A *B*IO-OPTICAL AQUATIC DATABASE FOR THE *R*EMOTE SENSING OF W*A*TER QUALITY IN COA*S*TAL AND INLAND WATERS OF BRAZ*IL*

Daniel Maciel[1], Claudio Barbosa[1], Evlyn Novo[1], Raianny Wanderley[1], Lohan Baia[2], Rafael Chasles[1], Aurea Ciotti[3], Rogério Flores Júnior[1], Fernando Lopes[4], Rogério Marinho[5], Vitor Martins[6], Marcela Miranda[7], Gabriel Moiano[1], David da Motta Marques[8], Maurício Noernberg[9], Fernanda Helena Oliveira[4], Felipe Pacheco, Waterloo Pereira Filho[10], Lino Augusto Sander de Carvalho[11], Marcio Sousa[12], Matheus Tavares[8], Marta Ummus[13], Aline Valério[1], Pedro Walfir[12], Fernanda Watanabe[14], Rafaela Zem[9]

[1]Earth Observation Coordination of National Institute for Space Research (INPE), São José dos Campos/SP/Brazil (daniel.maciel, claudio.barbosa, evlyn.novo, aline.valerio, raianny.leite@inpe.br, gabrielmoiano@gmail.com, rafaelchasles@gmail.com). [2]Federal University of Pará, Belém/PA/Brazil., lohanbaia@gmail.com [3]CEBIMAR/USP, São Sebastião/SP/Brazil, ciotti@usp.br [4]Federal University of Ceará, Fortaleza/CE/Brazil. lopesfb@ufc.br, fernepomuceno.helena@gmail.com [5]Federal University of Amazonas, Manaus/AM/Brazil. rogeo@ufam.edu.br. [6]Mississippi State University, US. vitorsm@msu.edu[7]Cornell University, US. mc2828, felipe.pacheco@cornell.edu[8] Federal University of Rio Grande do SUL, Porto Alegre/RS/Brazil. dmm@iph.ufrgs.br, tavaresmatheush@gmail.com [9]Federal University of Paraná (UFPR), Pontal do Paraná/PR/Brasil. rafaelazem@gmail.com, m.noernberg@ufpr.br [10]Federal University of Santa Maria (UFSM), Santa Maria/RS/Brasil. waterloopf@gmail.com. [11]Federal University of Rio de Janeiro, Rio de Janeiro/RJ/Brazil. lino.sander@igeo.ufrj.br [12]Vale Institute of Technology (ITV), Belém/PA/Brazil. marcio.sousa.silva@itv.org, pedrowalfir@icloud.com [13]Embrapa Pesca e Aquicultura, Palmas/TO/Brazil. marta.ummus@embrapa.br [14]São Paulo State University (UNESP), Presidente Prudente/SP/Brazil. fernanda.watanabe@unesp.br

ABSTRACT

Monitoring water quality (WQ) in a vast country as Brazil demands significant effort and financial resources. Remote sensing (RS) plays a crucial role in optimizing this monitoring by providing a synoptic view of aquatic environments with high spatiotemporal resolution. However, developing high-quality WQ products using RS requires a comprehensive and extensive database that includes concurrent aquatic reflectance data and WQ variables (e.g., Chlorophyll-a, Secchi Disk Depth). Such databases enable the application of advanced semi-analytical and machine learning models to capture the bio-optical complexity of these waters. Building this database requires a collaborative community effort. This paper presents the first compilation of bio-optical datasets for the Brazilian coastal and freshwater ecosystems. This initiative involves 14 institutions and currently comprises 3,058 unique stations across more than eighty lakes, rivers, reservoirs, and coastal waters in Brazil. Our dataset reveals that the Amazon Basin is well represented with over 1,600 stations; however, the central-west region, particularly the Pantanal, still lacks adequate representation in these bio-optical datasets. The final dataset is still under development and will be submitted to a *Data* journal soon.

Key words — Bio-optical dataset, tropical freshwater database, reflectance, water quality, remote sensing.

1. INTRODUCTION

Brazil covers 8.51 millions of km^2 and approximately 180.000 km^2 of water, including rivers, reservoirs, flooded forest, floodplain and coastal waters [1]. The freshwater resources in Brazil are abundant but unevenly distributed, with large areas in the Amazon having plenty of water while the northeast region faces scarcity. Moreover, the recent antrophic pressures such as climate change, illegal gold mining, deforestation, fires, extensive agricultural and industrial development are putting the quality of Brazilian aquatic resources at risk [2]. Therefore, it is crucial to establish a comprehensive system for monitoring water quality in Brazil to track the impacts of human activities. However, due to the country's large size, it is not feasible to conduct standard punctual in-situ monitoring by increasing the number of stations and the frequency of sampling the WQ variables. Nevertheless, several organizations (i.e., the Brazilian Water Agency – ANA) have several stations distributed in the country [3]. These stations only offer punctual description of the water bodies, with a limited temporal coverage (generally ~ 1 month) and spatial representativeness. With that in mind, RS plays a key role into expanding water quality monitoring possibilities by providing a synoptic view with a high spatiotemporal resolution of freshwater ecosystems [4].

Even though RS is an essential tool for large scale water quality monitoring, its application to Brazilian aquatic ecosystems still has some limitations. First, the large size of the Brazilian territory implies a diversity of environments, from the organic-matter rich flooded forests of the Amazon Basin, passing through eutrophic reservoirs in the São Paulo and Rio de Janeiro states, turbid waters of the Amazon River, clear waters of well-preserved reservoirs, and coastal waters [5]. This high variability implies in a mixture of optically active constituents (i.e., colored dissolved organic matter, suspended sediments, phytoplankton pigments), making it an optically complex environment, imposing challenges for algorithm development.

The retrieval of water quality parameters using RS data generally requires the use of empirical (e.g., band-ratios, machine learning) and semi-analytical methods, aligned with co-located measurements of remote sensing reflectance (R_{rs}) (i.e., ratio of water-leaving radiance by downwelling irradiance) and these parameters [4]. This allows the construction of algorithms that could be validated using satellite data. Large-scale applications or even local algorithms require a comprehensive sample size to allow the model to understand and adapt to the variability of optical parameters. Therefore, it is necessary to construct *in-situ* databases allowing the use of data collected in different locations. These initiatives are only possible through an inter-institutional collaboration network, based on the principles of free data sharing [6]. To construct a bio-optical database for Brazilian fresh and coastal ecosystems, this study describes the process used to collect, compile and organize the first bio-optical dataset for Brazilian waters. With more than ten institutions over Brazil (30 > researchers), we compiled by now ~3,000 *in-situ* measurements of R_{rs} and other water quality parameters over the five regions of Brazil.

2. MATERIAL AND METHODS

2.1. Study Area

The study area encompasses the five regions of Brazil. The dataset collected comprehends 87 different environments in Brazil, comprehending coastal waters, reservoirs, small and large rivers (e.g., Amazon, Negro), and floodplain lakes. Currently, our project partners are sharing data from 2,718 stations. Out of these, 1,545 contains valid measurements of Chlorophyl-*a*, 243 stations contain valid measurements of phycocyanin, 1,250, 1,109, and 1,046 contains valid measurements of total suspended sediments (TSS) and its organic (TSO) and inorganic portion (TSI). Secchi Disk Depth (Z_{sd}) is available now for 2,178 stations. Turbidity is available for 1,126 stations, and the absorption coefficient by colored dissolved organic matter (CDOM) at 440nm is available for 689 stations. A brief description of the methods used for each variable measurement is available in Section 2.2.

2.2. Database organization

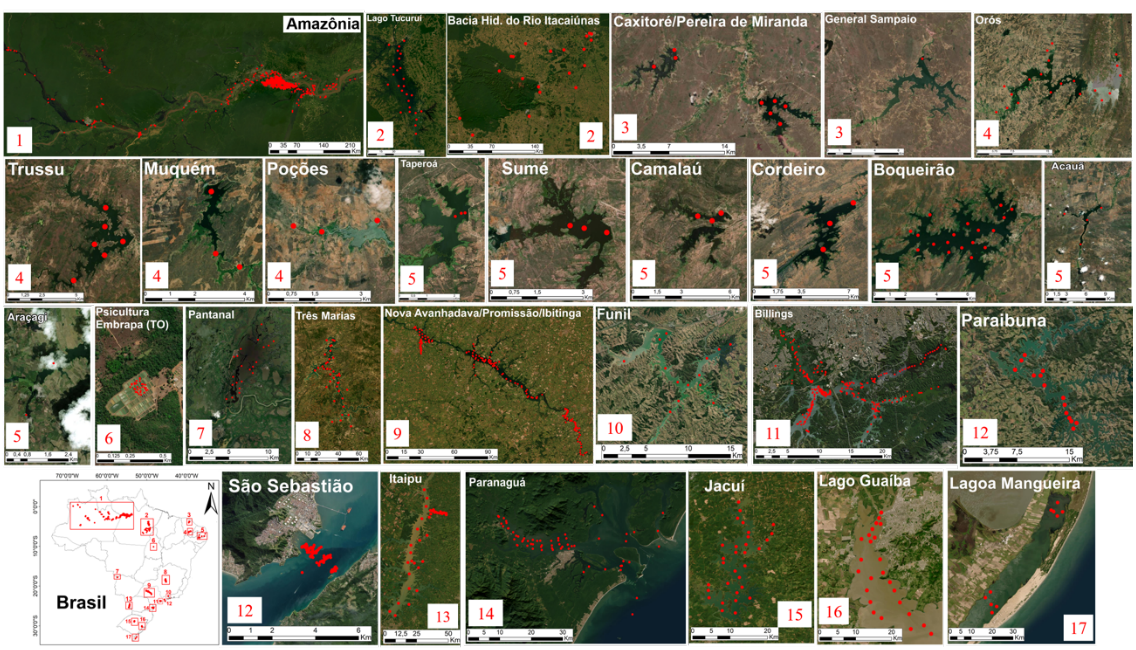

Figure 1. Study area. Each red-point correspond to *in-situ* measured station. Amazon box was grouped due to the large number of unique lakes. Numbers in each box refer to the location highlighted in the Brazil map.

The database organization used a database-oriented structure following what was adopted by LabISA team, which is also inspired in the GLORIA dataset. The RS data is available in three tables with radiometric quantities (See Section 2.2.1) and a table with R_{rs} data. A third table contains the metadata (i.e., station identifier) and other necessary information such as date and time, geographic coordinates, and water quality parameters. Quality-control for each dataset will be implemented after database finishing and will follow what was adopted for GLORIA dataset.

2.2.1 In-situ remote sensing reflectance

The *in-situ* R_{rs} data was collected using different types of radiometers as identified in the *Metadata* table and can follow different settings, as follows: **1) TRIOS-RAMSES** radiometers. In this setting, one radiometer is used for measuring water leaving radiance (L_w), another for sky radiance (L_{sky}) and another for downwelling irradiance just beneath the water surface (E_s). This setting follows the geometrical protocols of [7]. **2) ASD HandHeld-2**. In this setup, the HandHeld-2 instrument is used in radiance mode to collect water leaving radiance (L_w), sky radiance (L_{sky}) and the radiance of a Spectralon Lambertian plaque (L_s), which is then converted to irradiance by integrating it into the hemisphere (i.e., multiplying byπ). **3) ASD FieldSpec**. FieldSpec was also used in some collections with the same configuration of the HandHeld. **4) Satlantic HyperPRO II** measuring L_u (upwelling radiance below water surface) and E_s. L_u data were extrapolated to the surface by empirically calculating the KL_u [8] (buoy mode) or based on L_u profiles when measured in profile mode. Lu(0-) data were converted to L_w using the Fresnel and water refraction index (1.34). R_{rs} was calculated using Equation 01, with the ρ calculated using Mobley [7] Look-up-Table and auxiliary parameters, such as wind speed, sun-zenith and view angle.

$$R_{rs} = \frac{L_w - \rho\, L_{sky}}{E_s} \quad (1)$$

2.2.2 Water Quality parameters

Diverse water quality parameters were compiled in this study to construct the database. These parameters are: Chlorophyll-*a (*Chl-*a*); Pheophytin and phycocyanin (PC) concentration, Suspended Sediment (total, organic and inorganic – TSS, TSI and TSO) concentration; absorption coefficient by colored dissolved organic matter (aCDOM(440)), Secchi Disk Depth (Z_{sd}), Turbidity, Dissolved Organic Carbon and Dissolved Inorganic Carbon (DOC, DIC). These data were measured using different methods by each laboratory and these methods will be identified in the *Metadata* table.

3. RESULTS

3.1 Dataset variability

The compiled in-situ dataset represents the bio-optical properties across all five regions of Brazil. In the current dataset, the largest R_{rs} sample size is from Pará state (N = 1,213 points), primarily from Curuai Lake, which has been the subject of study for over 20 years by the LabISA team and other partners, involving more than ten field missions for data collection. Following Pará, São Paulo state has 874 unique stations, with data collected over more than 15 years in several reservoirs and coastal waters of São Sebastião Channel. The dataset for São Paulo encompasses very eutrophic reservoirs (e.g., Billings, Ibitinga, and Promissão), clear water environments (e.g., Paraíbuna Reservoir), and coastal waters (São Sebastião Channel). This wide diversity of environments (rivers, reservoirs, floodplains, and coastal waters) enables the construction of a database with high optical diversity. Considering all the available dataset, after PA and SP, the third most representative region is the Amazonas state, with 309 unique stations. It is followed by Ceará (N = 162), Paraná (N = 122), Tocantins (N = 92), Mato Grosso (N = 46), Minas Gerais (N = 42), Paraiba (N = 39), Rio Grande do Sul (N = 31) and Rio de Janeiro (N = 14) states. The variability in the R_{rs} spectra is evident and is further demonstrated by the variability of the optical significant constituents among the available R_{rs} data. Please note that these R_{rs} plots are preliminary, as not all partners have yet provided the complete dataset with R_{rs} and OACs data. Despite this, we can observe that the spectra capture high turbidity environments (e.g., the Amazon Region) eutrophic environments (e.g., RJ and SP states), as well as clearwater locations, such as the Três Marias reservoir in MG state. High sediment concentrations were observed in the Amazon Basin stations (e.g., Pará state, mean TSS = 25.31 mgL^{-1}). Water transparency was lowest for the Pará stations (Z_{sd} = 0.39 m), and the higher was obtained for MG state (Z_{sd} = 3.08m). This is because in MG state only one reservoir (Três Marias) was sampled in 2013 and 2019, and this is a clear water reservoir [9]. The aCDOM(440) was higher for the AM state (3.07 m^{-1}), much connected to the waters of Mamirauá Sustainable Reservoir.

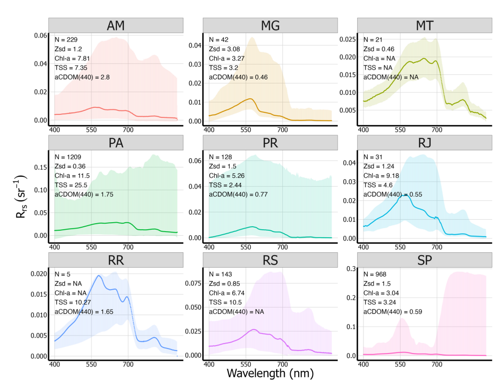

Figure 2. Variability of R_{rs} for the available data. Each box-title represents the state in which the R_{rs} came from.

3.2 Water quality parameters variation

The relationship between the observed water quality parameters (Chl-a, TSS, TSI, DTC, DOC, Zsd, aCDOM(440) and turbidity) also reflects the variation in the composition of aquatic ecosystems in Brazil (Figure 3). A strong correlation was observed between TSS and TSI concentrations in the Amazonian data (R = 0.99). Additionally, a high correlation was found between Chl-a and TSS in the datasets from São Paulo (R = 0.77) and Rio de Janeiro (R = 0.89), suggesting the prevalence of phytoplankton cells in the suspended sediments. In the Minas Gerais (MG) dataset, aCDOM(440) showed a strong correlation with water transparency, indicating that organic matter plays a significant role in the absorption budget of the Três Marias reservoir.

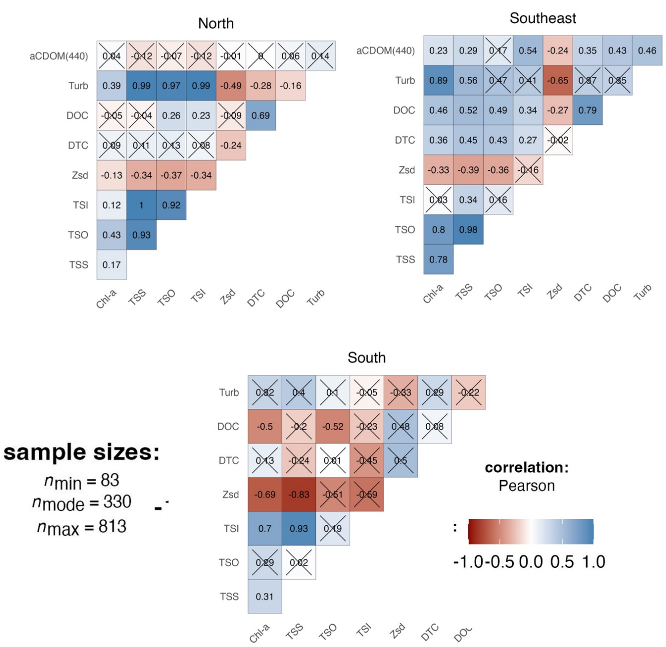

Figure 3. Pearson correlation coefficient between the OACs for the current available data.

4. DISCUSSION AND CONCLUSIONS

In this study, we present the first comprehensive organization of a bio-optical database focused on Brazilian inland and coastal ecosystems. When finished, the dataset will comprise > 3,100 stations with concomitant measurements of R_{rs} and water quality parameters. Numerous partners across Brazil have been invited to contribute to the dataset, significantly enhancing its geographic scope and comprehensiveness.

The establishment of this dataset for Brazil will greatly facilitates the development of various algorithms using both semi-analytical and machine learning approaches. Despite its extensive spatiotemporal coverage, there are still gaps in bio-optical measurements across several Brazilian states. For instance, there is a lack of data for the Pantanal biome, a critical region under severe pressure from fires and deforestation. In Northeast Brazil, the dataset remains limited relative to the vast size of the region. The most substantial data coverage is in the Amazon basin, followed by São Paulo state, reflecting the emphasis on Amazonian water quality studies over the past 20 years. The dataset will allow future developments of large-scale algorithms for Brazilian freshwater ecosystems, as well as to ensure a high-quality dataset for validation of RS products (i.e., atmosphere correction).

4. ACKNOWLEDGMENTS

This study was supported by FAPESP projects 2018/12083-1, 2020/14613-8, 2021/13367-6 and 2023/13904-7.

8. REFERENCES

[1] MAPBIOMAS, "MAPBIOMAS v. 4.1."

[2] K. de Mello *et al.*, "Multiscale land use impacts on water quality: Assessment, planning, and future perspectives in Brazil," *J Environ Manage*, vol. 270, no. October 2019, p. 110879, 2020, doi: 10.1016/j.jenvman.2020.110879.

[3] CETESB, "InfoAguas."

[4] V. Sagan *et al.*, "Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing," *Earth Sci Rev*, vol. 205, no. August 2019, p. 103187, 2020, doi: 10.1016/j.earscirev.2020.103187.

[5] E. F. F. da Silva *et al.*, "Optical water types found in Brazilian waters," *Limnology (Tokyo)*, vol. 22, no. 1, pp. 57–68, Jan. 2021, doi: 10.1007/s10201-020-00633-z.

[6] M. K. Lehmann *et al.*, "GLORIA - A globally representative hyperspectral in situ dataset for optical sensing of water quality.," *Sci Data*, 2023.

[7] C. D. Mobley, "Polarized reflectance and transmittance properties of windblown sea surfaces," *Appl. Opt.*, vol. 54, no. 15, pp. 4828–4849, 2015, doi: 10.1364/AO.54.004828.

[8] R. W. Austin and T. J. Petzold, "The determination of the diffuse attenuation coefficient of sea water using the coastal zone color scanner," *Austin, R. W Petzold, T. J.*, vol. 112, no. 483, pp. 211–212, 1981, doi: 10.1192/bjp.112.483.211-a.

[9] V. P. Curtarelli *et al.*, "Diffuse attenuation of clear water tropical reservoir : A remote sensing semi-analytical approach," *Remote Sens (Basel)*, pp. 1–23, 2020.