maximum depths for the reservoir are 5.7 and 13 meters. Therefore, the values encountered by this

study correlated well with data collected by CPLEA and COBRAPE (2006). It is important to note that the reservoir volume at the field campaign time was 70% of total volume. Therefore, the difference between the depth values encountered by this study and values provided by CPLEA and COBRAPE (2006) could be linked to that percentage difference.

Figures 5.2 and 5.3 depicted the overall reservoir morphology which dictates its hydraulics. It is noted how the input streams flow from east, south, and southwest into the main water body. Annex B showed how the input streams were under the influence of major socioeconomic areas. Each area was described in section 3. These areas are major sources of the pollution that flows into the reservoir, except area number four which is considered an area of better socioeconomic status and therefore maintains a better sanitary infrastructure, not allowing sewage overflow towards the reservoir.

Four main thalwegs drain stream waters into the reservoir main body (Annex B). These areas cover all five groups suggested by *SABESP* except group four. Summarising the groups already described in section 3, they can be characterised as: group 1, highly polluted lower-urban areas; 2, lower-urban areas; 3, non-urban areas; 4, higher-urban areas; 5, mostly unaltered rural areas. By the description of each group, one can note how all groups act as major pollution dischargers via the four main thalwegs. Then, water is carried from these input locations inwards and northwards. The reservoir body tends to reach a flatter structure at around 6 km from the southernmost part (Figure 5.2) from where it moves towards the dam. Besides the main input streams thalwegs, minor streams occur all the way from the southernmost part up until the northwesternmost locations of the reservoir, especially areas on the western regions.

Therefore, Guarapiranga reservoir dynamic hydraulic system allow for a varied water type at different locations in the reservoir. This characteristic affected the WQP concentration values range distribution. The hydraulics also indicate how

one type of constituent can be generated or found at the input locations where the main input streams thalwegs become the water body itself. And further into the reservoir, the water limnological characteristics can vary according to water treatment strategies and flow dynamics towards the dam at the northernmost limit.

Figure 5.2 – Guarapiranga north-south bathymetry cross-section. Leftmost part of figure represents the southernmost area, and the rightmost part, the northernmost. Different shades of purple represent the various regions along the cross section where the reservoir has more than one depth variation for that latitude.

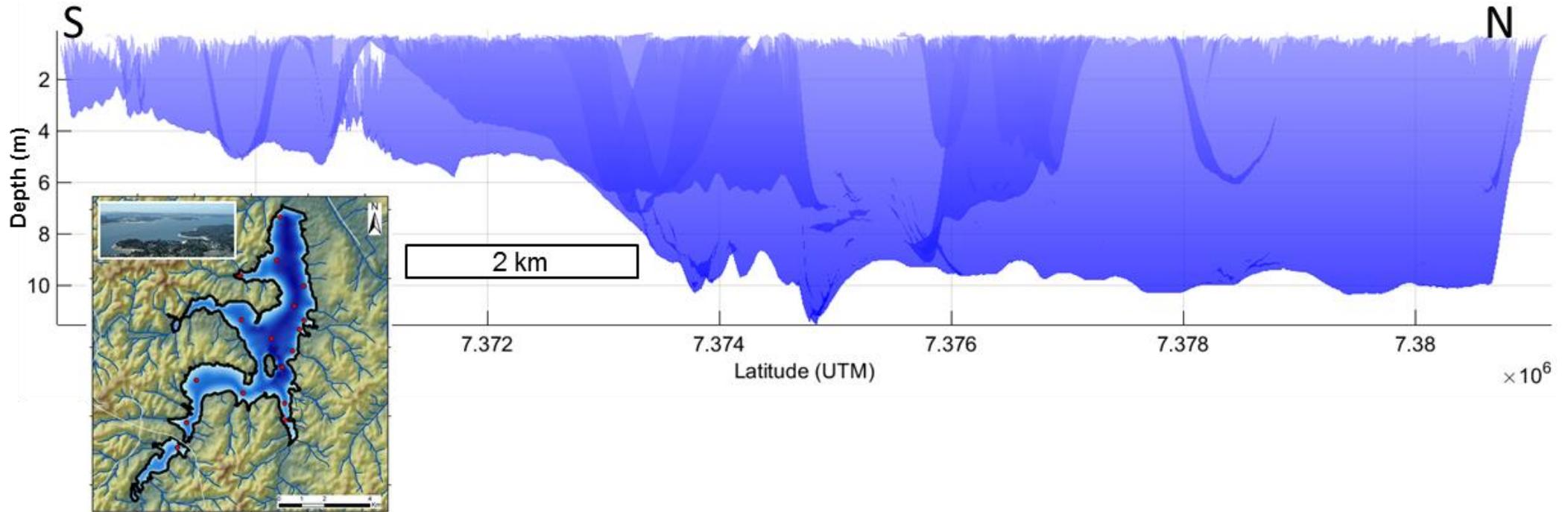
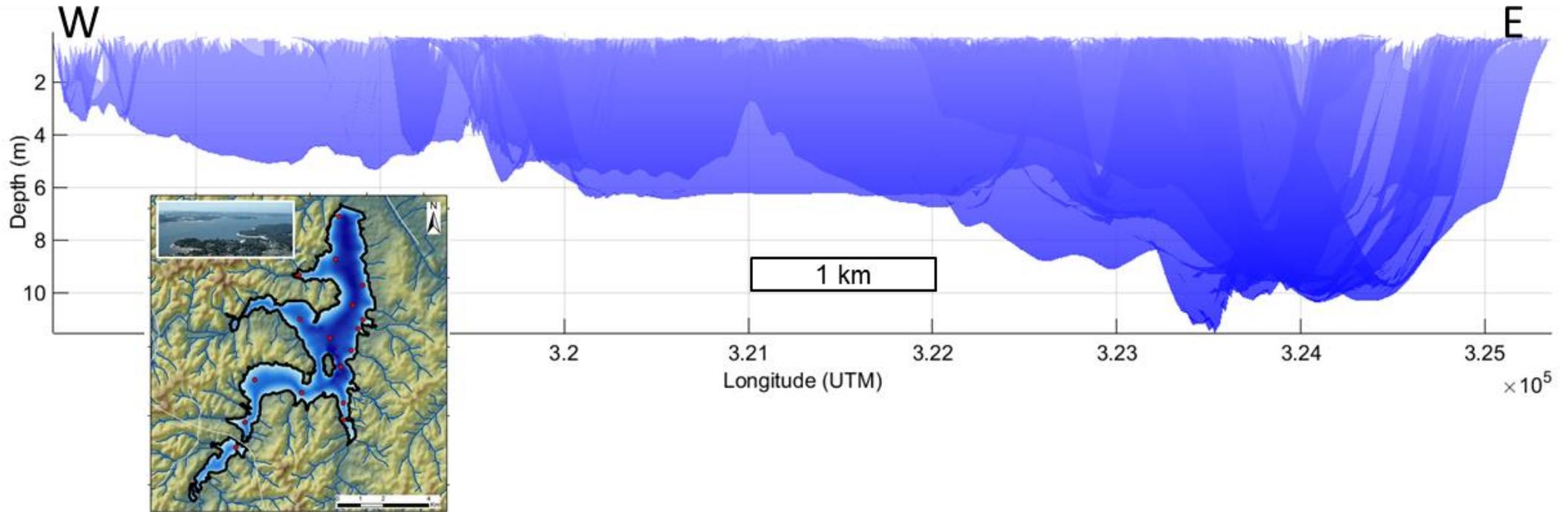


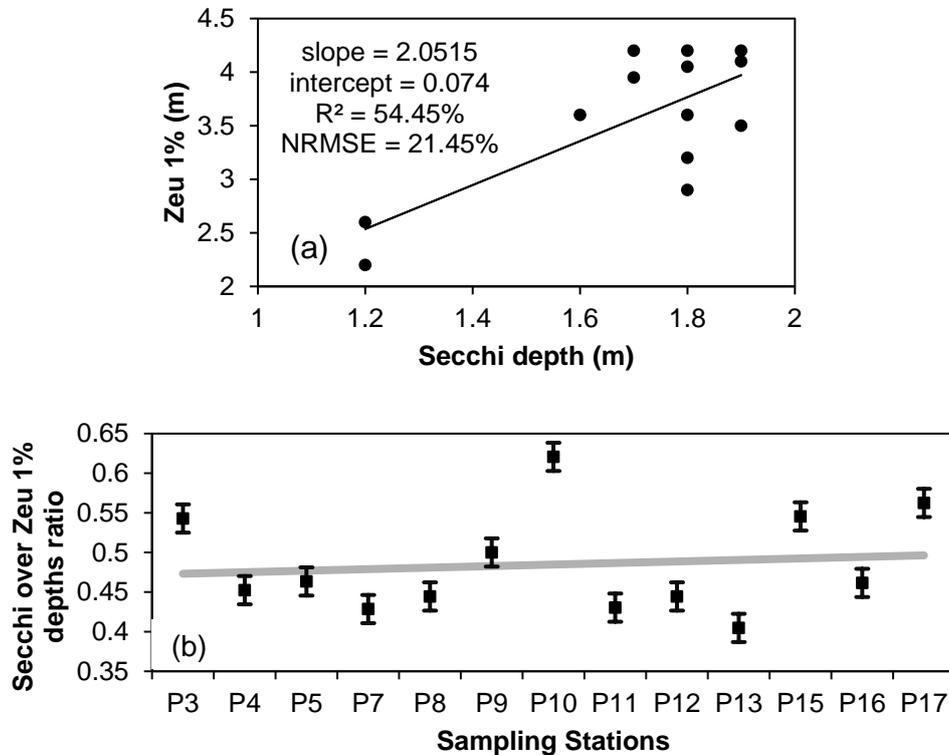
Figure 5.3 – Guarapiranga west-east bathymetry cross-section. Leftmost part of figure represents the westernmost area, and the rightmost part, the easternmost. Different shades of purple represent the various regions along the cross section where the reservoir has more than one depth variation for that longitude.



### 5.1.1. Underwater Light Field

Overall, secchi depth for points measured correlated well with  $Z_{eu1}$  (Figure 5.4a). Applying Beer's Law to the Secchi depth over  $Z_{eu1}$  depth ratio, and assuming a homogeneous water column for those depths, one gets an average of 48.47% (half) of the light energy that reaches  $Z_{eu1}$  being backscattered to above-surface level from Secchi depth, and being perceivable to human sight (Figure 5.4b).

Figure 5.4 – (a) Regression analysis between secchi depth and one-per-cent euphotic zone depth  $Z_{eu1}$ . (b) Secchi depth and  $Z_{eu1}$  ratio for all points with standard error bars. Trend line indicates the average ratio value of 48.47%.



Underwater light field (ULF) depicted how the EMR available at the above-surface ( $0^+$ ) behaves after it penetrates the water column. ULF expresses the amount of radiation or energy per unit wavelength or integrated over the PAR EMR region available within the water body. Overall, the depth reaching  $Z_{eu1}$  by this available radiation over the PAR EMR region is expressed as a yellow line and shaded-grey bars in Figures 5.5 and 5.1, respectively. The lower bottom of

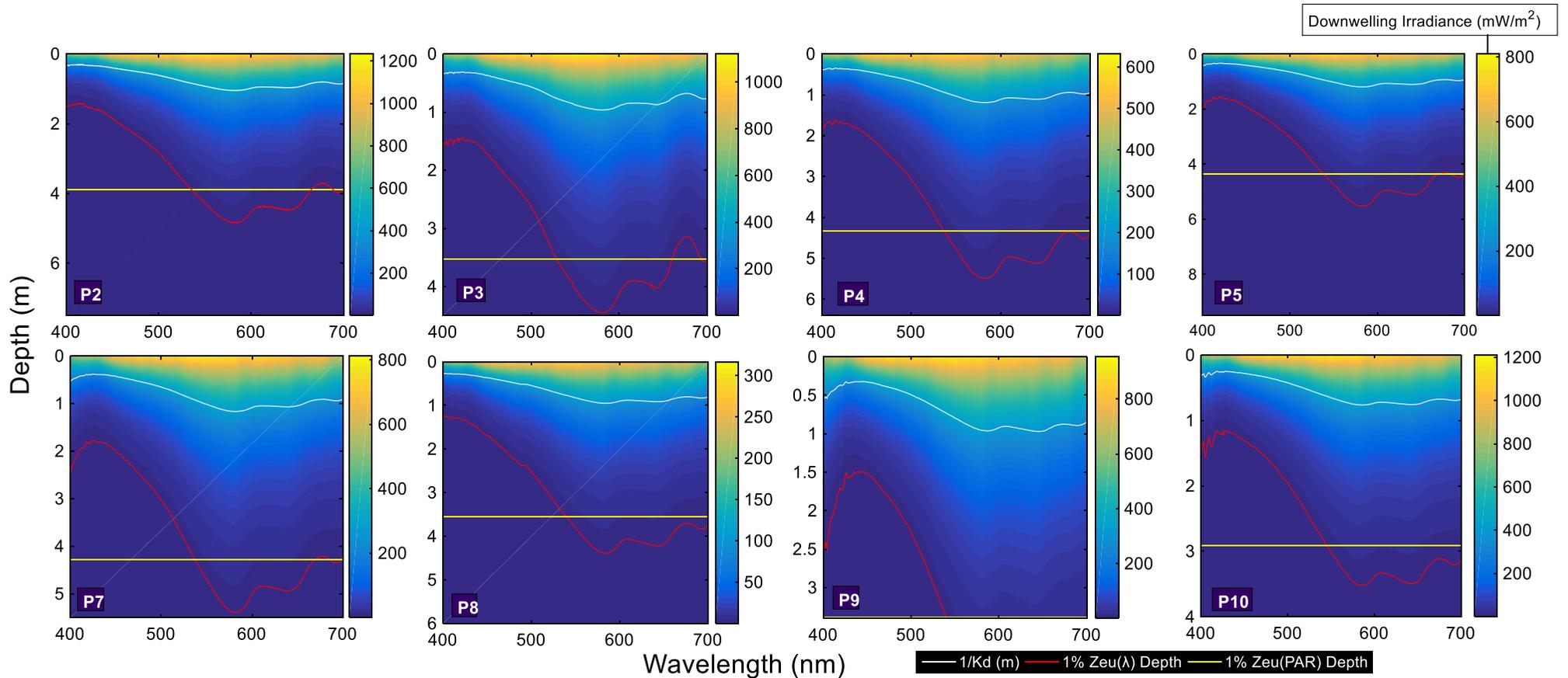
each sampling point coloured-shaded subplot in Figure 5.5 represented the maximum depth for that point.

Higher initial irradiance values do not imply a higher  $Z_{eu1}$  depth.  $K_d$  determines how quickly light gets extinct down the water column.  $K_d$  varies for each sampling point according to the OACs concentrations, and an investigation on how IOPs correlate with WQPs is carried out in section 5.7. These correlations act as proxy for  $K_d$  variability over all sampled points.

Taking each point's average  $K_d$  calculated from each irradiance-measured depth, the ULF for each point was calculated per unit wavelength, and integrated over the PAR region for  $Z_{eu1}$  (Figure 5.5). Also, the reciprocal of  $K_d$  in units of meters was calculated and plotted per sampling point per unit wavelength as the white line in Figure 5.5. This line provides information on the average photosynthetic pigment concentration – IOP-related - for the upper 22% of the homogeneous euphotic zone depth (KIRK, 2011); such depth comprising 36.8% of surface irradiance value (STRAMSKA; STRAMSKI, 2005); and being the depth reached by remote sensors ( $Z_{90}$ ) – AOP-related - (GORDON; MCCLUNEY, 1975b; HIDALGO-GONZÁLEZ; ALVAREZ-BORREGO, 2001).

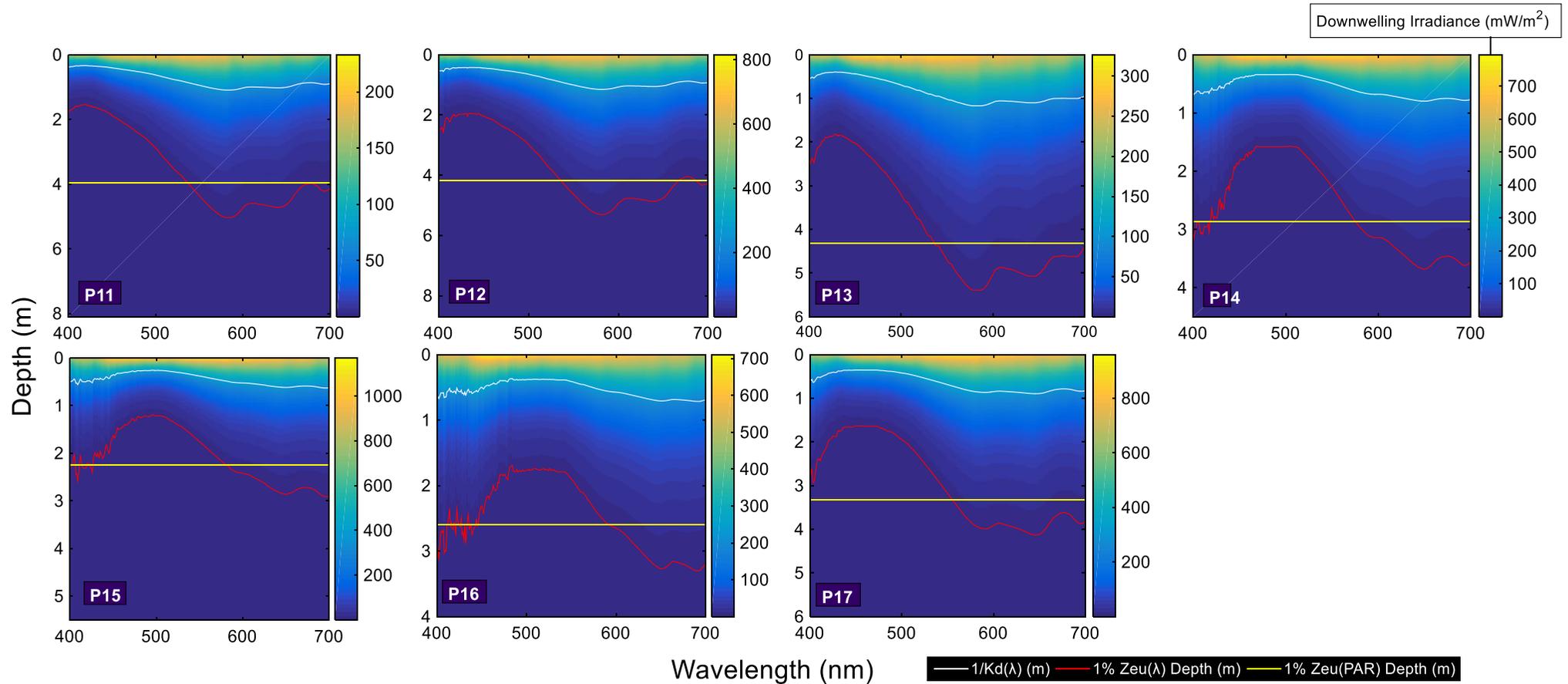
Red lines in Figure 5.5 accounted for  $Z_{eu1}$  depth per unit wavelength. Integrating over the PAR EMR spectrum region, one observes what is depicted as yellow lines in Figure 5.5. Sampling station nine has its yellow line equal to its maximum bathymetric depth. One can identify by this feature that some wavelengths can reach bottom substrate and some other wavelengths cannot. Therefore, more of useful photosynthetic-active wavelengths (e.g., red-wavelength region) might be available for lake bottom vegetation at distances from reservoir margin such as that of point nine (southeasternmost sampled point, at 150 m from reservoirs west and eastern margins). However, it must be considered the reservoir bottom slope. Varying bottom slopes might make same distances from margin have different maximum depths.

Figure 5.5 – Irradiance extinction depth per unit wavelength and integrated over the PAR region per sampling point. White line represents the reciprocal of averaged  $K_d$  per unit wavelength for all irradiance-measured depths. Red line represents the irradiance per unit wavelength extinguished by one per cent of above-water ( $0^+$ ) irradiance values. Yellow line represents the integrated  $Z_{eu1}$  values.



Note: Two samples (sampling points 1 and 6) presented error during field campaign acquisition and were exclude from data processing.

Figure 5.5 – Irradiance extinction depth per unit wavelength and integrated over the PAR region per sampling point. White line represents the reciprocal of averaged  $K_d$  per unit wavelength for all irradiance-measured depths. Red line represents the irradiance per unit wavelength extinguished by one per cent of above-water ( $0^+$ ) irradiance values. Yellow line represents the integrated  $Z_{eu1}$  values – Continuation.

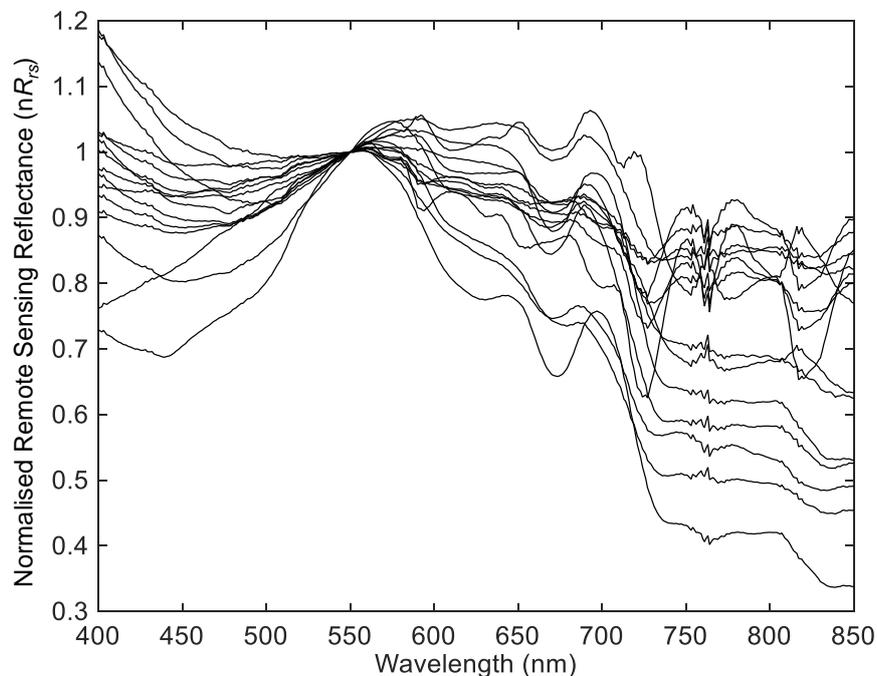


Note: Two samples presented error during field campaign acquisition and were excluded from data processing.

## 5.2. AOP dataset

The original  $R_{rs}$  spectra calculated for this study had a highly varying magnitude. Therefore, the spectra were normalised at the 550 nm in order to assess its spectral behaviour or shape more efficiently (Figure 5.6). A notable aspect of these spectra is how the slope from 400 nm towards the 550 nm varies significantly. Some spectra have a negative slope, some others have a positive slope, and others have a zero-value (flat) slope. However, one can note that the higher the slope, either positive or negative (the more negative the more ocean-like shape it seemed to be), the greater the absorption local maximum within the 450-490 nm. The closer to zero the slope, the shallower this absorption maximum – this is true except for one spectrum with a straight positive slope from 400 till 550 nm. This slope feature can be an indication of a constituent (e.g., chl-*a*) absorption in this range, and it could be considered a confusion factor in the estimation of other constituents such as PC.

Figure 5.6 – Normalised remote sensing reflectance ( $nR_{rs}$ ) spectra at 550 nm for fifteen samples.



Note: Two samples presented error during field campaign acquisition and were excluded from further analysis.

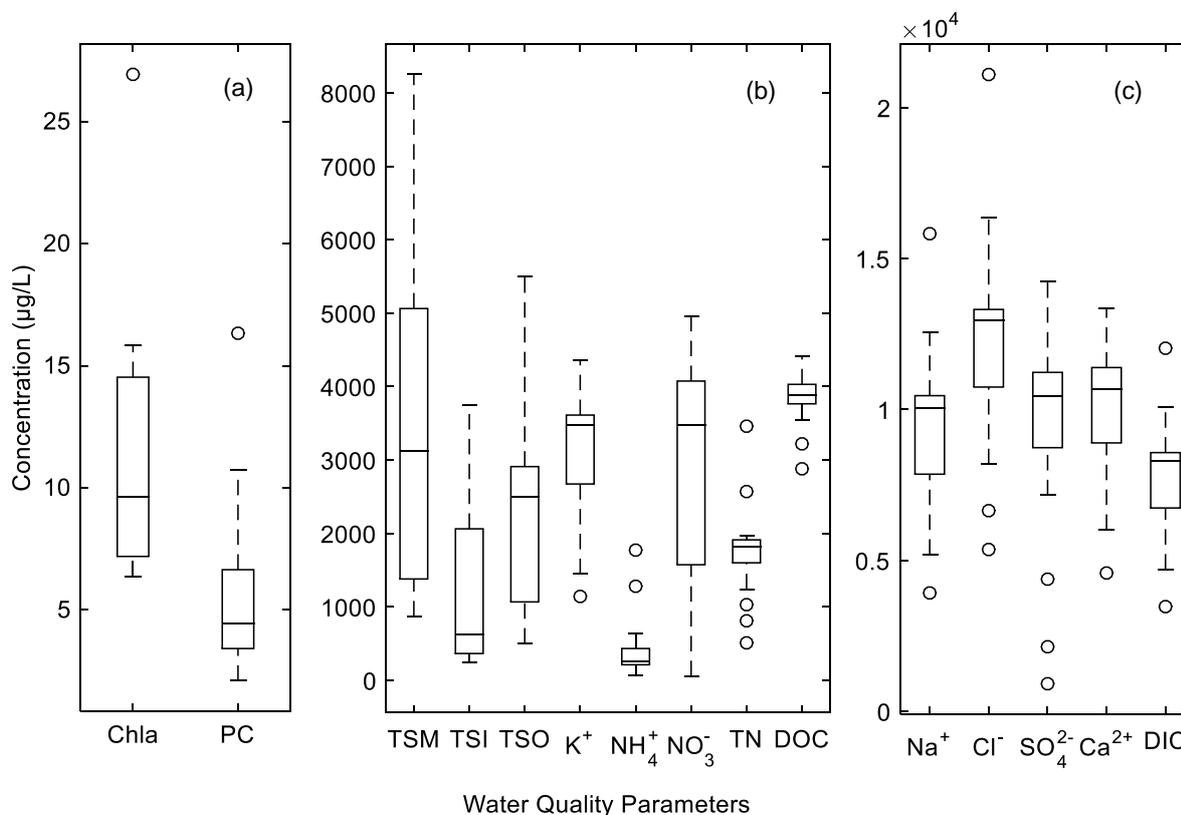
Further down the wavelength range from the minimum absorption at 550 nm, the spectra presented features such as a plateau within the red part of the spectrum for points in which the overall TSM concentration was higher to features. But also, the spectra presented features resembling lower magnitudes in the red part of spectrum. This was so for spectra with 675-nm and 620-nm absorption maximums quite high, indicating the presence of both chl-a and PC. Beyond the need to identify specific features within the spectra's behaviour, the normalisation was carried out in order to minimise the  $R_{rs}$  magnitude variation over the DWTCA processing. This will be further discussed in section 5.9.

### 5.3 Limnological and Fluorometric Dataset

Some limnological WQPs are major water agents which characterise OACs. WQPs samples distribution for this study was graphed as boxplots (Figure 5.7). Boxplots condense most of useful information regarding samples characteristics such as minimum, maximum values, median, outliers and kurtoses (KIRK, 2008). Also, distribution patterns in samples concentrations could also be identified more easily via these plots.

While considering the general statistics of all variables sample distribution, one notes that only three variables, ammonium ( $\text{NH}_4^+$ ), TN, and DOC, presented higher kurtoses. Even so, they presented outliers Figure 5.7b. Further investigating Figure 5.7b one notes that all variables present a sparse sample distribution but only those with higher kurtoses generated statistical outliers. Outliers and kurtoses for variables within Figure 5.7a presented similar patterns between each other and seemed to have proportional value range. This feature indicates some level of correlation between PC and chl-a concentrations which suggests the need for an appropriate de-correlation procedure while predicting PC via bio-optical modelling. Figure 5.7c variables presented lower kurtoses and some outliers unlike Figure 5.7b. Such feature is most likely due to the higher concentration range associated with Figure 5.7c variables, increasing the likelihood of extreme values development.

Figure 5.7 – Limnological water quality parameters boxplots for all samples.



Notes: Phosphates were measured in laboratory for this study, however the concentrations were so low that the IC machine did not account for it in an appropriate manner and the results were discarded. Total phosphorus was not measured due to technical difficulties.

An important feature in Figure 5.7 was the fact it generated three groups separated as Figure 5.7a, b, and c in an increasing concentration order. Figure 5.7a group was linked to two major OACs (chl-a and PC). Figure 5.7c group, diametrically opposed to Figure 5.7a concentrationwise, contained mostly ions linked to water treatment and DIC. Figure 5.7b contained TSM, TSO, total TSI, TN, and DOC.

*SABESP* treats the water stored within the reservoir in order to control HABs and other pathogens that might contaminate the water supply. Algaecides (sulphate-based and chloride-based compounds) are very commonly used in water treatment due to their oxidising properties (MOSCHINI-CARLOS; FREITAS, 2009; MOSCHINI-CARLOS et al., 2010). Accordingly, this study

identified high concentration levels of sulphates ( $\text{SO}_4^{2-}$ ) and chloride ( $\text{Cl}^-$ ) as well as the metals calcium ( $\text{Ca}^{2+}$ ) and sodium ( $\text{Na}^+$ ) which are cations commonly associated with the previously mentioned anions. And as a consequence, low concentrations of chl-a and PC. Despite the fact that water is being treated locally, toxins can be released as an aftermath of the rapid decay of algal blooms takes place. Therefore, the most appropriate measures to maintain high levels of water quality would be to reduce nutrient inputs to water catchments (KALLIS, 2001). This can be achieved by applying phosphate removal processes into sewage-loaded stream branches.

Figure 5.7b variables could be associated with organic compounds since both dissolved organic carbon and organic solids are in this group. As a stepwise way to consolidate this affirmation, one can take the following inductive-deductive approach: the higher the organic matter concentration, the more food for bacteria to feed on and convert ammonium to nitrate (INGESTAD; ÅGREN, 1995). Therefore, higher levels of nitrate and lower levels of ammonium should occur, following that process. And that is what has been identified by data from this study. It is important to note that this bacterial activity most likely occurs in soil due to agricultural practices. Then, these activities by-products (including potassium ( $\text{K}^+$ ) and nitrates ( $\text{NO}_3^-$ ) as fertilisers) are washed away into the reservoir.

Another account on limnological parameters should be done by exposing the correlation between total nitrogen to total phosphorus ratio (N:P) among chl-a and PC concentrations. Literature on the subject (BULGAKOV; LEVICH, 1999; DE HOYOS et al., 2004) shows that the lower the N:P the higher the cyanobacterial dominance. This study found the reciprocal of an equivalent ratio (higher  $\text{N}:\text{PO}_4^{3-}$ ) with low levels of phosphates and high levels of TN. A higher  $\text{N}:\text{PO}_4^{3-}$  ratio can be a proxy for the cyanobacteria non-dominance in that environment. Again, low measured PC concentrations sustain this affirmation since PC acts as a known proxy for cyanobacteria presence. Total phosphorus could not be measured due to technical difficulties on laboratory.

Following Carlson (1977) and Lamparelli (2004) the chl-a TSIx averaged out for the entire reservoir was 62.25. This value places the reservoir in a eutrophic water body category. Instead, the Secchi TSIx rated 52.63 placing the reservoir in the oligotrophic category but right about the mesotrophic threshold. Usually, a third TSIx for phosphorus would be considered and a mean value calculated for the overall TSIx. However, very low levels of phosphates (phosphorus-proxy) could not even be accounted for in laboratory measurements. And this could indicate that the reservoir phosphorus TSIx was ultraoligotrophic (lowest possible). A mean TSIx value for Guarapiranga reservoir for April, 2015 would, therefore, place it somewhere between the oligotrophic and mesotrophic categories. This means the reservoir is characterised as a clean or intermediate productivity water body, with none to possible implications for the reservoir quality (CETESB, 2007).

### 5.3.1 Correlation Matrices

The basis of variable dependence is to identify positive and negative correlations that indicate these variables might have a common relation. A basic statistics practice is not to infer causation, but identify variable correlations over mutual influential drivers. Skewness is an important statistical sample distribution parameter for correlation analyses since it may influence the data correlation linearity. That was also considered for the correlation coefficients chosen in this study. Thence, this section evaluated the degree of correlation among all WQP variables.

Figure 5.8 presented a shaded-colour correlogram among all WQPs, and Figure 5.9 depicts the correlation between  $\text{NH}_4^+$  and all other WQP for each correlation coefficient. Three different correlation coefficients were used (Figure 5.8a, b, and c). The choice for applying the different correlation coefficients was based on the elaboration given in section 4.3. Overall, one notes that Figures 5.8a, Figures 5.8b, and Figures 5.8c had the same general correlation patterns. However, one also notes that Kendall ( $\tau$ ), Spearman ( $\rho$ ), and Pearson (R)

coefficients have increasing correlation strengths in that order respectively.  $\tau$ ,  $\rho$  are ranked coefficients; moreover,  $\tau$  is a non-parametric test that does not make any assumptions on the variables distributions. R is not a ranked coefficient and has parametric assumptions. These qualities make R capable of generating more robust results if compared to  $\tau$ , and  $\rho$ . However, two assumptions need to be met for that to be true: robustness (evaluated by inspecting the scatterplot) and outlier resistance (evaluated by inspecting the histogram) (WILCOX, 2005). So, the variables, in this study, that matched those assumptions (see Figure C.1 scatterplots) yielded stronger R than  $\tau$  and  $\rho$ .

Figure 5.8 – Kendall (a), Spearman (b), and Pearson (c) correlation coefficients for limnological water quality parameters.  $\text{NH}_4^+$  was removed from this figure due to difference in sample numbers.

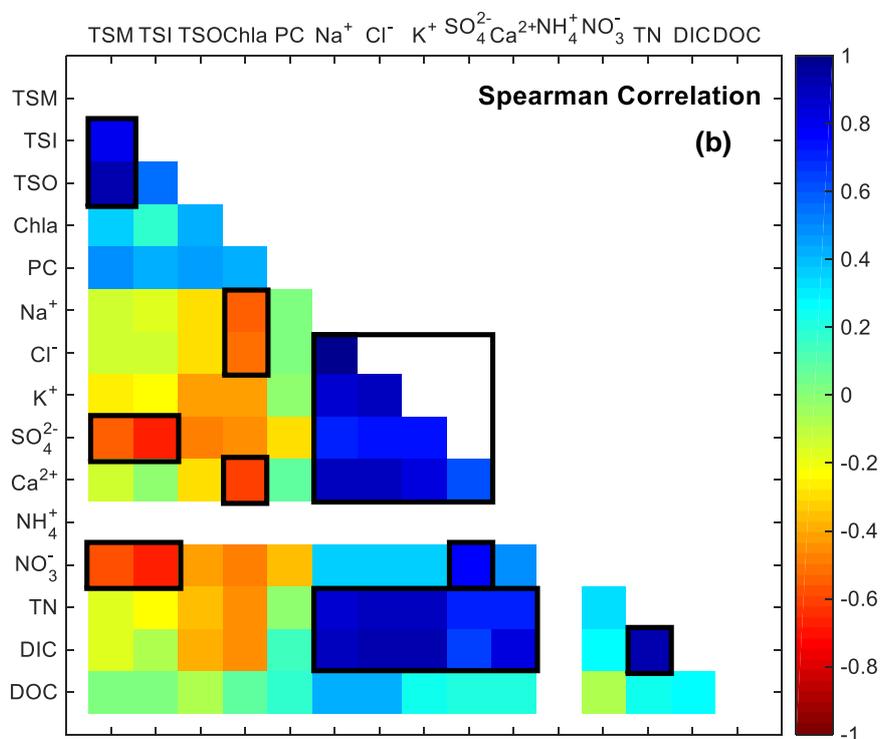
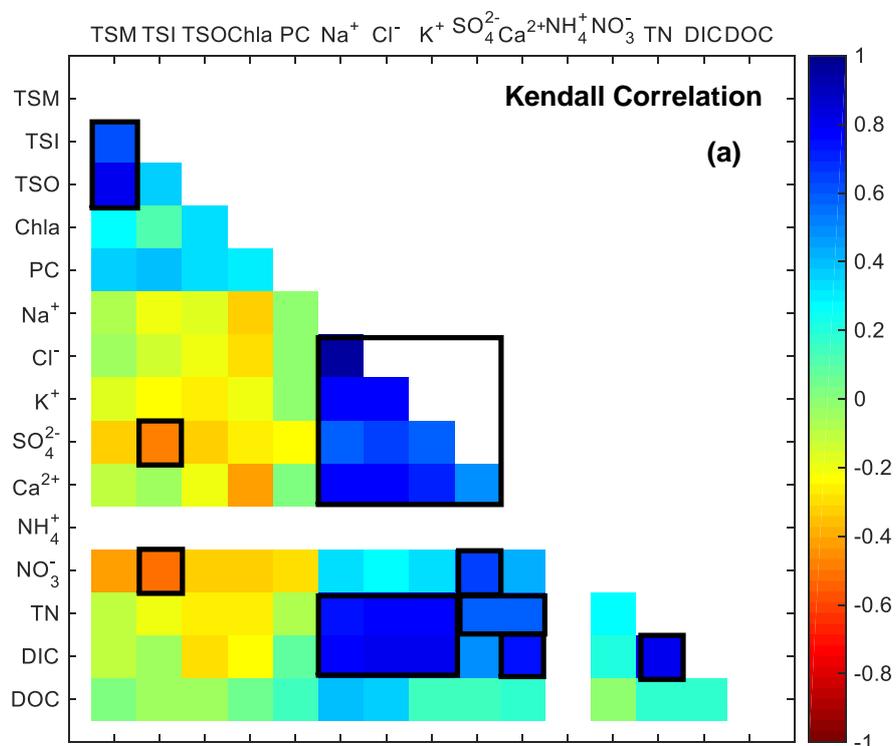


Figure 5.8 – Kendall (a), Spearman (b), and Pearson (c) correlation coefficients for limnological water quality parameters.  $\text{NH}_4^+$  was removed from this figure due to difference in sample numbers. – Continuation

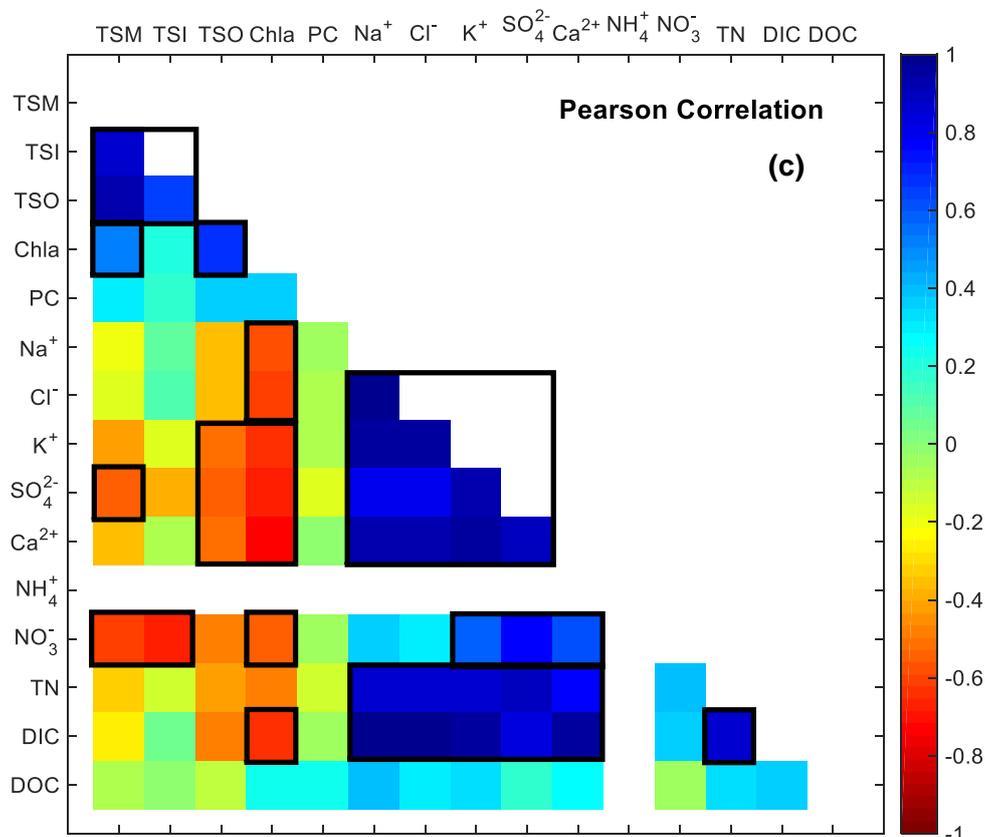
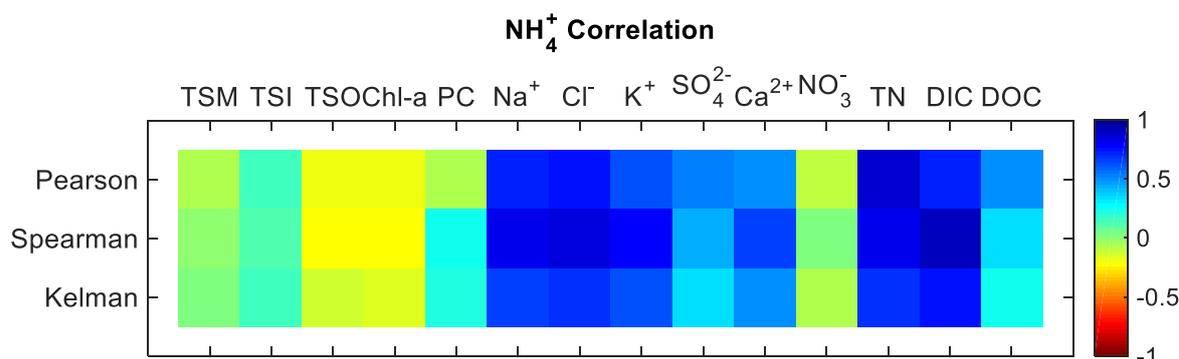


Figure 5.9 – Kendall, Spearman, and Pearson correlation coefficients for  $\text{NH}_4^+$  against all other parameters.  $\text{NH}_4^+$  had three inappropriate samples; therefore, the whole dataset was adjusted to the  $\text{NH}_4^+$  sample range (fourteen rather than seventeen samples).



An important point regarding the apparent overall monotonic increase in correlation robustness from  $\tau$  through  $\rho$  till R is that  $\rho$  for  $\text{SO}_4^{2-}$  and TSI

produced a better negative correlation than R, reversing the monotonic trend. That happened as an aftermath of the non-linear relation between the two variables samples distribution. Despite the fact, R yielded overall more robust correlations,  $\rho$  also presented better results for variables correlations such those for  $\text{NH}_4^+$  and  $\text{Cl}^-$ ,  $\text{Na}^+$ ,  $\text{K}^+$ , DIC which also presented a non-linear relation, with some degree of curve saturation trend (Figure C.1). Once again, this is a finding in accordance with the literature background information.

Considering positive dependencies, two main groups were found: first, TSM and phytoplankton-related pigments; second, inorganic carbon, nitrogen-based ions and TN, and water treatment ions. The first group further indicates as pointed in the previous section that most suspended particles are primarily organic in nature. As a consequence chl-*a* has a high correlation with TSO. PC follows the same pattern as chl-*a* but with weaker correlation, most likely due to low concentrations which allow for more random sampling distribution which in turn affects the scatterplot correlation.

Chl-*a* and PC presented a weaker correlation ( $<0.4$ ). This result indicated that despite their sampling distribution being similar, neither of them comes from a same biological nor location source, and perhaps from neither biological nor location source. This further corroborated the fact that the hydraulic dynamics of the reservoir carries the OACs within it from one place to another as mentioned previously in section 5.1, imbuing some degree of randomization in the analysed variables correlation.

PC presented very clear variations in  $\tau$ ,  $\rho$ , and R values against TSM, TSI, and TSO. Figure C.1 allowed one to infer that PC against TSM and TSO scatterplots presented a non-linear behaviour. PC against TSI presented outliers but nevertheless an overall linear relation. Should these statements well represent the samples distributions, one can infer that  $\tau$  yielded a better correlation for PC and TSI since this scatterplot had a clear outlier and a linear relation; also, one can infer that  $\rho$  yielded a better correlation for PC against TSM and TSO since

their scatterplots presented a non-linear behaviour with no extremely influential outlier, and that sample distribution feature could be better represented by  $\rho$ . R has some strong parametric assumptions, which could not be met by these samples distributions, and therefore, presented a weaker correlation value.

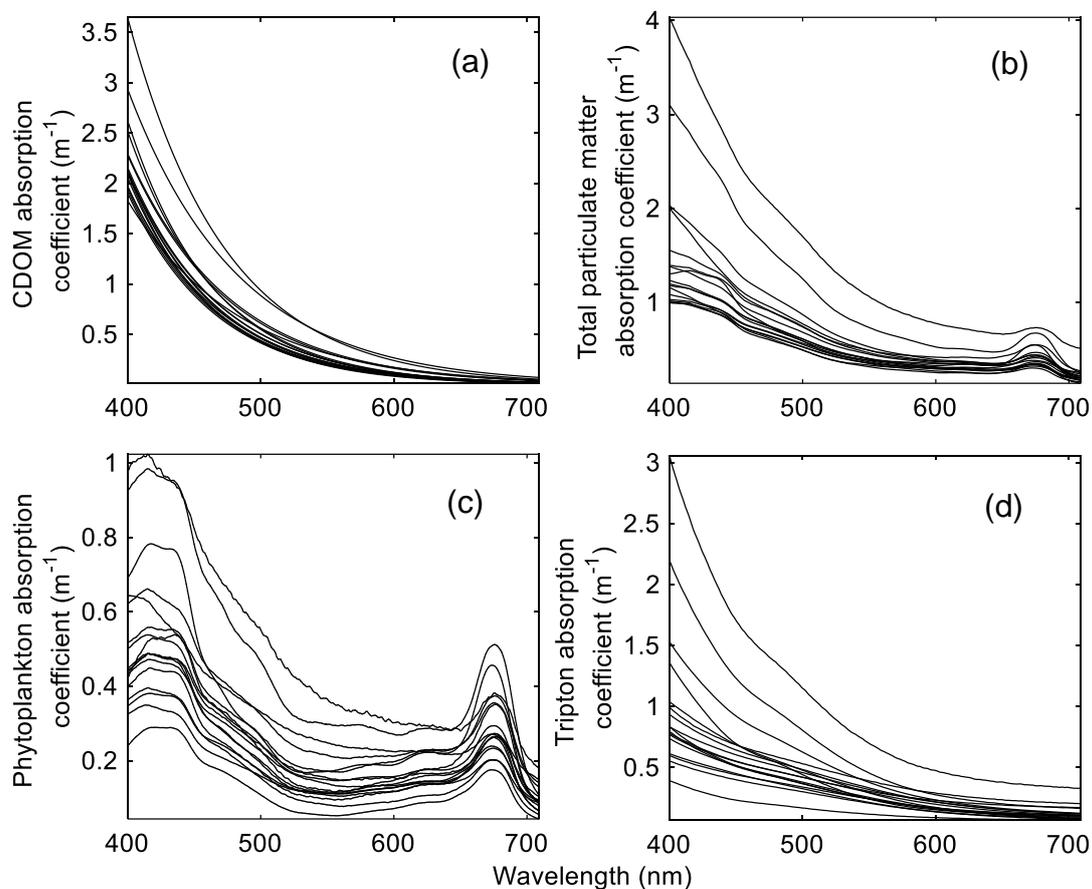
The second group is linked to water treatment compounds, TN and nitrogen ions, and inorganic carbon. The water treatment ions are highly correlated because they come from same chemical sources to treat the water. Nitrogen-based compounds also correlated well with  $\text{Na}^+$ ,  $\text{Cl}^-$ ,  $\text{K}^+$ ,  $\text{SO}_4^{2-}$ , and  $\text{Ca}^{2+}$ . Since  $\text{K}^+$  and  $\text{NO}_3^-$  originate from fertilisers and are washed into the water body, these ions might come from agricultural sources. Also, DIC correlates well with these ions, indicating that carbonate minerals, along with fertiliser-based and sewage-generated compounds, make their way into the water body, getting dissolved once in it (WETZEL, 2001; CPLEA; COBRAPE, 2006).

Considering the negative dependencies, the higher correlation values are those among water treatment compounds and phytoplankton-related pigments. If a lake has a high photosynthetic rate, a decrease occurs in  $\text{CO}_2$  (TSI proxy) which is used up in the process. The less  $\text{CO}_2$ , the more alkaline and less acidic the water body (higher pH). The higher pH favours the decomposition of organic matter (MCKINLEY; VESTAL, 1982). And as mentioned previously, the higher the decomposition, the less organic matter for bacteria to carry out  $\text{NH}_4^+$  conversion into  $\text{NO}_3^-$ . Therefore, the higher the TSI concentration in a photosynthetic-active lake, the less  $\text{NO}_3^-$  generated within the lake biogeochemical dynamics. And that sort of negative dependence was also encountered in this study. Finally, individual scatterplots for all correlations can be found in Figure C.1, and tables for all correlations can be found in Appendix D containing correlation values as well as their p-values.

## 5.4 Laboratory spectrophotometric dataset

Figure 5.10a showed the  $a_{CDOM}$  spectra for Guarapiranga Reservoir. The spectra seemed similar to other  $a_{CDOM}$  spectra from literature, following the typical shape with high absorption coefficients in the blue spectral region, followed by a rapid decay in the green spectral region, and continuing to decrease towards longer wavelengths but at lower rates. Figure 5.10c showed  $a_{phy}$  spectra in which it was observed not only a high absorption peak around 675 nm but also a short peak around 620 nm because of PC absorption. Figure 5.10d showed the  $a_{nap}$  spectra. As described in the literature the  $a_{nap}$  spectra follows the same shape of  $a_{CDOM}$  backing the idea that blue and green spectral regions are influenced by these two components. Figure 5.10b was the total particulate matter absorption coefficient and represents the sum of  $a_{phy}$  and  $a_{nap}$ .

Figure 5.10 – Laboratory-calculated CDOM (a), total (b), phytoplankton (c), and non-algal particles (tripton) (d) absorption coefficients.

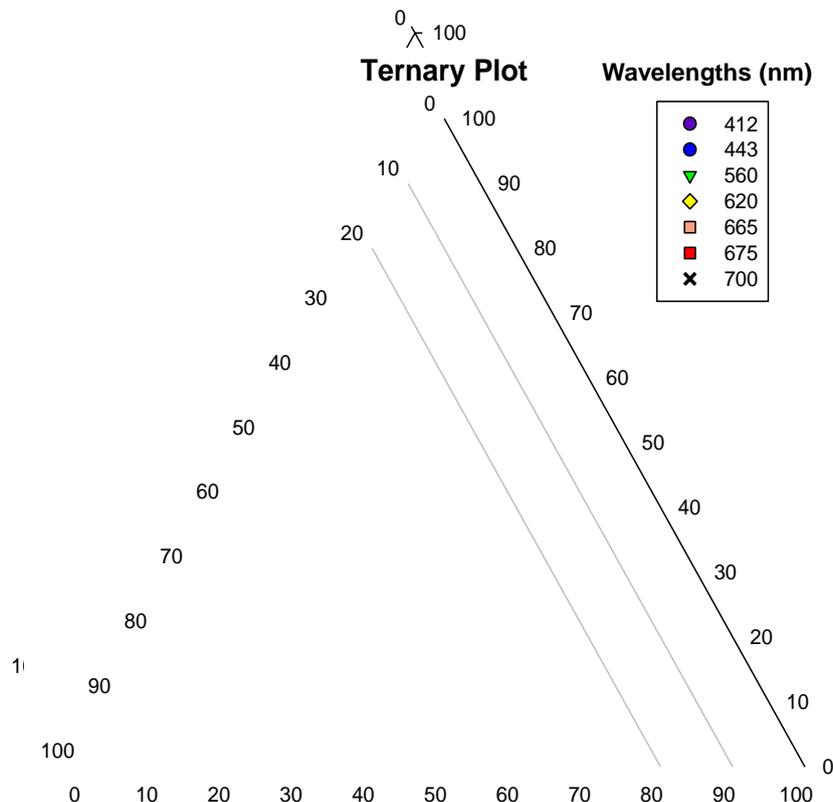


### 5.4.1 Ternary Plot

Ternary plot is a graphical representation of three variables being evaluated for one or more aspects. Its purpose is to correlate the relative magnitude of each of the variables aspects as to how each variable - in this study, three out of the four major water bodies OACs relative absorption coefficients ( $a_{phy}$ ,  $a_{CDOM}$ , and  $a_{nap}$ ) - contributes to the total relative sum. This study's ternary plot shows seven different variable aspects. They are the wavelengths from the EMR visible-NIR range (BABIN et al., 2003).

The wavelengths used in the ternary plot (Figure 5.11) were closely linked to the OACs considered in this study. Each colour in the ternary plot was approximately associated with its respective EMR wavelength, in order to facilitate visual interpretation of the graph.

Figure 5.11 – Ternary plot representing the relative absorption coefficients for phytoplankton ( $a_{phy}$ ), tripton ( $a_{nap}$ ), and CDOM ( $a_{CDOM}$ ). Wavelengths (nm) represent the ones most used in bio-optical algorithms, imaging satellite sensors bands, and OAC-related minimum/maximum absorption bands.



The shorter 412-443 nm wavelengths showed the relative dominance of CDOM, and samples presented themselves with a reasonably circular cohesive distribution. The latter meaning all samples having approximately the same distance from the mean value within the ternary plot sample distribution. 665 and 675 nm presented a tendency to fall under phytoplankton absorption dominance. 700 nm started a distribution that ranged from phytoplankton absorption dominance to a no-OAC dominance. The fact the 620 nm did not resemble entirely the phytoplankton dominance indicated - as can be observed by closely investigating Figure 5.10c - that waters bearing low PC concentrations can shift the actual reference wavelength to a different location (in this study it shifted slightly forward near the 630 nm position) in the wavelength domain.

Literature on this topic matches the OAC dominance on the wavelengths depicted in Figure 5.11. CDOM and phytoplankton tended to absorb EMR more intensely in wavelengths 443 and 675 nm, respectively. However, it is a relative absorption rather than the absolute one linked to the actual OACs concentration (Figure 5.7). High CDOM absorption relative dominance at 443 nm has been observed by (BRICAUD et al., 1981; KUTSER et al., 2005; BREZONIK et al., 2015). This absorption occurs as a consequence of optically-active dissolved organic matter originated from natural and industrial land discharges (BRICAUD et al., 1981).

High phytoplankton absorption relative dominance at 675 nm is associated with high chl-a pigment concentration, which absorbs most of the incoming EMR if compared to tripton (non-algal particles) and CDOM in that wavelength. It is important to note how in the shorter wavelengths (412 and 443 nm) relative proportion was fairly distributed around the ternary plot mean value for those wavelengths; tripton and phytoplankton relative absorption contributions were 10% to 30% and 10% to 30%, respectively, for the 412 and 443 nm wavelengths. Considering the lower frequency wavelengths (665 and 675 nm), relative proportion was not as homogeneously distributed around mean value;

they ranged from 10% to 40% and from 40% to 85%, respectively, for tripton and phytoplankton. It is important to remember the relative absorption ternary plot does not include the fourth major OAC: pure water. And this means that a high contribution from a certain ternary plot OAC does not necessarily imply that sample water is highly loaded with that specific OAC.

As mentioned in the previous paragraph, absorption in the ternary plot for 443 nm presented a more homogeneous distribution around the mean. This homogeneity cohesiveness instigated this study on the possibility of using the higher frequency wavelengths in the re-parameterisation proposed in this study. And, thence, evaluate whether this spectral range absorptivity cohesiveness, indeed could be a graphical representation of the more precise OAC confusion improver coefficient for discriminating between chl-a and PC.

## 5.5 AC-S plots

Laboratory measurements of absorption coefficients can be further decomposed into the OACs contribution. Instead, *in situ* absorption coefficients measurements via AC-S instrument acquire only total absorption coefficients via the absorption tube (WETLABS, 2013) and the total attenuation coefficients via the attenuation tube (WETLABS, 2013). Both these coefficients allowed the absorption and attenuation (Figures 5.12 and 5.13) vertical profiles generation for Guarapiranga reservoir. These profiles were created for five wavelengths within visible-NIR range (i.e., blue, cyan, green, red, and black (NIR)). Also, absorption plots (Figure 5.14) for the 400-700 nm PAR range for five different scattering correction methods were generated for the first measurement depth (approximately 40-cm depth).

Figure 5.12 depicted total absorption coefficients for the fifteen available sampling stations measurements. Overall, total absorption for each station had larger values for shorter wavelengths. As wavelengths got longer, absorption values decreased exponentially, which is a known feature in the literature. The absorption magnitudes did not vary significantly over the vertical profile for all points. Only a few sampling points presented abrupt variations or non-linear variations on absorption magnitude over depth. Also, these variations seemed to be weaker for longer wavelengths.

P13, P15, P16 represented the points with fairly higher variations over larger depths (Figure 5.12). P13 and P16 had slightly opposite variation in which absorption magnitudes moved to higher and lower values, respectively. P15 presented a slightly more unstable variation on absorption magnitudes. These three points are close to EMBUMIRIM and EMBUGUAÇU input streams thalwegs. And as such, they receive the fastest-moving amount of input water, due to the fact the water takes longer to reach maximum bottom depth, as EMBUMIRIM and EMBUGUAÇU input streams belong to the longest reservoir branches. This depth-varying water speed might change the OACs

concentrations over the gradient depth causing the absorption to vary. This phenomenon is clearer in P16 where absorption quickly decreases as the measurements are carried out at higher depths where, possibly, the input flow carries clearer waters with lower OAC concentrations, and therefore, lower total absorption per unit volume.

P6 and P13, both located over the drainage basin in-reservoir thalweg, presented an increase in absorption magnitude, unlike P16. These again, demonstrated how the hydraulic dynamics of the reservoir together with OACs concentration variation altered on a point-by-point basis the water IOP characteristics. Figure 5.13 depicted the attenuation coefficient magnitudes over depth which was similar to the absorption magnitudes. Attenuation magnitude, though, was higher since it includes both total absorption and total scattering.

Figure 5.14 depicted the total absorption per unit wavelength for the available points for each correction method. Overall, individual subplots for each point presented the same exponential decrease shape towards lower frequency wavelengths. A slight absorption high could be noted at 675 nm due to phytoplankton presence. Overall, average absorption values from all points from Figure 5.14 (not accounting extreme values) matched average values summation from Figure 5.10a and Figure 5.10b. Extreme values from Figures 5.10a and 5.10b were underestimated in comparison with Figure 5.14. This most likely happened due to inconsistencies associated with CDOM adjustment curve for Figure 5.10a; or due to possible data loss from laboratory CDOM storage via rapid decay; or even, due to possible water samples filtering inconsistencies for particulate matter laboratory data.

Finally, P15 (Figure 5.14) presented a highly variable absorption magnitude in comparison with all other points among all correction methods. Furthermore, Appendix E three-dimensional scatterplots for five different wavelengths showed a similar behaviour to that of Figure 5.14-P15. Kirk and temperature-only corrections yielded similar results considering the absorption spectral

shape as yielded proportional 715 and 750 corrections. Flat correction was the main discrepant one in Figure 5.14-P15. One notes how flat correction started at the same magnitude level as Kirk and Temperature-only corrections at the 400 nm wavelength and ended at the 750 nm at the same magnitude as the proportional corrections. Therefore, Flat correction presented a steeper slope than the other corrections. The three-dimensional scatterplots in Appendix E will be further discussed in section 5.8.

Figure 5.12 – ACS-measured total absorption coefficient per unit depth for fifteen sampling stations. Individual figures for each station are standardised by the point with maximum depth (y-axis) and absorption magnitude (x-axis). Stations with fewer measurements represent points with shorter maximum depth. Data colours are linked to their respective spectrum wavelength colour, black being the NIR range. Two sampling stations failed to appropriately collect data and were discarded.

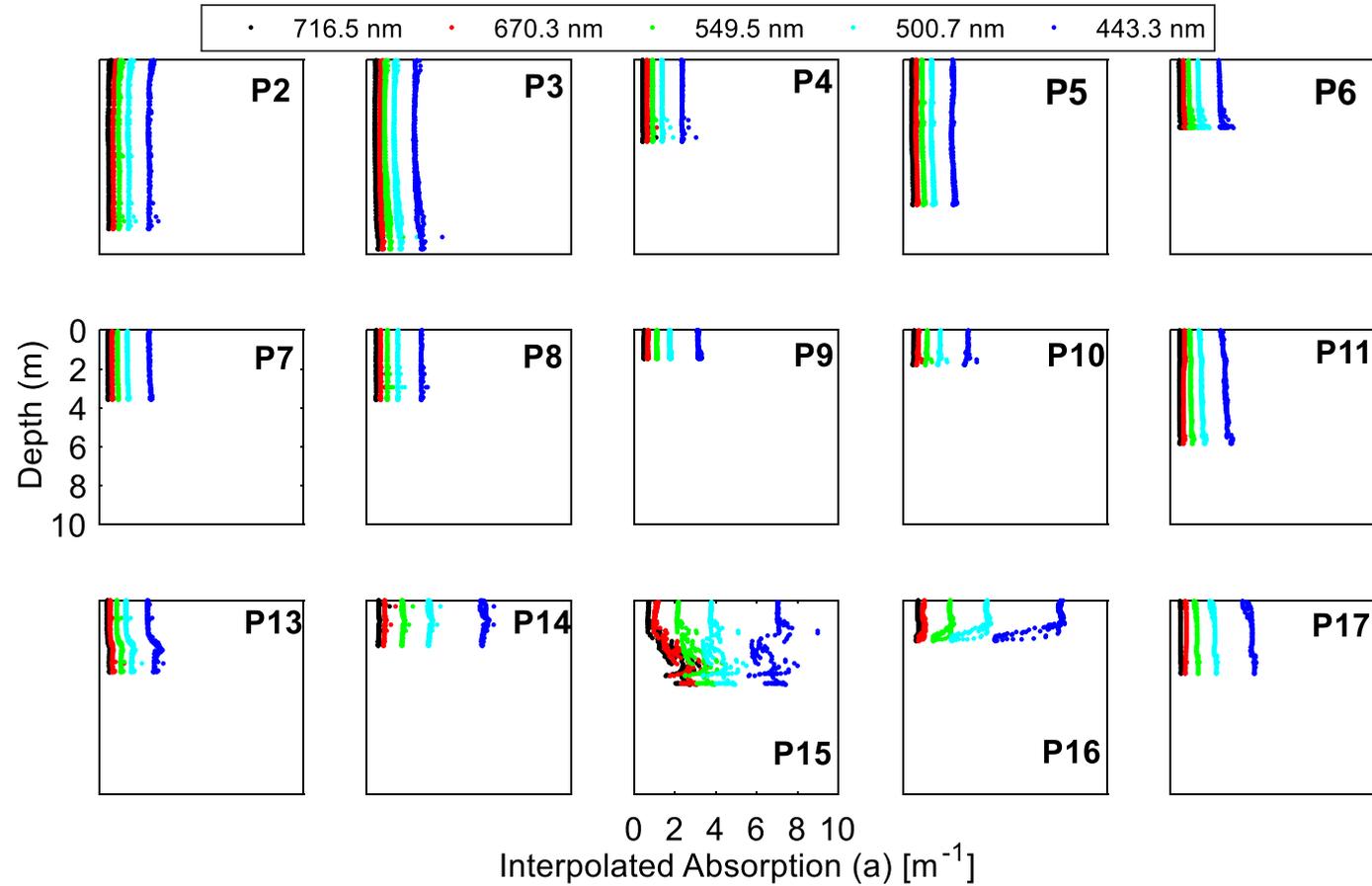


Figure 5.13 – ACS-measured total attenuation coefficient per unit depth for fifteen sampling stations. Individual figures for each station are standardised by point with maximum depth (y-axis). P9, P15, and P16 absorption magnitudes have their own plot ranges, and all other points' absorption magnitudes are standardised by P14's magnitude range. Stations with fewer measurements represent points with shorter maximum depth. Data colours are linked to their respective spectrum wavelength colour, black being the NIR range. Two sampling stations failed to appropriately collect data and were discarded.

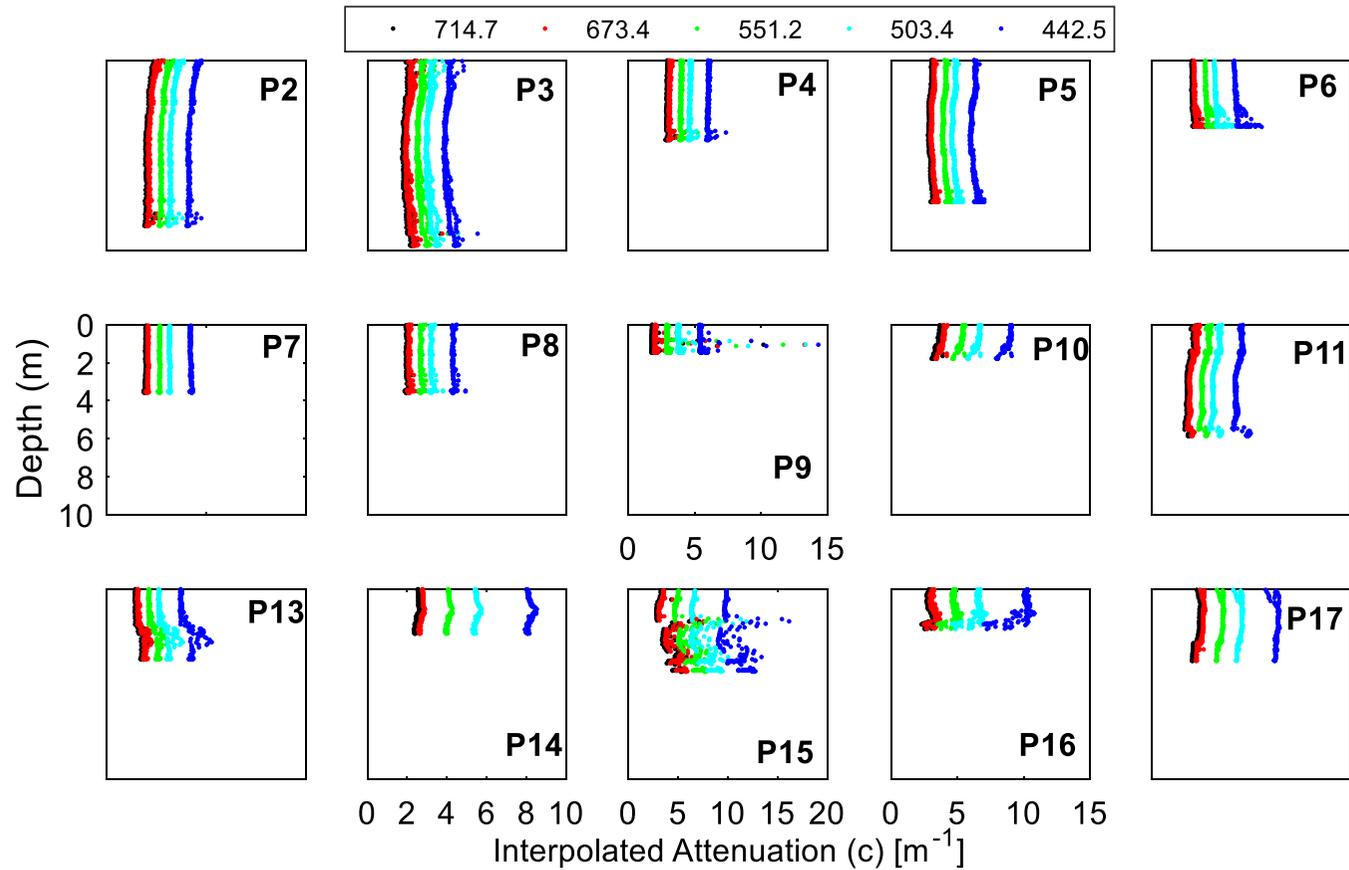
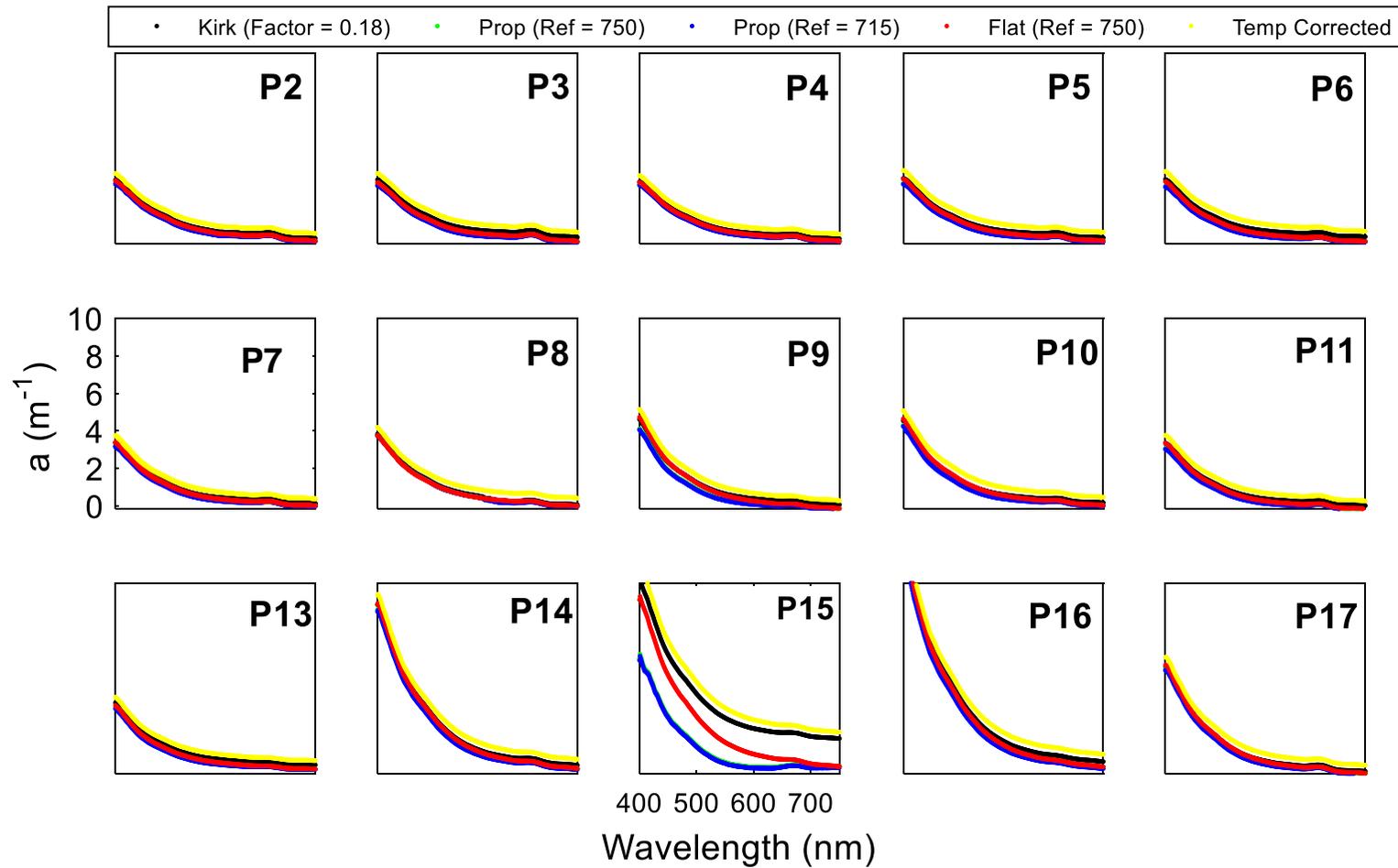


Figure 5.14 – ACS-measured total absorption coefficient per unit wavelength for fifteen sampling stations. Individual figures for each station are standardised by the wavelength (x-axis) range (400-750 nm), and by  $10 \text{ m}^{-1}$  absorption magnitude. Data colours are linked to the different correction methods applied to the dataset. Two sampling stations (P1 and P12) failed to appropriately collect data and were discarded.



### 5.5.1. Total Least Squares Regression

In the present study, results differed from Riddick et al. (2015) and were similar to Leymarie et al. (2010), with increasing underestimation results the longer the wavelengths. Riddick et al. (2015) correlated absorption data from spectrophotometry laboratory analysis and *in situ* ACS measurements in three different wavelengths (440 nm, 555 nm, and 676 nm). The authors found an agreement improvement in the red portion of the spectrum. However, Leymarie et al. (2010), identified that while using the proportional correction, an underestimation occurred, with the errors highest in the red portion of the spectrum.

Figures E.1, E.2, E.3, E.4, and E.5 depicted the features found in this study. However, this study only evaluated wavelengths in the blue and green region. The reason for that was that a total least squares regression (TLS; fitting the first principal component (PCt1) to data) was carried out on the dataset rather than a simple linear regression. And for such, TLS also correlated the results estimated from a QAA algorithm (LEE et al., 2002) with laboratory and *in situ* AC-S data, composing a 3-D regression plot (Figures E.1, E.2, E.3, E.4, E.5, and Table 5.1).

Regarding the correlation between AC-S and laboratory absorption data, AC-S Flat 750 scattering correction yielded results that varied from overestimating at the shorter wavelengths to more stable one-to-one estimation at longer wavelengths. Kirk and temperature-only corrections showed better correlations at shorter wavelengths and overestimation at longer wavelengths (Figures E.1, E.2, E.3, E.4, and E.5). Appendix E also indicated that TLS in the 412 nm region for all correction methods was more consistent than the other wavelengths. They displayed a better correlation slope between laboratory-calculated and AC-S-measured absorptions with weaker data over- or underestimation. Once again, the shorter wavelengths of the EMR spectrum presented a more enhanced result, furthering the possibility of applying this

spectral region into a re-parameterisation algorithm for PC retrieval via remote sensing techniques.

QAA estimation of total absorption yielded a highly underestimated result. An adjusted QAA for tropical waters proposed by (OGASHAWARA, 2014) was also applied to the dataset, in order to evaluate whether it could yield better results. However, it did not yield satisfactory results as all absorption data yielded negative values. Therefore, the original QAA proposed by Lee et al. (2002) was kept for the dataset results. An important aspect of this TLS approach is that results like the one found in this study can support the development of an enhanced semi- and quasi-analytical algorithm which can further improve the results of this 3-D orthogonal regression. This can be achieved by evaluating whether the main fit axis (PCt1) revolves around the data attribute space until it finds itself in a proper one-to-one position in which no over- or under-estimation is encountered; and also, until a robust sum of squared errors (SSE) and RMSE (Table 5.1) is yielded.

Table 5.1 presented the results of SSE, RMSE, and total variances explained by each principal component (PCt1, 2, and 3) from the TLS. These results were computed for each AC-S correction method. The SSE is a statistical quantity proportional to the  $R^2$ . In the case of this 3-D regression, it was only possible to generate SSE rather than  $R^2$ . However, the larger the SSE value, the more robust the correlation just as with  $R^2$ . SSE-PCt1 represented the robustness for the highest data variability fit curve. Therefore, it sought to represent the direction of greater data continuity. PCt2 and PCt3 represented second and third principal components for the correlation. Table 5.1 depicted the total variances for each principal component for each wavelength for each correction method. However, the main discussion for the total absorption correlation from AC-S data, laboratory data and QAA estimation revolved around PCt1 for its representation of the total least square regression. The evaluation of PCt2 and 3 in fitting for this dataset is beyond the scope of this section. However, as they represented the remaining variance not explained by PCt1, their results were

Table 5.1 – Principal-component (PCt) fitting total least squares regression sum of squared errors and root mean squared error (RMSE) for PCt one, two, and three for each wavelength and absorption correction method. PCt explanatory variances for each wavelength and correction method are also presented.

| <b>SSE PCt1</b>     | F 750 | F 715 | P 750 | Kirk  | Temp  | <b>RMSE PCt1</b>    | F 750 | F 715 | P 750 | Kirk | Temp  |
|---------------------|-------|-------|-------|-------|-------|---------------------|-------|-------|-------|------|-------|
| 412                 | 51.09 | 33.91 | 34.94 | 57.52 | 66.03 | 412                 | 2.06  | 1.68  | 1.71  | 2.19 | 2.35  |
| 430                 | 21.58 | 13.70 | 14.24 | 27.14 | 31.96 | 430                 | 1.34  | 1.07  | 1.09  | 1.50 | 1.63  |
| 489                 | 7.39  | 4.59  | 4.83  | 11.52 | 14.16 | 489                 | 0.78  | 0.62  | 0.63  | 0.98 | 1.09  |
| 530                 | 2.60  | 1.62  | 1.68  | 6.05  | 7.75  | 530                 | 0.47  | 0.37  | 0.37  | 0.71 | 0.80  |
| 555                 | 1.44  | 0.99  | 0.98  | 4.75  | 6.11  | 555                 | 0.35  | 0.29  | 0.29  | 0.63 | 0.71  |
| <b>SSE PCt2</b>     | F 750 | F 715 | P 750 | Kirk  | Temp  | <b>RMSE PCt2</b>    | F 750 | F 715 | P 750 | Kirk | Temp  |
| 412                 | 4.03  | 9.29  | 9.07  | 3.53  | 3.30  | 412                 | 0.58  | 0.88  | 0.87  | 0.54 | 0.52  |
| 430                 | 2.28  | 5.49  | 5.36  | 2.02  | 1.94  | 430                 | 0.44  | 0.68  | 0.67  | 0.41 | 0.40  |
| 489                 | 1.11  | 2.80  | 2.72  | 1.21  | 1.18  | 489                 | 0.30  | 0.48  | 0.48  | 0.32 | 0.31  |
| 530                 | 0.48  | 1.19  | 1.19  | 0.76  | 0.73  | 530                 | 0.20  | 0.31  | 0.31  | 0.25 | 0.25  |
| 555                 | 0.28  | 0.60  | 0.63  | 0.56  | 0.53  | 555                 | 0.15  | 0.22  | 0.23  | 0.22 | 0.21  |
| <b>SSE PCt3</b>     | F 750 | F 715 | P 750 | Kirk  | Temp  | <b>RMSE PCt3</b>    | F 750 | F 715 | P 750 | Kirk | Temp  |
| 412                 | 1.20  | 1.30  | 1.30  | 1.20  | 1.19  | 412                 | 0.32  | 0.33  | 0.33  | 0.32 | 0.31  |
| 430                 | 0.71  | 0.75  | 0.75  | 0.72  | 0.72  | 430                 | 0.24  | 0.25  | 0.25  | 0.24 | 0.25  |
| 489                 | 0.23  | 0.24  | 0.25  | 0.25  | 0.25  | 489                 | 0.14  | 0.14  | 0.14  | 0.14 | 0.14  |
| 530                 | 0.14  | 0.14  | 0.14  | 0.15  | 0.15  | 530                 | 0.11  | 0.11  | 0.11  | 0.11 | 0.11  |
| 555                 | 0.08  | 0.07  | 0.07  | 0.09  | 0.09  | 555                 | 0.08  | 0.08  | 0.08  | 0.08 | 0.08  |
| <b>Variance 412</b> | F 750 | F 715 | P 750 | Kirk  | Temp  | <b>Variance 430</b> | F 750 | F 715 | P 750 | Kirk | Temp  |
| PCt1                | 0.91  | 0.76  | 0.77  | 0.92  | 0.94  | PCt1                | 0.88  | 0.69  | 0.7   | 0.9  | 0.923 |
| PCt2                | 0.07  | 0.21  | 0.20  | 0.06  | 0.05  | PCt2                | 0.09  | 0.28  | 0.26  | 0.1  | 0.056 |
| PCt3                | 0.02  | 0.03  | 0.03  | 0.02  | 0.02  | PCt3                | 0.03  | 0.04  | 0.04  | 0    | 0.021 |
| <b>Variance 489</b> | F 750 | F 715 | P 750 | Kirk  | Temp  | <b>Variance 530</b> | F 750 | F 715 | P 750 | Kirk | Temp  |
| PCt1                | 0.85  | 0.6   | 0.62  | 0.89  | 0.908 | PCt1                | 0.81  | 0.55  | 0.56  | 0.9  | 0.897 |
| PCt2                | 0.13  | 0.37  | 0.35  | 0.09  | 0.076 | PCt2                | 0.15  | 0.4   | 0.39  | 0.1  | 0.085 |
| PCt3                | 0.03  | 0.03  | 0.03  | 0.02  | 0.016 | PCt3                | 0.04  | 0.05  | 0.05  | 0    | 0.018 |
| <b>Variance 555</b> | F 750 | F 715 | P 750 | Kirk  | Temp  |                     |       |       |       |      |       |
| PCt1                | 0.8   | 0.59  | 0.58  | 0.88  | 0.909 |                     |       |       |       |      |       |
| PCt2                | 0.16  | 0.36  | 0.37  | 0.1   | 0.078 |                     |       |       |       |      |       |
| PCt3                | 0.04  | 0.04  | 0.04  | 0.02  | 0.013 |                     |       |       |       |      |       |

also included in Table 5.1. Equal logic was applied to RMSE. SSE-PCt1 for the higher frequency wavelengths presented better results as mentioned previously. Also, regarding the correction methods, Kirk and temperature-only yielded the best results. RMSE, proportional to the absorption range magnitude, were similar to all dataset. The proportion between average RMSE for all corrections

for each wavelength equals the proportion of the average absorptions for all corrections for each wavelength.

## 5.6. Sonde measurements

Figure 5.15 depicted DO and chl-*a* (modelled via Figure 4.9 regression) concentrations over depth for sampling points with error-free measurements. DO and chl-*a* concurrent plots aimed to evaluate whether there could be a trend relation between chl-*a* and oxygen, the latter being a by-product of local production/consumption photosynthetic process. Overall, no special trend was observed to confirm this local source/sink process, indicating that the reservoir dynamics constantly rearranges the biogeochemical water components. However, some sampling points presented a degree of relation between the two variables; and some others did not show a specific pattern. DO presented an overall trend toward lower values (i.e., bottom hypoxia) at higher depths (LIN et al., 2006).

Chl-*a* data step-like behaviour was due to the sampling time during instrument descent in the water column as well as instrument measurements drift. P4, P5, and P13 surface values were excluded from plotting because they were extreme values. All other points represent chlorophyll estimations from the modelled curve extrapolated for all depths and, as such, they contain errors (e.g., P9 bottom values far from the statistically significant 5-15 µg/L range). However, P9 higher-depth chl-*a* values were most likely higher due to the fact that P9 is very close to the reservoir southeastern border at a shallow and narrow spot. The point location increased nutrient inflow, enhancing phytoplankton activity, which might be taking place still within the  $Z_{eu1}$ . P16 presented a very sharp decline in DO at about the 2-m depth near the zero-value concentration. This feature happened before the  $Z_{eu1}$  limit. Despite the low chl-*a* concentration level, the rapid increase in chl-*a* concentration (indicating higher biological activity) associated with DO depletion and with the higher nutrient input caused by the sampling point location at the southernmost part of the reservoir indicated a

highly active organic matter decomposition process which might have caused the rapid decrease in DO concentration (RAMSTEDT et al., 2003).

Figure 5.16 presented an overall trend of temperature decrease with depth. This decrease was smooth except for P6 which had a rapid decline at near-bottom depth. This feature in P6 was most likely due to the fact a local water basin thalweg was located right beneath that sampling point, injecting colder inflowing water. An important feature was observed in Figures 5.15 and 5.16. P5, P6, P10, P11, P15, and P16, all had significant an increase in conductivity levels (Figure 5.16). Conductivity can be regarded as a salinity proxy. Therefore, whenever this conductivity increase occurred, a halocline might be found. As noted by Ramsted et al. (2003), after a halocline (in this case, a rapid increase in salinity levels) appearance, DO levels dropped significantly (PINILLA, 2006), and such feature was observed in Figure 5.15.

Figure 5.15 – Dissolved Oxygen (DO) versus chlorophyll-a plot over depth for fourteen sampling stations. Red lines represent DO data, and green lines represent chlorophyll data. Black straight line across depth axis represents the one-percent euphotic zone depth for that point. Missing sampling points were excluded from data plotting due to faulty *in situ* measurements.

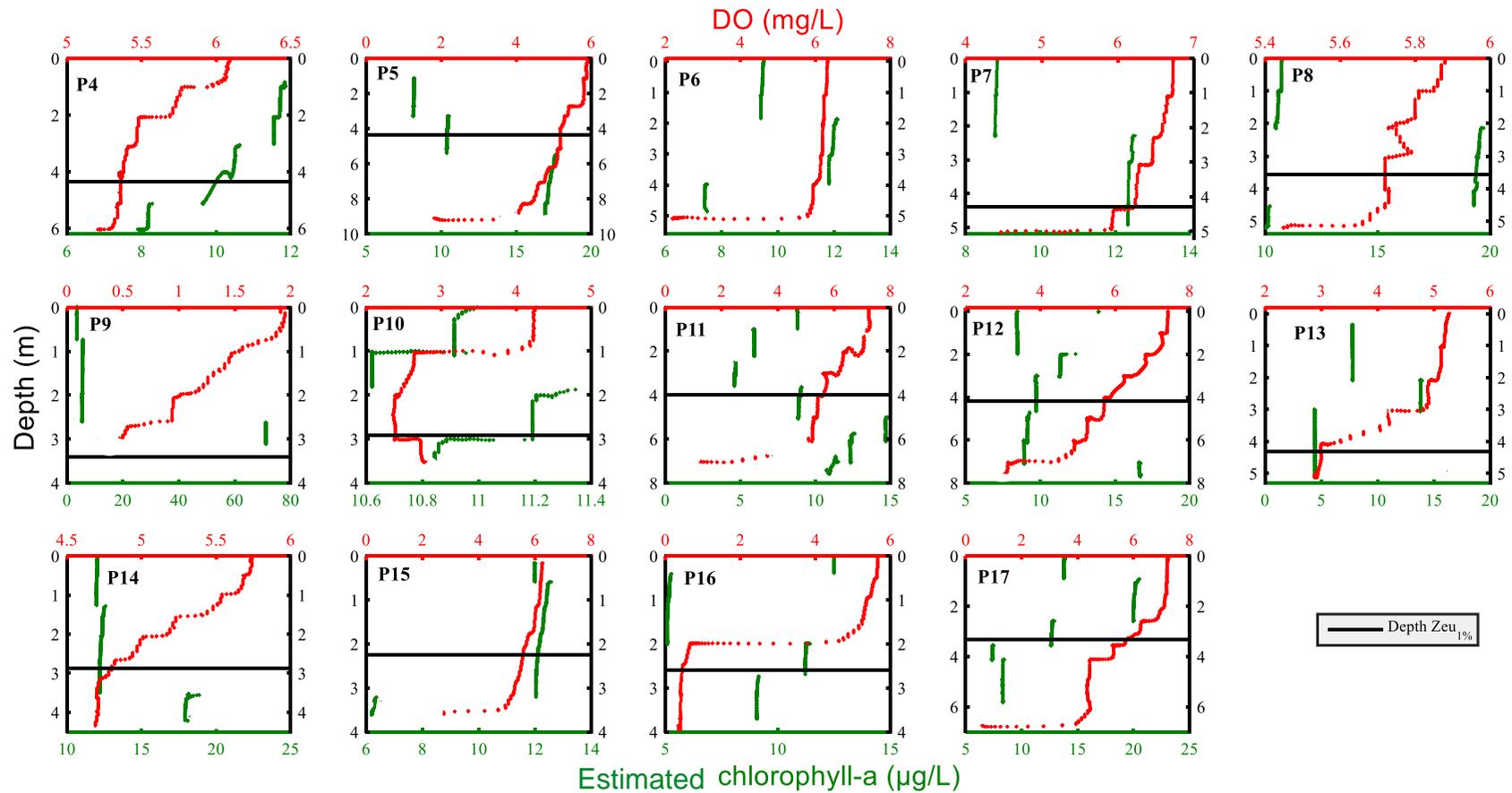
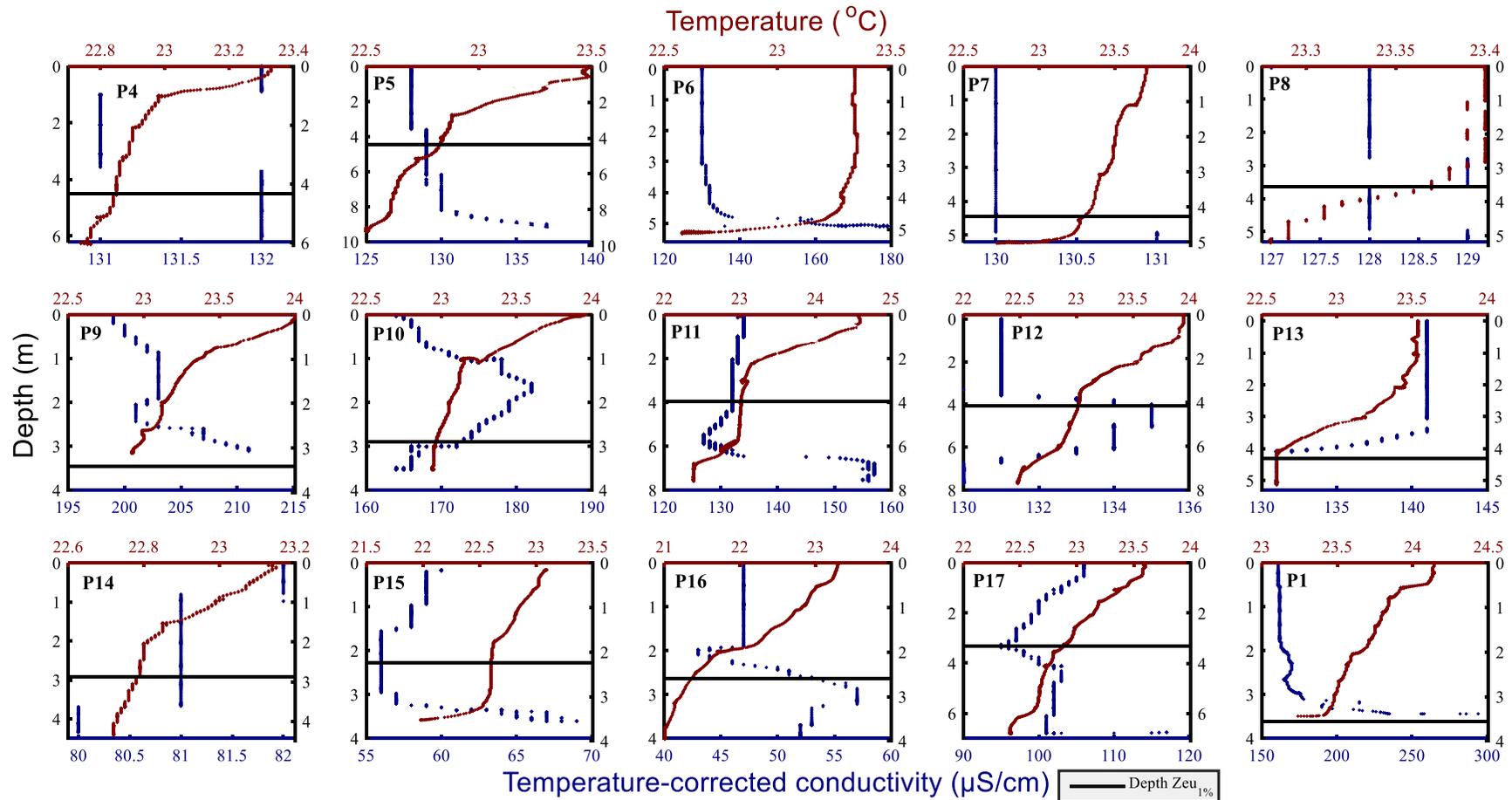


Figure 5.16 – Temperature versus conductivity plot over depth for fifteen sampling stations. Maroon lines represent temperature data, and blue lines represent conductivity data. Black straight line across depth axis represents the one-percent euphotic zone depth for that point. Missing sampling points were excluded from data plotting due to faulty *in situ* measurements.



## 5.7. Least Squares Regression

Figures 5.17, 5.18, 5.19, 5.20, 5.21, and 5.22 depicted the relationships between PC, chl-*a*, TSM, TSI, and DOC against the major IOPs. The regression relationships allowed the evaluation of whether the reservoir sampling distribution could provide any pattern in the reservoir OACs distribution. Generally, almost all correlations needed to have points excluded. They seemed clear outliers which did not represent the overall expected directly proportional relationships among the correlated variables.

Overall, the correlations varied significantly. Firstly, Figures 5.19b and 5.21a depicted the highest adjusted  $R^2$  ( $> 0.8$ ) which were found for chl-*a* correlations against  $b_{bp}(555)$  and  $a_{phy}(675)$  (BRICAUD et al., 1995; BABIN et al., 2003). This indicated higher phytoplankton dominance in Guarapiranga reservoir waters among tripton and CDOM. The 675 nm is a known absorption band for chl-*a* pigment; and backscattering at the 555 nm could represent that the backscattered radiation at the green region of the EMR spectrum is dominated by phytoplankton. However, the excluded points must be accounted for prior to this conclusion since five points were removed from the  $b_{bp}(555)$ -chl-*a* regression. Secondly,  $a_{phy}(620)$  against chl-*a*+PC (Figure 5.19a), and  $a_{CDOM}(440)$  (Figure 5.20a), and  $a_{phy}(440)$  (Figure 5.20b) against chl-*a* showed slightly lower, but still reasonable, adjusted  $R^2$  (between 0.56 and 0.72). Moreover,  $a_{phy}(620)$  against PC (Figure 5.18b) concentration actually presented a very low adjusted  $R^2$  (0.1826).

These features indicated that chl-*a* highly influenced the absorption at 620 nm which should be, primarily, a PC absorption band; and that chl-*a* concentration also positively correlated with absorption in the blue region of the spectrum for both CDOM and phytoplankton absorptions.  $a_{phy}(675)$  correlation with chl-*a* is a known feature in the literature. However, CDOM correlation against chl-*a* concentration only had a higher robustness due to the second-order fit curve, depicting an exponential relation between the two variables. The confusion that

chl-a applied onto the 620-nm PC-proxy absorption band indicated a necessity for the improvement in a PC-retrieval bio-optical algorithm for this study in order to minimise this confounding effect.

An important feature identified in all  $a_{CDOM}(440)$  correlations (Figures 5.17, 5.20a, and 5.22b) was that they all needed an exponential adjustment fit curve in order to improve the correlation robustness. This behaviour might have indicated some level of saturation tendency at the lower OACs concentration range with respect to the  $a_{CDOM}(440)$  not getting below the  $1 \text{ m}^{-1}$  absorption value. A keen assessment on Figure 5.10a allows one to notice that at the 440 nm no  $a_{CDOM}(440)$  value gets below  $1 \text{ m}^{-1}$ . This feature might be the reason for the exponential fitting among  $a_{CDOM}$  correlations.

Also, all  $b_{bp}(555)$  correlations (except that against chl-a (Figure 5.19b)) presented a strong underestimation behaviour (Figures 5.18a, and 5.22a). Such feature could be attributed to the minor contribution of PC and TSM on  $b_{bp}(555)$ , and the better contribution of chl-a on  $b_{bp}(555)$ . Also, an important consideration must be made over the results of  $b_{bp}(555)$  in comparison with the dataset in from which it was estimated. That is, AC-S total absorption coefficients overestimated laboratory-measured total absorption for the TLS results (Figure E.5), considering temperature-only correction for 555 nm. Instead,  $b_{bp}(555)$  calculated from AC-S data underestimated TSM and PC, and just about reached near-one-to-one slope against chl-a data. At last, a correlation between  $a_{nap}(440)$  against TSI (Figure 5.21b) concentration depicted a positive dependence but not a strong robustness ( $R^2 = 0.2352$ ). Nevertheless, this can be an indication of the strong influence by the mineral component of TSM towards tripton absorption (RIDDICK et al., 2015).

Figure 5.17 – Regression analysis plots for  $a_{CDOM}(440)$  as a function of dissolved organic matter (DOC) concentrations.

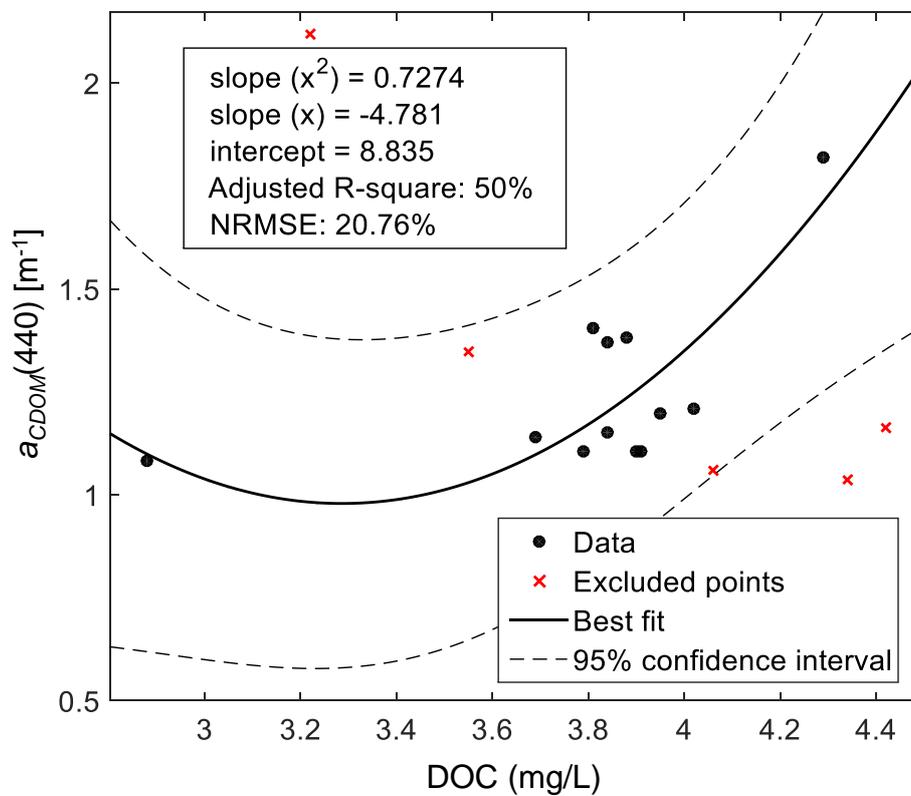


Figure 5.18 – Regression analysis plots for (a)  $b_{bp}(555)$  and (b)  $a_{phy}(620)$  as a function of phycocyanin (PC) concentrations.

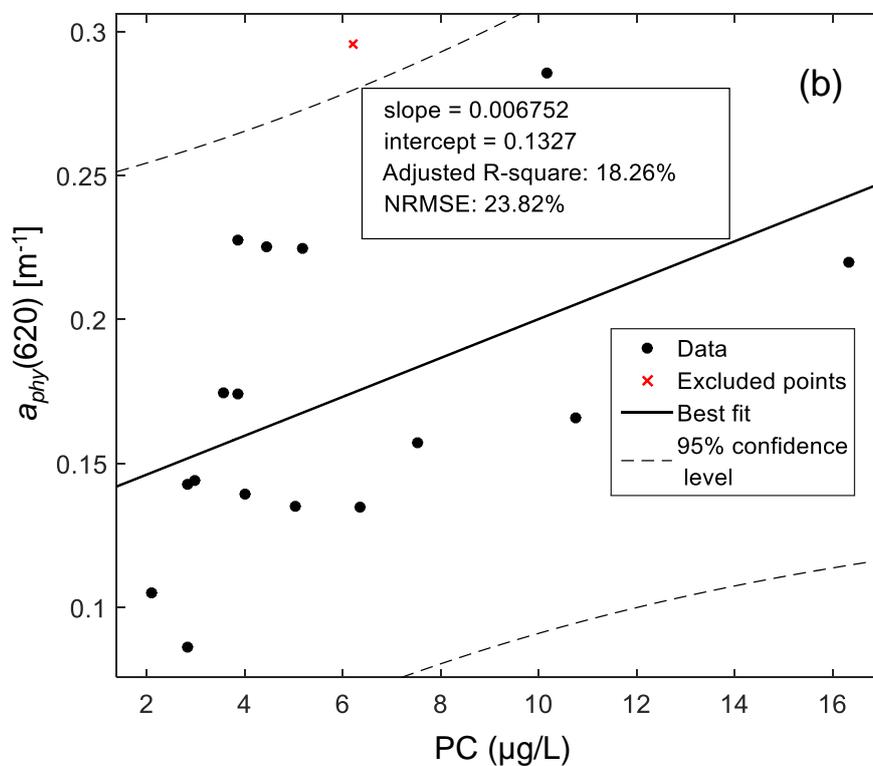
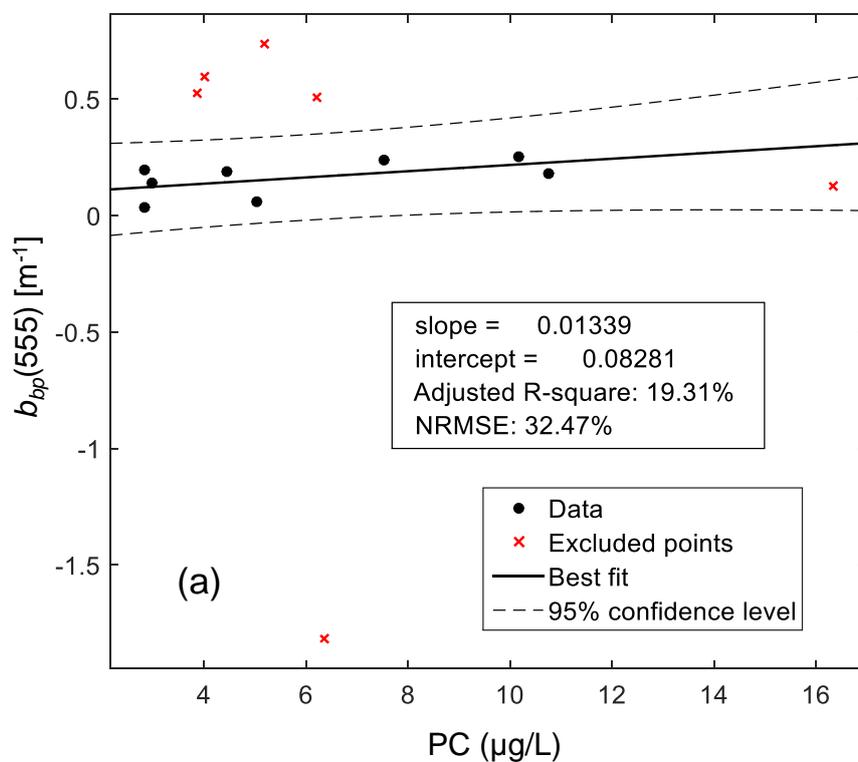


Figure 5.19 – Regression analysis plots for (a)  $a_{phy}(620)$  as a function of chlorophyll-a (chl-a) plus phycocyanin (PC) concentrations, and (b)  $b_{bp}(555)$  as a function of only chl-a.

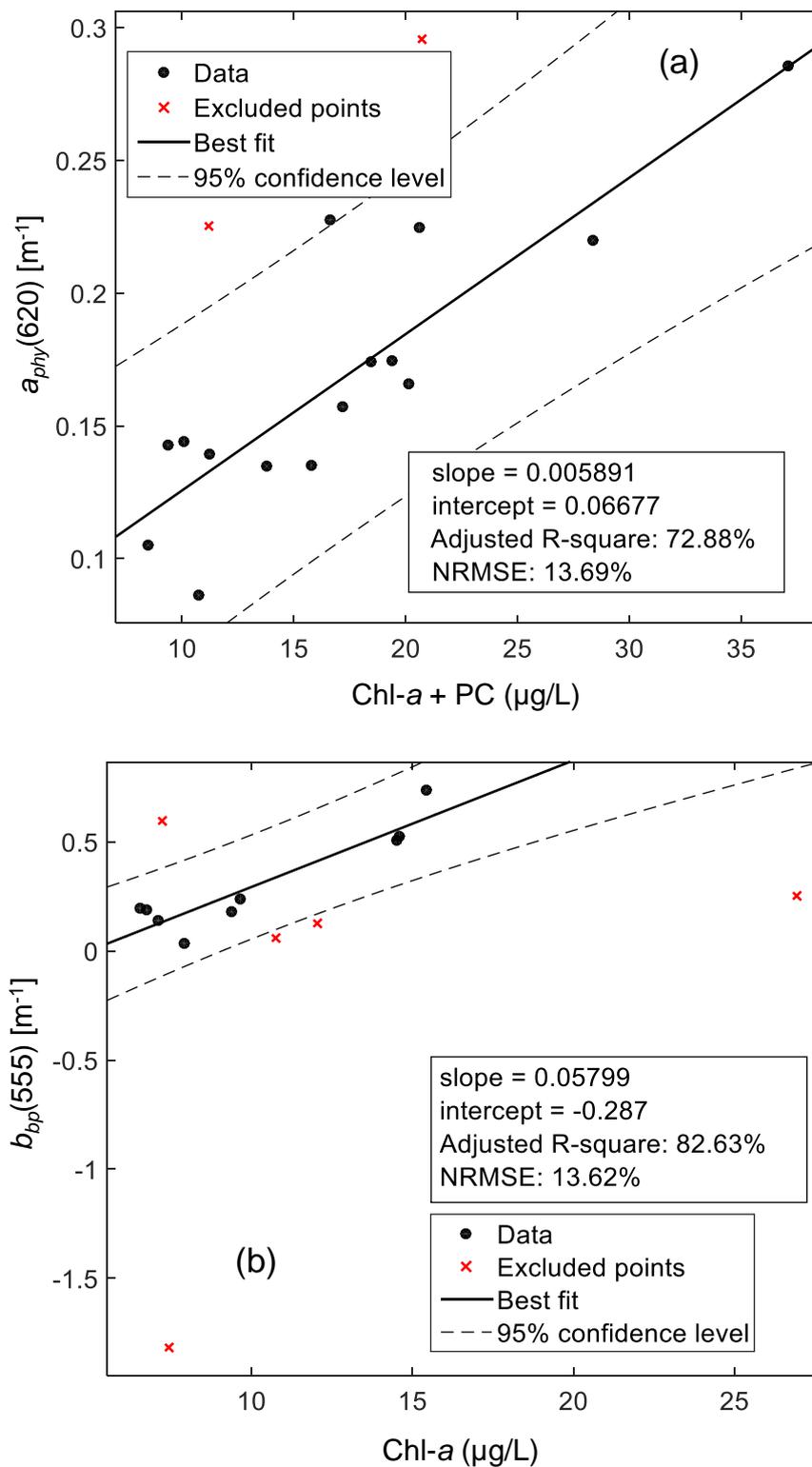


Figure 5.20 – Regression analysis plots for (a)  $a_{CDOM}(440)$  and (b)  $a_{phy}(440)$  as a function of chlorophyll-a (chl-a) concentrations.

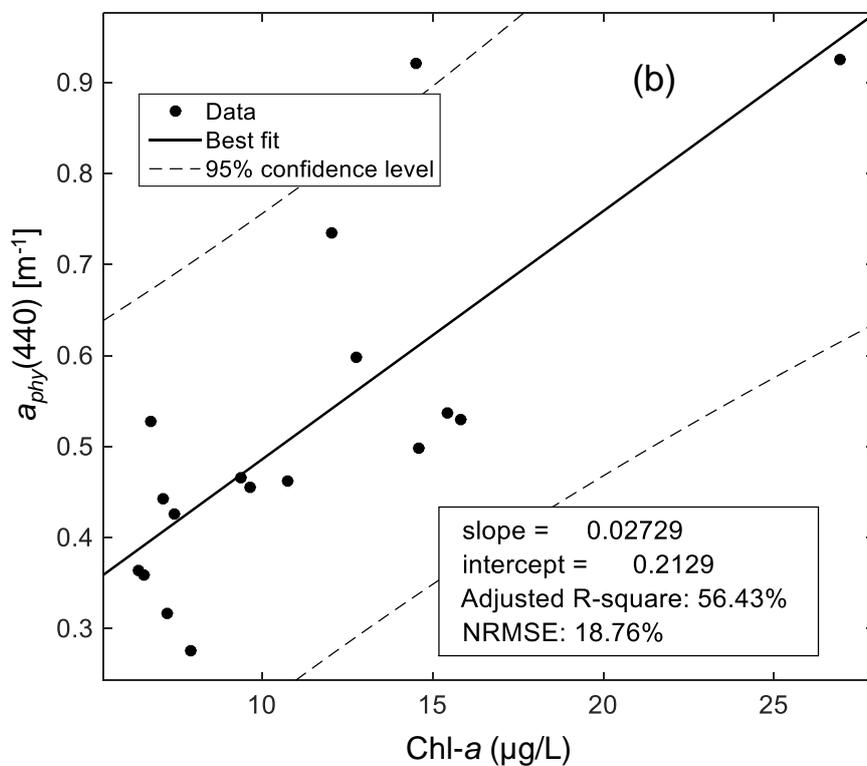
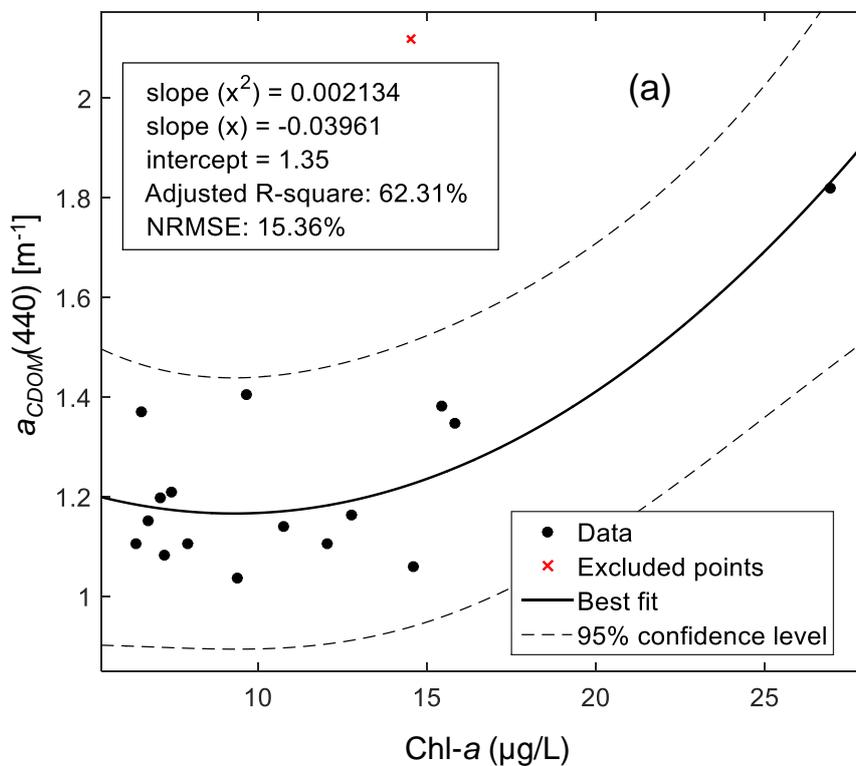


Figure 5.21 – Regression analysis plots for (a)  $a_{phy}(675)$  as a function of chlorophyll-a concentrations, and (b)  $a_{nap}(440)$  as a function of total suspended inorganics (TSI) concentrations.

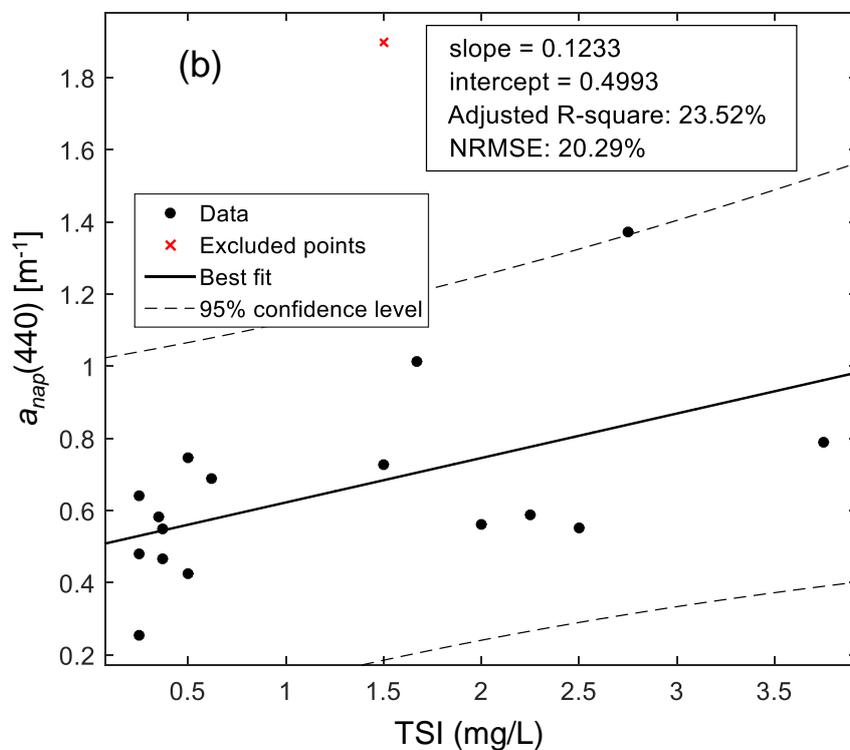
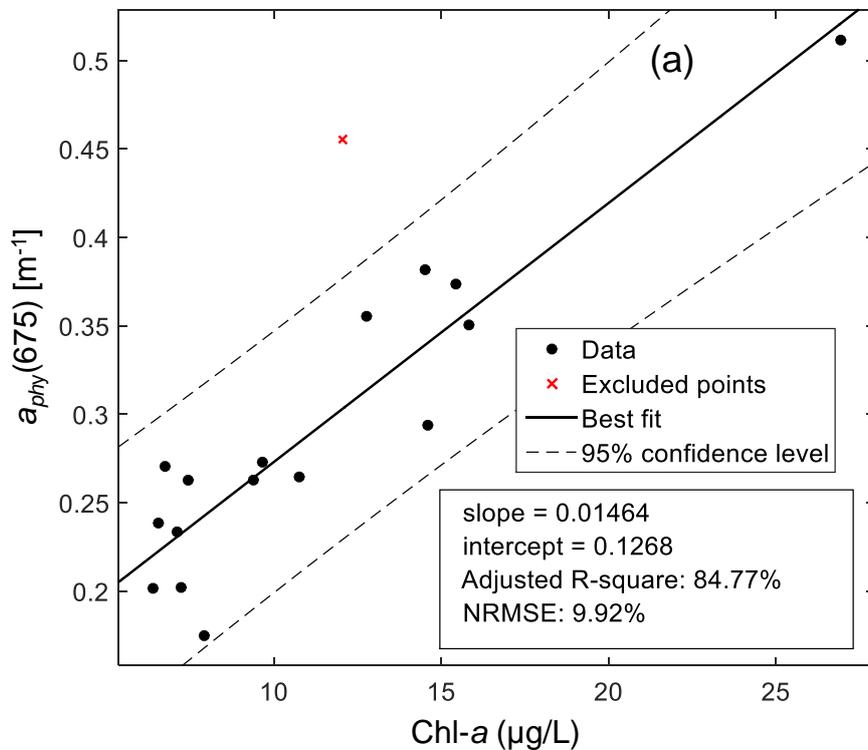
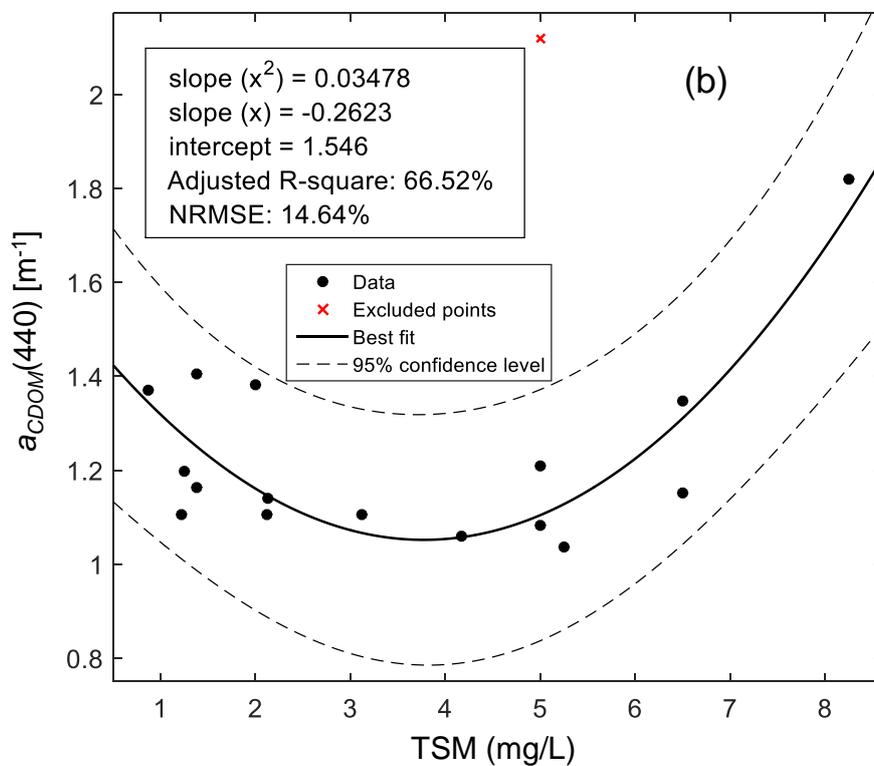
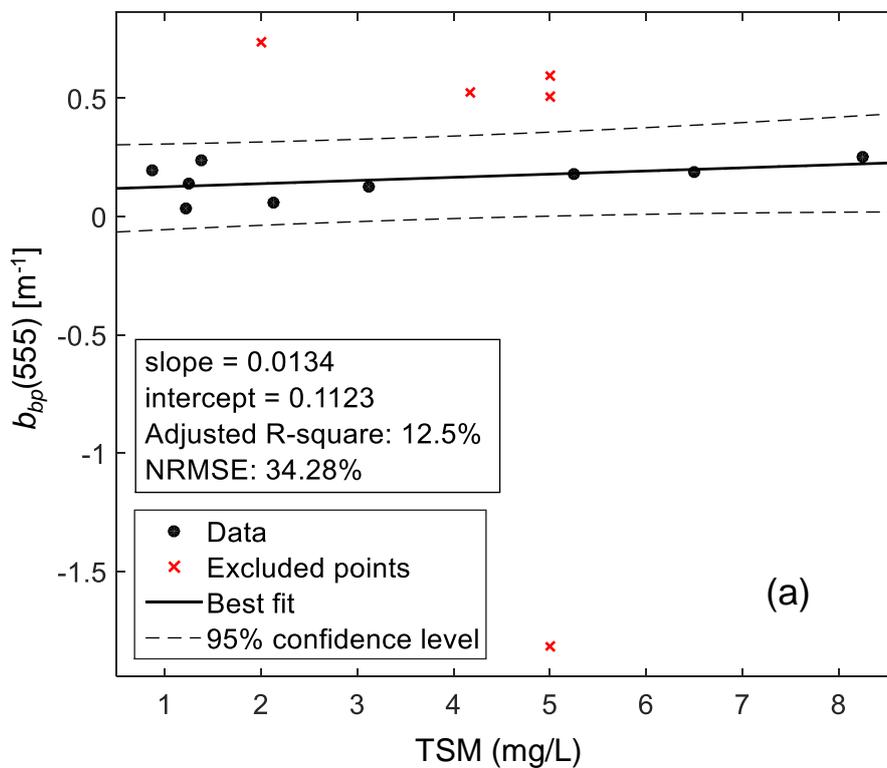


Figure 5.22 – Regression analysis plots for (a)  $b_{bp}(555)$  and (b)  $a_{CDOM}(440)$  as a function of total suspended matter (TSM) concentrations.



## 5.8. Signal processing techniques

### 5.8.1. DWTCA

This section covers the evaluation of DWTCA (Section 5.8.2 will cover SAM). These techniques serve as signal clustering tools in order to assess signal behaviour. Initial appraisal for DWTCA applied directly on  $R_{rs}$  data did not yield satisfactory results. Non-normalised spectra generated DWTCA results purely magnitude-based. Meaning, Spectra were clustered based on how large was the  $R_{rs}$  magnitude. In order to tackle that problem, the spectra were normalised (Figure 5.6). In this way, it was possible to evaluate the clustering process over the spectral shape only.

Both Haar and Dmeyer wavelets in DWTCA grouped the  $R_{rs}$  according to features identified by the wavelets (AMPE et al., 2014; 2015; Appendix F; Table 5.2). All grouped spectra in Table 5.2 could be compared to its graphic representation in Figure 4.12 in order for one to check visually the results.

Each decomposition level generated a particular grouping of spectra. However, the wavelets yielded results that differed from one another. This aspect was such due to the difference in nature of each wavelet (Appendix B). Haar wavelet was one of the first developed wavelets, therefore, a simpler one (DAUBECHIES, 1992). Such trait could allow a more inefficient OAC-related data clustering. Thereby, Dmeyer wavelet was also tested in order to evaluate which wavelet behaved more efficiently in discriminating the  $R_{rs}$  spectral features and grouping the spectra accordingly. The results were shown in Tables 5.3 and 5.4.

Table 5.2 – Groups allocated by the discrete wavelet transform clustering analysis for each sampling station for each decomposition level for both Haar and Discrete Meyer wavelets.

| <b>Haar</b>   | Signal | cd1 | cd2 | cd3 | cd4 | cd5 | cd6 | cd7 | cd8 |
|---------------|--------|-----|-----|-----|-----|-----|-----|-----|-----|
| P2            | 4      | 1   | 2   | 1   | 1   | 2   | 5   | 2   | 3   |
| P3            | 3      | 2   | 1   | 2   | 2   | 1   | 4   | 3   | 1   |
| P4            | 4      | 2   | 1   | 2   | 2   | 2   | 3   | 4   | 1   |
| P5            | 2      | 4   | 4   | 4   | 4   | 4   | 1   | 5   | 5   |
| P7            | 2      | 5   | 5   | 5   | 5   | 5   | 2   | 5   | 5   |
| P8            | 2      | 5   | 5   | 5   | 5   | 5   | 2   | 5   | 5   |
| P9            | 4      | 1   | 2   | 1   | 1   | 2   | 3   | 2   | 4   |
| P10           | 2      | 5   | 5   | 5   | 5   | 5   | 2   | 5   | 5   |
| P11           | 1      | 3   | 3   | 3   | 3   | 3   | 4   | 4   | 2   |
| P12           | 2      | 5   | 5   | 5   | 5   | 5   | 2   | 5   | 5   |
| P13           | 2      | 5   | 5   | 5   | 5   | 5   | 2   | 5   | 5   |
| P14           | 1      | 1   | 2   | 1   | 1   | 2   | 3   | 4   | 2   |
| P15           | 5      | 2   | 1   | 2   | 2   | 2   | 5   | 1   | 4   |
| P16           | 5      | 1   | 2   | 1   | 1   | 2   | 5   | 1   | 3   |
| P17           | 5      | 1   | 2   | 1   | 1   | 2   | 3   | 4   | 4   |
| <b>Dmeyer</b> | Signal | cd1 | cd2 | cd3 | cd4 | cd5 | cd6 | cd7 | cd8 |
| P2            | 4      | 4   | 5   | 5   | 2   | 5   | 5   | 4   | 1   |
| P3            | 3      | 4   | 5   | 5   | 2   | 5   | 4   | 1   | 5   |
| P4            | 4      | 3   | 5   | 5   | 2   | 5   | 4   | 4   | 5   |
| P5            | 2      | 5   | 1   | 3   | 3   | 3   | 1   | 5   | 2   |
| P7            | 2      | 5   | 2   | 5   | 2   | 5   | 2   | 4   | 1   |
| P8            | 2      | 5   | 1   | 4   | 4   | 4   | 2   | 5   | 2   |
| P9            | 4      | 4   | 5   | 5   | 2   | 5   | 4   | 4   | 5   |
| P10           | 2      | 2   | 3   | 4   | 4   | 4   | 2   | 5   | 1   |
| P11           | 1      | 1   | 3   | 1   | 5   | 1   | 3   | 2   | 3   |
| P12           | 2      | 2   | 2   | 5   | 2   | 5   | 2   | 4   | 1   |
| P13           | 2      | 5   | 2   | 5   | 2   | 5   | 2   | 5   | 2   |
| P14           | 1      | 2   | 3   | 5   | 2   | 5   | 4   | 4   | 3   |
| P15           | 5      | 4   | 5   | 5   | 2   | 5   | 5   | 3   | 4   |
| P16           | 5      | 3   | 4   | 5   | 2   | 5   | 5   | 3   | 4   |
| P17           | 5      | 3   | 5   | 2   | 1   | 2   | 4   | 4   | 3   |

Notes: decomposition levels are described as: signal, representing the actual signal scale level, and coefficients of detail (cd) ranging from one till eight. The number of cds is determined directly by the processing routine according to the levels identified in the input data.

Despite the fact that Tables 5.3 and 5.4 represented the statistical coefficients for Haar and Dmeyer wavelets results, respectively, for all decomposition levels, dendrograms in Figures 5.23 and 5.24 were not based on these tables. Instead of generating nine dendrograms, one for each decomposition level containing the linkages of each sampling station, one dendrogram (based on Table 5.2), per wavelet, was created as a summary of the relationship found by each wavelet clustering among all spectra. As such, the dendrograms are not representation of how the actual spectra relate to one another over the wavelength domain. Instead, Figures 5.23 and 5.24 correlated how those

spectra were classified in similar clusters over the  $cd$  range domain (Table 5.2). It is noted how Dmeyer's dendrogram yielded more balanced links among all sampling points than Haar's, meaning it could sieve through spectral details more robustly.

Cophenetic correlation coefficients were a measure of how faithfully the linkage tree represented the dissimilarities among observations. They quantitatively represented statistics of  $R_{rs}$  plots in Appendix F. Better results for Haar's wavelet cophenetic correlation were observed for  $cds$  seven and eight (Table 5.3) than for Dmeyer's (Table 5.4). Dmeyer's clustering was more efficient from  $cds$  one till six. Signal level (i.e., actual  $R_{rs}$  signal level considered by the wavelets before decomposing the  $R_{rs}$  into several levels) presented the same correlation value for both wavelets. Figures F.1c1 and F.9c3 depicted how signal level and  $cd$  8 grouped the spectra associated with the spectra with greatest slope between the blue to green region of the spectrum plus a plateau in the green to red region. Also, Figures from F.2c1 and F.2c2 until F.5c1 and F.5c2 depicted the groups where variations in slope within blue to green, and then, from green to red were also accounted for during clustering. For conciseness, only Haar wavelet results were shown in Appendix F. The actual statistics appraisal was conducted via the correlation coefficients in Tables 5.3 and 5.4. Once again, variations in the blue region seemed just as relevant a variation as in the red part of the EMR spectrum. Therefore, this section also imbued the author towards attempting a semi-empirical algorithm re-parameterisation for PC prediction using information in the blue region of the spectrum.

Inconsistency coefficients correlated each of the fourteen links generated in Figures 5.23 and 5.24 for how they compared to other links at the same linkage height. Higher values represented a worse correlation; lower values, a better correlation. Monotonically decreasing numbers in the first column of Tables 5.3 and 5.4 indicated each dendrogram link. Smaller numbers represented links closest to zero-value linkage distance; higher values represented links closest

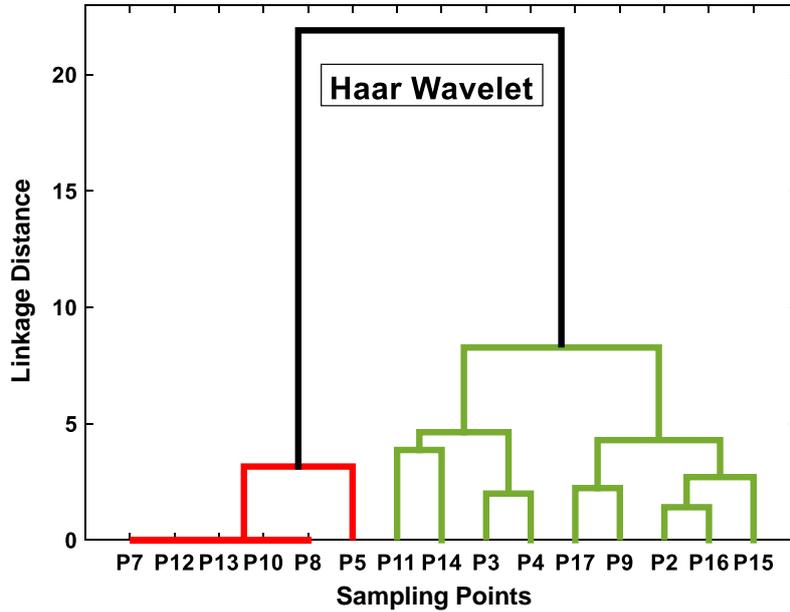
to the highest-value linkage distance. This aspect is the reason the inconsistency values are listed from top value of fourteen down to bottom value of one. Once clarified tables characteristics, one notes how bottom of tables have smaller values and top have higher values corresponding to the aforementioned aspect of this coefficient that higher values represent a worse correlation; and lower values, a better correlation. As such, lower values corresponded to lower linkage distances; and higher values corresponded to higher linkage distances.

Comparing Tables 5.3 and 5.4 (Haar and Dmeyer wavelets, respectively) one notes very subtle differences. For the longest linkage distance Dmeyer yielded slightly better correlation coefficients. For the second and third longest linkage distances, Haar yielded better results, especially for intermediate *cds*. For all the other linkage distances, both wavelets interchanged very heterogeneously on which performed best. This heterogeneity associated with Tables 5.3 and 5.4 results could be interpreted as a result of the lack of input data (i.e.,  $R_{rs}$  spectra). Should more data be used as input, perhaps clearer results - indicating which wavelet actually behaved best for each link of each *cd* - could be inferred from these correlation coefficients.

Despite the results generated so far, the actual intention for the utilisation of DWTCA was to use a large number of input data (i.e., a few hundred  $R_{rs}$  spectra). Then, cluster these signals with all possible wavelets able to be applied onto the dataset, and correlate the clusters results with as many as possible bio-optical algorithm-predicted OAC concentrations. And then evaluate results robustness and error and conclude which wavelets, decomposition levels, and bio-optical algorithms performed best in estimating OAC concentrations. However, as this work had only fifteen useful signals, and some clusters only had one  $R_{rs}$  signal associated with it, carrying out statistical analysis over these results was not a viable approach. Once this sort of results is acquired, comparisons with other works already carried out in this regard could be done. Examples of these works are those of Ampe et al. (2014) and

Ampe et al. (2015). Ampe et al. (2014) attempted to isolate the most informative wavelet regions via thresholding, and then, relate all five regions to known inherent optical properties. And also, Ampe et al. (2015) proposed a wavelet-enhanced inversion method, specifically designed for complex waters in which it integrated wavelet-transformed high-spectral resolution reflectance spectra in a multi-scale analysis tool.

Figure 5.23 – Haar wavelet clustering analysis dendrogram for fifteen sampling points.



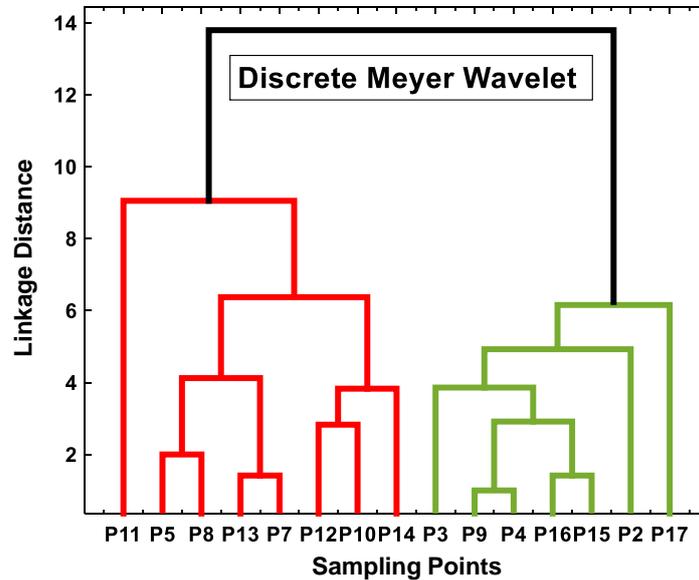
Notes: the links in the cluster tree were generated based on table 5.2, rather than via the  $R_{rs}$  spectra themselves.

Table 5.3 – Haar wavelet cophenetic and inconsistency coefficients.

| Cophenetic coefficient    | Signal | cd1   | cd 2  | cd 3  | cd 4  | cd 5  | cd 6  | cd 7  | cd8   |       |
|---------------------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                           |        | 0.806 | 0.520 | 0.509 | 0.499 | 0.517 | 0.566 | 0.511 | 0.691 | 0.781 |
| Inconsistency coefficient | Signal | cd 1  | cd 2  | cd 3  | cd 4  | cd 5  | cd 6  | cd 7  | cd 8  |       |
|                           | 14     | 1.135 | 1.150 | 1.151 | 1.152 | 1.153 | 1.147 | 0.850 | 1.023 | 0.765 |
|                           | 13     | 0.961 | 0.707 | 0.707 | 0.707 | 0.707 | 0.707 | 0.976 | 1.026 | 1.116 |
|                           | 12     | 0.707 | 0.707 | 0.707 | 0.707 | 0.707 | 0.707 | 0.832 | 0.707 | 1.118 |
|                           | 11     | 1.083 | 1.154 | 1.154 | 1.152 | 1.141 | 0.707 | 0.707 | 1.144 | 1.114 |
|                           | 10     | 0.707 | 1.146 | 1.136 | 1.133 | 0.707 | 1.151 | 0.000 | 1.058 | 0.707 |
|                           | 9      | 1.075 | 0.707 | 1.133 | 0.707 | 1.116 | 0.775 | 0.707 | 0.000 | 0.707 |
|                           | 8      | 0.000 | 1.126 | 0.707 | 1.104 | 1.145 | 1.032 | 0.707 | 0.856 | 0.000 |
|                           | 7      | 0.707 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.989 | 0.000 | 0.000 |
|                           | 6      | 0.000 | 0.707 | 0.707 | 0.707 | 0.707 | 0.707 | 0.707 | 0.707 | 0.707 |
|                           | 5      | 0.707 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 1.107 |
|                           | 4      | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.707 | 0.000 | 0.000 | 0.000 |
|                           | 3      | 0.000 | 0.707 | 0.707 | 0.707 | 0.707 | 0.000 | 0.707 | 0.000 | 0.000 |
|                           | 2      | 0.707 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.707 | 0.000 |
|                           | 1      | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Notes: decomposition levels are described as: signal, representing the actual signal scale level, and coefficients of detail (cd) ranging from one till eight. The number of cds was determined directly by the processing routine according to the levels identified in the input data.

Figure 5.24 – Discrete Meyer wavelet clustering analysis dendrogram for fifteen sampling points.



Notes: the links in the cluster tree were generated based on table 5.2, rather than via the  $R_{rs}$  spectra themselves.

Table 5.4 – Discrete Meyer cophenetic and inconsistency coefficients.

| Cophenetic coefficient    | Signal | cd 1  | cd 2  | cd 3  | cd 4  | cd 5  | cd 6  | cd 7  | cd 8  |       |
|---------------------------|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
|                           |        | 0.806 | 0.719 | 0.703 | 0.710 | 0.757 | 0.763 | 0.662 | 0.618 | 0.715 |
| Inconsistency coefficient | Signal | cd 1  | cd 2  | cd 3  | cd 4  | cd 5  | cd 6  | cd 7  | cd 8  |       |
|                           | 14     | 1.135 | 1.150 | 1.152 | 1.073 | 1.030 | 1.105 | 0.719 | 0.854 | 0.855 |
|                           | 13     | 0.961 | 1.151 | 0.707 | 1.146 | 0.707 | 1.154 | 1.085 | 1.117 | 1.117 |
|                           | 12     | 0.707 | 1.086 | 1.154 | 0.707 | 0.707 | 0.707 | 0.707 | 1.016 | 1.140 |
|                           | 11     | 1.083 | 0.707 | 1.109 | 0.000 | 0.707 | 0.000 | 0.707 | 0.000 | 1.151 |
|                           | 10     | 0.707 | 0.707 | 0.707 | 0.707 | 0.813 | 1.085 | 0.707 | 0.707 | 0.707 |
|                           | 9      | 1.075 | 0.707 | 1.114 | 0.711 | 0.972 | 0.814 | 0.707 | 1.066 | 0.000 |
|                           | 8      | 0.000 | 0.707 | 0.000 | 1.104 | 0.707 | 0.707 | 0.707 | 0.809 | 0.000 |
|                           | 7      | 0.707 | 0.707 | 0.707 | 0.707 | 0.707 | 0.000 | 0.000 | 0.707 | 0.707 |
|                           | 6      | 0.000 | 0.707 | 1.078 | 0.707 | 0.000 | 1.097 | 1.141 | 0.000 | 0.707 |
|                           | 5      | 0.707 | 0.707 | 0.000 | 0.000 | 0.707 | 0.000 | 0.707 | 0.707 | 0.707 |
|                           | 4      | 0.000 | 0.000 | 0.000 | 0.707 | 0.000 | 0.000 | 0.000 | 0.000 | 0.707 |
|                           | 3      | 0.000 | 0.000 | 0.000 | 0.000 | 0.707 | 0.000 | 0.000 | 0.000 | 0.000 |
|                           | 2      | 0.707 | 0.000 | 0.000 | 0.000 | 0.000 | 0.707 | 0.707 | 0.000 | 0.000 |
|                           | 1      | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Notes: decomposition levels are described as: signal, representing the actual signal scale level, and coefficients of detail (cd) ranging from one till eight. The number of cds was determined directly by the processing routine according to the levels identified in the input data.

### 5.8.2. Spectral Angle Mapping (SAM)

SAM analysis is a very useful signal processing technique commonly applied to signal data for the study of optically complex waters (SANDER DE CARVALHO et al., 2015; SHEN et al., 2015). SAM results were represented in Table 5.5 and in Figure 5.25. The Table 5.5 was the numerical representation of Figure 5.25b (colour shading matrix). Table 5.5 depicted the spectral angle error values. Also, bold numbers represent to which reference class each sampling point spectrum was associated. All points were allocated to classes one, three, or five. None was designated to classes two or four. However, sampling points seven (P7) and fifteen (P15) were nearly classified as classes three and one, respectively, rather than one and five, respectively.

Comparing P7 and P15 in Figure 4.12 against the reference classes (Figure 5.25a) to which they were allocated, one notes that P7 was classified as class one which had the shallowest blue-green slope. And it was almost classified as class three which had the second (paired with class five) shallowest blue-green slope; and a slight negative slope in the green-red spectral region. P15 was classified as class five which had the second (paired with class three) shallowest blue-green slope; and a slight plateau in the 550-600 nm region; and an absorption trough near where chl-*a* has a red absorption maximum (675 nm). And P15 was almost classified as class one, which had the shallowest blue-green slope; and a slightly negative slope plateau in the 550-600 nm.

One can identify by looking at these two sampling points (P7 and P15) classification results that the spectral angle error calculation, respectively, pondered over how each of these spectral features regions (i.e., blue-green and green-red regions) contributed to the final classification. Perhaps, nearly being classified as another class might be an indication that subtle variations in these regions accumulated more or less errors as the calculation took place. And this process might have led the final error to be a mesh of different regions contributions. Therefore, as yet again, the blue spectral region seemed as much

an important parameter as the red region for the  $R_{rs}$  signal analysis. These results further consolidated the re-parameterisation attempt carried out on the semi-empirical bio-optical algorithm using information from the higher frequency blue-green spectral region.

Figure 5.25b depicted a shaded coloured matrix in which the redder the cells the lesser the spectral angle error. Sampling points were mostly designated to classes one, three and five, as previously mentioned. Moreover, reference class two performed very badly in the classification process. This feature was most likely associated with the fact that this spectrum had a distinct type of phytoplankton (phyto5, Figure 5.25a) concentration compared to the other ones. The actual biology of this group is not of interest here mainly because it cannot be further evaluated as no cell count for a specific group had been performed for field samples. However, the fact it is a different group indicated that it would perform differently from the other groups; in this case, it did not match this study's signal data appropriately, whereas the other classes displayed some degree of concordance.

If the average error value for each reference class in Table 5.5 (bottom row) were taken for all sampling points, the following reference class rank order would be achieved from first place to fifth place: reference class five, one, four, three, and two. This aspect could be expected by investigating Figure 5.25b overall colour distribution. It could, then, be compared to the normalised spectra in Figure 5.6, thence, one can infer that Guarapiranga reservoir waters would fall somewhere in between the OAC concentration of water classes five and one. Mostly class five whose OAC heterogeneity was in fact the highest of all classes. Since the OACs concentration in all classes was only a representation of reality, the actual average OAC concentration for the reservoir was not considered for comparison.

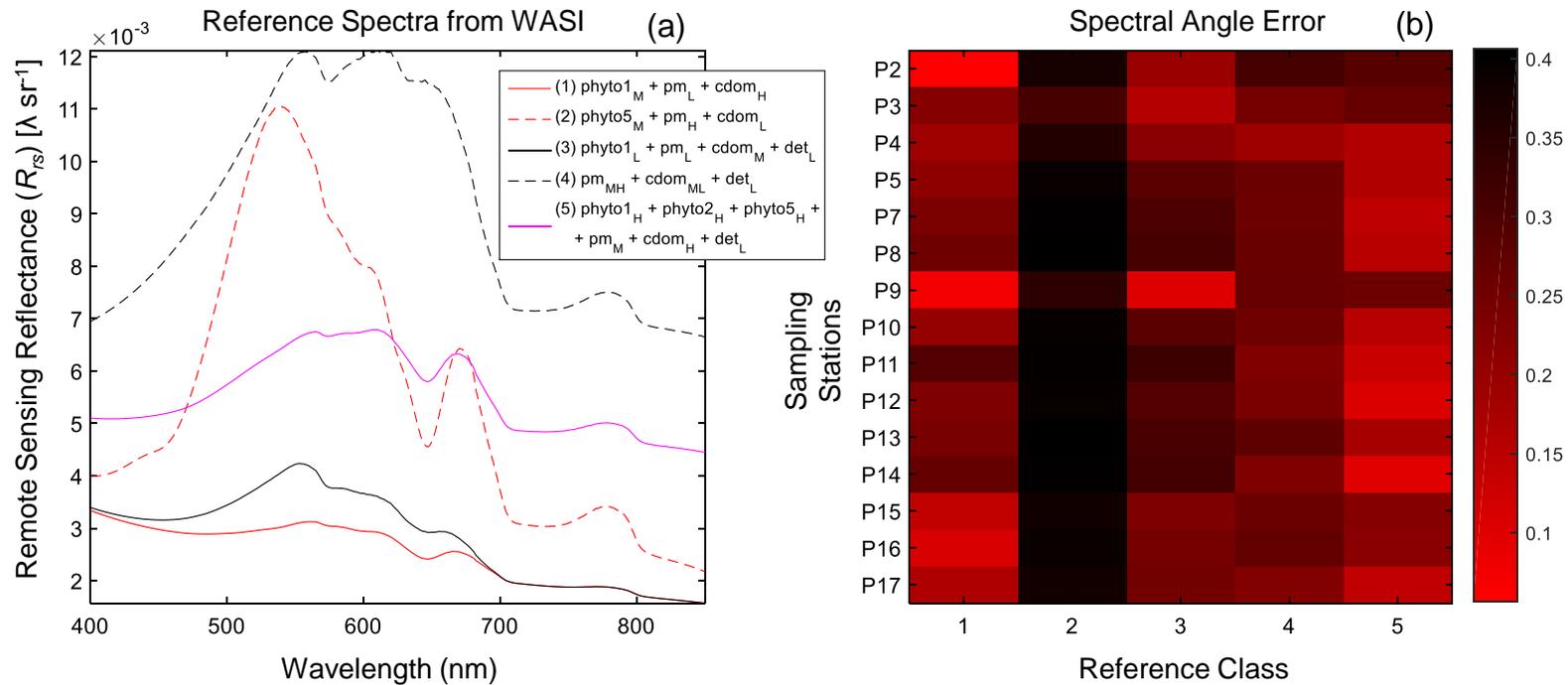
Lastly, an elaboration should account for the comparison and contrast of both DWTCA and SAM analyses. These techniques have fairly different theoretical

and mathematical frameworks. DWTCA can evaluate signal data and investigate the similarities they might possess among themselves. No reference signal is necessary, other than the actual test data. Further analyses, such as that suggested at the last paragraph of the previous chapter could then be carried out. On the other hand, SAM requires reference signal data to investigate test data. In this study WASI software was used to generate these reference data. These specificities of each technique are important to be considered. Mainly because the objective of this section was to demonstrate via the results acquired the dichotomy existing between these two techniques.

Table 5.5 – Spectral angle error from spectral angle mapping (SAM) analysis for each sampling point (column-title) against reference classes (row-title) generated in WASI. Bold numbers represent to which class each sampling point has been allocated.

|             | Class 1      | Class 2      | Class 3      | Class 4      | Class 5      |
|-------------|--------------|--------------|--------------|--------------|--------------|
| P1          | <b>0.056</b> | 0.337        | 0.122        | 0.235        | 0.206        |
| P2          | 0.144        | 0.235        | <b>0.100</b> | 0.164        | 0.183        |
| P3          | 0.118        | 0.307        | 0.139        | 0.119        | <b>0.102</b> |
| P4          | 0.134        | 0.377        | 0.200        | 0.176        | <b>0.103</b> |
| P5          | 0.157        | 0.396        | 0.224        | 0.175        | <b>0.092</b> |
| P6          | 0.171        | 0.401        | 0.236        | 0.180        | <b>0.097</b> |
| P7          | <b>0.061</b> | 0.288        | 0.073        | 0.181        | 0.174        |
| P8          | 0.127        | 0.381        | 0.199        | 0.173        | <b>0.097</b> |
| P9          | 0.210        | 0.391        | 0.253        | 0.149        | <b>0.086</b> |
| P10         | 0.151        | 0.378        | 0.211        | 0.157        | <b>0.074</b> |
| P11         | 0.159        | 0.406        | 0.231        | 0.194        | <b>0.112</b> |
| P12         | 0.185        | 0.396        | 0.239        | 0.150        | <b>0.071</b> |
| P13         | <b>0.091</b> | 0.354        | 0.153        | 0.177        | 0.142        |
| P14         | <b>0.076</b> | 0.371        | 0.164        | 0.190        | 0.138        |
| P15         | 0.106        | 0.351        | 0.166        | 0.146        | <b>0.091</b> |
| <i>Mean</i> | <i>0.130</i> | <i>0.358</i> | <i>0.181</i> | <i>0.171</i> | <i>0.118</i> |

Figure 5.25 – Reference spectra generated from WASI software used for spectral angle mapping (SAM) classification and spectral angle errors for fifteen sampling stations correlated with each reference class. The redder (smaller) the error value the more correlated the station spectrum is to the reference class. Reddest values represent the reference class to which the station spectrum was allocated during the SAM classification.



Notes: each reference spectrum is presented with their respective OACs concentrations. L, M, H indices indicated low, medium and high concentrations, according to WASI default range values. phyto1, phyto2, phyto5, pm, cdom, det stand, respectively, for cryptophyta type L, cryptophyta type H, green algae, particulate matter, coloured dissolved organic matter, and detritus. Table 5.5 has the numerical representation of Figure 5.25b.

## 5.9. Semi-empirical algorithms

Ogashawara (2015) presented a letter suggesting a novel bio-optical modelling classification method. Following the elaboration in that work and previous works on this subject (OGASHAWARA et al., 2013; AUGUSTO-SILVA et al., 2014; MISHRA; MISHRA, 2014; OGASHAWARA; MORENO-MADRINIÁN, 2014), this section tested several semi-empirical bio-optical algorithms (Table 4.3; Figure 5.26). An investigation was then applied to the results to assess which performed best at predicting chl-a, PC, and chl-a:TSM in Guarapiranga Reservoir. All algorithms in Figure 5.26 underwent adjustment by fitting a curve between their model results to this study's OAC concentration. Then, the best Pearson's correlation,  $R$ , among PC-predictive algorithms indicated onto which the re-parameterisation would be applied.

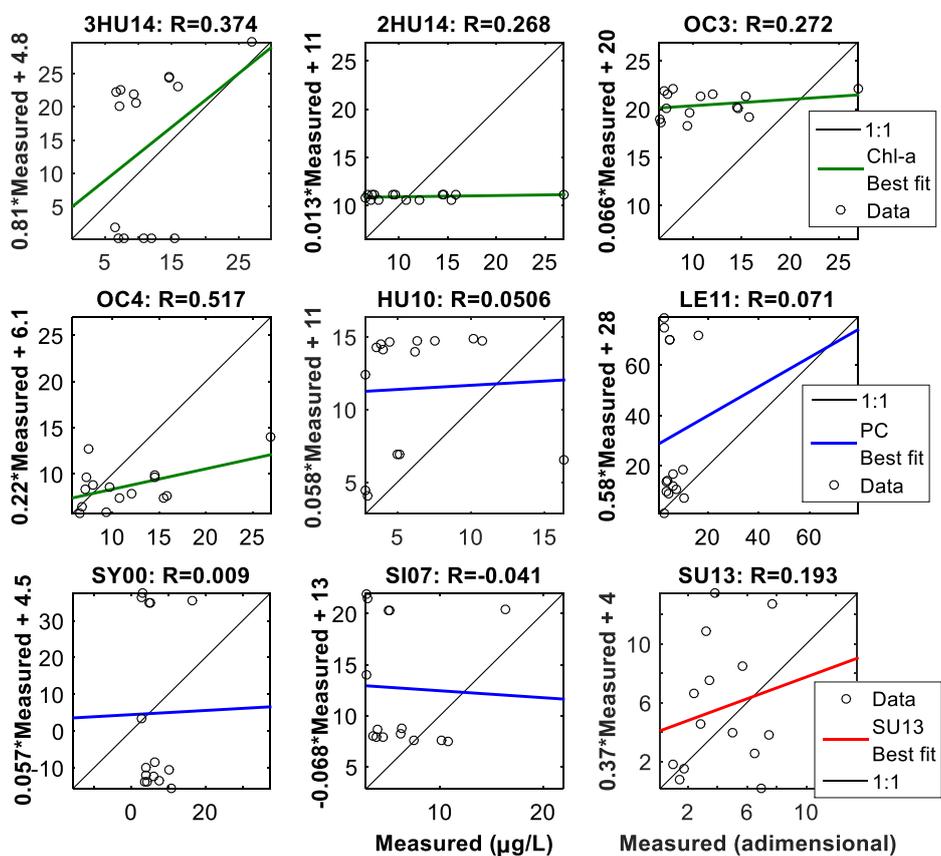
All four chl-a predictive algorithms presented  $R$  values ranging from 0.272-0.517, all being very weak correlation values. 3HU14  $R$  was 0.374. Thence, a slightly overestimating positive correlation was achieved. 2HU14 presented the weakest  $R$  (0.268), and a flat correlation slope, indicating no correlation whatsoever. OC3  $R$  value was also weak (0.272), and it did not yield a significant correlation slope. OC4 had the best chl-a prediction  $R$  (0.517) result, despite fairly underestimating the prediction results. Overall, chl-a estimation was not very efficient. SU13 presented a weak  $R$  value (0.193). However, this value was better than the initial results for this correlation. A very strong outlier was identified in this dataset and was removed from the scatterplot. The results presented an overestimation for lower ratio and an underestimation for higher ratio. Chl-a:TSM estimation, as chl-a results, did not present efficient prediction results for this study's dataset.

PC predictive algorithms yielded worse  $R$  values than chl-a and chl-a:TSM algorithms.  $R$  values for HU10, LE11, SY00, and SI07 were respectively, 0.0506, 0.0071, 0.009, and -0.041. SI07 negative correlation was not attempted to be fixed and was maintained negative since all other scatterplots also presented

very poor results. All these poor results probably had greater relation with the reservoir very low OAC concentration coupled with the reservoir hydraulic dynamics. Such characteristics are capable of creating confusion during the signal data acquisition due to constituents overwhelming one another. Another reason for the poor results is the fact such algorithms are not suitable for the water type encountered in Guarapiranga reservoir. However, as literature has shown over the past years of research in this topic, there has been a difficulty in standardise a global model for OAC prediction in any given water body and then, efficiently further extrapolate the results towards other regions.

Ogashawara et al., (2013) also found poor results in their model calibrations. The authors also tested HU10 algorithm and concluded the poor results from such algorithm demonstrated an inability to suppress specular reflection from wavy conditions. Their results support how such high water composition variability influence on bio-optical modelling of inland waters. Finally, in this study, MI14 had the best result among PC-predictive algorithms despite its low value. Its adjusted  $R^2$  value for the calibration dataset was 2.3% (0.023) (Figure 5.31). And, therefore, it was the chosen algorithm for the re-parameterisation. The following section will describe the re-parameterisation process.

Figure 5.26 – Chlorophyll-a (chl-a), phycocyanin (PC), and chl-a to total suspended matter (TSM) ratio bio-optical algorithms evaluation. Statistics for the correlations are given by Pearson-coefficient (R) for predicted against measured values.



Notes: green, blue, and red lines represent, respectively, the best fit lines for chl-a, PC, and chl-a to TSM ratio. Adimensional x label is only for SU13 algorithm.

### 5.10. Semi-empirical three-band PC-retrieval algorithm re-parameterisation

Figure 5.7 depicted the boxplots for all WQPs. Chl-*a* and PC concentrations had medians at around 9 and 4 µg/L, respectively. Overall, chl-*a* was found in higher concentrations than PC in Guarapiranga reservoir. Figure 5.27a, depicted exactly how many sampling stations had higher chl-*a* concentrations than PC for Guarapiranga. About 64.71% of samples had PC concentrations with less than half that of chl-*a* concentrations; only 11.76% of samples had PC concentrations higher than chl-*a*, with a magnitude of 20% to 40% of higher PC concentration than chl-*a*. Considering that Figure 5.27b described the PC to chl-*a* ratio (PC:chl-*a*) from Funil reservoir and a Catfish Pond (Ogashawara et al. 2013). The authors also described issues over the influence of chl-*a* on PC concentrations. Moreover, they found 56.52% of samples having PC concentrations higher than chl-*a*. They also found only 21.74% samples containing PC concentrations with less than half that of chl-*a*. These numbers indicated this study had samples even more influenced by chl-*a* presence than the previous study in Funil reservoir, and that a confusion-enhancer re-parameterisation needed to be sensitive to that level.

Figure 5.28 depicted the scatterplot between  $R_{rs620}:R_{rs709}$  ratio and PC concentration. Mishra and Mishra (2014) stated the non-linearity existing between PC-predictive algorithms and PC concentration, also found in this study (Figure 5.28). Also, they invested on the conclusion that their novel algorithm had a more robust linearity than its band ratio counterparts (MISHRA; MISHRA, 2014; Figure 4b for that study). Following on this premise, this study evaluated the results from previous sections, and identified the need for a re-parameterisation of a PC-predictive algorithm (MI14-Table 4.3), applying the prediction to a water body with low PC concentration, and suggested a new coefficient for PC prediction in MI14.

Results from previous sections that motivated the generation of this new re-parameterisation coefficient were:

- Least squares regression between  $a_{phy}(620)$  against chl-a+PC showed better correlation strength than that between  $a_{phy}(620)$  against PC only, indicating chl-a was masking PC;
- TLS results generally presented more robust correlation in the higher frequency wavelengths (412-443 nm);
- Ternary plot had the 412 and 443 nm results yielding a more cohesive data distribution, meaning the distance from each point to the mean value in those wavelengths was less variable than for the other wavelengths;
- Outliers and kurtoses for PC and chl-a concentration distribution seemed proportional and influential on one another, specially possible chl-a influence over PC due to higher chl-a concentration;
- DWTCA clusters (Appendix F) presented significant variation among groups in the blue region as well as in the red region;
- And lastly, Guarapiranga's normalised  $R_{rs}$  showed a highly varying slope between 400-550 nm, indicating some fertile spectral region for bio-optical investigation;

These reasons collaborated for the investigation of a possible association between the blue-green region of the spectrum and PC prediction.

The reasoning behind the coefficient generation was as follows: firstly, the non-linearity was encountered in the dataset as shown in Figure 5.28. Then, a thorough investigation of Figure 4.12 along with the last reason in the aforementioned previous sections results list, allowed the identification of dependence between the 460-556 nm slope (S1) and the PC-proxy 600-640 nm depth spectral band (phycocyanin line height, PLH). Thence, Figure 5.29 depicted the reasoning behind the dependence. If S1 to PLH ratio (S1:PLH) displayed a coefficient of proportionality, (i.e., a statistically significant best fit regression line (Figure 5.30a)) then S1 could act as a new re-parameterisation

coefficient replacement for PC-prediction into the bio-optical algorithm MI14 for PC estimation in a tropical reservoir bearing low PC concentration.

Figure 5.30a depicted the regression between S1 and PLH. Adjusted  $R^2$  and NRMSE were, respectively, 0.68 and 15% (0.15), indicating a reasonable linear agreement between variables. An important point is to notice that two sampling stations were excluded from regression in order to improve the correlation. This removal was carried out as to focus the regression results on the scatterplot regions between the variables that displayed the relevant linear correlation degree. Figure 5.30b depicted a surface sensitivity plot for measured PC concentration. It was created in order to evaluate the relation among S1, PLH, and PC as to how S1 and PLH correlated with PC. Lower values of both S1 and PLH were linked to larger PC concentrations; followed by a rapid decrease in PC, then again, an exponential increase in PC concentration as S1 and PLH reached larger values. This trend was similar to that of Figure 5.28, indicating that  $R_{rs620:R_{rs709}}$ , S1, and PLH, presented a non-linear correlation with PC concentration.

Sampling stations P8, P10, P15, and P16 were the ones increasing the PC concentration at lower PLH values in Figure 5.30b. P15 and P16 had high concentrations of both TSM and chl-*a*. Also, these four sampling points had the highest PC concentration in this study. As PC concentration measured in this study was low, other OACs might have overwhelmed PC response to incoming radiation in those sampling points.  $R_{rs}$  spectra used to calculate S1 and PLH might have generated spectral artefacts such that of placing higher PC concentration values at the lower end of S1 and PLH. Therefore, a highly sensitive PC-retrieval algorithm, for optically complex inland waters with low PC concentration, in order to reduce confounding factors and spectral artefacts was again demonstrated as needed. As S1 and PLH were considered having a linear correlation (Figure 5.30a), PLH as a function of S1 was then applied into MI14 (MISHRA; MISHRA, 2014), replacing the originally suggested coefficient ( $\psi$ ), and creating the adjusted MI14 by adding the new coefficient, *PLH*, instead.

Figure 5.27 – PC to chl-a ratio (PC:chl-a) monotonically increasing for all sampling stations from this study’s dataset and from Ogashawara et al. (2013). Horizontal line at the ratio value of one indicates how many samples have higher concentrations of PC (above line) or chl-a (beneath line).

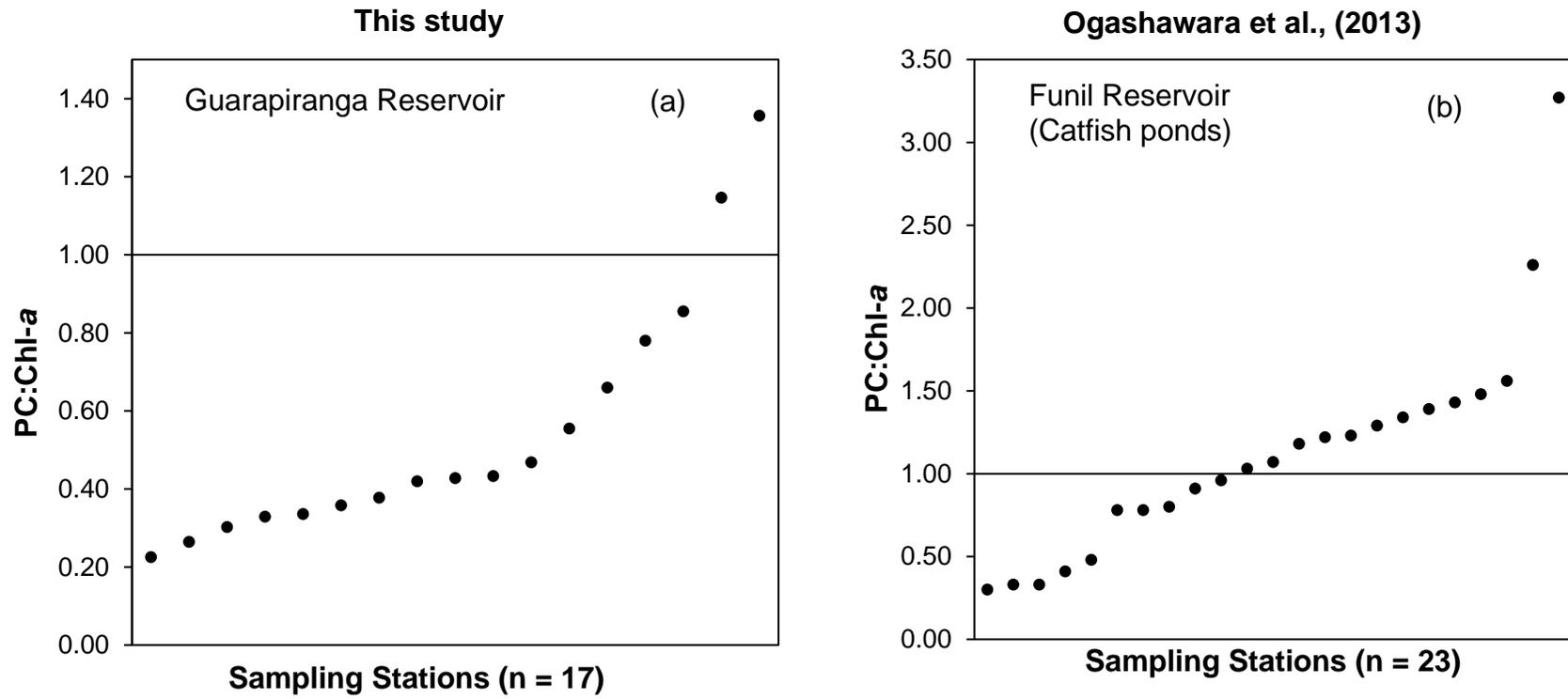


Figure 5.28 – Scatterplot between a reflectance band ratio [ $R_{rs}(620):R_{rs}(709)$ ] and measured PC concentration.

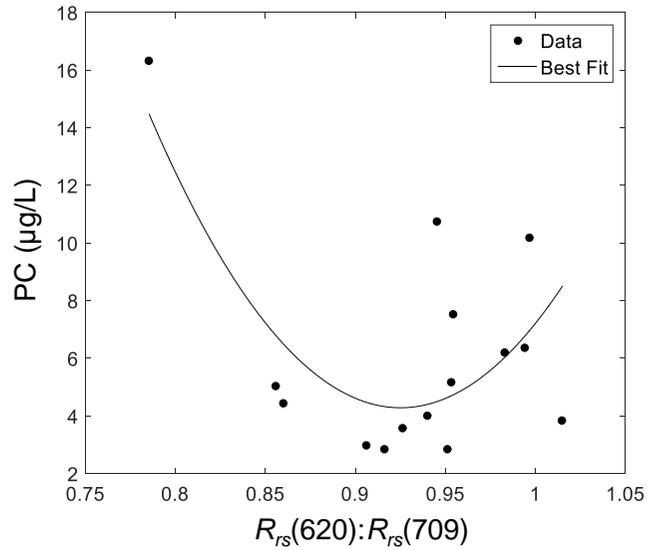
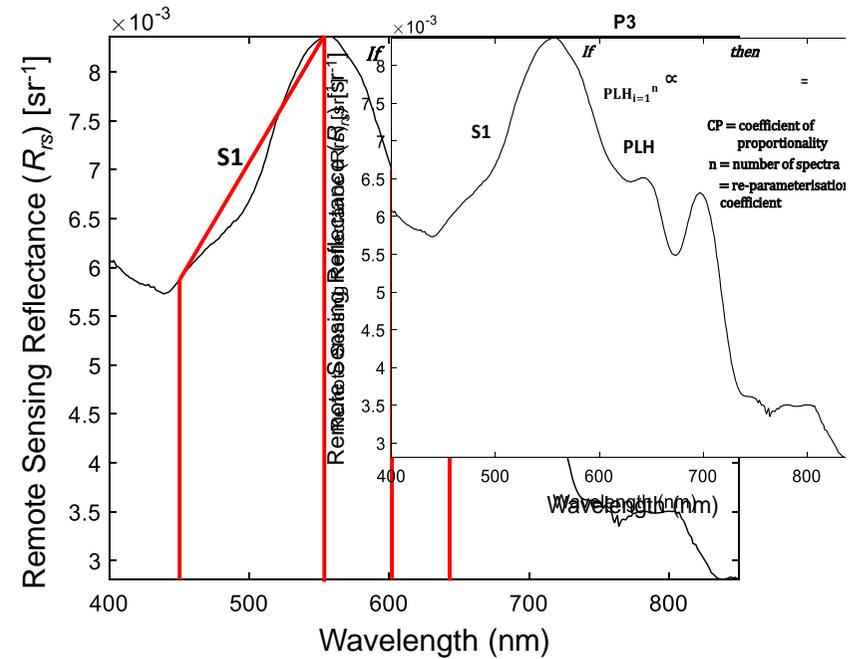
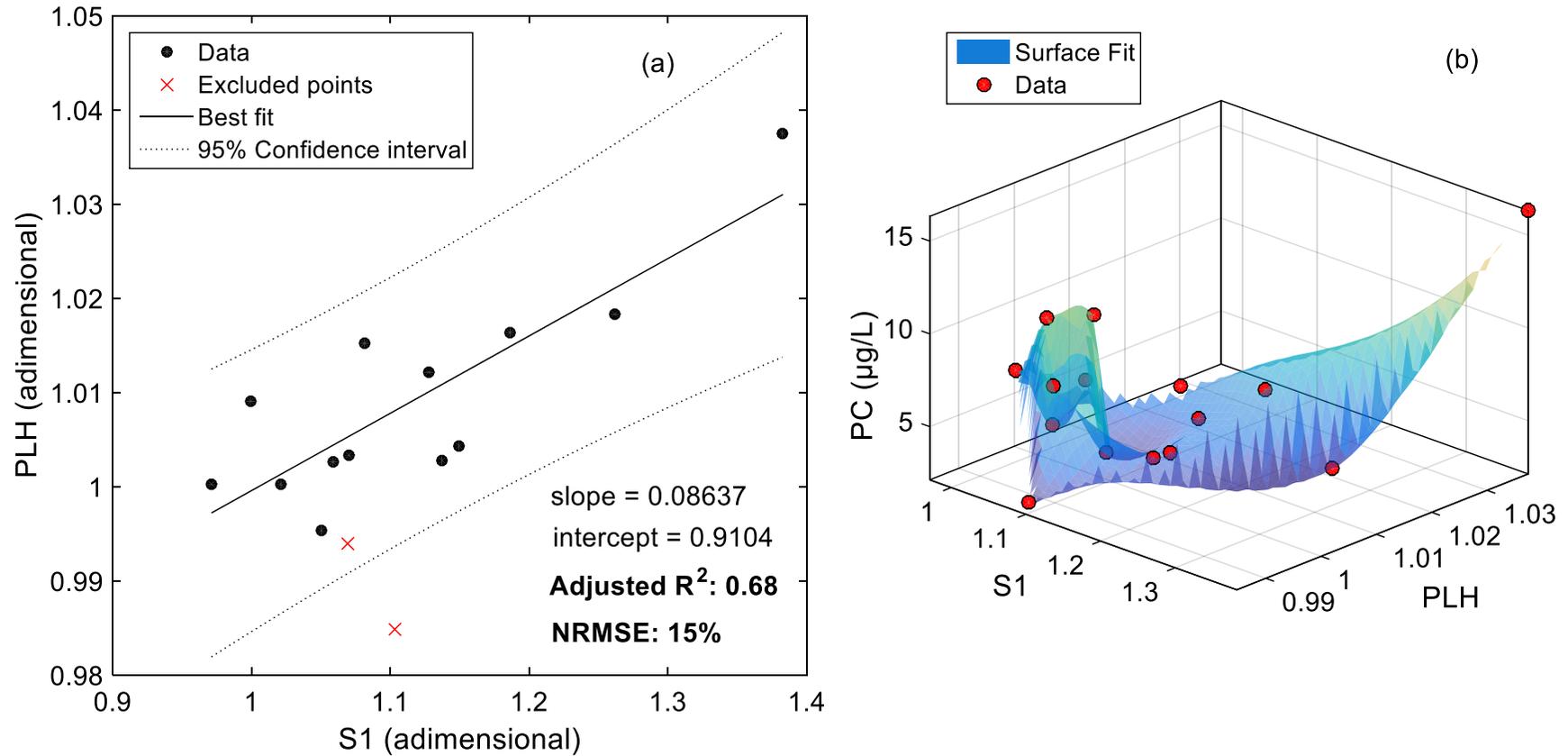


Figure 5.29 – Remote sensing reflectance ( $R_{rs}$ ) spectrum for sampling station three, slope one (S1), and phycocyanin line height (PLH). The coefficient of proportionality represents the correlation between S1 and PLH for all sampling points. The slope between  $PLH$  and S1 (Figure 30a) represents the re-parameterisation coefficient.



Notes: PLH represents the ratio between the value of the adjusted fit line between 600 and 640 nm and the actual  $R_{rs}$  value at 620 nm; S1 represents the ratio between the  $R_{rs}(556)$  over  $R_{rs}(460)$ . CP stands for coefficient of proportionality between  $S1$  and  $PLH$ ; and  $PLH$  stands for the new coefficient applied to the adjusted algorithm.

Figure 5.30 – (a) Regression analysis between phycocyanin (PC) line height (PLH) and slope one (S1). (b) Sensitivity plot between PLH, S1 and PC concentration.



Notes: PLH represents the ratio between the value of the adjusted fit line between 600 and 640 nm and the actual  $R_{rs}$  value at 620 nm; S1 represents the ratio between the  $R_{rs}(556)$  over  $R_{rs}(460)$ .

Dash et al. (2011) evaluated whether wavelengths less than 620 nm were applicable for the estimating PC. This study has reached a bit further and attempted to stipulate whether the blue-green region slope could improve PC prediction. Higher frequency regions had already been used by Dong et al. (2013). Moreover, Yacobi et al. (2015) attempted to minimise chlorophylls influence over PC on the basis of chlorophylls relative absorption at 625 nm, using 675 nm as reference wavelength.

In order to further the scientific frontier associated with this area of study, this master's thesis capitalised on the initial attempts proposed by the aforementioned studies. Unlike Yacobi et al. (2015), this thesis has not used specific absorptions coefficients for eliminating the confusion between chl-*a* and PC. Nonetheless, this thesis followed on Dash et al. (2011)'s work, altogether with the endeavour to further the semi-empirical algorithm proposed by Gitelson et al. (2003), adapted by Hunter et al. (2010), and enhanced by Mishra and Mishra (2014). And peremptorily, it aimed to excelling in bettering PC retrieval at very low concentrations.

Figure 5.31 depicted the results for the new algorithm calibration, using this study's dataset, and validation, using dataset from Augusto-Silva et al. (2014). Calibration of original algorithm (MI14) and adjusted algorithm (MI14-Adjst.) were represented by blue best fit lines; and validation results were represented by red best fit lines. MI14 algorithm, for both calibration and validation, kept the coefficient fit curve ( $\psi = 0.709 * [R_{rs}(560)/R_{rs}(665)]^{0.874}$ ) proposed by Mishra and Mishra (2014). The MI14 Adjst. used the new re-parameterising coefficient slope and intercept depicted in Figure 5.30a correlating S1 and PLH. Optimal bands for both MI14 and MI14-Adjst. were selected based on the previous works of Hunter et al. (2010) and Mishra and Mishra (2014).

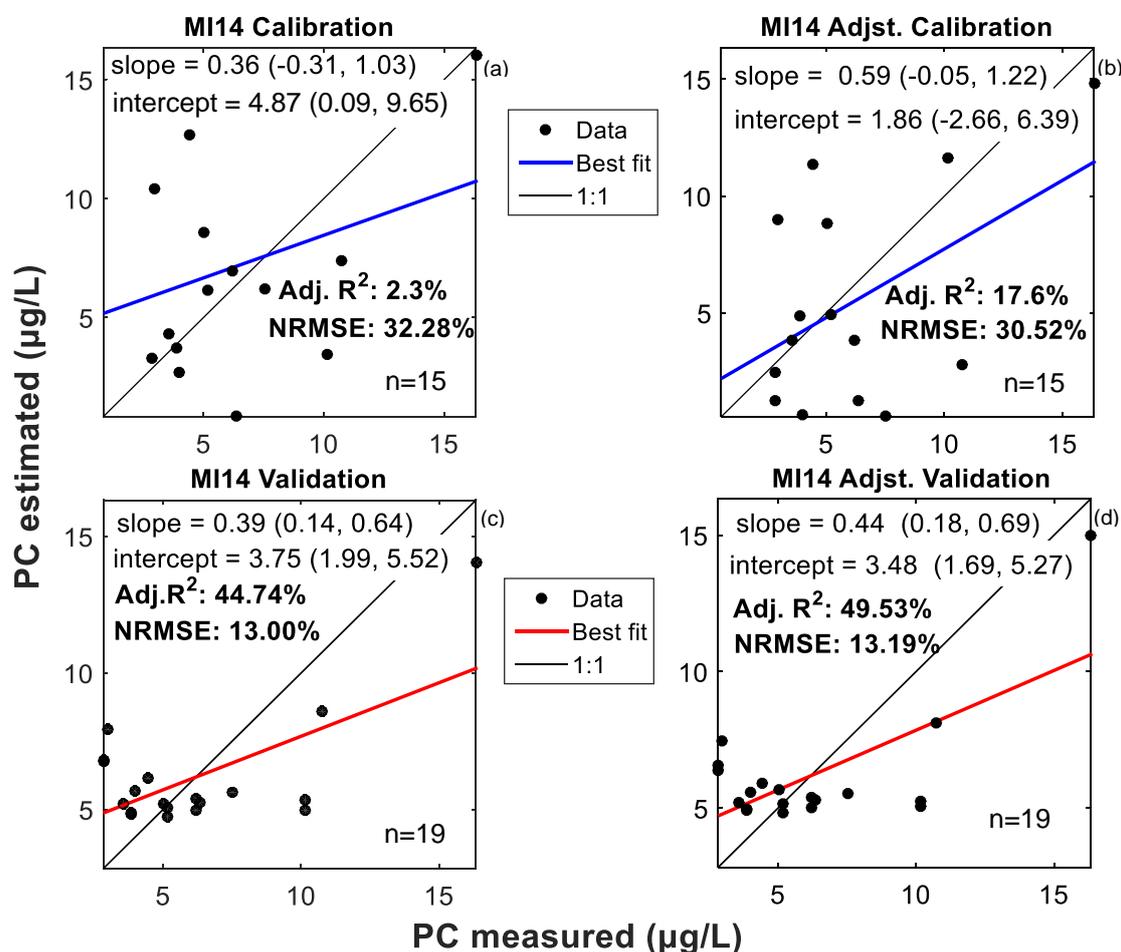
MI14 and MI14-Adjst. calibrations yielded adjusted  $R^2$  of 2.3%, 17.6%, respectively; and NRMSE of 32.28%, and 30.52%, respectively. Also, there was a slight improvement in slope prediction angles from MI14 to MI14-Adjst.

shifting towards the one-to-one position, however, all datasets presented overestimation trend at lower concentrations and underestimation trend at higher concentrations. Both algorithms performed poorly overall, mostly because of the very low PC concentration range appraised in the study which, as mentioned previously, might have generated spectral artefacts increasing uncertainty in OAC predictions. At very low concentrations, confounding effects could negatively contribute by neglecting PC interaction with the incoming radiation. However, it is clear from the results that the adjusted algorithm calibration dataset improved the correlation robustness ( $R^2$ ), and slightly minimised the prediction error.

It is important to note that the validation (Funil reservoir) dataset must not be assessed for its results statistics magnitude. Funil's dataset did not have available measured PC concentration, but only the  $R_{rs}$  dataset for applying the prediction algorithm. Therefore, a randomisation procedure was applied to Guarapiranga's measured PC concentration, and then, it was used as measured PC data into the validation dataset. Therefore, regression analysis for the validation dataset should only be evaluated for the difference between the original and adjusted algorithms rather than the actual results magnitudes.

One can observe that for validation dataset, both algorithms presented corresponding NRMSE. However,  $R^2$  from MI14-Adjst. was still better than that of MI14, in spite of displaying a narrower  $R^2$  gap value between the two algorithms. Such feature indicated that the re-parameterisation tended to reach a saturation difference. This might be an indication that, somehow, the blue spectral region improved PC prediction in Guarapiranga better than it did for Funil reservoir which had slight more homogeneous blue-green  $R_{rs}$  slope than Guarapiranga reservoir. Perhaps, higher-complexity waters might benefit most from this re-parameterisation.

Figure 5.31 – Calibration (this study’s dataset) and validation (dataset from Augusto-Silva, 2014) of the adjusted MI14 algorithm, and comparison with the original MI14 algorithm. Brackets indicate 95% confidence level.



Notes: measured PC concentrations for the validation dataset were randomly generated from the calibration dataset. Four extra samples were added in the randomisation in order to match the calibration dataset sample size with the validation dataset sample size (i.e., nineteen values). Therefore, R<sup>2</sup> and NRMSE values for the validation dataset must not be assessed for their actual magnitude. These magnitudes must only be investigated as a comparison between the original and the adjusted algorithms is carried out.

### 5.11. OLCI/Sentinel 3 simulation

OLCI simulation was carried out in order to investigate how this new sensor would have behaved acquiring signal data from Guarapiranga reservoir waters. Firstly, a simulated dataset of OLCI spectral bands was created following its spectral response function showed on section 4.5 (Figure 4.14). A comparison between hyperspectral and the synthetic datasets was shown on Figure 5.32 which showed both the hyperspectral *in situ* data, overlapped by the synthetic dataset for each successfully measured sampling station.

The simulated dataset of the 16 spectral bands of OLCI could estimate well the hyperspectral data, and could retrieve efficiently some important OAC-related features from the  $R_{rs}$  spectra. The absorption feature of chl-*a* at 665 nm was not well described in the simulated dataset. That happened due to the fact the 675 nm was more clearly representative of maximum absorption in Guarapiranga's spectra. Moreover, the three narrowly related 666, 674, and 682 nm bands presented varying positions over all spectra.

P2, P3, P10 and P16 in Figure 5.32 showed a clear absorption trough at the 674 nm position. It was an indication these sampling points matched the spectral characteristics from data used to generate conclusions over the choice for OLCI's 666, 674, and 682 nm bands. However, other stations did not so equivalently matched that chl-*a* absorption trough. One point had that trio band left-shifted (P9). Some others, had it right-shifted (P4, P7, P11, P12, P14, P15, and, P17). And some had it flat-lined (P5 and P8). These features indicated how one study site might have had its spectral data slightly altered by fixedly-selected narrow bands. And due to the complexity found in inland water environments, great care must be taken while generating universal algorithms for OAC prediction.

The same reasoning could be applied to the 620 nm region where PC maximum absorption is located. Some of the synthetic spectra matched this region's features from the hyperspectral *in situ* dataset, and some others did not. Again,

while developing algorithms for PC prediction using OLCI data, one must be careful as to note how subtle variations in OLCI data in comparison with *in situ* data could imbue uncertainties in the results. Overall, this study evaluated optimal bands for the re-parameterisation, closely identifying the maximum absorption happening near the 620 nm region. Therefore, the re-parameterisation included the 630 nm instead of 629 nm for the PC band suggested by Mishra and Mishra (2014); also, it was used 670 nm, instead of 659 nm for chl-*a* band; and the 730 nm instead of the 724 nm band for high water absorption region, by noting some shifts in the over-700 nm region in Guarapiranga dataset.

Lastly, although OLCI is a MERIS heritage sensor, a difficulty has always existed for inland water scientists to universally apply bio-optical remote-sensing algorithms for OAC prediction. This fact has been due to the high complexity associated with inland complex waters. And such complexity might appear in signal data in such a way that even sensors like OLCI may collect signal variation subtleties capable of hindering scientists from excelling in OCW research.

Figure 5.32 – Remote sensing reflectance ( $R_{rs}$ ) (red line) for each sampling station calculated from *in situ*-collected radiance and irradiance data via Ramses TRioS; and resampled  $R_{rs}$  data to the first sixteen OLCI sensor bands (400- 779 nm) (blue line).

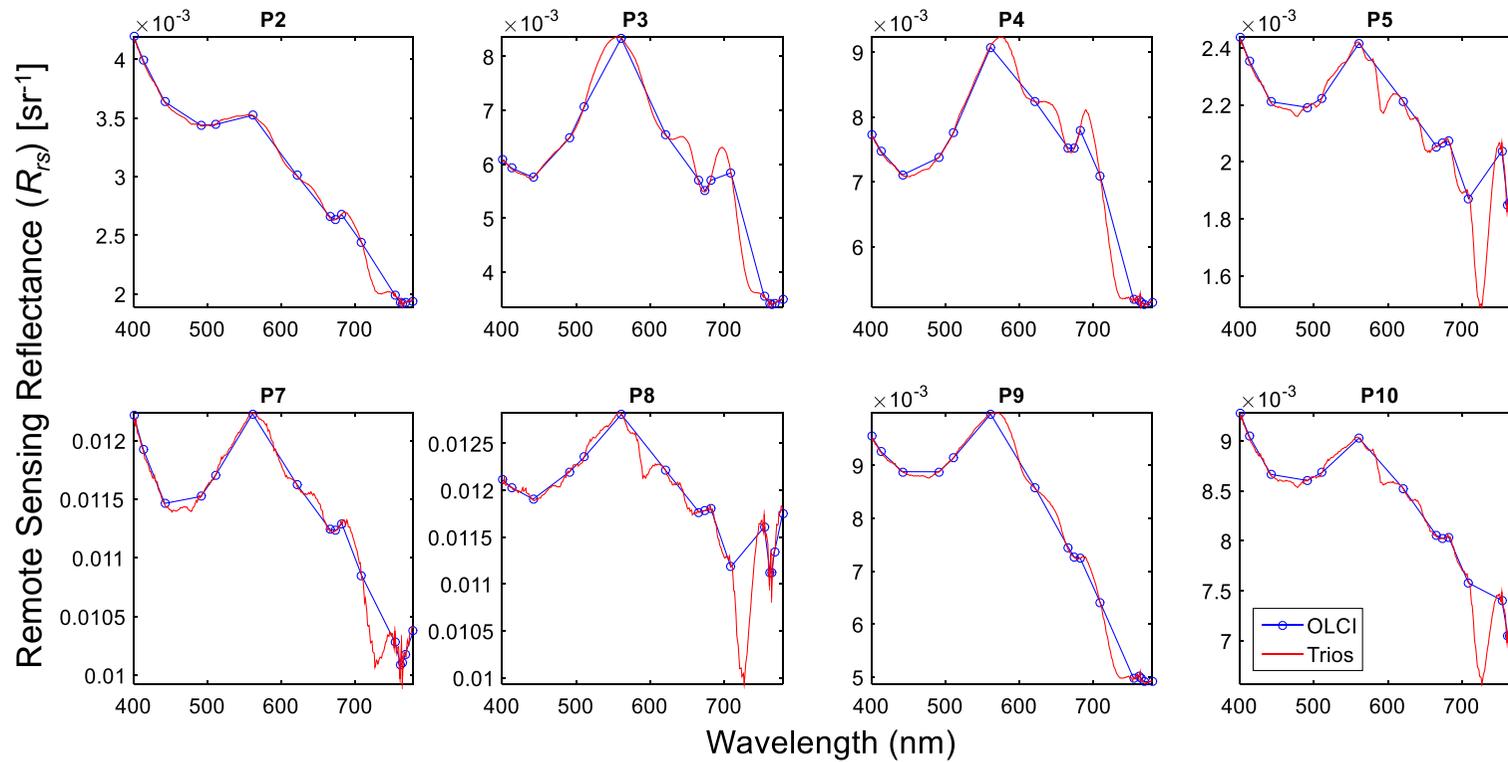
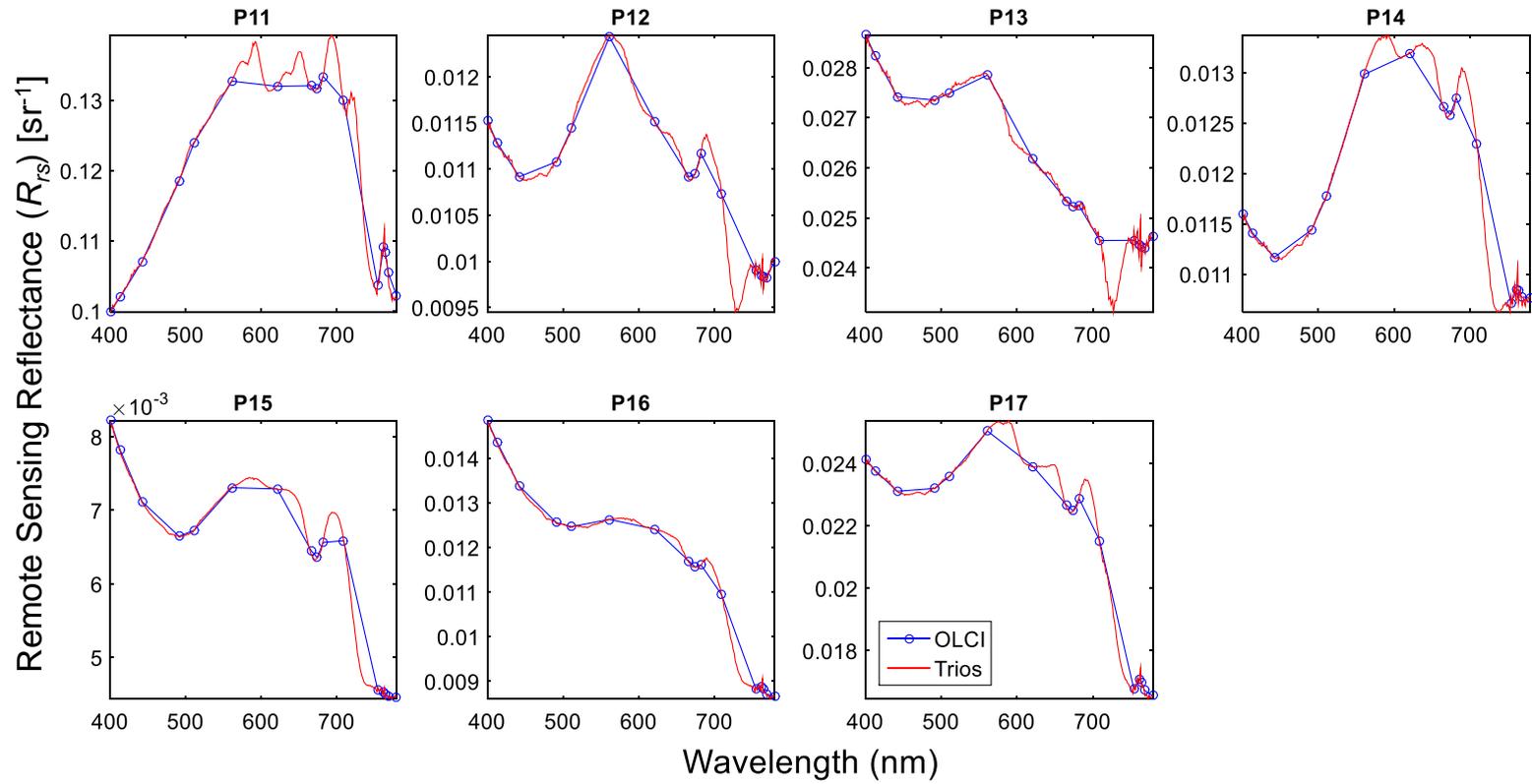


Figure 5.32 – Remote sensing reflectance ( $R_{rs}$ ) (red line) for each sampling station calculated from *in situ*-collected radiance and irradiance data via Ramses TRioS; and resampled  $R_{rs}$  data to the first sixteen OLCI sensor bands (400- 779 nm) (blue line) – Continuation.





## 6 FINAL CONSIDERATIONS AND CONCLUDING REMARKS

In this research, a three-band semi-empirical PC-predictive algorithm, MI14, was improved, MI14-Adjst., to accurately retrieve PC concentration in an inland water body bearing low PC concentration. Best bands were selected for the re-parameterised algorithm input wavelengths at 630 nm, 670 nm, and 730 nm for wavelengths that are most sensitive to PC absorption, least sensitive to PC absorption (via the inclusion of enhancing coefficient) and, a wavelength region where pigment absorption is negligible, respectively. The optimal adjusted algorithm improved prediction from the original MI14. It showed an improvement from 2.3% to 17.6% in the adjusted  $R^2$  and a decrease from 32.28% to 30.52% in the NRMSE for the calibration dataset. As for the validation dataset, adjusted  $R^2$  was less improved but still better (from 44.74% to 49.53%) and NRMSE reached an equalised value (from 13.00% to 13.19%).

For practical applicability of the algorithm, OLCI spectral bands were considered while re-parameterising the algorithm. However, application of OLCI-specific bands into the algorithms has not been made because bands (630 nm, 670 nm, and 730 nm) used in MI14-Adjst. showed different  $R_{rs}$  magnitude and shape for the simulated dataset, and would have not shown optimal results just as Mishra and Mishra (2014) found for their study. This was true because optimal bands vary from study site to study site, and for one sort of OAC concentration to another. The new coefficient introduced in this study had the green band (556) centred at 5-nm distance from OLCI's band location (561). And the blue band (460) was located halfway through from the 442-nm band and the 491-nm band allowing it to be averaged out from those two values. Nevertheless, the coefficient and main algorithm bands in this study were very proximal to OLCI's bands, and for the purposes of remote sensing studies, OLCI sensor is still the best opportunity to further inland water research. And above all, the new improved algorithm seemed to improve prediction for waters with low PC concentration range as a shortcoming pointed out by Mishra and Mishra (2014) for future research investigations.

It is evident from the results that: (a) MI14-Adjst. can be applied to retrieve low-concentration PC in inland waters with slight higher accuracy than an existing algorithm; (b) inclusion of *PLH* in the model compensates better for the confounding effect of Chl-*a* absorption at the PC absorption band - i.e. 620 nm – than MI14 coefficient ( $\psi$ ); and (c) unlike reflectance band ratios, MI14-Adjst. addressed the non-linearity issue also found in lower PC concentrations. However, it is a must for other studies to be carried out using this improved algorithm under similar PC concentrations for further validation. Development of remote sensing algorithms for quantifying cyanobacteria PC is a challenging task primarily because specific absorption coefficient of PC and chl-*a* varies significantly with change in phytoplankton species composition. In addition, chl-*a* absorption at the PC absorbance maxima causes significant interference that adds to the model complexities (MISHRA et al., 2013; SONG et al., 2013a).

The aforementioned conclusions reinforce that the hypothesis set for this study has been accepted. Also, as for the three specific aims set-out, the first was fairly descriptively concluded, and supported the third one which led to the algorithm re-parameterisation. The first specific aim managed to explain the high complexity associated with Guarapiranga water reservoir which is used for a major city supply such as the city of São Paulo. Such complexity exists mostly because of the highly varying land cover and land use present in the city's surroundings. This complexity makes it more difficult the understanding of pollution sources into the reservoir but also overwhelms the remote sensing techniques used for this very same pollution evaluation.

The second specific aim had a minor contribution for the final re-parameterisation as this objective was far less extensive than the first one. However, differently from the first specific aim, the second one intended to further the state-of-art associated with the most-wanted analytical bio-optical OAC-prediction algorithms, which is an all-encompassing bio-optical model. The second specific aim initial intention (not achieved in this work) was for using the DWTCAs coupled with many spectra. And, then, set empirical relations that

described more efficiently every OAC present in water bodies via only  $R_{rs}$  signal data without the need for all the IOPs used in semi-analytical algorithms. This feat could generate results which would greatly contribute for future satellite missions.

Local indices for environmental parameters estimation can be highly useful for practical purposes such as real-time monitoring of water quality. This study's results aimed to better future inland waters monitoring systems as cyanobacterial bloom prevention systems in waters bearing low OAC concentration. It suggests OCW not to be treated simplistically, as their variation complexity over the spatial-temporal domain is similar to the complexity of signals datasets investigated by a wavelet technique. Such technique can investigate much information inside intricate signal levels that are not easily perceivable at certain scales. All this complexity requires much effort from scientific community, and collective work. This way, perhaps, future inland water research might achieve more productive results than those brought forth to this very day.



## BIBLIOGRAPHIC REFERENCES

- ALCAIDE, R. L. M.; FORTI, M. C. **Manual de procedimentos do analisador de carbono total Toc-Vcpn**. 1. ed. São José dos Campos: INPE, 2012. v. 1, 31 p. (sid.inpe.br/mtc-m19/2012/01.20.16.33-MAN). Disponível em: <http://urlib.net/8JMKD3MGP7W/3B86M5L>. Acesso em: 03 mar. 2016.
- AMPE, E. M.; HESTIR, E. L.; BRESCIANI, M.; et al. A wavelet approach for estimating chlorophyll-a from inland waters with reflectance spectroscopy. **IEEE Geoscience and Remote Sensing Letters**, v. 11, n. 1, p. 89–93. doi: 10.1109/LGRS.2013.2247021, 2014.
- AMPE, E. M.; RAYMAEKERS, D.; HESTIR, E. L.; et al. A wavelet-enhanced inversion method for water quality retrieval from high spectral resolution data for complex waters. **IEEE Transactions on Geoscience and Remote Sensing**, v. 53, n. 2, p. 869–882. doi: 10.1109/TGRS.2014.2330251, 2015.
- AUGUSTO-SILVA, P.; OGASHAWARA, I.; BARBOSA, C.; et al. Analysis of MERIS reflectance algorithms for estimating chlorophyll-a concentration in a brazilian reservoir. **Remote Sensing**, v. 6, n. 12, p. 11689–11707. doi: 10.3390/rs61211689, 2014.
- AUSTIN, R. W.; HALIKAS, G. **The index of refraction of seawater SIO**. Ref. 76-1. Retrieved from <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA024800>, 1976.
- BABIN, M.; STRAMSKI, D.; FERRARI, G.; et al. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. **Journal of Geophysical Research**, v. 108, n. C7, p. 3211. doi: 10.1029/2001JC000882, 2003.
- BATIUK, R. A.; BERGSTROM, P.; KEMP, M. **Chesapeake bay submerged aquatic vegetation water quality and habitat-based requirements and restoration targets: a second technical synthesis**. Chesapeake Bay: Environmental Protection Agency, 2000.
- BLOCH, J. R. Using and interpreting statistics: a practical text for the health, behavioral, and social sciences. **Nursing Education Perspectives**, v. 28, n. 6, p. 346. National League for Nursing, Inc. Retrieved November 20, 2015, from <https://www.questia.com/library/journal/1P3-1401240311/using-and-interpreting-statistics-a-practical-text>, 2007.
- BOLZAN, M. J. A. Análise da transformada em ondeletas aplicada em sinal geofísico. **Revista Brasileira de Ensino de Física**, v. 26, n. 1, p. 37–41.

Sociedade Brasileira de Física. doi: 10.1590/S1806-11172004000100006, 2004.

BREZONIK, P. L.; OLMANSON, L. G.; FINLAY, J. C.; BAUER, M. E. Factors affecting the measurement of CDOM by remote sensing of optically complex inland waters. **Remote Sensing of Environment**, v. 157, p. 199–215. doi: 10.1016/j.rse.2014.04.033, 2015.

BRICAUD, A.; BABIN, M.; MOREL, A.; CLAUSTRÉ, H. Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization. **Journal of Geophysical Research**, v. 100, n. C7, p. 13321. doi: 10.1029/95JC00463, 1995.

BRICAUD, A.; MOREL, A.; PRIEUR, L. 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. doi: 10.4319/lo.1981.26.1.0043, 1981.

BUKATA, R. P. **Satellite monitoring of inland and coastal water quality**. [S.l.]: CRC Press, 2005. ISBN 9780849333569 - CAT# 3356.

BUKATA, R. P. Retrospection and introspection on remote sensing of inland water quality: “Like Déjà Vu All Over Again.” **Journal of Great Lakes Research**, v. 39, p. 2–5. doi: 10.1016/j.jglr.2013.04.001, 2013.

BUKATA, R. P.; BRUTON, J. E.; JEROME, J. H.; JAIN, S. C.; ZWICK, H. H. Optical water quality model of Lake Ontario. 2: Determination of chlorophyll a and suspended mineral concentrations of natural waters from submersible and low altitude optical sensors. **Applied optics**, v. 20, n. 9, p. 1704–14. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/20309372>, 1981b.

BUKATA, R. P.; JEROME, J. H.; BRUTON, J. E.; JAIN, S. C. Determination of inherent optical properties of Lake Ontario coastal waters. **Applied optics**, v. 18, n. 23, p. 3926–32. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/20216727>, 1979.

BUKATA, R. P.; JEROME, J. H.; BRUTON, J. E.; JAIN, S. C.; ZWICK, H. H. Optical water quality model of Lake Ontario. 1: determination of the optical cross sections of organic and inorganic particulates in Lake Ontario. **Applied optics**, v. 20, n. 9, p. 1696–703. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/20309371>, 1981a.

BULGAKOV, N.; LEVICH, A. The nitrogen:phosphorus ratio as a factor regulating phytoplankton community structure. **Archiv für Hydrobiologie**, v. 146, n. 1, p. 3–32. Retrieved from <http://cat.inist.fr/?aModele=afficheN&cpsidt=1954645>, 1999.

CARLSON, R. E. A trophic state index for lakes. **Limnology and Oceanography**, v. 22, n. 2, p. 361–369. doi: 10.4319/lo.1977.22.2.0361, 1977.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Censo 2000. Rio de Janeiro: Instituto Brasileiro de Geografia e Estatística, 2000.

COMPANHIA DE TECNOLOGIA DE SANEAMENTO AMBIENTAL (CETESB). **Relatório de qualidade das águas interiores no estado de São Paulo: 2006**. São Paulo, 2007.

CHAMON, M. A. **The SABIA-Mar Mission**. Retrieved from [http://www.iocccg.org/sensors/Chamon\\_SABIA\\_MAR.pdf](http://www.iocccg.org/sensors/Chamon_SABIA_MAR.pdf). Accessed at: 10th January 2016, 2013. 14 transparências.

ERDOS, P. Frequency analysis: the fourier series. In: CHAPARRO, L. (ed.). **Signals and systems using MATLAB**. 2. ed. Amsterdam: Elsevier, 1015. p.263–332. Elsevier. doi: 10.1016/B978-0-12-394812-0.00004-8.

CLEVERS, J. G. P. W.; GITELSON, A. A. Remote estimation of crop and grass chlorophyll and nitrogen content using red-edge bands on Sentinel-2 and -3. **International Journal of Applied Earth Observation and Geoinformation**, v. 23, n. 2012, p. 344–351. Elsevier B.V. doi: 10.1016/j.jag.2012.10.008, 2013.

COX, C.; MUNK, W. Measurement of the roughness of the sea surface from photographs of the sun's glitter. **Journal of the Optical Society of America**, v. 44, n. 11, p. 838–850, 1954.

CPLEA; COBRAPE. **Atualização do plano de desenvolvimento e proteção ambiental da bacia hidrográfica do Guarapiranga**. São Paulo, 2006.

CURTARELLI, M.; LEÃO, J.; OGASHAWARA, I.; LORENZZETTI, J.; STECH, J. Assessment of Spatial Interpolation Methods to Map the Bathymetry of an Amazonian Hydroelectric Reservoir to Aid in Decision Making for Water Management. **ISPRS International Journal of Geo-Information**, v. 4, n. 1, p. 220–235. doi: 10.3390/ijgi4010220, 2015.

DASH, P.; WALKER, N. D.; MISHRA, D. R.; et al. Estimation of cyanobacterial pigments in a freshwater lake using OCM satellite data. **Remote Sensing of Environment**, v. 115, n. 12, p. 3409–3423. Elsevier Inc. doi: 10.1016/j.rse.2011.08.004, 2011.

DAUBECHIES, I. **Ten lectures on wavelets**. v. 61. Society for Industrial and Applied Mathematics. doi: 10.1137/1.9781611970104, 1992.

DEKKER, A. G. **Detection of optical water quality parameters for eutrophic waters by high resolution remote sensing.** 222 p. PhD. Thesis. Vrije Universiteit, 1993.

DEKKER, A. G.; HESTIR, E. L. **Evaluating the feasibility of systematic inland water quality monitoring with satellite remote sensing.** CSIRO, 2012. Water for a Healthy Country National Research Flagship.

DEUTSCH, C. V.; JOURNEL, A. G. **GSLIB: Geostatistical software library and user's guide**Oxford. Oxford: University Press, 1992. doi: 10.1016/0098-3004(94)90041-8.

DIETZ, T.; OSTROM, E.; STERN, P. C. The struggle to govern the commons. **Science**, v. 302, n. 5652, p. 1907–1912, 2003. doi: 10.1126/science.1091015.

DODDS, W. K.; WHILES, M. R. Aquatic chemistry and factors controlling nutrient cycling. **Freshwater Ecology**. p.289–321, 2010. Elsevier. doi: 10.1016/B978-0-12-374724-2.00012-X.

DONG, Q.; SHANG, S.; LEE, Z. An algorithm to retrieve absorption coefficient of chromophoric dissolved organic matter from ocean color. **Remote Sensing of Environment**, v. 128, p. 259–267, 2013. doi: 10.1016/j.rse.2012.10.013.

DONLON, C.; BERRUTI, B.; BUONGIORNO, A.; et al. The Global Monitoring for Environment and Security (GMES) sentinel-3 mission. **Remote Sensing of Environment**, v. 120, n. 2012, p. 37–57, 2012. doi: 10.1016/j.rse.2011.07.024.

ESA, E. S. A. Spectral Response Function Data. . Retrieved from <https://sentinel.esa.int/web/sentinel/technical-guides/sentinel-3-olci/olci-instrument/spectral-response-function-data>, 2016.

Fletcher, K. (ed). **Sentinel-3: ESA's global land and ocean mission for GMES operational services.** Noordwijk, NL: ESA Communications, 2012.

FALCONER, I. R.; HUMPAGE, A. R. Cyanobacterial (blue-green algal) toxins in water supplies: Cylindrospermopsins. **Environmental Toxicology**, v. 21, n. 4, p. 299–304, 2006. doi: 10.1002/tox.

FALKOWSKI, P. G.; LAROCHE, J. Acclimation of spectral irradiance in algae. **Journal of Phycology**, v. 27, p. 8–14, 1991. doi: 10.1111/1529-8817.ep10868738.

FISCHER, M. M.; GETIS, A. **Handbook of applied spatial analysis** software tools, methods and applications. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. v. 3. doi: 10.1007/978-3-642-03647-7.

FLORES, A. P.; SORRINI, E.; REAL, T. Monitoramento em tempo real da qualidade da água dos mananciais da região metropolitana de São Paulo - RMSP. In: CONGRESSO BRASILEIRO DE ENGENHARIA SANITÁRIA E AMBIENTAL VI-004, 21., 2001, João Pessoa. **Anais...** João Pessoa: ABES, 2001. n. 1, p. 1–8.

FORTI, M. C.; LEE, R.; ALCAIDE, M. **Manual de procedimentos de cromatografia iônica do laboratório de soluções aquosas e tecnologias - Laquatec**. 1. ed. São José dos Campos: INPE, 2011. v. 1, 52 p. (sid.inpe.br/mtc-m19/2011/06.03.13.41-MAN). Disponível em:<<http://urlib.net/8JMKD3MGP7W/39QJ77E>>. Acesso em: 03 mar. 2016.

FREITAS, R. M.; SHIMABUKURO, Y. E. Combining wavelets and linear spectral mixture model for MODIS satellite sensor time-series analysis. **Journal of Computational Interdisciplinary Sciences**, v. 1. n. 1, p. 51-56, 2008. doi: 10.6062/jcis.2008.01.01.0005.

GEGE, P. WASI-2D: A software tool for regionally optimized analysis of imaging spectrometer data from deep and shallow waters. **Computers and Geosciences**, v. 62, p. 208–215, 2014. doi: 10.1016/j.cageo.2013.07.022.

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

GIARDINO, C.; BRANDO, V. E.; DEKKER, A. G.; STRÖMBECK, N.; CANDIANI, G. Assessment of water quality in Lake Garda (Italy) using Hyperion. **Remote Sensing of Environment**, v. 109, p. 183–195, 2007. doi: 10.1016/j.rse.2006.12.017.

GITELSON, A. A.; DALL'OLMO, G.; MOSES, W.; et al. A simple semi-analytical model for remote estimation of chlorophyll-a in turbid waters: Validation. **Remote Sensing of Environment**, v. 112, n. 9, p. 3582–3593, 2008. doi: 10.1016/j.rse.2008.04.015.

GITELSON, A. A.; GRITZ, Y.; MERZLYAK, M. N.; et al. 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, n. 3, p.271-282, 2003.

GÓMEZ, J. A. D.; ALONSO, C. A.; GARCÍA, A. A. Remote sensing as a tool for monitoring water quality parameters for Mediterranean Lakes of European Union water framework directive (WFD) and as a system of surveillance of cyanobacterial harmful algae blooms (SCyanoHABs). **Environmental**

**Monitoring and Assessment**, v. 181, n. 1-4, p. 317–334, 2011. doi: 10.1007/s10661-010-1831-7.

GONS, H. J. Optical teledetection of chlorophyll a in turbid inland waters. **Environmental Science & Technology**, v. 33, n. 7, p. 1127–1132, 1999. doi: 10.1021/es9809657.

GORDON, H. R.; BROWN, O. B.; JACOBS, M. M. Computed relationships between the inherent and apparent optical properties of a flat homogeneous ocean. **Applied Optics**, v. 14, n. 2, p. 417, 1975a. doi: 10.1364/AO.14.000417.

GORDON, H. R.; MCCLUNEY, W. R. Estimation of the depth of sunlight penetration in the sea for remote sensing. **Applied Optics**, v. 14, n. 2, p. 413, 1975b. doi: 10.1364/AO.14.000413.

GORDON, H. R.; MOREL, A. Y. Remote assessment of ocean color for interpretation of satellite visible imagery: a review. **Lect. Notes Coastal Estuarine Stud.**, 1983.v. 4. doi: 10.1029/LN004.

GORDON, R.; BROWN, O. B.; EVANS, H.; et al. A semianalytic radiance model of ocean color. **Journal of Geophysical Research**, v. 93, n. D9, p. 10909–10924, 1988.

GROEN, P. P. N. **An introduction to total least squares.** , v. 14, n. 2. Retrieved November 24, 2015, from [http://www.researchgate.net/publication/2139664\\_An\\_Introduction\\_to\\_Total\\_Least\\_Squares](http://www.researchgate.net/publication/2139664_An_Introduction_to_Total_Least_Squares), 1998.

HARRISON, W. G.; PLATT, T. Photosynthesis-irradiance relationships in polar and temperate phytoplankton populations. **Polar Biology**, v. 5, n. 3, p. 153–164, 1986. doi: 10.1007/BF00441695.

HELSEL, D. R.; HIRSCH, R. M. Simple linear regression. In: \_\_\_\_\_. **Statistical methods in water resources**, studies in environmental science. p.221–263, 1992. v. 49. Elsevier. doi: 10.1016/S0166-1116(08)71108-7, 1992.

HESTIR, E. L.; BRANDO, V. E.; BRESCIANI, M.; et al. Measuring freshwater aquatic ecosystems: the need for a hyperspectral global mapping satellite mission. **Remote Sensing of Environment**, v. 167, p. 181–195, 2015. doi: 10.1016/j.rse.2015.05.023.

HIDALGO-GONZÁLEZ, R. M.; ALVAREZ-BORREGO, S. Chlorophyll profiles and the water column structure in the Gulf of California. **Oceanologica Acta**, v. 24, n. 1, p. 19–28, 2001. doi: 10.1016/S0399-1784(00)01126-9.

HOGUE, F. E.; LYON, P. E. Satellite retrieval of inherent optical properties by linear matrix inversion of oceanic radiance models: An analysis of model and radiance measurement errors. **Journal of Geophysical Research**, v. 101, n. C7, p. 16631- 16648, 1996. doi: 10.1029/96JC01414.

HOYOS, C. DE; NEGRO, A. I.; ALDASORO, J. J. Cyanobacteria distribution and abundance in the Spanish water reservoirs during thermal stratification. **Limnetica**, v. 23, n. 1-2, p. 119–132. Retrieved from <http://www.scopus.com/scopus/inward/record.url?eid=2-s2.0-21044431919&partnerID=40&rel=R8.2.0>, 2004.

HUANG, C.; ZOU, J.; LI, Y.; et al. Assessment of NIR-red algorithms for observation of chlorophyll-a in highly turbid inland waters in China. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 93, p. 29–39. International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). doi: 10.1016/j.isprsjprs.2014.03.012, 2014.

HUNTER, P. D.; TYLER, A. N.; CARVALHO, L.; CODD, G. A.; MABERLY, S. C. Hyperspectral remote sensing of cyanobacterial pigments as indicators for cell populations and toxins in eutrophic lakes. **Remote Sensing of Environment**, v. 114, n. 11, p. 2705–2718. Elsevier Inc. doi: 10.1016/j.rse.2010.06.006, 2010.

HUNTER, P. D.; TYLER, A. N.; GILVEAR, D. J.; WILLBY, N. J. Using Remote Sensing to Aid the Assessment of Human Health Risks from Blooms of Potentially Toxic Cyanobacteria. **Environmental Science & Technology**, v. 43, n. 7, p. 2627–2633, 2009.

HUNTER, P.; TYLER, A.; PRESING, M.; KOVACS, A.; PRESTON, T. Spectral discrimination of phytoplankton colour groups: the effect of suspended particulate matter and sensor spectral resolution. **Remote Sensing of Environment**, v. 112, n. 4, p. 1527–1544. doi: 10.1016/j.rse.2007.08.003, 2008.

HUOT, Y.; MOREL, A.; TWARDOWSKI, M. S.; STRAMSKI, D.; REYNOLDS, R. A. Particle optical backscattering along a chlorophyll gradient in the upper layer of the eastern South Pacific Ocean. **Biogeosciences Discussions**, v. 4, n. 6, p. 4571–4604. doi: 10.5194/bgd-4-4571-2007, 2007.

INGESTAD, T.; ÅGREN, G. I. Plant nutrition and growth: Basic principles. **Plant and Soil**, v. 168-169, n. 1, p. 15–20. doi: 10.1007/BF00029309, 1995.

JAIN, A. K.; DUBES, R. C. **Algorithms for clustering data**. Englewood Cliffs: Prentice Hall. v. 355, 1988. doi: 10.1126/science.311.5762.765.

JENSEN, J. R. **Remote sensing of the environment: an earth resource perspective**. Upper Saddle River, NJ: Prentice Hall, 2000. p. 333–378.

- JERLOV, N. G. **Optical oceanography**. Amsterdam: Elsevier, 1968. doi: 10.1016/0031-8663(69)90017-9 Elsevier oceanography series.
- KALLIS, G. The EU water framework directive: measures and implications. **Water Policy**, v. 3, n. 2, p. 125–142, 2001. doi: 10.1016/S1366-7017(01)00007-1.
- KASINAK, J.-M. E.; HOLT, B. M.; CHISLOCK, M. F.; WILSON, A. E. Benchtop fluorometry of phycocyanin as a rapid approach for estimating cyanobacterial biovolume. **Journal of Plankton Research**, v. 37, n. 1, p. 248–257. doi: 10.1093/plankt/fbu096, 2015.
- KHAMIS, H. Measures of Association: How to Choose? **Journal of Diagnostic Medical Sonography**, v. 24, n. 3, p. 155–162, 2008. doi: 10.1177/8756479308317006.
- KHORRAM, S.; CHESHIRE, H.; GERACI, A. L.; ROSA, G. LA. Water quality mapping of Augusta Bay, Italy from Landsat-TM data. **International Journal of Remote Sensing**, v. 12, n. 4, p. 803–808, 1991. doi: 10.1080/01431169108929696.
- KIRK, J. T. O. Monte Carlo modeling of the performance of a reflective tube absorption meter. **Applied Optics**, v. 31, n. 30, p. 6463-8, 1992. doi: 10.1364/AO.31.006463.
- KIRK, J. T. O. **Light and photosynthesis in aquatic ecosystems**. 3. ed. Cambridge: University Press, 2011.
- KIRK, R. E. **Statistics: An Introduction** (T. Wadsworth, Ed.). Belmont, 2008.
- KRUSE, F. A.; LEFKOFF, A. B.; BOARDMAN, J. W.; et al. The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. **Remote Sensing of Environment**, v. 44, n. 2-3, p. 145–163, 1993. doi: 10.1016/0034-4257(93)90013-N.
- KUTSER, T. Quantitative detection of chlorophyll in cyanobacterial blooms by satellite remote sensing. **Limnology and Oceanography**, v. 49, n. 6, p. 2179–2189, 2004.
- KUTSER, T.; PIERSON, D. C.; KALLIO, K. Y.; REINART, A.; SOBEK, S. Mapping lake CDOM by satellite remote sensing. **Remote Sensing of Environment**, v. 94, n. 4, p. 535–540, 2005. doi: 10.1016/j.rse.2004.11.009.
- LABATE, D.; CECCHERINI, M.; CISBANI, A.; et al. The PRISMA payload optomechanical design, a high performance instrument for a new hyperspectral

mission. **Acta Astronautica**, v. 65, n. 9-10, p. 1429–1436, 2009. Elsevier. doi: 10.1016/j.actaastro.2009.03.077.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). **LABISA-** Laboratório de Instrumentação de Sistemas Aquáticos: equipamentos. Aão José dos Campos, 2016. Retrieved January 10, 2016, from <http://www.dpi.inpe.br/labisa/equipamento.html>.

LAMPARELLI, M. C. **Graus de trofia em corpos d'água do Estado de São Paulo: avaliação de métodos e monitoramento**. 2004. 235p Tese (Doutorado em Ciências na Área de Ecossistemas Terrestres e Aquáticos) - Universidade de São Paulo, USP, São Paulo.

LAU, K. M.; WENG, H. Climate signal detection using wavelet transform: How to make a time series sing. **Bulletin of the American Meteorological Society**, v. 76, n. 12, p. 2391–2402, 1995. doi: 10.1175/1520-0477(1995)076<2391:csduwt>2.0.co;2.

LE, C.; LI, Y.; ZHA, Y.; et al. A four-band semi-analytical model for estimating chlorophyll a in highly turbid lakes: the case of Taihu Lake, China. **Remote Sensing of Environment**, v. 113, n. 6, p. 1175–1182, 2009. Elsevier B.V. doi: 10.1016/j.rse.2009.02.005.

LE, C.; LI, Y.; ZHA, Y.; et al. Remote sensing of phycocyanin pigment in highly turbid inland waters in Lake Taihu, China. **International Journal of Remote Sensing**, v. 32, n. 23, p. 8253–8269, 2011a. doi: 10.1080/01431161.2010.533210.

LEE, Z.; CARDER, K. L.; ARNONE, R. A. Deriving inherent optical properties from water color : a multiband quasi-analytical algorithm for optically deep waters. **Applied Optics**, v. 41, n. 27, p. 23–30, 2002.

LEE, Z.; CARDER, K. L.; MOBLEY, C. D.; STEWARD, R. G.; PATCH, J. S. Hyperspectral remote sensing for shallow waters: 2 Deriving bottom depths and water properties by optimization. **Applied Optics**, v. 38, n. 18, p. 3831- 3843, 1999. doi: 10.1364/AO.38.003831.

LEGENDRE, P.; LEGENDRE, L. Interpretation of ecological structures. **Developments in Environmental Modelling**, developments in environmental modelling. v. 24, p.521–624, 2012. Elsevier. doi: 10.1016/B978-0-444-53868-0.50010-1.

LELONG, C. C. .; PINET, P. C.; POILVÉ, H. Hyperspectral imaging and stress mapping in agriculture. **Remote Sensing of Environment**, v. 66, n. 2, p. 179–191. doi: 10.1016/S0034-4257(98)00049-2.

LEYMARIE, E.; DOXARAN, D.; BABIN, M. Uncertainties associated to measurements of inherent optical properties in natural waters. **Applied optics**, v. 49, n. 28, p. 5415–5436, 2010. doi: 10.1364/AO.49.005415.

LI, L.; LI, L.; SONG, K. Remote sensing of freshwater cyanobacteria: An extended IOP Inversion Model of Inland Waters (IIMIW) for partitioning absorption coefficient and estimating phycocyanin. **Remote Sensing of Environment**, v. 157, p. 9–23, 2015. Elsevier Inc. doi: 10.1016/j.rse.2014.06.009.

LIEBETRAU, A. M. **Measures of association**. 32. ed. Retrieved November 20, 2015, from [https://books.google.com.br/books/about/Measures\\_of\\_Association.html?id=7RNHAAAAMAAJ&pgis=1](https://books.google.com.br/books/about/Measures_of_Association.html?id=7RNHAAAAMAAJ&pgis=1), 1983.

LIN, J.; XIE, L.; PIETRAFESA, L. J.; et al. Dissolved oxygen stratification in two micro-tidal partially-mixed estuaries. **Estuarine, Coastal and Shelf Science**, v. 70, n. 3, p. 423–437, 2006. doi: 10.1016/j.ecss.2006.06.032.

LORENZEN, C. Determination of chlorophyll and pheopigments: spectrophotometric equations. **Limnology and Oceanography**, v. 12, n. 1961, p. 343–346, 1967. doi: 10.4319/lo.1967.12.2.0343.

LORENZI, A. S. **Abordagens moleculares para detectar cianobactérias e seus genótipos produtores de microcistinas presentes nas represas Billings e Guarapiranga, São Paulo, Brasil**. 2004. 92p. Dissertação (Mestrado em Ciência) - Universidade de São Paulo, São Paulo, 2004.

MACINTYRE, H. L.; LAWRENZ, E.; RICHARDSON, T. L. Taxonomic Discrimination of Phytoplankton by Spectral Fluorescence. In: Suggett, D. J.; Prášil, O.; Borowitzka, M. A. (Eds.). **Chlorophyll a fluorescence in aquatic sciences: methods and applications**. Dordrecht: Springer Netherlands, 2010. p.129–169. doi: 10.1007/978-90-481-9268-7.

MALENOVSKÝ, Z.; ROTT, H.; CIHLAR, J.; et al. Sentinels for science: Potential of Sentinel-1, -2, and -3 missions for scientific observations of ocean, cryosphere, and land. **Remote Sensing of Environment**, v. 120, p. 91–101, 2012. doi: 10.1016/j.rse.2011.09.026.

MALLAT, S. G. A theory for multiresolution signal decomposition: the wavelet representation. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 11, n. 7, p. 674–693, 1989. doi: 10.1109/34.192463.

MARITORENA, S.; SIEGEL, D. A.; PETERSON, A. R. Optimization of a semianalytical ocean color model for global-scale applications. **Applied Optics**, v. 41, n. 15, p. 2705, 2002. doi: 10.1364/AO.41.002705.

MARKOVSKY, I.; HUFFEL, S. VAN. Overview of total least-squares methods. **Signal Processing**, v. 87, n. 10, p. 2283–2302, 2007. doi: 10.1016/j.sigpro.2007.04.004.

MCKINLEY, V. L.; VESTAL, J. R. Effects of acid on plant litter decomposition in an Arctic Lake. **Applied and Environmental Microbiology**, v. 43, n. 5, p. 1188–1195, 1982. Retrieved from <http://aem.asm.org/content/43/5/1188.abstract>.

METSAMAA, L.; KUTSER, T.; STRÖMBECK, N. Recognising cyanobacterial blooms based on their optical signature : a modelling study. **Boreal Environment Research**, n. 11, Nov, p. 493–506, 2006.

MISHRA, D. R.; NARUMALANI, S.; RUNDQUIST, D.; LAWSON, M. Characterizing the vertical diffuse attenuation coefficient for downwelling irradiance in coastal waters: Implications for water penetration by high resolution satellite data. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 60, n. 1, p. 48–64, 2005. doi: 10.1016/j.isprsjprs.2005.09.003.

MISHRA, S.; MISHRA, D. R. A novel remote sensing algorithm to quantify phycocyanin in cyanobacterial algal blooms. **Environmental Research Letters**, v. 9, n. 11, p. 114003, 2014. doi: 10.1088/1748-9326/9/11/114003.

MISHRA, S.; MISHRA, D. R.; LEE, Z.; TUCKER, C. S. Quantifying cyanobacterial phycocyanin concentration in turbid productive waters: a quasi-analytical approach. **Remote Sensing of Environment**, v. 133, p. 141–151, 2013. Elsevier Inc. doi: 10.1016/j.rse.2013.02.004.

MISHRA, S.; MISHRA, D. R.; SCHLUCHTER, W. M. A novel algorithm for predicting phycocyanin concentrations in cyanobacteria: a proximal hyperspectral remote sensing approach. **Remote Sensing**, v. 1, n. 4, p. 758–775, 2009. doi: 10.3390/rs1040758.

MOBLEY, C. D. Estimation of the remote-sensing reflectance from above-surface measurements. **Applied optics**, v. 38, n. 36, p. 7442–7455, 1999. doi: 10.1364/AO.38.007442.

MOBLEY, C. D. Radiative transfer in the ocean. **Encyclopedia of Ocean Sciences**. p.2321–2330, 2001. Elsevier. doi: 10.1006/rwos.2001.0469,.

MOBLEY, C. D.; STRAMSKI, D.; BISSETT, W. P.; BOSS, E. Optical modeling of ocean waters: is the case 1-case 2 classification still useful? **Oceanography-Washington DC-Oceanography Society-**, v. 17, n. 2, p. 60–67. doi: 10.5670/oceanog.2004.48, 2004.

MOREL, A. Optical properties of pure water and pure sea water. In: JERLOV, N. G.; NIELSON, E. S. (Eds.); **Optical Aspects of Oceanography**. New York: Academic, 1974. p.1–24.

MOREL, A. Bio-optical models. In: Steele, J. H. (Ed.). **Encyclopedia of ocean sciences**. London: Academic Press Elsevier, 2001. p.317–326.

MOREL, A.; GORDON, H. R. Report of the working group on water color. **Boundary-Layer Meteorology**, v. 18, n. 4, p. 343–355, 1980.

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

MOSCHINI-CARLOS, V.; BORTOLI, S.; PINTO, E. et al. Cyanobacteria and cyanotoxin in the billings reservoir (São Paulo, SP, Brazil). **Limnetica**, v. 28, n. 2, p. 273–282, 2009.

MOSCHINI-CARLOS, V.; FREITAS, L. G. Limnological evaluation of water in the Rio Grande and Taquacetuba branches of the Billings Complex (São Paulo, Brazil) and management implications. **Ambiente e Agua - An Interdisciplinary Journal of Applied Science**, v. 5, n. 3, p. 47–59, 2010. doi: 10.4136/ambiente-agua.153.

MOUW, C. B.; GREB, S.; AURIN, D.; et al. Aquatic color radiometry remote sensing of coastal and inland waters: Challenges and recommendations for future satellite missions. **Remote Sensing of Environment**, v. 160, p. 15–30. Elsevier Inc. doi: 10.1016/j.rse.2015.02.001, 2015.

MUELLER, J. L. In-water radiometric profile measurements and data analysis protocols. In: fargion, g. s.; mueller, J. L. (Eds.). **Ocean optics protocols for satellite ocean color sensor validation**. Greenbelt, Maryland : Goddard Space Flight Center: NASA, 2000. p.87–97. Tech. Memo. 2000209966/Rev2.

MURCHIE, E. H.; HUBBART, S.; CHEN, Y.; PENG, S.; HORTON, P. Acclimation of rice photosynthesis to irradiance under field conditions. **Plant physiology**, v. 130, n. 4, p. 1999–2010. doi: 10.1104/pp.011098, 2002.

MURCHIE, E. H.; LAWSON, T. Chlorophyll fluorescence analysis: A guide to good practice and understanding some new applications. **Journal of Experimental Botany**. doi: 10.1093/jxb/ert208, 2013.

NEUMANN, A.; ZIMMERMANN, G.; KRAWCZYK, H.; WALZEL, T. Test of specific methods for reduction of spectral dimension of ocean remote sensing data. **Acta Astronautica**, v. 29, n. 1, p. 61–66, 1993. doi: 10.1016/0094-5765(93)90070-D.

NIEVERGELT, Y. Total Least Squares: State-of-the-Art Regression in Numerical Analysis. **SIAM Review**, v. 36, n. 2, p. 258–264, 1994. Society for Industrial and Applied Mathematics. doi: 10.1137/1036055.

NRC. **The drama of commons**. Washington, DC: National Academy Press, 2002.

NRC. **Earth science and applications from space: national imperatives for the next decade and beyond**. The National Academies Press. v. 8. doi: ISBN: 0-309-66714-3, 2007.

NUSH, E. A. Comparison of different methods for chlorophyll and phaeopigment determination. **Archiv für Hydrobiologie–Beiheft Ergebnisse der Limnologie**, v. 14, p. 14–36, 1980.

O'REILLY, J. E.; MARITORENA, S.; MITCHELL, B. G.; et al. Ocean color chlorophyll algorithms for SeaWiFS. **Journal of Geophysical Research**, v. 103, n. C11, p. 24937, 1998. doi: 10.1029/98JC02160.

O'REILLY, J.; MARITORENA, S.; MITCHELL, B. G.; STRUTTON, P. Ocean color chlorophyll a algorithms for SeaWiFS, OC2, and OC4: Version 4. **SeaWiFS postlaunch ...** 2000.p. 8–22. Retrieved from [http://www.spg.ucsd.edu/People/Mati/2000\\_OReilly\\_et\\_al\\_OC4v4\\_11-Chapt2.pdf](http://www.spg.ucsd.edu/People/Mati/2000_OReilly_et_al_OC4v4_11-Chapt2.pdf), 2000.

OGASHAWARA, I. **Re-parameterization of a quasi analytical algorithm and phycocyanin estimation in a tropical reservoir**. 2014. 147 p. (sid.inpe.br/mtc-m21b/2014/02.13.12.56-TDI). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2014. Disponível em: <<http://urlib.net/8JMKD3MGP5W34M/3FNS6NE>>. Acesso em: 04 mar. 2016.

OGASHAWARA, I. Terminology and classification of bio-optical algorithms. **Remote Sensing Letters**, v. 6, n. 8, p. 613–617, 2015. doi: 10.1080/2150704X.2015.1066523.

OGASHAWARA, I.; MISHRA, D.; MISHRA, S.; CURTARELLI, M.; STECH, J. A Performance Review of Reflectance Based Algorithms for Predicting Phycocyanin Concentrations in Inland Waters. **Remote Sensing**, v. 5, n. 10, p. 4774–4798, 2013. doi: 10.3390/rs5104774.

OGASHAWARA, I.; MORENO-MADRILAN, M. Improving Inland Water Quality Monitoring through Remote Sensing Techniques. **ISPRS International Journal of Geo-Information**, v. 3, n. 4, p. 1234–1255, 2014. doi: 10.3390/ijgi3041234.

OGASHAWARA, I.; ZAVATTINI, J.; TUNDISI, J. The climatic rhythm and blooms of cyanobacteria in a tropical reservoir in São Paulo, Brazil. **Brazilian journal of biology**, v. 74, n. 1, p. 72–8, 2014. doi: 10.1590/1519-6984.17412.

OHDE, T.; SIEGEL, H. Derivation of immersion factors for the hyperspectral TriOS radiance sensor. **Journal of Optics A: Pure and Applied Optics**, v. 5, n. 3, p. L12–L14, 2003. doi: 10.1088/1464-4258/5/3/103.

PAL, S. K.; MAJUMDAR, T. J.; BHATTACHARYA, A. K. ERS-2 SAR and IRS-1C LISS III data fusion: A PCA approach to improve remote sensing based geological interpretation. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 61, n. 5, p. 281–297, 2007. doi: 10.1016/j.isprsjprs.2006.10.001.

PALMER, S. C. J.; KUTSER, T.; HUNTER, P. D. Remote sensing of inland waters: Challenges, progress and future directions. **Remote Sensing of Environment**, v. 157, p. 1–8. doi: 10.1016/j.rse.2014.09.021, 2015.

PETZOLD, T. J. **Volume scattering functions for selected ocean waters.** - scripps institute of oceanography, San Diego: University of California, 1972. Ref. 72-28.

PINILLA, G. A. Vertical distribution of phytoplankton in a clear water lake of Colombian Amazon (Lake Boa). **Hydrobiologia**, v. 568, p. 79–90, 2006. doi: 10.1007/s10750-006-0038-y.

PLATT, T.; HOEPPFNER, N.; STUART, V.; BROWN, C. (eds.). **Why ocean colour?** The societal benefits of ocean-colour technology. , Dartmouth, Canada: IOCCG, n. 7, 2008. Reports and Monographs of the International Ocean-Colour Coordinating Group.

PREISENDORFER, R. W. **Hydrologic optics.** v. II. Honolulu, Hawaii: Department of Commerce, 1976.

RAMSTEDT, M.; CARLSSON, E.; LÖVGREN, L. Aqueous geochemistry in the Udden pit lake, northern Sweden. **Applied Geochemistry**, v. 18, n. 1, p. 97–108, 2003. doi: 10.1016/S0883-2927(02)00068-9.

RANDOLPH, K.; WILSON, J.; TEDESCO, L.; et al. Hyperspectral remote sensing of cyanobacteria in turbid productive water using optically active pigments , chlorophyll a and phycocyanin. **Remote Sensing of Environment**, v. 112, p. 4009–4019, 2006. doi: 10.1016/j.rse.2008.06.002.

REINART, A.; KUTSER, T. Comparison of different satellite sensors in detecting cyanobacterial bloom events in the Baltic Sea. **Remote Sensing of Environment**, v. 102, n. 1-2, p. 74–85, 2006. doi: 10.1016/j.rse.2006.02.013.

REYNOLDS, C. S. **Ecology of phytoplankton**. 1. ed. Cambridge: University Press, 2006.

RIDDICK, C. A. L.; HUNTER, P. D.; TYLER, A. N.; et al. Spatial variability of absorption coefficients over a biogeochemical gradient in a large and optically complex shallow lake. **Journal of Geophysical Research: Oceans**, v. 120, n. 10, p. 7040–7066, 2015. doi: 10.1002/2015JC011202.

ROBERTS, D. A.; QUATTROCHI, D. A.; HULLEY, G. C.; HOOK, S. J.; GREEN, R. O. Synergies between VSWIR and TIR data for the urban environment: An evaluation of the potential for the Hyperspectral Infrared Imager (HyspIRI) Decadal Survey mission. **Remote Sensing of Environment**, v. 117, n. 2012, p. 83–101, 2012. doi: 10.1016/j.rse.2011.07.021.

ROHLF, F.; FISHER, D. Tests for hierarchical structure in random data sets. **Systematic Zoology**, v. 17, p. 407–412, 1968. doi: 10.1093/sysbio/17.4.407.

RUIZ-VERDÚ, A.; SIMIS, S. G. H.; HOYOS, C. DE; GONS, H. J.; PEÑA-MARTÍNEZ, R. An evaluation of algorithms for the remote sensing of cyanobacterial biomass. **Remote Sensing of Environment**, v. 112, p. 3996–4008, 2008. doi: 10.1016/j.rse.2007.11.019.

SAITO, H.; GOOVAERTS, P. Geostatistical Interpolation of Positively Skewed and Censored Data in a Dioxin-Contaminated Site. **Environmental Science & Technology**, v. 34, n. 19, p. 4228–4235, 2000. doi: 10.1021/es991450y.

SANDER DE CARVALHO, L. A.; FARIA BARBOSA, C. C.; LEÃO DE MORAES NOVO, E. M.; MORAES RUDORFF, C. DE. Implications of scatter corrections for absorption measurements on optical closure of Amazon floodplain lakes using the Spectral Absorption and Attenuation Meter (AC-S-WETLabs). **Remote Sensing of Environment**, v. 157, p. 123–137, 2015. doi: 10.1016/j.rse.2014.06.018.

SCHALLES, J. F.; YACOBI, Y. Z. Remote detection and seasonal patterns of phycocyanin, carotenoid, and chlorophyll pigments in eutrophic waters. **Archives Hydrobiologica Special Issues on Advanced Limnology**, v. 55, p. 153–168, 2000.

SCHLUMPBERGER, S.; LU, N. B.; SUSS, M. E.; BAZANT, M. Z. Scalable and Continuous Water Deionization by Shock Electrodialysis. **Environmental Science & Technology Letters**, v. 2, n. 12, p. 367–372, 2015. doi: 10.1021/acs.estlett.5b00303.

SCHMIDT, K. S.; SKIDMORE, A. K. Smoothing vegetation spectra with wavelets. **International Journal of Remote Sensing**, v. 25, n. 6, p. 1167–1184, 2004. doi: 10.1080/0143116031000115085.

SHEN, Q.; LI, J.; ZHANG, F.; et al. Classification of Several Optically Complex Waters in China Using in Situ Remote Sensing Reflectance. **Remote Sensing**, v. 7, n. 11, p. 14731–14756, 2015. doi: 10.3390/rs71114731.

SHENG, Y. Wavelet transform. **The Transforms and Applications Handbook**. p.1–88, 2000. doi: 10.1201/9781420036756.ch10.

SIEGEL, S. Nonparametric Statistics. **The American Statistician**, v. 11, n. 3, p. 13, 1957. doi: 10.2307/2685679.

SIMIS, S. G. H.; PETERS, S. W. M.; GONS, H. J. Remote sensing of the cyanobacterial pigment phycocyanin in turbid inland water. **Limnology and Oceanography**, v. 50, n. 1, p. 237–245, 2005. doi: 10.4319/lo.2005.50.1.0237.

Simis, S. G. H.; Ruiz-verdú, A.; Domínguez-Gómez, J. A.; Peña-Martinez, R.; Peters, S. W. M.; Gons, H. J. Influence of phytoplankton pigment composition on remote sensing of cyanobacterial biomass. **Remote Sensing of Environment**, v. 106, p. 414–427, 2007a. doi: 10.1016/j.rse.2006.09.008.

SIVONEN, K.; JONES, G. Cyanobacterial toxins. **Toxic Cyanobacteria in Water: A guide to their public health consequences, monitoring and management**. p.55–124, 1999. doi: 10.1046/j.1365-2427.2003.01107.x.

SMITH, C.; BAKER, K. S. The bio-optical state of ocean waters and remote sensing. **Limnology and Oceanography**, v. 23, n. 2, p. 247-259, 1978. DOI: 10.4319/lo.1978.23.2.0247.

SMITH, R. C.; BAKER, K. S. Optical properties of the clearest natural waters (200–800 nm). **Applied Optics**, v. 20, n. 2, p. 177, 1981. doi: 10.1364/AO.20.000177.

SMYTH, T. J.; MOORE, G. F.; HIRATA, T.; AIKEN, J. Semianalytical model for the derivation of ocean color inherent optical properties: description, implementation, and performance assessment. **Applied optics**, v. 45, n. 31, p. 8116–8131, 2006. doi: 10.1364/AO.46.000429.

SONG, K.; LI, L.; LI, S.; et al. Hyperspectral retrieval of phycocyanin in potable water sources using genetic algorithm – partial least squares ( GA – PLS ) modeling. **International Journal of Applied Earth Observations and Geoinformation**, v. 18, p. 368–385, 2012. Elsevier B.V. doi: 10.1016/j.jag.2012.03.013.

SONG, K.; LI, L.; LI, Z.; et al. Remote detection of cyanobacteria through phycocyanin for water supply source using three-band model. **Ecological Informatics**, v. 15, p. 22–33, 2013b. Elsevier B.V. doi: 10.1016/j.ecoinf.2013.02.006.

SONG, K. et al. Remote estimation of phycocyanin (PC) for inland waters coupled with YSI PC fluorescence probe. **Environmental Science and Pollution Research**, v. 20, n. 8, p. 5330–5340, 2013a. doi: 10.1007/s11356-013-1527-y.

SONG, K. et al. Remote estimation of chlorophyll-a in turbid inland waters: Three-band model versus GA-PLS model. **Remote Sensing of Environment**, v. 136, p. 342–357, 2013c. Elsevier Inc. doi: 10.1016/j.rse.2013.05.017.

SPRENT, P.; SMEETON, N. C. **Applied nonparametric statistical methods**. 3. ed. 2015. Retrieved November 20, 2015, from [https://books.google.com.br/books/about/Applied\\_Nonparametric\\_Statistical\\_Method.html?id=M59rMICM4FIC&pgis=1](https://books.google.com.br/books/about/Applied_Nonparametric_Statistical_Method.html?id=M59rMICM4FIC&pgis=1), 2000.

STOŃ-EGIERT, J.; MAJCHROWSKI, R.; DARECKI, M.; KOSAKOWSKA, A.; OSTROWSKA, M. Influence of underwater light fields on pigment characteristics in the Baltic Sea – results of statistical analysis. **Oceanologia**, v. 54, n. 1, p. 7–27, 2012. Elsevier Masson SAS. doi: 10.5697/oc.54-1.007.

STRAMSKA, M.; STRAMSKI, D. Effects of a nonuniform vertical profile of chlorophyll concentration on remote-sensing reflectance of the ocean. **Applied Optics**, v. 44, n. 9, p. 1735, 2005. doi: 10.1364/AO.44.001735.

STUFFLER, T.; FÖRSTER, K.; HOFER, S.; et al. Hyperspectral imaging—An advanced instrument concept for the EnMAP mission (Environmental Mapping and Analysis Programme). **Acta Astronautica**, v. 65, n. 7-8, p. 1107–1112, 2009. Elsevier. doi: 10.1016/j.actaastro.2009.03.042.

SUN, D.; LI, Y.; LE, C.; et al. A semi-analytical approach for detecting suspended particulate composition in complex turbid inland waters (China). **Remote Sensing of Environment**, v. 134, p. 92–99, 2013b. Elsevier Inc. doi: 10.1016/j.rse.2013.02.024.

SUN, D.; LI, Y.; WANG, Q.; et al. Hyperspectral remote sensing of the pigment C-phycocyanin in turbid inland waters, based on optical classification. **IEEE Transactions on Geoscience and Remote Sensing**, v. 51, n. 7, p. 3871–3884, 2013a.

TASSAN, S.; FERRARI, G. M. A sensitivity analysis of the “Transmittance-Reflectance” method for measuring light absorption by aquatic particles. **J. Plankton Res.**, v. 24, n. 8, p. 757–774, 2002. doi: 10.1093/plankt/24.8.757.

TRIOS. **TriOS** - optical sensors - Ramses. Oldenburg, Germany, 2009. Retrieved from ([http://www.trios.de/index.php?option=com\\_content&view=category&id=47&layout=trios&Itemid=76#item196\\_top](http://www.trios.de/index.php?option=com_content&view=category&id=47&layout=trios&Itemid=76#item196_top)). Accessed on 10th January 2016.

TUNDISI, J. G.; MATSUMURA-TUNDISI, T.; ARANTES JÚNIOR, J. D.; et al. The response of Carlos Botelho (Lobo, Broa) Reservoir to the passage of cold fronts as reflected by physical, chemical, and biological variables. **Brazilian journal of Biology**, v. 64, n. 1, p. 177–86, 2004. doi: /S1519-69842004000100020.

TURNER DESIGNS. **Instrumentation**. Sunnyvale, CA: Turner Designs, 2008.

TURNER DESIGNS. **MODEL 10-AU-005-CE fluorometer user's manual**. Sunnyvale, CA: Turner Designs, 2009. 162p.

UNEP, U. N. E. P. **Planning and management of lakes and reservoirs: an integrated approach to eutrophication**. Osaka/Shiga, Japan: UNEP/International Environmental Technology Center, 2000. ISBN 92-807-1810-X.

VINCENT, R. K.; QIN, X.; MCKAY, R. M. L.; et al. Phycocyanin detection from LANDSAT TM data for mapping cyanobacterial blooms in Lake Erie. **Remote Sensing of Environment**, v. 89, n. 3, p. 381–392, 2004. doi: 10.1016/j.rse.2003.10.014.

WENG, H.; LAU, K.-M. Wavelets, Period Doubling, and Time–Frequency Localization with Application to Organization of Convection over the Tropical Western Pacific. **Journal of the Atmospheric Sciences**, v. 51, n. 17, p. 2523–2541, 1994. doi: 10.1175/1520-0469(1994)051<2523:WPDATL>2.0.CO;2.

WET LABS INC. **Spectral absorption and attenuation meter - user's guide**. Philomath, OR, 2013.

WETZEL, R. G. **Limnology**. Retrieved from <http://store.elsevier.com/Limnology/Robert-Wetzel/isbn-9780127447605/>, 2001.

WETZEL, R. G.; LIKENS, G. E. **Limnological analysis**. New York: Springer-Verlag, 2000. v. 3. doi: 10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C.

WHEELER, S. M.; MORRISSEY, L. A.; LEVINE, S. N.; LIVINGSTON, G. P.; VINCENT, W. F. Mapping cyanobacterial blooms in Lake Champlain's Missisquoi Bay using QuickBird and MERIS satellite data. **Journal of Great Lakes Research**, v. 38, p. 68–75, 2012. Elsevier B.V. doi: 10.1016/j.jglr.2011.06.009.

WHITMIRE, A. L.; PEGAU, W. S.; KARP-BOSS, L.; BOSS, E.; COWLES, T. J. Spectral backscattering properties of marine phytoplankton cultures. **Optics Express**, v. 18, n. 14, p. 15073–15093, 2010. doi: 10.1029/2003RG000148.D.WILCOX, R. **Introduction to robust estimation**

**and hypothesis testing technometrics**. Academic Press, 2005. v. 47. doi: 10.1198/tech.2005.s334.

WORLD HEALTH ORGANIZATION. **Toxic cyanobacteria in water**: a guide to their public health consequences, monitoring and management. 1999. WHO Report. Retrieved March. doi: 10.1046/j.1365-2427.2003.01107.x,.

WU, J.; NORVELL, W. A.; WELCH, R. M. Kriging on highly skewed data for DTPA-extractable soil Zn with auxiliary information for pH and organic carbon. **Geoderma**, v. 134, p. 187–199, 2006. doi: 10.1016/j.geoderma.2005.11.002.

WYNNE, T. T.; STUMPF, R. P.; TOMLINSON, M. C.; et al. Relating spectral shape to cyanobacterial blooms in the Laurentian Great Lakes. **International Journal of Remote Sensing**, v. 29, n. December 2014, p. 3665–3672, 2008. doi: 10.1080/01431160802007640.

WYNNE, T. T.; STUMPF, R. P.; TOMLINSON, M. C.; DYBLE, J. Characterizing a cyanobacterial bloom in western Lake Erie using satellite imagery and meteorological data. **Limnology and Oceanography**, v. 55, n. 5, p. 2025–2036, 2010. doi: 10.4319/lo.2010.55.5.2025.

YACOBI, Y. Z.; KOHLER, J.; LEUNERT, F.; GITELSON, A. Phycocyanin-specific absorption coefficient: Eliminating the effect of chlorophylls absorption. **Limnol. Oceanogr.: Methods**, v. 13, p. 157–168, 2015. doi: 10.1002/lom3.10015.

YSI. **6-Series: Multiparameter water quality sondes**. Ohio (USA): YSI Incorporated, 2011.

ZABALZA, J.; REN, J.; YANG, M.; et al. Novel Folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 93, p. 112–122, 2014. doi: 10.1016/j.isprsjprs.2014.04.006.

ZHANG, X.; LI, P. Lithological mapping from hyperspectral data by improved use of spectral angle mapper. **International Journal of Applied Earth Observation and Geoinformation**, v. 31, p. 95–109, 2014. doi: 10.1016/j.jag.2014.03.007.

ZHOU, L.; ROBERTS, D. A.; MA, W.; ZHANG, H.; TANG, L. Estimation of higher chlorophylla concentrations using field spectral measurement and HJ-1A hyperspectral satellite data in Dianshan Lake, China. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 88, p. 41–47, 2014. doi: 10.1016/j.isprsjprs.2013.11.016.

## APPENDIX A

Figure A.1– List of some bio-optical models found in literature covering from empirical to QAA algorithms

| Model           | Reference                   | Sensor                       | Algorithm   |
|-----------------|-----------------------------|------------------------------|---|
| Empirical       | Vincent (2004)              | TM and ETM+                  | $PC = 47.7 - 9.21(TM R3/TMR1) + 29.7(TM R4/TMR1) - 118(TM R4/TMR3) - 6.819.21(TM R5/TMR3) + 41.9(TM R7/TMR3) - 14.7(TM R7/TMR4)$        |
| Empirical       | Song, Kaishan et al. (2012) | AISA                         | GA-PLS of (713/647; 704/628; 694/638; FDR713; FDR713; $R_{rs}$ 628)   |
| Empirical       | Song, Kaishan et al. (2013) | ASD FieldSpec / Ocean Optics | $PC \propto [R_{rs}^{-1}(622) - R_{rs}^{-1}(691)] \cdot R_{rs}(740)$  |
| Empirical       | Sun et al. (2012)           | FieldSpec spectroradiometer  | Type 3 waters (765/268; 765/623; 804/628; 820/628; 765/609)   |
| Semi-Empirical  | Dekker (1993)               | CAESAR / CASI                | $PC \propto [R_{rs}(600) + R_{rs}(648)] - R_{rs}(624)$  |
| Semi-Empirical  | Schalles and Yacobi (2000)  | Ocean Optics                 | $PC \propto R_{rs}(650)/R_{rs}(625)$  |
| Semi-Empirical  | Simis et al. (2005)         | MERIS                        | $PC \propto \left\{ \left[ \frac{R_{rs}(709)}{R_{rs}(620)} \times (a_w(709) + b_b) - b_b - a_w(620) \right] - 0.24a_{ph}(665) \right\}$ |
| Semi-Empirical  | Wynne et al. (2008)         | MERIS                        | $SS = nL_w(681) - nL_w(665) - \{nL_w(709) - nL_w(665)\} \frac{(681 - 665)}{(709 - 665)}$  |
| Semi-Empirical  | Hunter et al. (2008)        | ASD FieldSpec                | $PC \propto [R_{rs}^{-1}(630) - R_{rs}^{-1}(660)] \cdot R_{rs}(750)$  |
| Semi-Empirical  | Mishra et al. (2009)        | Ocean Optics                 | $PC \propto R_{rs}(700)/R_{rs}(600)$  |
| Semi-Empirical  | Hunter et al. (2010)        | AISA / CASI-2                | $PC \propto [R_{rs}^{-1}(615) - R_{rs}^{-1}(600)] \cdot R_{rs}(725)$  |
| Semi-Empirical  | Le et al. (2011)            | ASD FieldSpec                | $PC \propto [R_{rs}^{-1}(630) - R_{rs}^{-1}(645)][R_{rs}^{-1}(730) - R_{rs}^{-1}(694)]^{-1}$  |
| Semi-Empirical  | Dash et al. (2011)          | Ocean Color Monitor          | $PC \propto OCMR_{rs}(556.4) - OCMR_{rs}(669)/[556.4 - 669]$  |
| Semi-Empirical  | Dominguez et al. (2011)     | MERIS/CHRIS                  | $PC \propto (MB9 - MB7)/(MB9 + MB7)$  |
| Semi-Empirical  | Wheeler et al. (2012)       | QuickBird                    | $PC \propto QBNIR/QBRed$  |
| Semi-Analytical | Li, Linhai et al. (2012)    | OceanOpticsUSB4000           | $a_{pc}(624) = a_{ph}(624) - 0.5 \times [a_{ph}(600) + a_{ph}(648)]$  |
| QAA             | Mishra et al. (2013)        | Ocean Optics                 | $a_{pc}(620) = \frac{\varphi_1 a_{ph}(620) - a_{ph}(665)}{\varphi_1 - \varphi_2}$   |

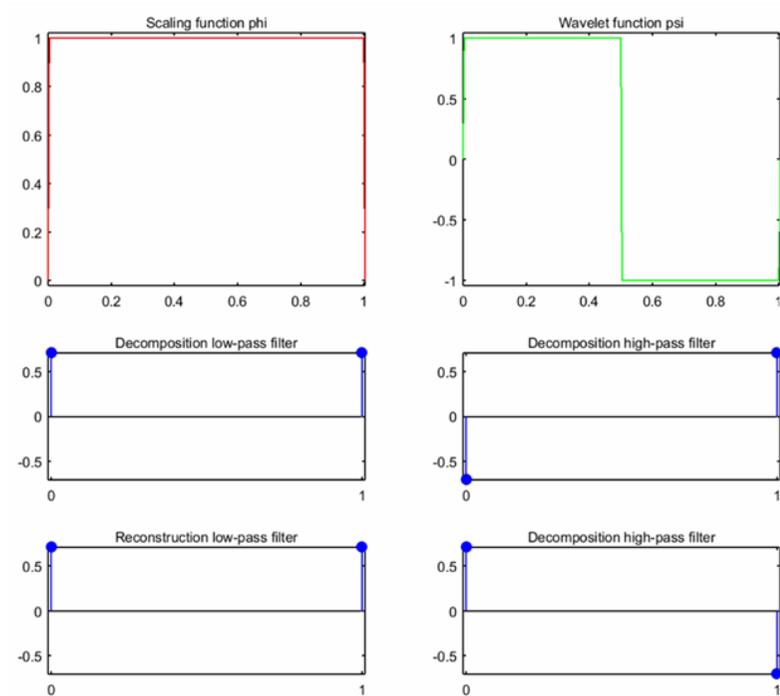
Where: TRM is Landsat TM Radiance; FDR is the first derivative;  $nL_w$  is the normalized water leaving radiance;  $OCMR_{rs}$  is the  $R_{rs}$  from OCM sensor; MB is MERIS band; QBNIR is the Quickbird NIR channel and QBRed is the Quickbird red channel;  $\varphi$  is the ratio between  $a_{ch}(665)$  and  $a_{ch}(620)$  and  $\varphi_2$  is the ratio between  $a_{pc}(665)$  and  $a_{pc}(620)$ .



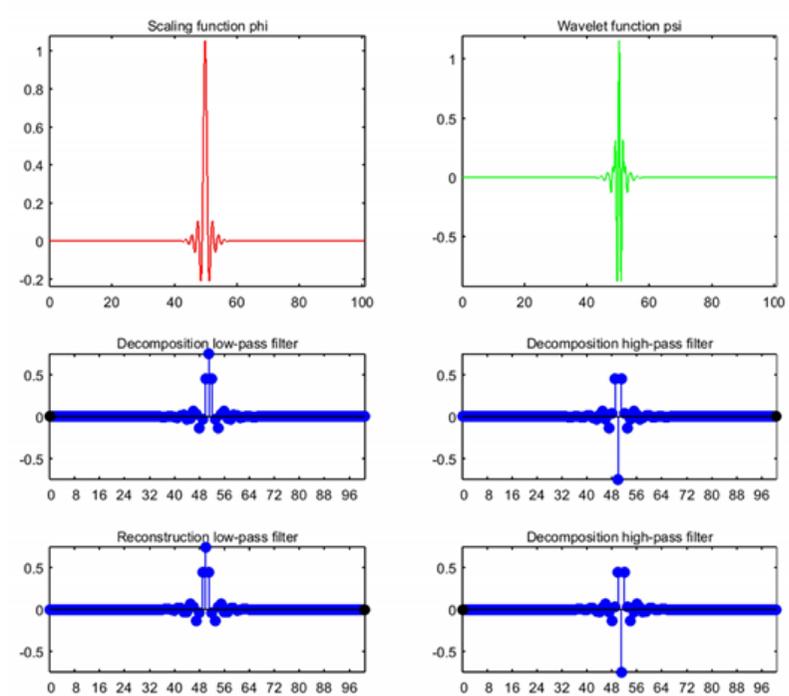
## APPENDIX B

Figure B.1– Haar and Discrete Meyer wavelets scaling and wavelet functions

Haar Wavelet



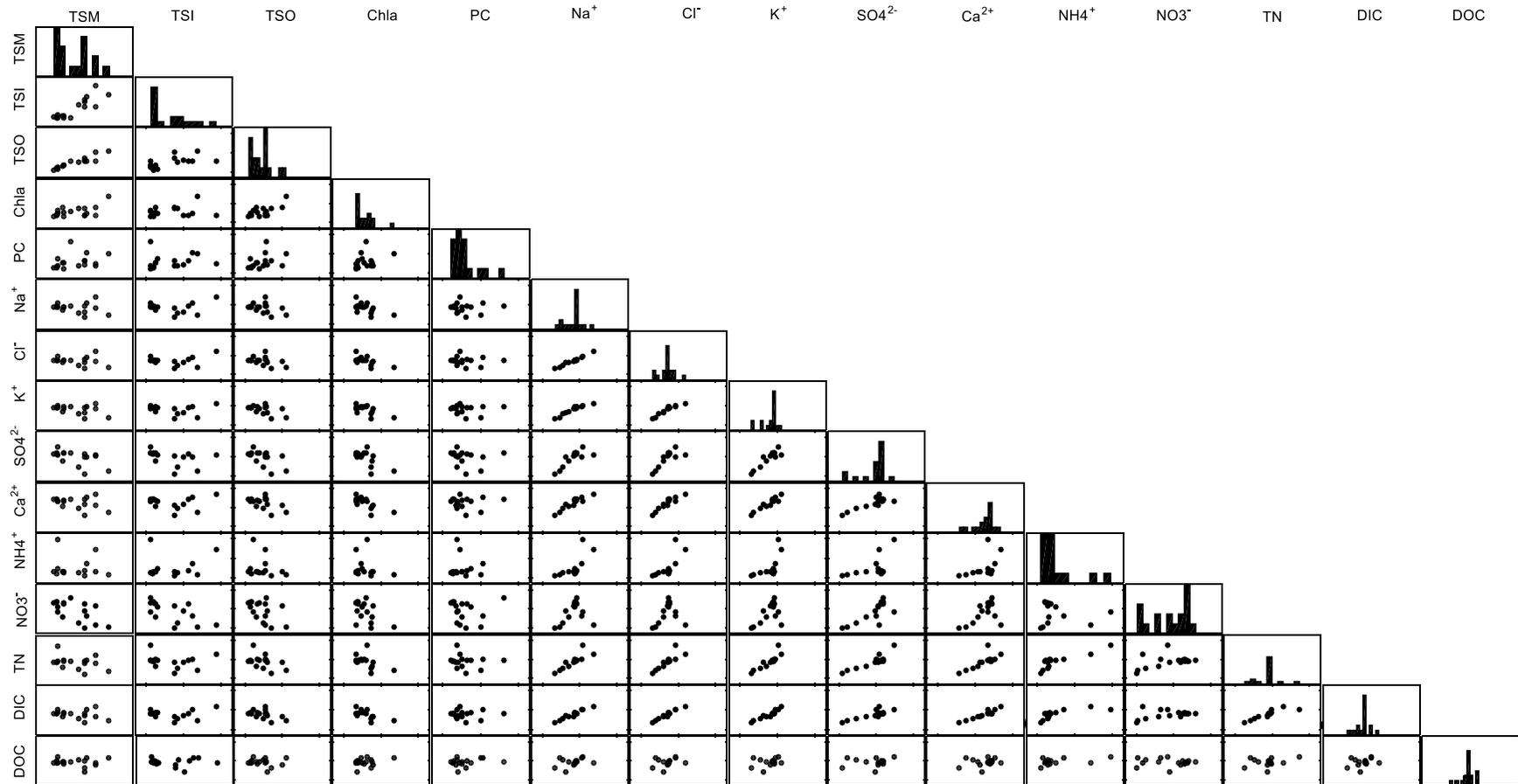
Discrete Meyer Wavelet





# APPENDIX C

Figure C.1– Water Quality Parameters Cross Correlation Scatterplots and histogram





**APPENDIX D**

Table D.1 - Kendall correlation coefficients for the limnological dataset

|       | TSM  | TSI  | TSO  | Chla | PC   | Na+  | Cl-  | K+   | SO42- | Ca2+ | NH4+ | NO3-  | TN   | IC   | NPOC |
|-------|------|------|------|------|------|------|------|------|-------|------|------|-------|------|------|------|
| TSM   | 1    |      |      |      |      |      |      |      |       |      |      |       |      |      |      |
| TSI   | 0.61 | 1    |      |      |      |      |      |      |       |      |      |       |      |      |      |
| TSO   | 0.80 | 0.37 | 1    |      |      |      |      |      |       |      |      |       |      |      |      |
| Chla  | 0.28 | 0.12 | 0.32 | 1    |      |      |      |      |       |      |      |       |      |      |      |
| PC    | 0.35 | 0.38 | 0.32 | 0.30 | 1    |      |      |      |       |      |      |       |      |      |      |
| Na+   | 0.07 | 0.20 | 0.19 | 0.32 | 0.03 | 1    |      |      |       |      |      |       |      |      |      |
| Cl-   | 0.05 | 0.15 | 0.20 | 0.31 | 0.01 | 0.96 | 1    |      |       |      |      |       |      |      |      |
| K+    | 0.19 | 0.23 | 0.28 | 0.21 | 0.01 | 0.76 | 0.78 | 1    |       |      |      |       |      |      |      |
| SO42- | 0.34 | 0.50 | 0.33 | 0.26 | 0.22 | 0.59 | 0.63 | 0.59 | 1     |      |      |       |      |      |      |
| Ca2+  | 0.10 | 0.06 | 0.22 | 0.41 | 0.01 | 0.76 | 0.75 | 0.71 | 0.47  | 1    |      |       |      |      |      |
| NH4+  | NaN   | NaN  | 1    |       |      |      |      |
| NO3-  | 0.43 | 0.53 | 0.32 | 0.32 | 0.28 | 0.32 | 0.28 | 0.32 | 0.65  | 0.41 | NaN  | 1     |      |      |      |
| TN    | 0.11 | 0.21 | 0.26 | 0.25 | 0.08 | 0.74 | 0.76 | 0.76 | 0.58  | 0.58 | NaN  | 0.25  | 1    |      |      |
| IC    | 0.11 | 0.06 | 0.29 | 0.25 | 0.07 | 0.78 | 0.79 | 0.81 | 0.49  | 0.72 | NaN  | 0.19  | 0.80 | 1    |      |
| NPOC  | 0.03 | 0.05 | 0.04 | 0.05 | 0.14 | 0.38 | 0.36 | 0.14 | 0.15  | 0.17 | NaN  | -0.02 | 0.17 | 0.18 | 1    |

Notes: NaN represent not-available-number values

Table D.2 - Kendall p-value for the limnological dataset

|       | TSM   | TSI   | TSO   | Chla  | PC    | Na+   | Cl-   | K+    | SO42- | Ca2+ | NH4+ | NO3- | TN | IC | NPOC |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|------|------|----|----|------|
| TSM   | 0     |       |       |       |       |       |       |       |       |      |      |      |    |    |      |
| TSI   | 0.001 | 0     |       |       |       |       |       |       |       |      |      |      |    |    |      |
| TSO   | 0.000 | 0.046 | 0     |       |       |       |       |       |       |      |      |      |    |    |      |
| Chla  | 0.136 | 0.534 | 0.083 | 0     |       |       |       |       |       |      |      |      |    |    |      |
| PC    | 0.056 | 0.042 | 0.082 | 0.108 | 0     |       |       |       |       |      |      |      |    |    |      |
| Na+   | 0.741 | 0.300 | 0.321 | 0.076 | 0.901 | 0     |       |       |       |      |      |      |    |    |      |
| Cl-   | 0.804 | 0.431 | 0.283 | 0.091 | 0.967 | 0.000 | 0     |       |       |      |      |      |    |    |      |
| K+    | 0.321 | 0.230 | 0.137 | 0.271 | 0.967 | 0.000 | 0.000 | 0     |       |      |      |      |    |    |      |
| SO42- | 0.069 | 0.007 | 0.069 | 0.151 | 0.231 | 0.001 | 0.000 | 0.001 | 0     |      |      |      |    |    |      |

Table D.2 - Kendall p-value for the limnological dataset - Continuation

|      |       |       |       |       |       |       |       |       |       |       |     |       |       |       |   |
|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-----|-------|-------|-------|---|
| Ca2+ | 0.619 | 0.772 | 0.247 | 0.022 | 0.967 | 0.000 | 0.000 | 0.000 | 0.008 | 0     |     |       |       |       |   |
| NH4+ | NaN   | 0   |       |       |       |   |
| NO3- | 0.020 | 0.004 | 0.083 | 0.076 | 0.127 | 0.076 | 0.129 | 0.076 | 0.000 | 0.022 | NaN | 0     |       |       |   |
| TN   | 0.562 | 0.263 | 0.159 | 0.173 | 0.679 | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 | NaN | 0.173 | 0     |       |   |
| IC   | 0.562 | 0.772 | 0.116 | 0.177 | 0.710 | 0.000 | 0.000 | 0.000 | 0.006 | 0.000 | NaN | 0.308 | 0.000 | 0     |   |
| NPOC | 0.901 | 0.804 | 0.836 | 0.805 | 0.457 | 0.039 | 0.048 | 0.458 | 0.410 | 0.364 | NaN | 0.934 | 0.364 | 0.322 | 0 |

Notes: NaN represent not-available-number values

Table D.3 – Spearman correlation coefficient for the limnological dataset

|       | TSM   | TSI   | TSO   | Chla  | PC    | Na+  | Cl-  | K+   | SO42- | Ca2+ | NH4+ | NO3-  | TN   | IC   | NPOC |
|-------|-------|-------|-------|-------|-------|------|------|------|-------|------|------|-------|------|------|------|
| TSM   | 1     |       |       |       |       |      |      |      |       |      |      |       |      |      |      |
| TSI   | 0.79  | 1     |       |       |       |      |      |      |       |      |      |       |      |      |      |
| TSO   | 0.91  | 0.55  | 1     |       |       |      |      |      |       |      |      |       |      |      |      |
| Chla  | 0.36  | 0.16  | 0.43  | 1     |       |      |      |      |       |      |      |       |      |      |      |
| PC    | 0.49  | 0.44  | 0.44  | 0.43  | 1     |      |      |      |       |      |      |       |      |      |      |
| Na+   | -0.13 | -0.17 | -0.29 | -0.53 | 0.01  | 1    |      |      |       |      |      |       |      |      |      |
| Cl-   | -0.13 | -0.14 | -0.30 | -0.51 | 0.02  | 0.99 | 1    |      |       |      |      |       |      |      |      |
| K+    | -0.25 | -0.24 | -0.41 | -0.41 | -0.01 | 0.87 | 0.89 | 1    |       |      |      |       |      |      |      |
| SO42- | -0.54 | -0.68 | -0.49 | -0.46 | -0.31 | 0.72 | 0.73 | 0.73 | 1     |      |      |       |      |      |      |
| Ca2+  | -0.13 | -0.03 | -0.29 | -0.61 | 0.07  | 0.88 | 0.88 | 0.84 | 0.62  | 1    |      |       |      |      |      |
| NH4+  | NaN   | NaN   | NaN   | NaN   | NaN   | NaN  | NaN  | NaN  | NaN   | NaN  | 1    |       |      |      |      |
| NO3-  | -0.58 | -0.69 | -0.44 | -0.48 | -0.36 | 0.36 | 0.35 | 0.37 | 0.77  | 0.47 | NaN  | 1     |      |      |      |
| TN    | -0.17 | -0.25 | -0.35 | -0.46 | -0.01 | 0.87 | 0.88 | 0.90 | 0.71  | 0.72 | NaN  | 0.32  | 1    |      |      |
| IC    | -0.18 | -0.08 | -0.40 | -0.45 | 0.14  | 0.89 | 0.92 | 0.92 | 0.65  | 0.83 | NaN  | 0.26  | 0.92 | 1    |      |
| NPOC  | 0.02  | 0.00  | -0.09 | 0.08  | 0.18  | 0.43 | 0.41 | 0.22 | 0.19  | 0.22 | NaN  | -0.09 | 0.24 | 0.26 | 1    |

Notes: NaN represent not-available-number values

Table D.4 - Spearman p-value for the limnological dataset

|       | TSM   | TSI   | TSO   | Chla  | PC    | Na+   | Cl-   | K+    | SO42- | Ca2+  | NH4+ | NO3-  | TN    | IC    | NPOC |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|-------|-------|-------|------|
| TSM   | 1     |       |       |       |       |       |       |       |       |       |      |       |       |       |      |
| TSI   | 0.000 | 1     |       |       |       |       |       |       |       |       |      |       |       |       |      |
| TSO   | 0.000 | 0.022 | 1     |       |       |       |       |       |       |       |      |       |       |       |      |
| Chla  | 0.160 | 0.540 | 0.087 | 1     |       |       |       |       |       |       |      |       |       |       |      |
| PC    | 0.047 | 0.080 | 0.079 | 0.088 | 1     |       |       |       |       |       |      |       |       |       |      |
| Na+   | 0.608 | 0.524 | 0.259 | 0.028 | 0.963 | 1     |       |       |       |       |      |       |       |       |      |
| Cl-   | 0.621 | 0.594 | 0.242 | 0.038 | 0.937 | 0.000 | 1     |       |       |       |      |       |       |       |      |
| K+    | 0.326 | 0.348 | 0.100 | 0.103 | 0.966 | 0.000 | 0.000 | 1     |       |       |      |       |       |       |      |
| SO42- | 0.024 | 0.003 | 0.044 | 0.063 | 0.223 | 0.001 | 0.001 | 0.001 | 1     |       |      |       |       |       |      |
| Ca2+  | 0.608 | 0.910 | 0.263 | 0.009 | 0.775 | 0.000 | 0.000 | 0.000 | 0.008 | 1     |      |       |       |       |      |
| NH4+  | NaN   | 1    |       |       |       |      |
| NO3-  | 0.014 | 0.002 | 0.079 | 0.052 | 0.159 | 0.158 | 0.171 | 0.141 | 0.000 | 0.057 | NaN  | 1     |       |       |      |
| TN    | 0.504 | 0.339 | 0.168 | 0.066 | 0.978 | 0.000 | 0.000 | 0.000 | 0.001 | 0.001 | NaN  | 0.205 | 1     |       |      |
| IC    | 0.484 | 0.760 | 0.111 | 0.068 | 0.579 | 0.000 | 0.000 | 0.000 | 0.005 | 0.000 | NaN  | 0.319 | 0.000 | 1     |      |
| NPOC  | 0.950 | 0.994 | 0.743 | 0.754 | 0.496 | 0.087 | 0.105 | 0.387 | 0.459 | 0.400 | NaN  | 0.743 | 0.361 | 0.323 | 1    |

Notes: NaN represent not-available-number values

Table D.5 – Pearson correlation coefficient for the limnological dataset

|       | TSM   | TSI   | TSO   | Chla  | PC    | Na+  | Cl-  | K+   | SO42- | Ca2+ | NH4+ | NO3- | TN | IC | NPOC |
|-------|-------|-------|-------|-------|-------|------|------|------|-------|------|------|------|----|----|------|
| TSM   | 1     |       |       |       |       |      |      |      |       |      |      |      |    |    |      |
| TSI   | 0.87  | 1     |       |       |       |      |      |      |       |      |      |      |    |    |      |
| TSO   | 0.93  | 0.63  | 1     |       |       |      |      |      |       |      |      |      |    |    |      |
| Chla  | 0.51  | 0.20  | 0.66  | 1     |       |      |      |      |       |      |      |      |    |    |      |
| PC    | 0.30  | 0.16  | 0.36  | 0.35  | 1     |      |      |      |       |      |      |      |    |    |      |
| Na+   | -0.20 | 0.07  | -0.37 | -0.57 | -0.05 | 1    |      |      |       |      |      |      |    |    |      |
| Cl-   | -0.17 | 0.11  | -0.36 | -0.60 | -0.07 | 0.99 | 1    |      |       |      |      |      |    |    |      |
| K+    | -0.42 | -0.19 | -0.53 | -0.64 | -0.09 | 0.95 | 0.94 | 1    |       |      |      |      |    |    |      |
| SO42- | -0.54 | -0.39 | -0.56 | -0.66 | -0.17 | 0.80 | 0.79 | 0.92 | 1     |      |      |      |    |    |      |

Table D.5 - Pearson correlation coefficient for the limnological dataset – Continuation

|      |       |       |       |       |       |      |      |      |      |      |     |       |      |      |   |
|------|-------|-------|-------|-------|-------|------|------|------|------|------|-----|-------|------|------|---|
| Ca2+ | -0.36 | -0.08 | -0.51 | -0.73 | -0.02 | 0.91 | 0.91 | 0.95 | 0.88 | 1    |     |       |      |      |   |
| NH4+ | NaN   | NaN   | NaN   | NaN   | NaN   | NaN  | NaN  | NaN  | NaN  | NaN  | 1   |       |      |      |   |
| NO3- | -0.61 | -0.66 | -0.48 | -0.55 | -0.05 | 0.35 | 0.31 | 0.58 | 0.77 | 0.59 | NaN | 1     |      |      |   |
| TN   | -0.34 | -0.15 | -0.43 | -0.49 | -0.15 | 0.86 | 0.86 | 0.87 | 0.88 | 0.77 | NaN | 0.40  | 1    |      |   |
| IC   | -0.28 | 0.04  | -0.48 | -0.63 | -0.04 | 0.97 | 0.98 | 0.94 | 0.82 | 0.94 | NaN | 0.35  | 0.87 | 1    |   |
| NPOC | -0.08 | -0.01 | -0.12 | 0.23  | 0.24  | 0.38 | 0.31 | 0.33 | 0.19 | 0.26 | NaN | -0.04 | 0.32 | 0.35 | 1 |

Notes: NaN represent not-available-number values

Table D.6 - Pearson p-value for the limnological dataset

|       | TSM   | TSI   | TSO   | Chla  | PC    | Na+   | Cl-   | K+    | SO42- | Ca2+  | NH4+ | NO3-  | TN    | IC    | NPOC |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|------|-------|-------|-------|------|
| TSM   | 1     |       |       |       |       |       |       |       |       |       |      |       |       |       |      |
| TSI   | 0.000 | 1     |       |       |       |       |       |       |       |       |      |       |       |       |      |
| TSO   | 0.000 | 0.006 | 1     |       |       |       |       |       |       |       |      |       |       |       |      |
| Chla  | 0.037 | 0.446 | 0.004 | 1     |       |       |       |       |       |       |      |       |       |       |      |
| PC    | 0.236 | 0.542 | 0.151 | 0.163 | 1     |       |       |       |       |       |      |       |       |       |      |
| Na+   | 0.438 | 0.798 | 0.142 | 0.017 | 0.856 | 1     |       |       |       |       |      |       |       |       |      |
| Cl-   | 0.503 | 0.664 | 0.151 | 0.012 | 0.792 | 0.000 | 1     |       |       |       |      |       |       |       |      |
| K+    | 0.092 | 0.472 | 0.029 | 0.005 | 0.744 | 0.000 | 0.000 | 1     |       |       |      |       |       |       |      |
| SO42- | 0.027 | 0.123 | 0.020 | 0.004 | 0.509 | 0.000 | 0.000 | 0.000 | 1     |       |      |       |       |       |      |
| Ca2+  | 0.157 | 0.765 | 0.036 | 0.001 | 0.933 | 0.000 | 0.000 | 0.000 | 0.000 | 1     |      |       |       |       |      |
| NH4+  | NaN   | 1    |       |       |       |      |
| NO3-  | 0.009 | 0.004 | 0.054 | 0.023 | 0.859 | 0.171 | 0.230 | 0.015 | 0.000 | 0.012 | NaN  | 1     |       |       |      |
| TN    | 0.183 | 0.572 | 0.087 | 0.048 | 0.569 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | NaN  | 0.116 | 1     |       |      |
| IC    | 0.279 | 0.870 | 0.053 | 0.006 | 0.888 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | NaN  | 0.167 | 0.000 | 1     |      |
| NPOC  | 0.763 | 0.983 | 0.642 | 0.379 | 0.353 | 0.137 | 0.230 | 0.200 | 0.472 | 0.319 | NaN  | 0.871 | 0.206 | 0.163 | 1    |

Notes: NaN represent not-available-number values

## APPENDIX E

Figure E.1 – TLS scatterplots for laboratory-calculated, ACS-measured, and QAA-estimated absorption coefficients

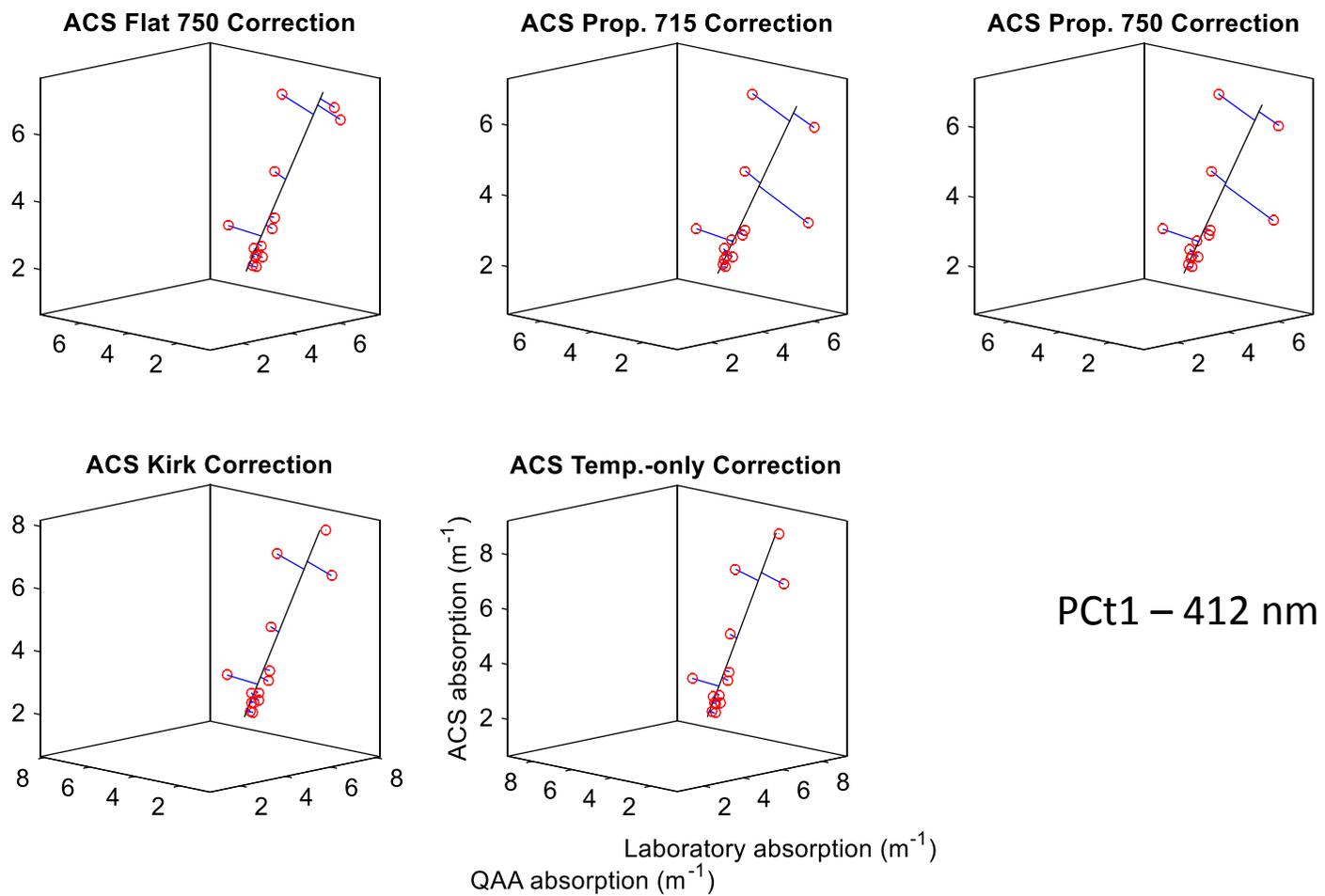


Figure E.2 – TLS scatterplots for laboratory-calculated, ACS-measured, and QAA-estimated absorption coefficients

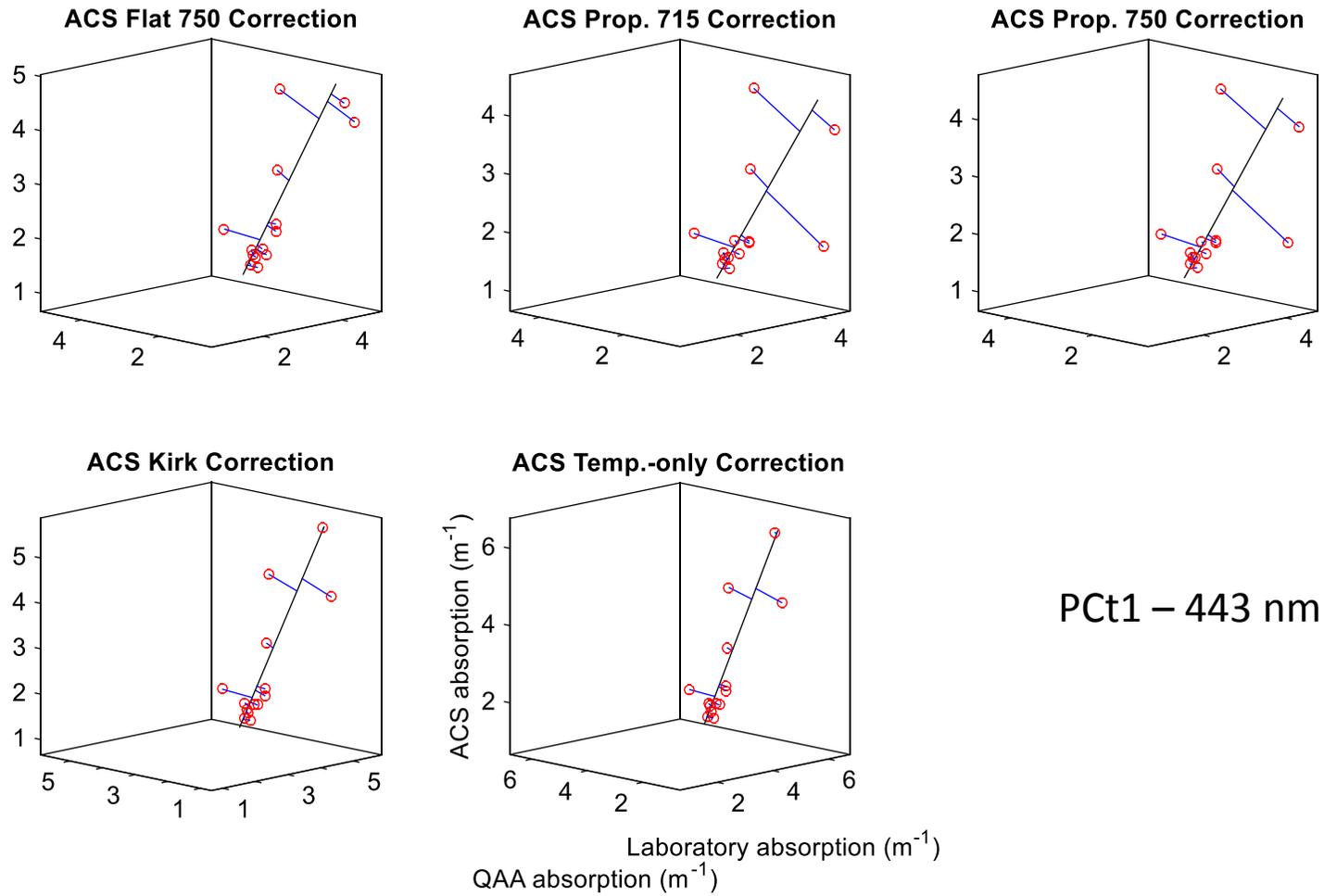


Figure E.3 – TLS scatterplots for laboratory-calculated, ACS-measured, and QAA-estimated absorption coefficients

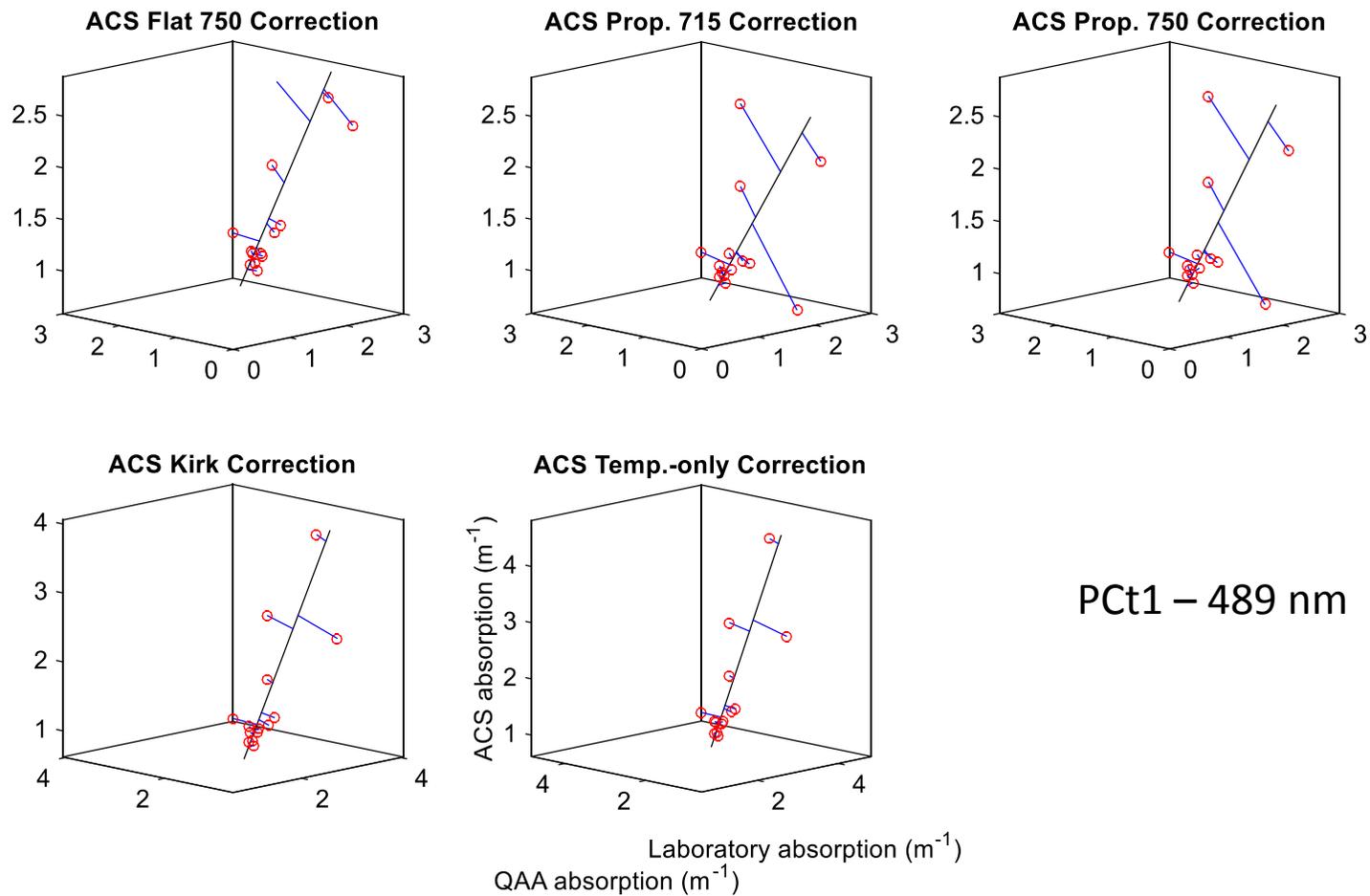


Figure E.4 – TLS scatterplots for laboratory-calculated, ACS-measured, and QAA-estimated absorption coefficients

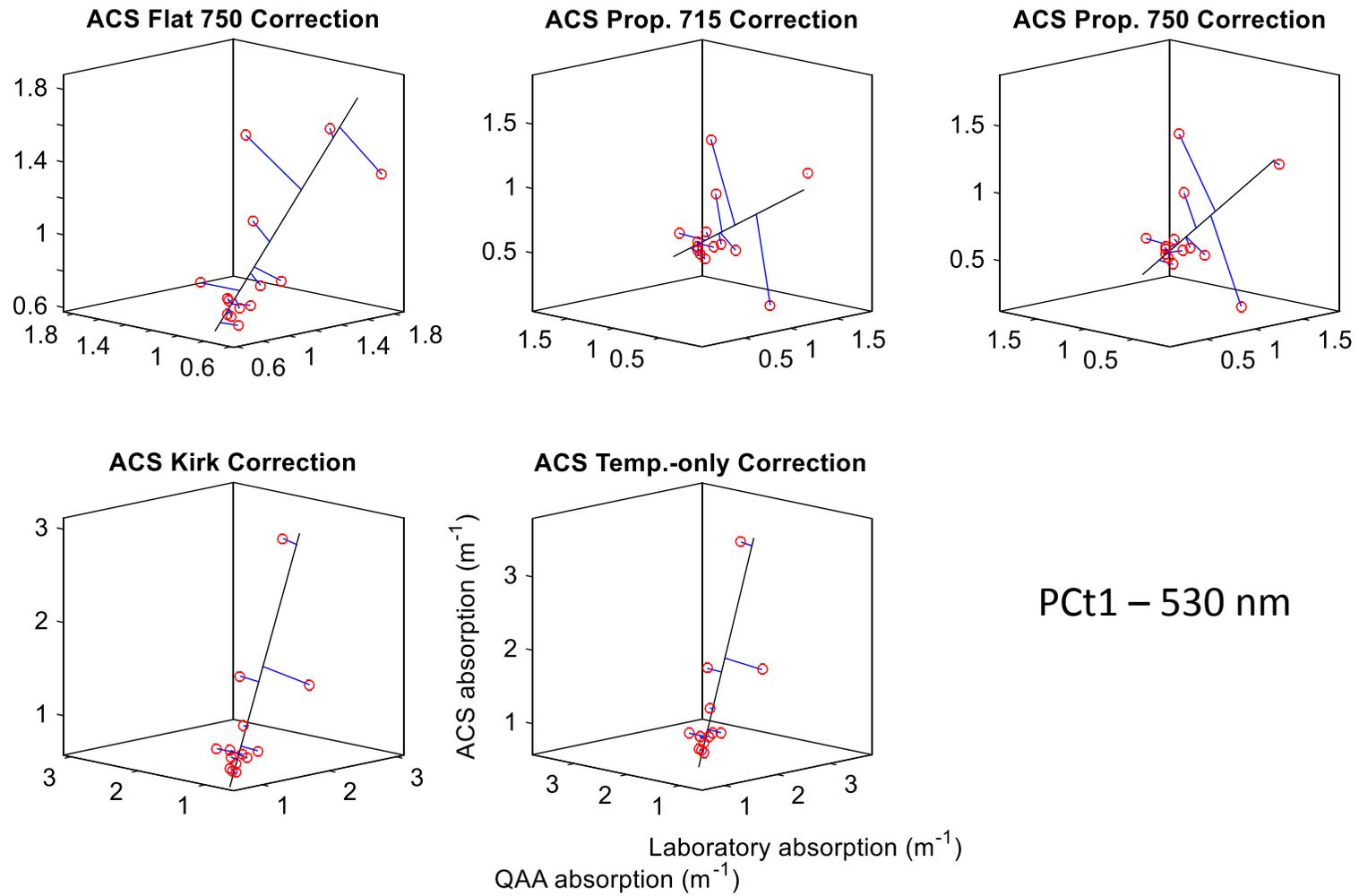
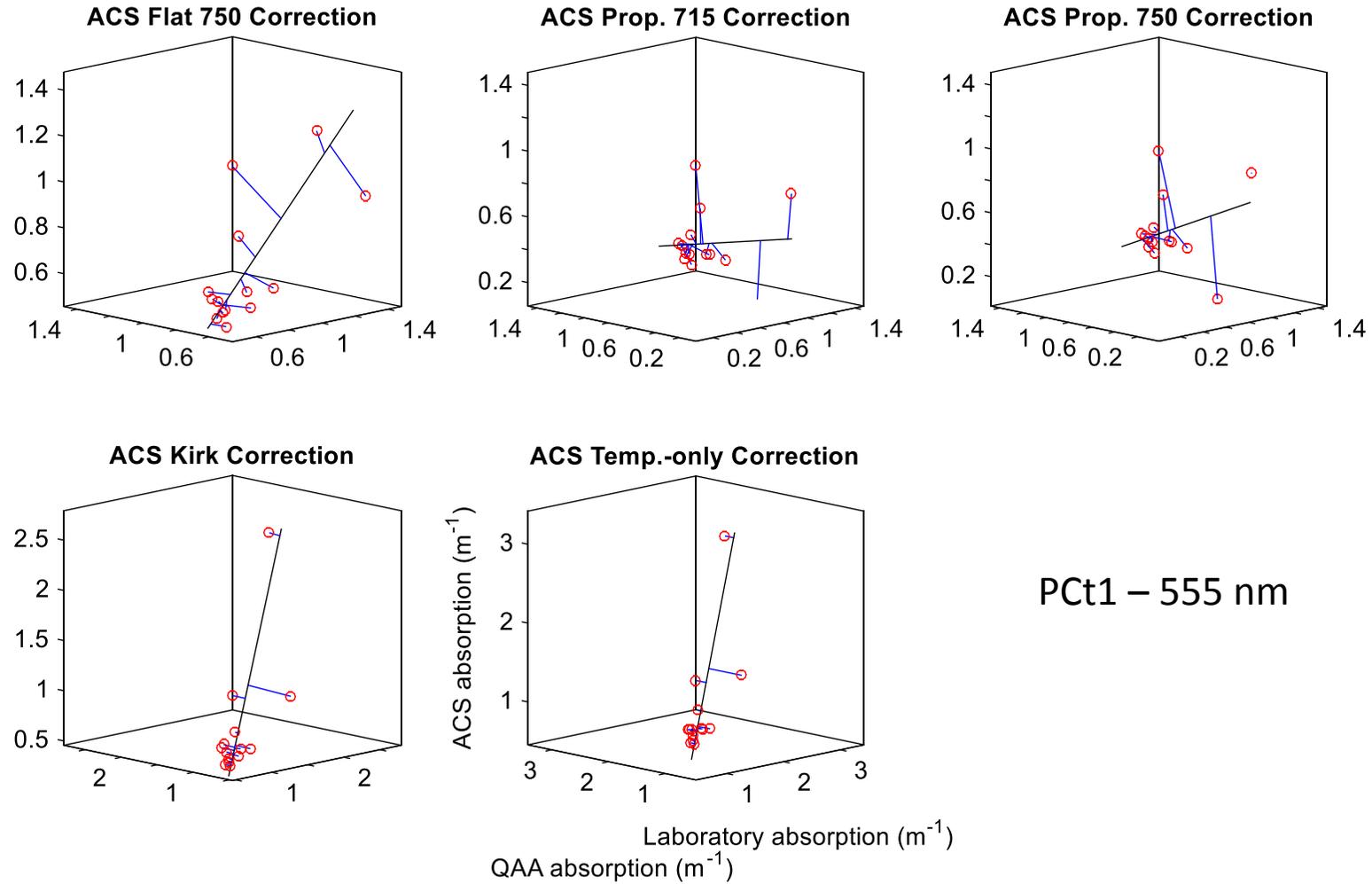


Figure E.5 – TLS scatterplots for laboratory-calculated, ACS-measured, and QAA-estimated absorption coefficients





## APPENDIX F

Figure F.1 – DWTCA-Clustered  $R_{rs}$  Spectra for signal decomposition level (Haar wavelet).

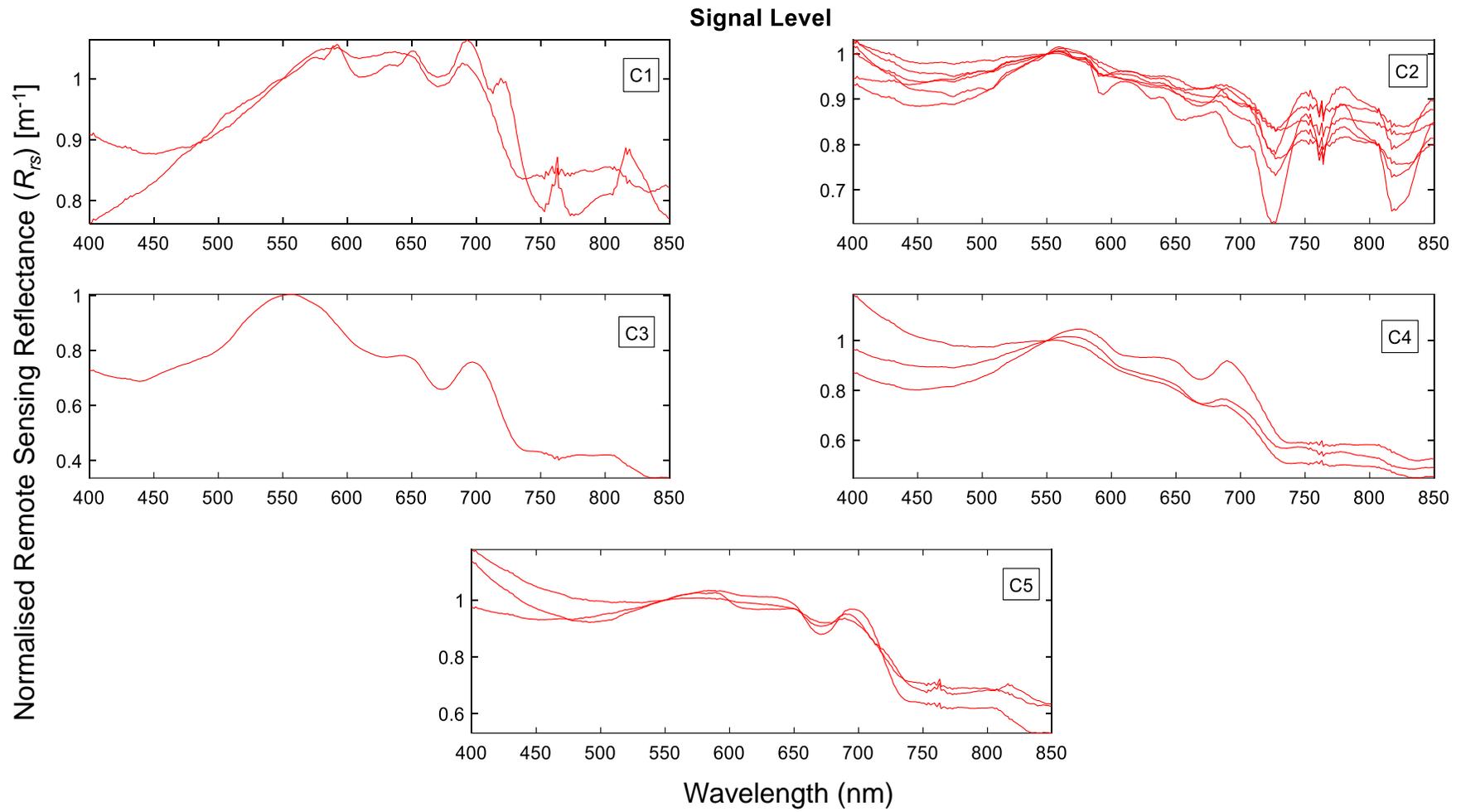


Figure F.2 – DWTC-Clustered  $R_{rs}$  Spectra for decomposition level one (Haar wavelet).

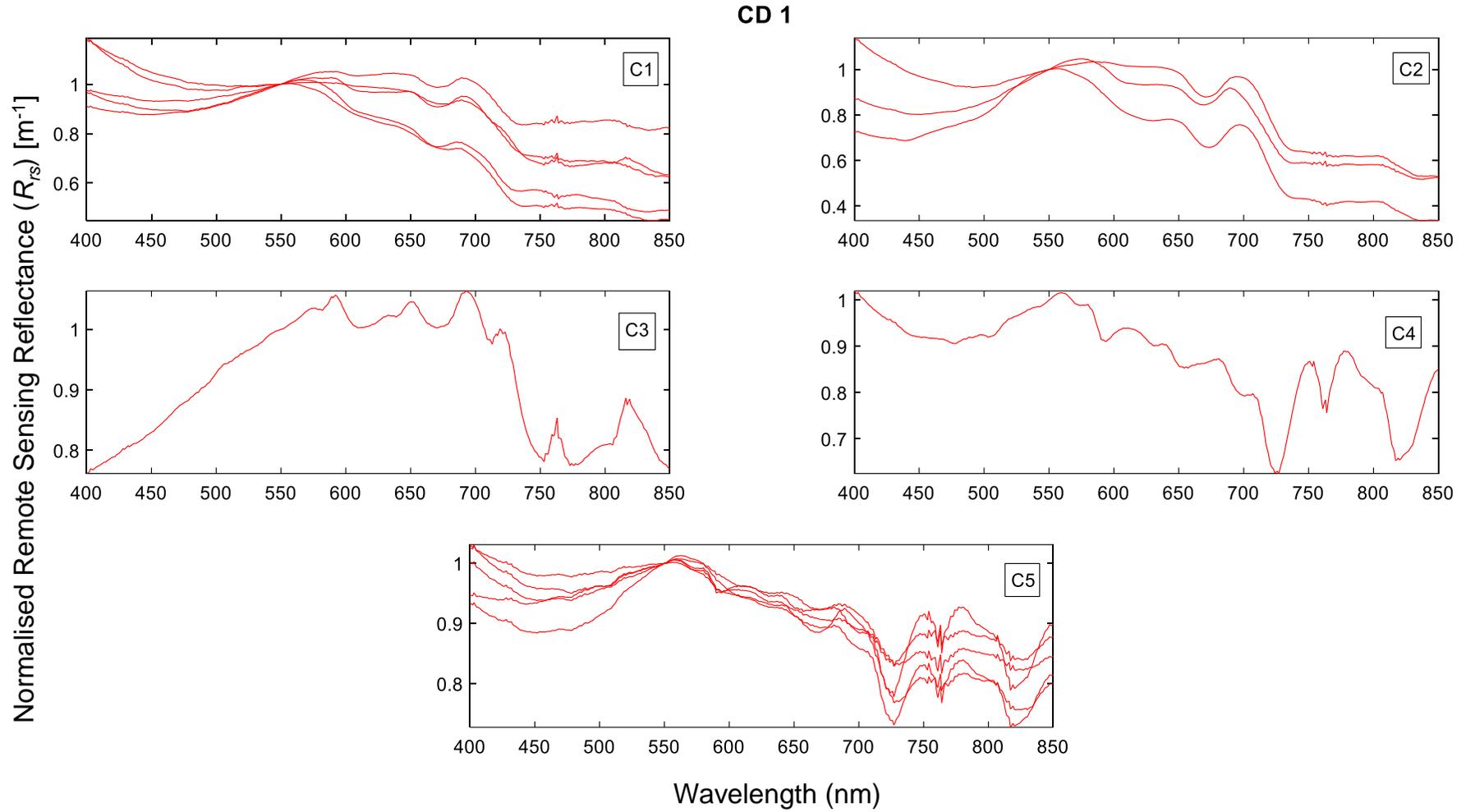


Figure F.3 – DWTCA-Clustered  $R_{rs}$  Spectra for decomposition level two (Haar wavelet).

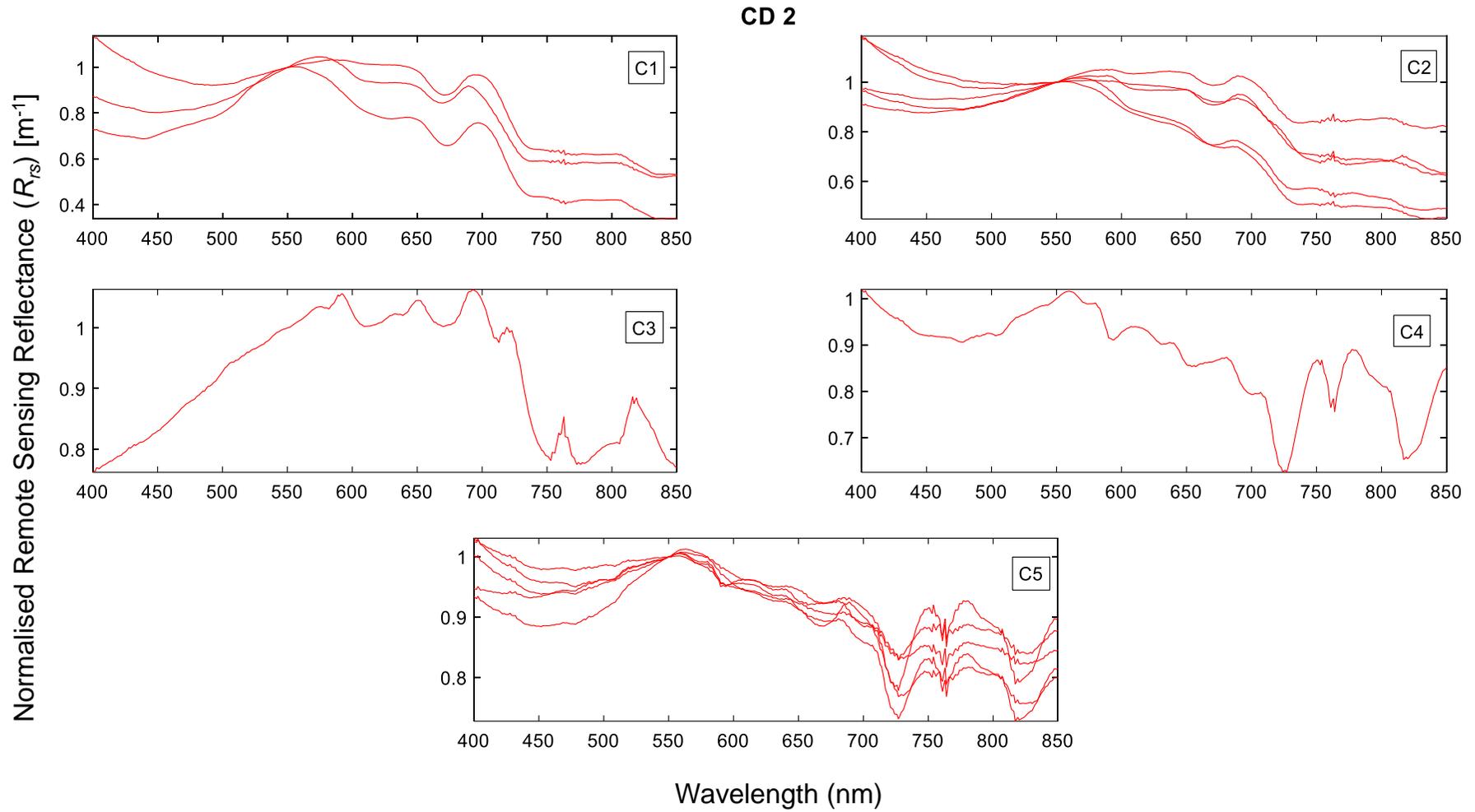


Figure F.4 – DWTCA-Clustered  $R_{rs}$  Spectra for decomposition level three (Haar wavelet).

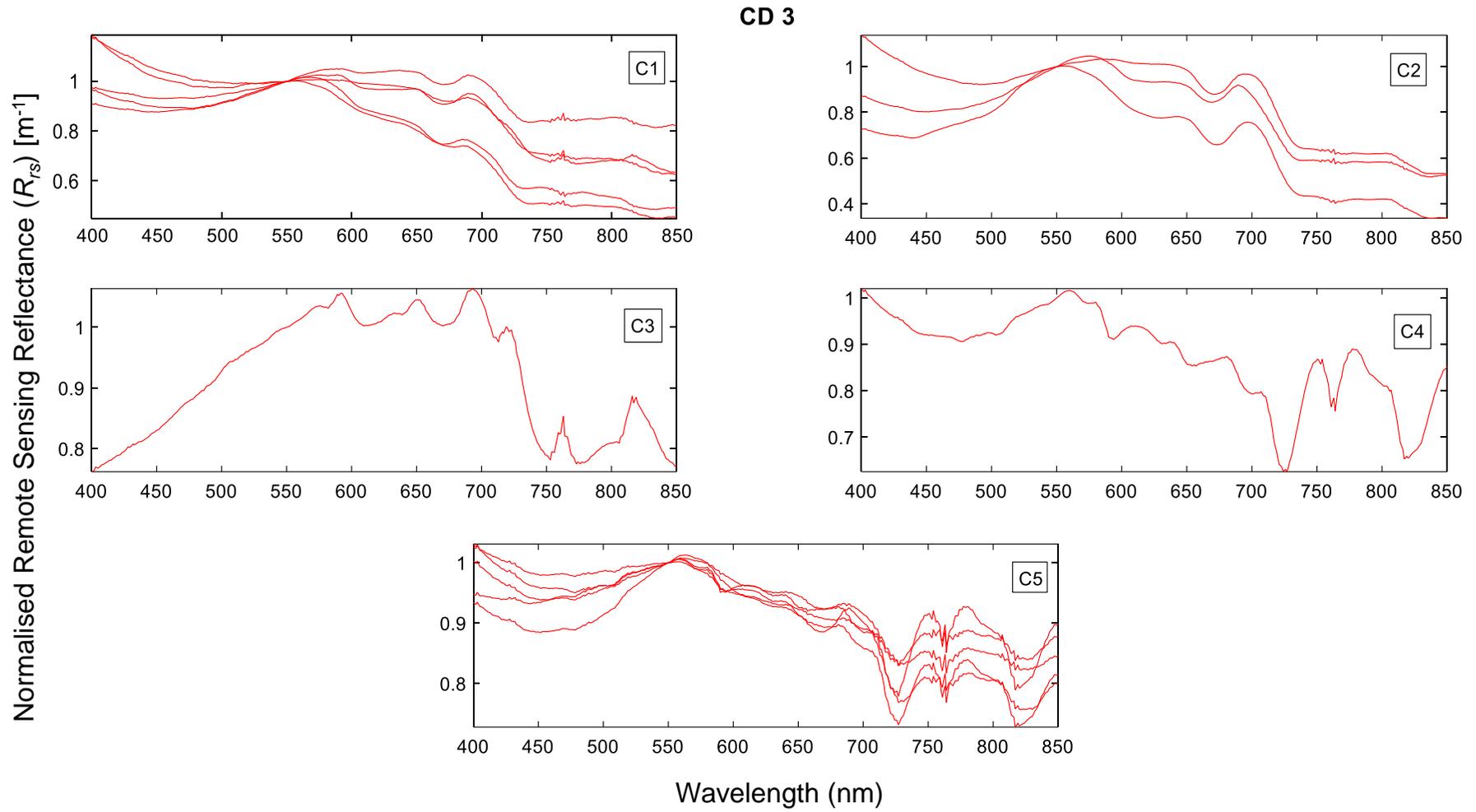


Figure F.5 – DWTC-Clustered  $R_{rs}$  Spectra for decomposition level four (Haar wavelet).

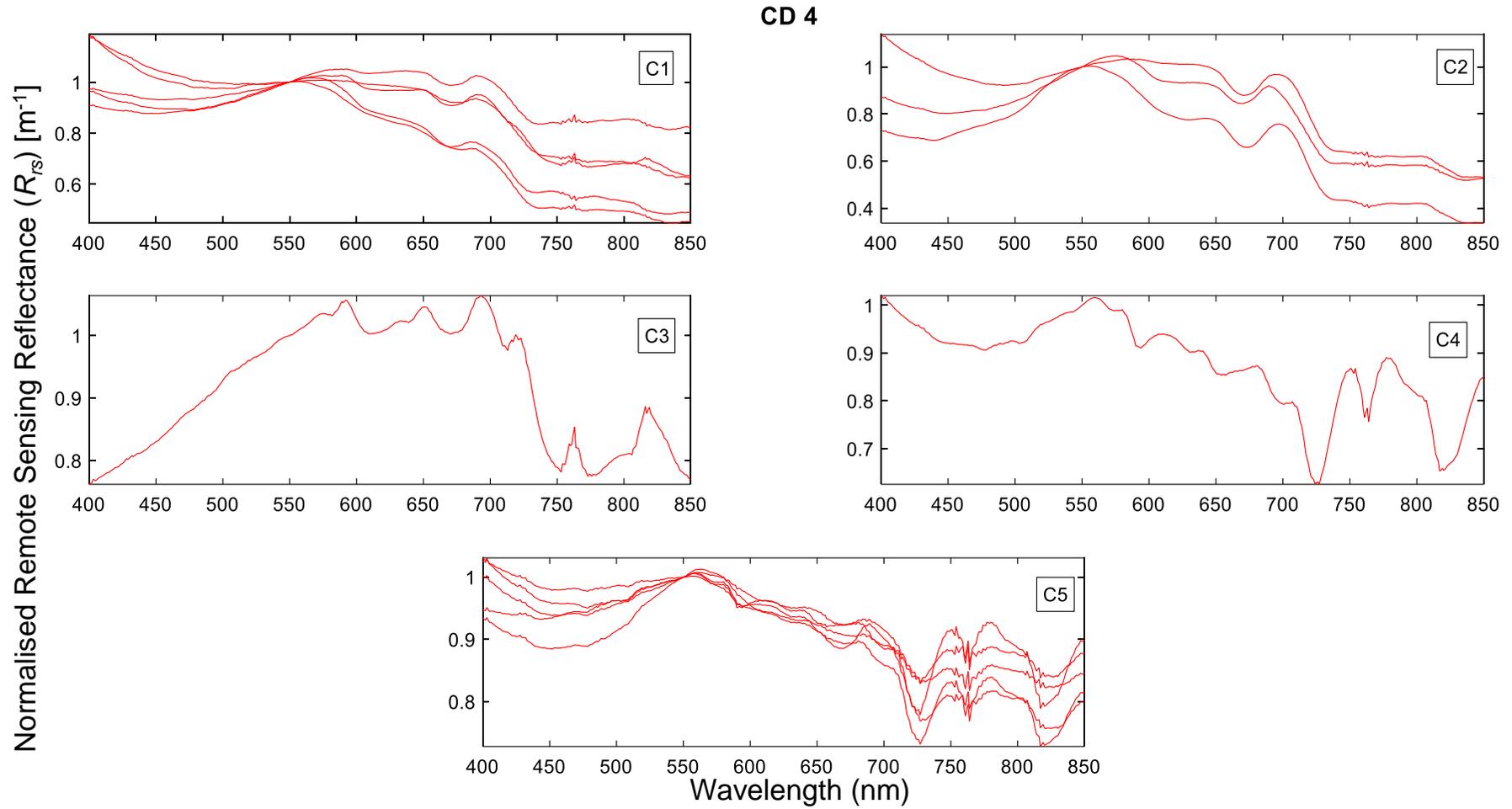


Figure F.6 – DWTCA-Clustered  $R_{rs}$  Spectra for decomposition level five (Haar wavelet).

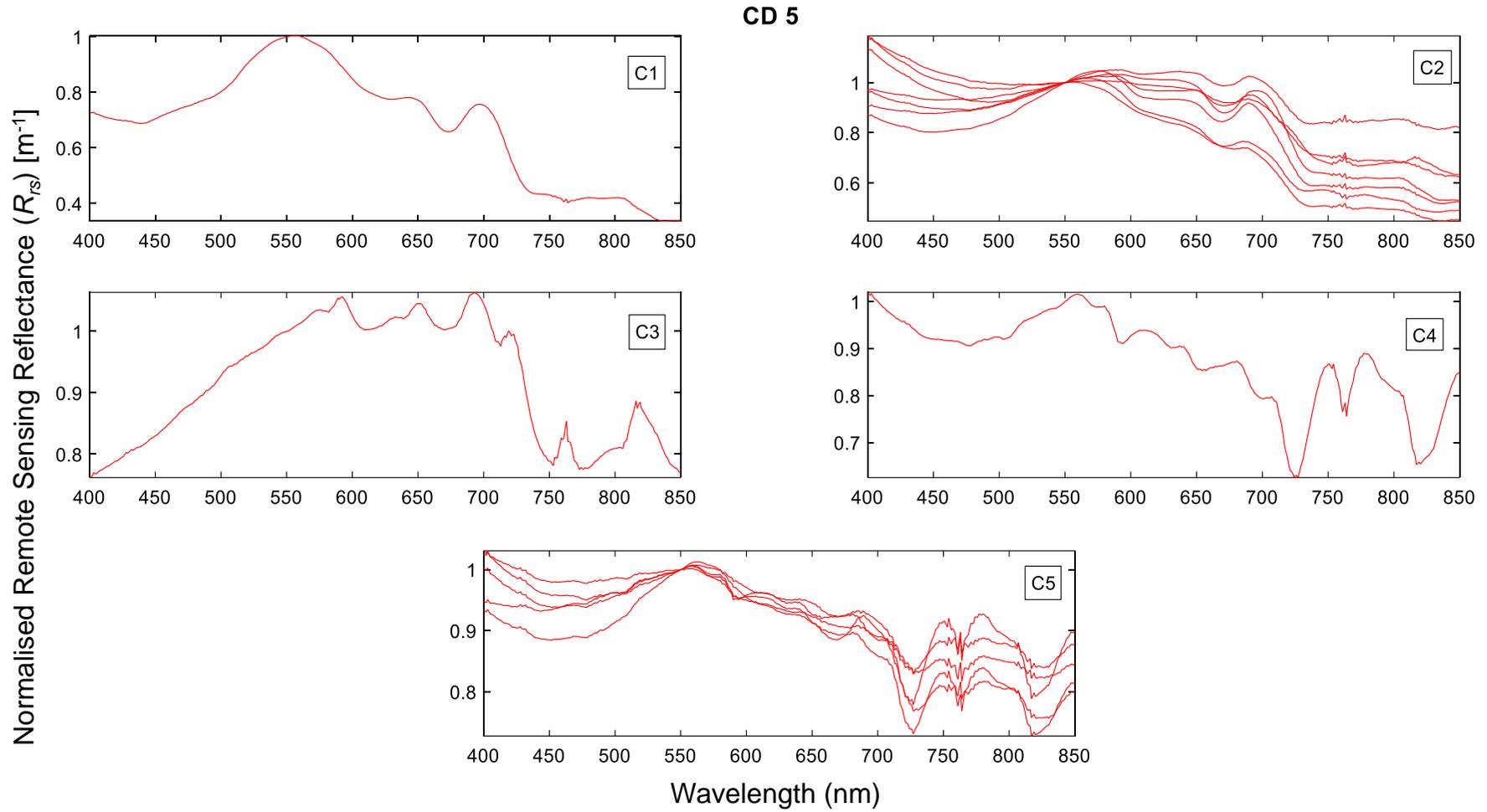


Figure F.7 – DWTCA-Clustered  $R_{rs}$  Spectra for decomposition level six (Haar wavelet).

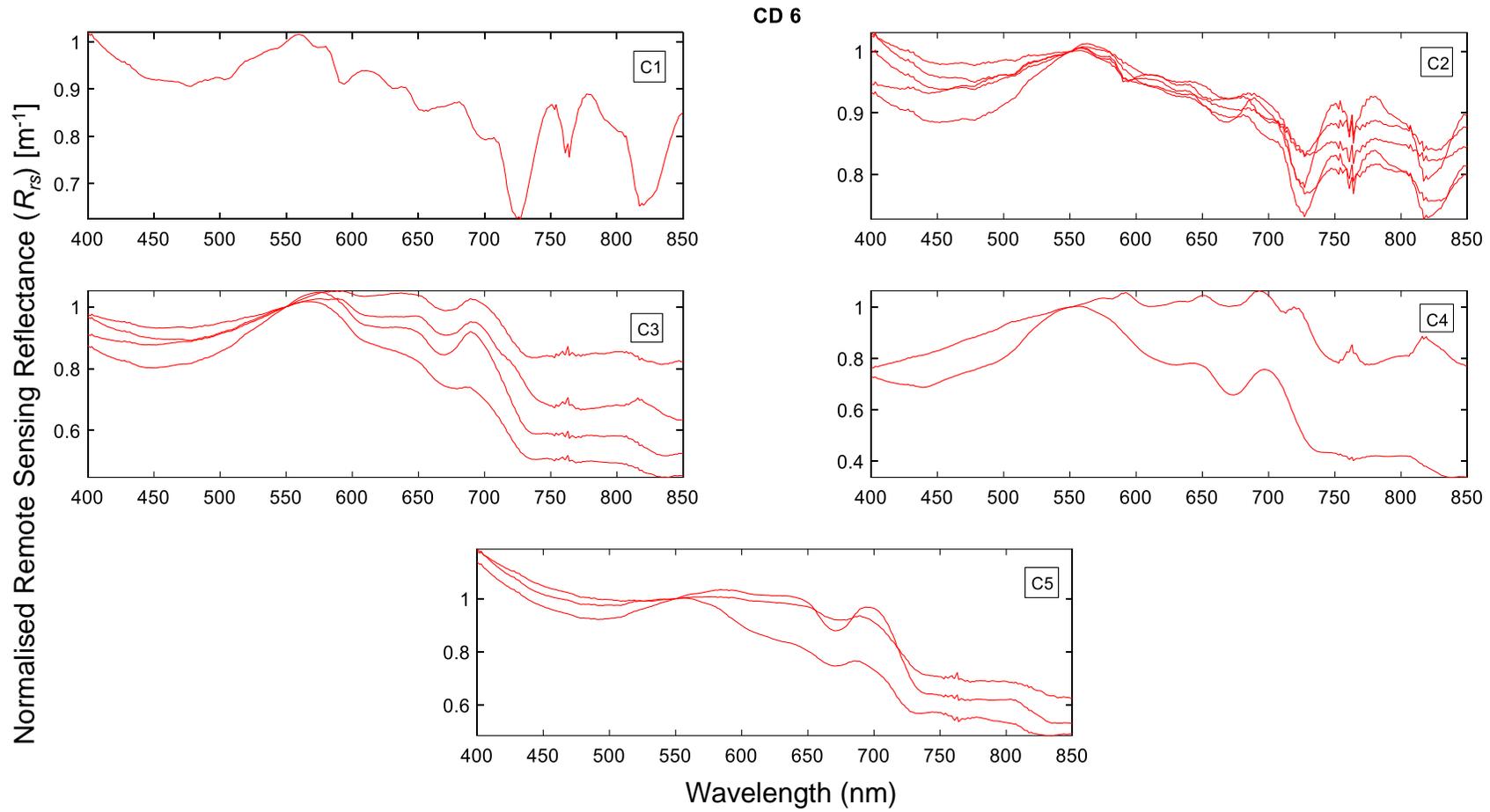


Figure F.8 – DWTCA-Clustered  $R_{rs}$  Spectra for decomposition level seven (Haar wavelet).

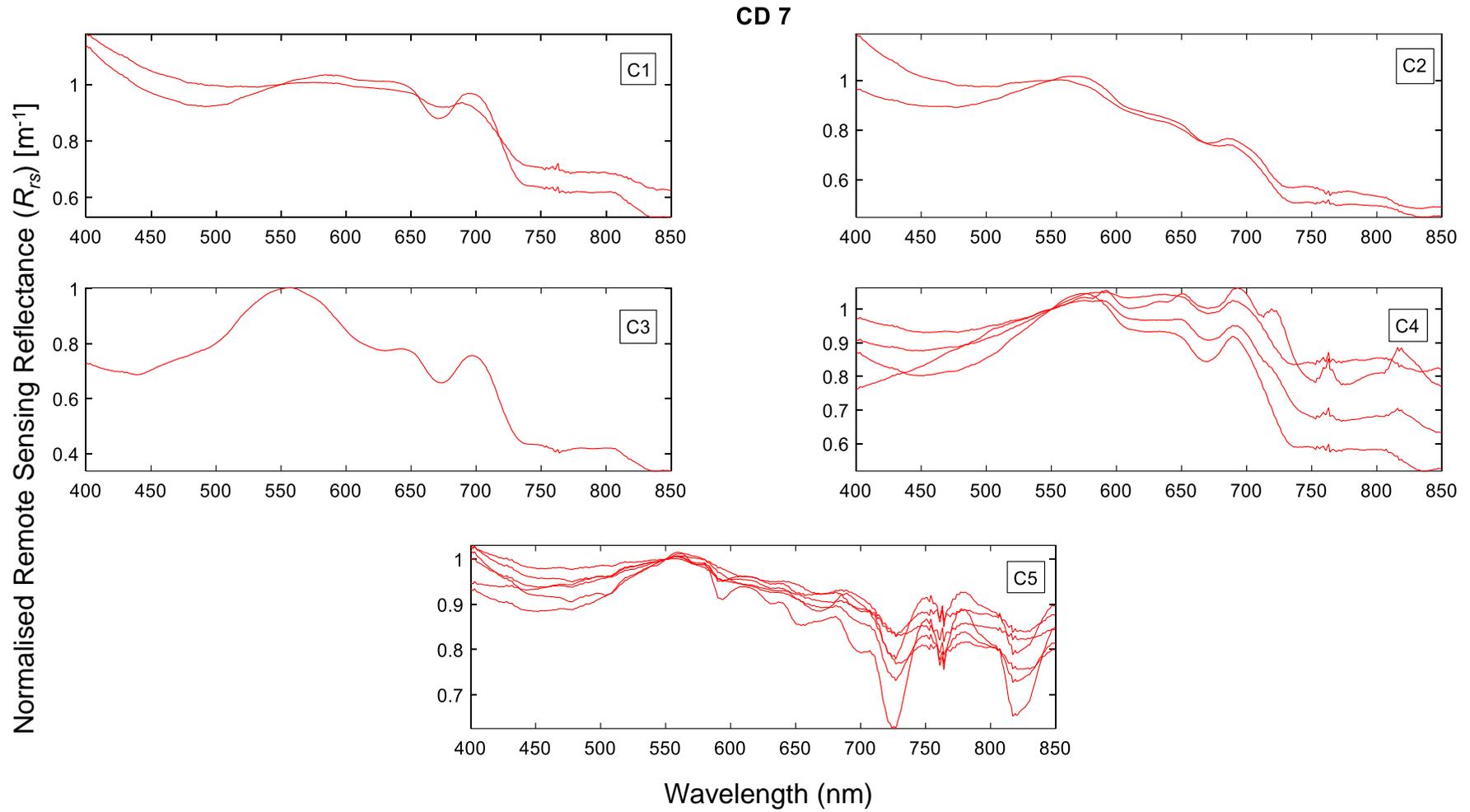
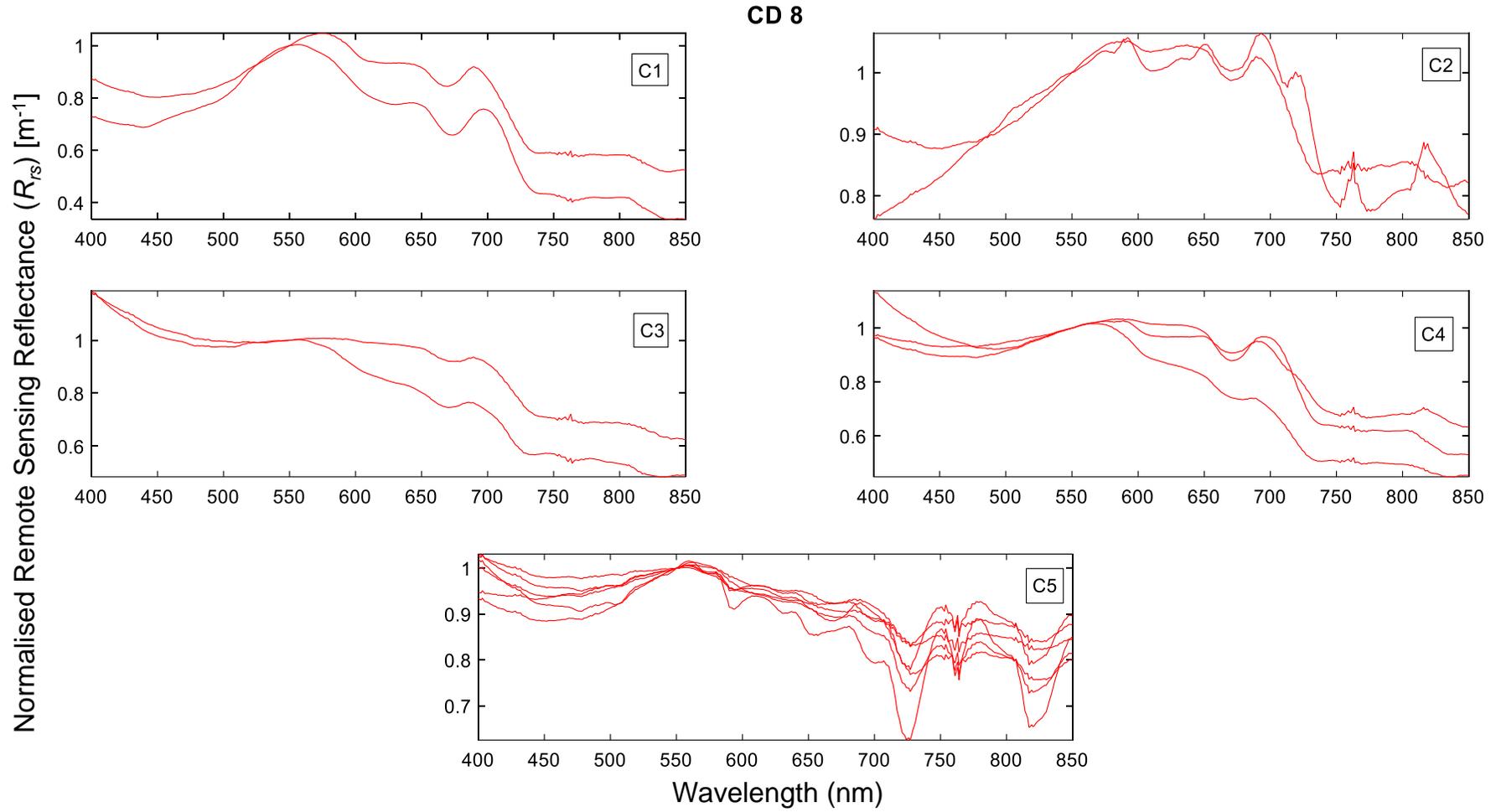


Figure F.9 – DWTCA-Clustered  $R_{rs}$  Spectra for decomposition level eight (Haar wavelet).





**ANNEX A. UTM Coordinate Table Location of All Water Quality Sampling Stations Monitored by SABESP and CETESB at Guarapiranga Reservoir**



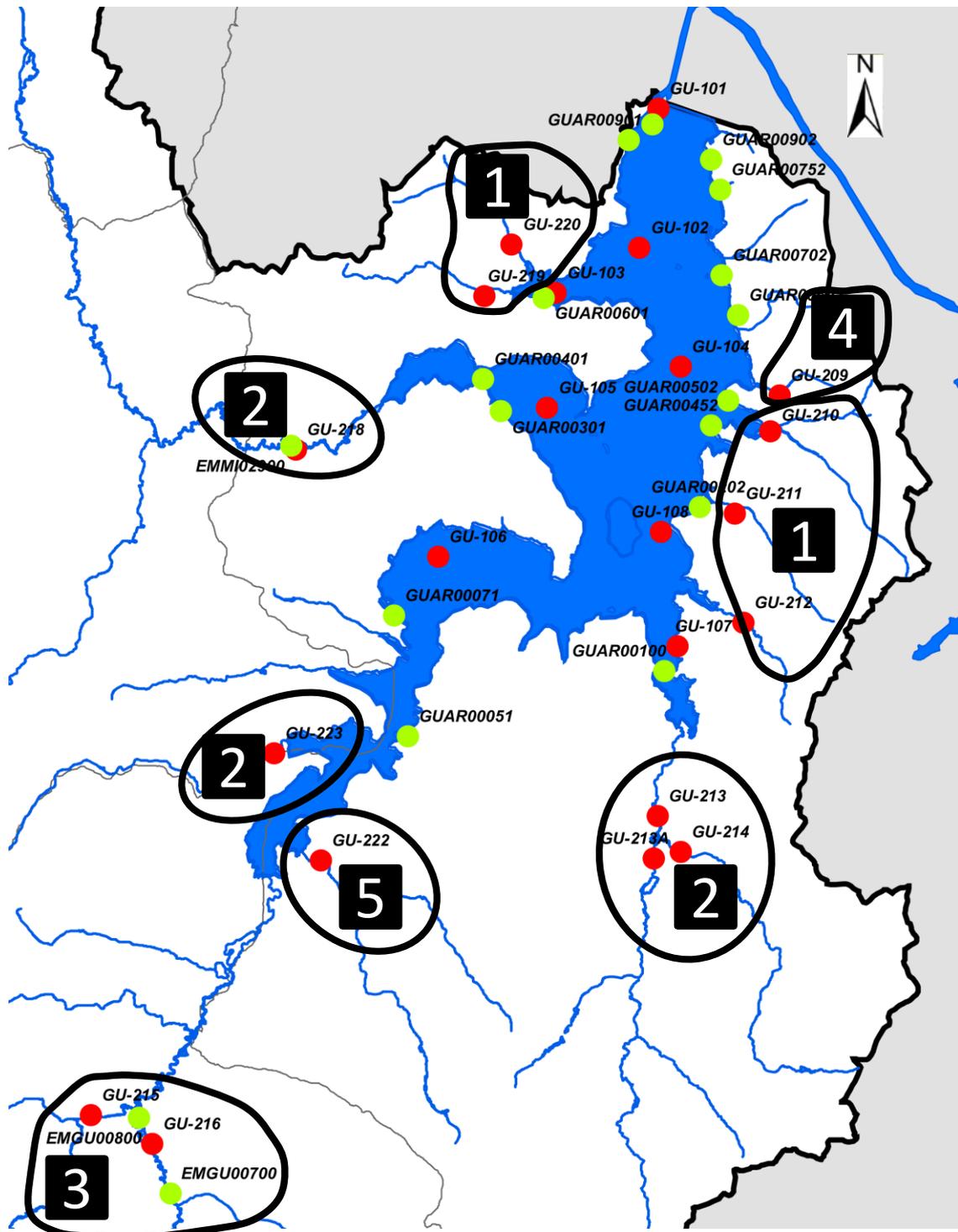
**GOVERNO DO ESTADO DE SÃO PAULO  
SMA/CPLA**

| Ponto de monitoramento | Rede de monitoramento   | Localização   | Coordenadas UTM                                       |           |         |
|------------------------|---|---|---|-----------|---------|
|                        |   |   | Norte   | Leste     |         |
| GU-101                 | SABESP  | Próximo à barragem Guarapiranga   | 7.381.023   | 323.900   |         |
| GU-102                 |   | Meio do corpo d'água, antes da zona de captação                         | 7.378.770   | 323.596   |         |
| GU-103                 |   | Braço dos córregos Itupu e Guavirutuba                                  | 7.378.037   | 322.254   |         |
| GU-104                 |   | Meio do corpo d'água, próximo ao desemboque do Rio Bonito               | 7.376.837   | 324.260   |         |
| GU-105                 |   | Meio do braço do rio Embu Mirim   | 7.376.179   | 322.114   |         |
| GU-106                 |   | Meio do braço do rio Embu Guaçu   | 7.373.766   | 320.360   |         |
| GU-107                 |   | Meio do braço do rio Parelheiros  | 7.372.307   | 324.205   |         |
| GU-108                 |   | Meio do corpo d'água, próximo à Ilha dos Eucaliptos                     | 7.374.162   | 323.939   |         |
| GU-209                 |   | Córrego 2   | 7.376.379   | 325.848   |         |
| GU-210                 |   | Córrego Bonito / Córrego das Pedras                                     | 7.375.799   | 325.696   |         |
| GU-211                 |   | Córrego São José  | 7.374.463   | 325.123   |         |
| GU-212                 |   | Córrego Tanquinho   | 7.372.696   | 325.264   |         |
| GU-213                 |   | Rio Parelheiros, à jusante da transferência, ponte da Estrada Jaceguava | 7.369.565   | 323.891   |         |
| GU-213A                |   | Rio Parelheiros, à montante da transferência                            | 7.368.874   | 323.822   |         |
| GU-214                 |   | Ribeirão Itaim  | 7.368.978   | 324.255   |         |
| GU-215                 |   | Ribeirão Santa Rita   | 7.364.511   | 314.843   |         |
| GU-216                 |   | Rio Embu Guaçu  | 7.364.246   | 315.772   |         |
| GU-218                 |   | Rio Embu Mirim  | 7.375.499   | 318.074   |         |
| GU-219                 |   | Córrego Itupu   | 7.377.989   | 321.102   |         |
| GU-220                 |   | Córrego Guavirutuba / Córrego Talamanca                                 | 7.378.818   | 321.544   |         |
| GU-222                 |   | Córrego Sem Nome, Golfe Clube Guarapiranga                              | 7.368.837   | 318.482   |         |
| GU-223                 |   | SABESP  | Córrego do Bairro Crispim                             | 7.370.577 | 317.726 |
| EMGU00700              |   | CETESB<br>Rede Básica   | Rio Embu Guaçu (a jusante da confluência do Rio Cipó) | 7.363.444 | 315.632 |
| EMGU00800              | Rio Embu Guaçu (Ponte na estrada que liga Embu Guaçu à Fazenda da Ilha)               |   | 7.363.841   | 315.748   |         |
| EMMI02900              | Rio Embu Mirim (Ponte na Estrada do M'Boi Mirim – SP-214)                             |   | 7.375.561   | 318.007   |         |
| GUAR00100              | Reservatório Guarapiranga (Braço do Rio Parelheiros, no bairro do Balneário São José) |   | 7.371.912   | 324.000   |         |
| GUAR00900              | CETESB<br>Rede Básica   | Reservatório Guarapiranga (Captação da SABESP, junto à casa de bombas)  | 7.380.771   | 323.807   |         |
|                        | CETESB<br>Sedimentos  |   |   |           |         |

| Ponto de monitoramento | Rede de monitoramento    | Localização   | Coordenadas UTM |         |
|------------------------|--------------------------|---|-----------------|---------|
|                        |                          |   | Norte           | Leste   |
| GUAR00051              | CETESB<br>Balneabilidade | Bairro do Crispim   | 7.370.847       | 319.878 |
| GUAR00071              |                          | Associação dos Funcionários Públicos do Estado de São Paulo | 7.372.813       | 319.656 |
| GUAR00202              |                          | Clube de Campo São Paulo                                    | 7.374.565       | 324.562 |
| GUAR00301              |                          | Bairro Miami Paulista                                       | 7.376.126       | 321.371 |
| GUAR00401              |                          | Marina Guaraci  | 7.376.646       | 321.081 |
| GUAR00452              |                          | Prainha do Jardim Represa                                   | 7.375.890       | 324.745 |
| GUAR00502              |                          | Clube de Campo Castelo                                      | 7.376.294       | 325.023 |
| GUAR00601              |                          | Restaurante do Odair  | 7.377.950       | 322.057 |
| GUAR00602              |                          | Restaurante Interlagos                                      | 7.377.680       | 325.176 |
| GUAR00702              |                          | Marina Guarapiranga   | 7.378.323       | 324.914 |
| GUAR00752              |                          | Marina Jd. 3 Marias   | 7.379.707       | 324.897 |
| GUAR00901              |                          | Parque Guarapiranga   | 7.380.512       | 323.426 |
| GUAR00902              |                          | Yacht Club Santo Amaro                                      | 7.380.198       | 324.749 |



**ANNEX B. Input Feeding Streams by Socio-Economic-Environmental Areas and Water Quality Sampling Stations by SABESP and CETESB**



Note: group 1, highly polluted lower-urban areas; 2, lower-urban areas; 3, non-urban areas; 4, higher-urban areas; 5, mostly unaltered rural areas.



## ANNEX C. Sentinel-3 OLCI instrument technical characteristics

|                                |  |
|--------------------------------|--|
| <b>Swath</b>                   | 1440 km  |
| <b>SSI at SSP (km)</b>         | 300 m  |
| <b>Calibration</b>             | MERIS type calibration arrangement with spectral calibration using a doped Erbium diffuser plate, PTFE diffuser plate and dark current plate viewed ~ every 2 weeks at the South Pole ecliptic. Spare diffuser plate viewed ~ periodically for calibration degradation monitoring. |
| <b>Detectors</b>               | ENVISAT MERIS heritage back illuminated CCD55-20 frame-transfer imaging device (780 columns by 576 row array of 22.5 $\mu\text{m}$ square active elements).  |
| <b>Optical scanning design</b> | Push-broom sensor. 5 cameras recurrent from MERIS dedicated Scrambling Widow Assembly supporting 5 Video Acquisition Modules (VAM) for analogue to digital conversion.   |
| <b>Spectral resolution</b>     | 1.25 nm (MERIS heritage), 21 bands.  |
| <b>Radiometric accuracy</b>    | < 2% with reference to the sun for the 400-900 nm waveband and < 5% with reference to the sun for wavebands > 900 nm<br>0.1% stability for radiometric accuracy over each orbit and 0.5% relative accuracy for the calibration diffuser BRDF.                                      |
| <b>Radiometric resolution</b>  | < 0.03 W m <sup>-2</sup> sr <sup>-1</sup> mm <sup>-1</sup>   |
| <b>Mass</b>                    | 150 kg   |
| <b>Size</b>                    | 1.3 m <sup>3</sup>   |
| <b>Design lifetime</b>         | 7.5 years  |

Source: Donlon et al. (2012)