WATER TRANSPARENCY TRENDS IN AMAZON BASIN FROM LONG-TERM LANDSAT DATA

Daniel Maciel[1], Claudio Barbosa[1], Evlyn Novo[1], Nima Pahlevan[2], Marie-Paule Bonnet[3], Rejane Paulino[1], Vitor Martins[4]

[1]National Institute for Space Research, Av. dos Astronautas, 1.758 - Jardim da Granja, São José dos Campos – SP, 12227,[2]NASA Goddard Space Flight Center, Greenbelt, MD, USA, [2]Science Systems and Applications Inc. (SSAI), Lanham, MD, USA, [3]Espace-DEV, IRD, Université de Montpellier, France, [4]Department of Agricultural & Biological Engineering, Mississippi State University, Starkville, MS 39762;

ABSTRACT

Water transparency is an important water quality measurement as a proxy to the amount of suspended and dissolved constituents in the water. In the Amazon basin, the water bodies are mainly controlled by annual flooding pulse, which drives the suspended sediments and phytoplankton dynamics in white, organic matter in black waters, and phytoplankton dynamics in clear water. However, anthropogenic impacts, such as massive deforestation and unrestricted mining, may change water transparency patterns. This study aims to develop a Mixture Density Networks Secchi Disk Depth (Z_{sd}) based on Landsat long-term data and evaluate its trends. *In-situ* data and Landsat surface reflectance product allowed the algorithm validation. Then, after applying the algorithm to the 1985-2022 Landsat time series, this study computed time Z_{sd} trends at key Amazon Basin locations based on water types. Results showed positive and significant changes for Madeira, Amazon Rivers, and Monte Alegre Lake.

Key words **—** Secchi Disk Depth, Water Quality, Machine Learning, Mixture Density Networks

1. INTRODUCTION

Amazon basin is one of the most important ecosystems in the world, draining about 20% of the global freshwater and hosting almost 15% of the worldwide biodiversity. The basin covers more than 6 million km² and is home to the world's largest rivers, such as Amazon and Madeira [1]. However, anthropogenic actions such as global warming, gold mining, deforestation, fires, and dam construction risk its diversity [2,3]. These effects could alter the normal sediment fluxes that occur in the Amazon, leading to changes in water transparency and aquatic biodiversity [4]. However, there is a lack of *in-situ* water quality monitoring programs in the Amazon Basin due to access limitations to remote areas and the lack of systematic monitoring programs [1]. For this reason, there is a need for an easy and cheap way to measure water quality, such as water transparency via Secchi Disk depth (Z_{sd}). (Pitarch, 2020).

The Z_{sd} is a long-lasting method for measuring water transparency. It allows infer sediment variability, underwater light field disturbance, water quality changes, phytoplankton variability, and taxonomic biodiversity. In addition, this variable can be estimated from remote sensing data enabling the improvement of Zsd monitoring at remote locations such as floodplain lakes and rivers in the Amazon Basin. Z_{sd} algorithms are usually divided into empirical and semi-analytical approaches. Semi-analytical algorithms provide higher accuracy but requires *in-situ* measurements of water inherent optical proprieties (IOPs). However, IOPs measurements are scarce and unavailable for several lakes in Amazon Region. On the other hand, empirical algorithms only require measures of Z_{sd} and water Apparent optical proprieties (AOPs), such as the Remote Sensing Reflectance (R_{rs}). These empirical algorithms could be developed based on regressions and machine learning methods. Recently, the water remote sensing community has demonstrated the power of Machine Learning methods for predicting water parameters, such as Chlorophyll-a and Suspended Sediments [6]. Recently, a Sentinel-2 Secchi Disk Depth algorithm based on machine learning was also developed for Brazilian Waters [7]. However, long-term time series are mandatory for unraveling the signal of anthropogenic changes on water transparency. This way, assessing Zsd time trends only might be possible for satellite missions covering at least 30 years, such as Landsat. Therefore, this paper aims to evaluate the trends in Secchi Disk Depth in selected locations of Amazon Basin based on time-series of Landsat-family sensors. For that, Machine Learning algorithms (Mixture Density Networks, Random Forest, Support Vector Machine) were calibrated and validated with *in-situ* and satellite data. The best-performed algorithm was used to create the time-series.

2. MATERIAL AND METHODS

2.1. Study Area

The selected study areas are floodplain lakes and rivers of the Amazon Basin which are broadly characterized by their color [1,8]: Black waters are those with low sediment

concentrations and extremely high concentration of colored dissolved organic matter (such as the Negro River); White waters are those with high concentration of suspended sediments as the Amazon and Madeira rivers, and Clear waters are the ones with low concentrations of suspended sediments, organic matter and dissolved organic matter as the Tapajós River. This way, the trends in water transparency was assessed for the time-series at the following locations: Negro, Solimões, Madeira, Amazon, and Tapajós Rivers and Curuai and Monte Alegre Lakes (Figure 1). These locations were defined based on the local/region importance and difference in optical water types.

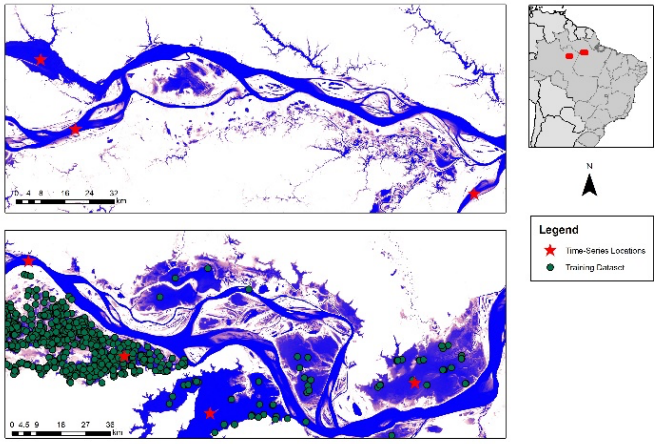

Figure 1. Study Area. Red-stars refers to the locations where the time-series were extracted. Green points refers to *in-situ* data used for training the algorithms.

2.2. Dataset

The dataset available for this study are *in-situ* measurements of Secchi Disk Depth and Remote Sensing Reflectance acquired between 2003 and 2022. For Z_{sd} measurements, a circular 30cm diameter black-and-white disk (Secchi Disk) was used at the boat's shady side, lowered into the water column to its complete disappearance. R_{rs} measurements, were acquired with different radiometers according to the acquisition date. Rrs is generally calculated as the ratio of water leaving radiance above water downwelling irradiance, corrected for glint effects as in Mobley et al. (2015). All samples were collected between 10 AM to 2 PM to avoid the impact of low sun-zenith angles. More informations about the R_{rs} measurements are available in Maciel et al. [7].

2.3 Algorithm Development and validation

For the development of the Z_{sd} algorithms based on Landsat data, three Machine Learning methods were evaluated. These algorithms were: Mixture Density Networks, Random Forest, and Support Vector Machines. The input features were: R_{rs} in bands: coastal, blue, green, and red R_{rs}, and the ratio between green/red and green/blue bands. Coastal bands were used only for OLI and OLI2.

For selecting the best algorithm for Z_{sd} retrieval, a Monte Carlo (MC) simulation (N = 100) was applied, using 70% of the dataset for training and 30% for validation. The dataset selected for training and validation was based on the data for each campaign separately. For example: at each round of MC, 70% of the campaigns were selected for training and 30% for validation. It ensures that points with high spatiotemporal dependence will not be simultaneously included in training and validation samples. At each round of the MC, the statistical metrics error (%), bias, and slope were calculated.

The algorithm was validated using satellite data to observe how the impacts of radiometric resolution and atmospheric correction influences the model results. For that, the best performed algorithm between the three evaluated was applied to Landsat Level-2 Collection-2 data. This data uses the algorithms Land Surface Reflectance Code (LaSRC) for Landsat-8/OLI and Landsat-9/OLI, and Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) for Landsat-5/TM and Landsat-7/ETM+ atmospheric correction. The data were extracted considering a 5-day time window, which is a reasonable time delay for the Amazon Region [9,10]. Clouds and Cloud shadows were masked based on the QA band. In addition, the Modified Normalized Difference Water Index with a threshold of zero was used to eliminate possible false negative locations with cloud/cloud shadows.

2.4 Trends evaluation

For trends evaluation in Z_{sd}, the best-performed algorithm was applied to Landsat 2 Collection 2 Products R_{rs} data between 1985 to 2022 to retrieve the time-series in the regions of interest described in Section 2.1. From the results of the Z_{sd} time-series, the magnitude of the trend was estimated by Thein-Sen's slope whereas the significance of this trend being estimated by the Mann-Kendall trend test.

3. RESULTS

The Secchi Disk depth algorithms performed well for the Amazon Region. As expected for both, TM and ETM+ data presented similar results due to the high similarity of their Spectral Response Functions (Table 1) .MDN showed errors of about 33% for TM and 34% for ETM+ with a BIAS of 2.64% for TM and 1.59 for ETM+, despite the high variability (e.g., the standard deviation of BIAS was 20.43% for TM and 19.03 for ETM+). Again, the results for OLI also follow the same pattern, with better outcomes for MDN. In

addition, improved results were observed for OLI in relation to ETM+ and TM due to the inclusion of the coastal bands.

Satellite	Model	BIAS(%)	E(%)	Slope
Landsat-5/TM	Random Forest	16.5 ± 26.31	38.01 ± 18.01	0.66 ± 0.13
	SVM	10.63 ± 18.11	35.31 ± 11.52	**0.72 ± 0.21**
	MDN	**2.64 ± 20.43**	**33.62 ± 16.32**	0.71 ± 0.22
Landsat-7/ETM+	Random Forest	17.13 ± 25.11	37.12 ± 16.66	0.66 ± 0.15
	SVM	10.34 ± 17.95	35.94 ± 12.27	**0.71 ± 0.22**
	MDN	**1.59 ± 19.03**	**34.24 ± 16.54**	0.7 ± 0.24
Landsat-8/OLI	Random Forest	6.08 ± 22.57	31.82 ± 11.39	0.73 ± 0.1
	SVM	**3.47 ± 17.91**	33.78 ± 13.86	0.79 ± 0.27
	MDN	-6.78 ± 15.51	**28.88 ± 7.67**	**0.81 ± 0.1**

Table 1. Monte Carlo results for each algorithm and satellite.

After the calibration of the algorithms with *in-situ* data, the MDN model was applied to Landsat data to validate the proposed algorithm (Figure 2). A total of 236 matchups with +- 5 days were obtained. The error was of 37.46%, with a BIAS of -22.45, indicating a slight underestimation of the algorithm.

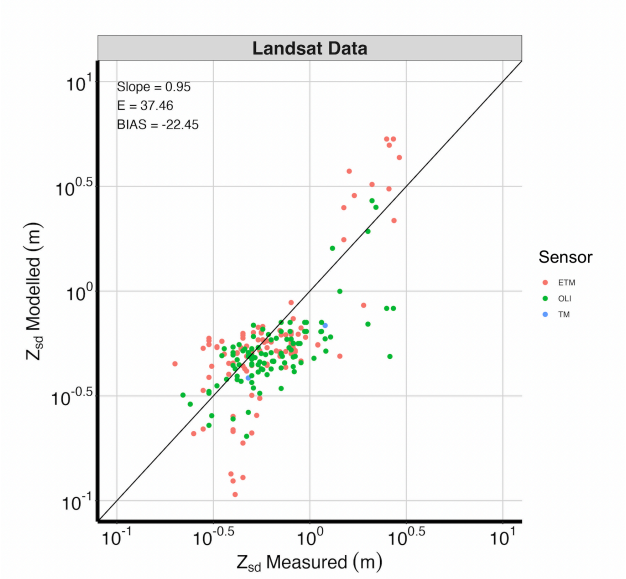

Figure 2. Model Validation with satellite R_{rs} data. Colors refer to each satellite.

The Zsd value trends obtained by MDN model (from 1985 to 2022) for each water body are shown in Table 1. Table 2 provides the Sen's Slope and P-Value results from the Mann-Kendall T-test for each location. From the results, it is possible to observe three different patterns in these locations: the first is an increase in Z_{sd} values over time for the white-water rivers of Madeira and Amazonas. In addition, Monte Alegre Lake, a sediment-dominated location with a high contribution from the Amazon River, also underwent a significant increase of the Z_{sd} values over time. For Negro and Solimões Rivers, a negative slope was observed, although a non-significant p-value indicates that it is not possible to attribute a trend in increase/decrease of Z_{sd}. Finally, for the Curuai lake and Tapajós River, a positive slope is observed, whereas it is also non-significant according to the Mann-Kendall test.

Location	Sen's Slope	P-Value
Negro River	-0.00073	> 0.05
Solimões River	-0.00028	> 0.05
Madeira River	0.00069	< 0.05 *
Amazon River	0.00112	< 0.05 *
Tapajós River	0.00012	> 0.05
Curuai	0.00007	> 0.05
Monte Alegre	0.00013	< 0.05 *

Table 2. Mann-Kendal and Sen's Slope results for selected locations.

4. DISCUSSION

Evaluating water clarity in inland, coastal, and oceanic waters is one of the most important and useful water quality measurements. It is easy to measure and provides a good description of the whole underwater light field. In the Amazon Region, changes in Z_{sd} are primarily associated with the floodpulse, which governs how the OACs will varies along the hydrological year. Therefore, changes in Z_{sd} could be a good proxy for understanding time changes in the OACs

In this study, we developed a Machine Learning algorithm based on Mixture Density Networks for predicting Secchi Disk Depth in Amazon Basin for long-term Landsat data. Errors of approximately 33% were obtained, indicating that MDN can produce accurate results. After the algorithm calibration with *in-situ* data, we applied the algorithm to Landsat-derived R_{rs}, considering matchups with +- 5 days. From this validation, we observed that the results were close to the obtained by *in-situ* data. It indicates that the atmospheric correction provided by the USGS in the Landsat Collection 2 Level 2 products, aligned with the MDN, well predicts the Z_{sd} in Amazon Region, in a range of 0 – 3.15m.

The calibrated algorithm was then applied to satellite a time-series of R_{rs} extracted from the Landsat Collection 2 Level products. This allowed us to construct a time series of predicted Z_{sd} and use statistical analysis to examine the trends in Secchi Disk Depth. From those results, we observed that for the white-water rivers of Madeira and Amazonas, Z_{sd} presented a positive trend. These results agree with recent studies in the Amazon region: [11] observed a decrease in sediment concentration trends in the Óbidos Gauge station. In addition, [12] observed that, for the Madeira River, there is also a trend for decreasing sediment concentration over time. These studies relate this decreasing trend to dam

construction. These dams are expected to trap the sediment and reduce the sediment transport downstream of the river. For example, [11] pointed out that there are 98 dams constructed or in construction upstream Óbidos station, which could also increase this decrease trend. In that region, the sediment concentration, the main regulator of transparency, is of extreme importance to the biodiversity of floodplain lakes and flooded forest. This sediment comes from the Andean mountains and is rich in nutrients, such as phosphorous, flourishing the downstream floodplains and being one of the important factors of high biodiversity in Amazon Lakes and flooded forest [13].

5. CONCLUSIONS

In this study, we evaluated satellite remote sensing algorithms to predict water transparency in Amazon Basin. From a Machine Learning algorithm, we estimated Secchi Disk depth (Z_{sd}) for this region using long-term Landsat data. In addition, we also evaluated the time series of Z_{sd} to understand trends of this variable along time. From the assessed algorithms, the best results were from the Mixture Density Networks, with errors of approximately 33%, although the other evaluated (Random Forest and Support Vector Machine) presented good performance. With the best-performed algorithm, we applied it to satellite data using Landsat Collection 2 Level 2 products to perform validation. This validation achieved proximal results observed by the *in-situ* algorithm (errors of ~33%). After that, MDN was applied to a time series of R_{rs} to obtain Z_{sd} trends in Amazon Region. Three of the seven locations evaluated presented a significant and positive trend in Z_{sd} values (Rio Amazonas, Rio Madeira, and Monte Alegre Lake), indicating an increase in water transparency. On the other hand, for the other evaluated locations, non-significant p-values were obtained. These results could be useful to understand, on a large spatiotemporal scale, the impacts of land use and climate change, dam constructions, mining, and other anthropic threats to the Amazonian environment.

8. REFERENCES

[1] Fassoni-Andrade, et al. Amazon Hydrology from Space: Scientific Advances and Future Challenges. Reviews of Geophysics 2021, 1–97, doi:10.1029/2020rg000728.

[2] Tundisi, J.G.; Goldemberg, J.; Matsumura-Tundisi, T.; Saraiva, A.C.F. How Many More Dams in the Amazon. Energy Policy 2014, 74, 703–708, doi:10.1016/j.enpol.2014.07.013.

[3] Marengo, J.A et al.. Changes in Climate and Land Use Over the Amazon Region: Current and Future Variability and Trends. Front Earth Sci (Lausanne) 2018, doi:10.3389/feart.2018.00228.

[4] Kraus, C.N., et al. Unraveling Flooding Dynamics and Nutrients' Controls upon Phytoplankton Functional Dynamics in Amazonian Floodplain Lakes. Water (Switzerland) 2019, 11, doi:10.3390/w11010154.

[5] Pitarch, J. A Review of Secchi's Contribution to Marine Optics and the Foundation of Secchi Disk Science. Oceanography 2020, 33, 26–37, doi:10.5670/OCEANOG.2020.301.

[6] Pahlevan, N., et al. Simultaneous Retrieval of Selected Optical Water Quality Indicators from Landsat-8, Sentinel-2, and Sentinel-3. Remote Sens Environ 2022, 270, 112860, doi:10.1016/J.RSE.2021.112860.

[7] Maciel, D.A.; et al. Water Clarity in Brazilian Water Assessed Using Sentinel-2 and Machine Learning Methods. ISPRS Journal of Photogrammetry and Remote Sensing 2021, 182, 134–152, doi:10.1016/j.isprsjprs.2021.10.009.

[8] Sioli, H. Alguns Resultados e Problemas Da Limnologia Amazônica. Boletim Técnico IPEAM, v. 24 1951, 3–44.

[9] Barbosa, C.C.F, et al. Remote Sensing of the Water Properties of the Amazon Floodplain Lakes: The Time Delay Effects between *in-situ* and Satellite Data Acquisition on Model Accuracy. In Proceedings of the International Symposium on Remote Sensing of Environment: Sustaining the Millennium Development Goals; 2009; Vol. 33, pp. 1–4.

[10] Martins, V.S., et al. Assessment of Atmospheric Correction Methods for Sentinel-2 MSI Images Applied to Amazon Floodplain Lakes. Remote Sens (Basel) 2017, 9, doi:10.3390/rs9040322.

[11] Montanher, O.C. et al. Temporal Trend of the Suspended Sediment Transport of the Amazon River (1984 – 2016). Hydrological Sciences Journal 2018, 63, 1901–1912, doi:10.1080/02626667.2018.1546387.

[12] Li, T. et al. Reversal of the Sediment Load Increase in the Amazon Basin Influenced by Divergent Trends of Sediment Transport from the Solimões and Madeira Rivers. Catena (Amst) 2020, 195, 104804, doi:10.1016/j.catena.2020.104804.

[13] Almeida, R, et al. Phosphorus Transport by the Largest Amazon Tributary (Madeira River, Brazil) and Its Sensitivity to Precipitation and Damming. Inland Waters 2015, 5, 275–282, doi:10.5268/IW-5.3.815.