MAPPING THE HYDROLOGICAL CONNECTIVITY OF FLOODPLAIN LAKES USING PLANET MOSAICS: AN EXPLORATORY ANALYSIS

Rejane Paulino[1], Mikhail Urbazaev[2], Antje Uhde[3], Daniel Maciel[4], Vitor Martins[5], Laura Hess[6], Marie Bonnet[7], Evlyn Novo[8]

[1,4,8]Instrumentation Laboratory for Aquatic Systems (LabISA), INPE, São José dos Campos 12227-010, SP, Brazil. E-mail: (rejane.paulino; daniel.maciel; evlyn.novo)@inpe.br, [2]University of Maryland, MD 20742, USA. E-mail: urbazaev@umd.edu, [2,3]Friedrich Schiller University Jena, 07743 Jena, Germany. E-mail: antje.uhde@uni-jena.de, [5]Mississippi State University, MS 39762, USA. E-mail: vmartins@abe.msstate.edu, [6]University of California, CA 93106, USA. E-mail: lauralhess@ucsb.edu, [7]UMR Espace-DEV, Institut de Recherche pour le Développement (IRD), 34000 Montpellier, France. E-mail: paule.bonnet@ird.fr.

ABSTRACT

Hydrological Connectivity (HC) of small floodplain lakes is rarely investigated using remote sensing techniques. This paper explores the application of planet mosaics to estimate the HC between the middle Juruá River and its floodplain lakes. Remote Sensing features from 41 floodplain lakes and related main river reaches were provided by a time series (2020-2022) of Planet mosaics. These features and river water level were input to a machine learning algorithm. The trained HC-model had a total accuracy of 82% but showed a trend to overestimate connected lakes. Model performance was compromised by planet mosaic spectral information and the quality of the connectivity data used for the training. Finally, this paper encourages new applications related to the HC of floodplain lakes using water color data as a proxy.

Key words — Connectivity, Water Color, Satellite Imagery, Machine Learning, Juruá River.

1. INTRODUCTION

Hydrological Connectivity refers to the transfer of matter, energy, and organisms within or among the water systems [1]. The HC information is essential to protect the species biodiversity (e.g., fishes and plants) and to manage water sources [2]. Conventionally, hydrological connectivity quantification needs several climatic and environmental parameters based on *in-situ* measurements [3], which prevents its application in large and remote areas such as Brazilian Amazon floodplains. When floodplain lakes are connected to their main river, permanent surface water sources connect these two systems from small channels [4]. At this site, the lakes commonly present optical water type [5] similar to that of the river water, suggesting that the water color could be applied to estimate the hydrological connectivity of these lakes.

This paper investigated the feasibility of estimating the hydrological connectivity of floodplain lakes using water color as a proxy. The challenge was to identify when and for how long small lakes located in Juruá River floodplain were connected to the Juruá river using monthly Planet mosaics to measure water color similarity between them. This information was extracted from 2020 to 2022 and combined with *in-situ* data for training a machine learning model.

2. MATERIAL AND METHODS

2.1. Study Area

The study area encompasses small floodplain lakes (N=41) from the Middle Juruá River in the Amazon Region (Figure 1). The Juruá is a large white-water river with a flood wave increasing about 12-15 m between low-water in September-October to high water in March. The lakes are inserted in well preserved areas with a dense flooded forest cover around them. These systems have their bio-optical composition influenced by the Juruá River flood pulse [6], which favors the application of water color for estimating the HC.

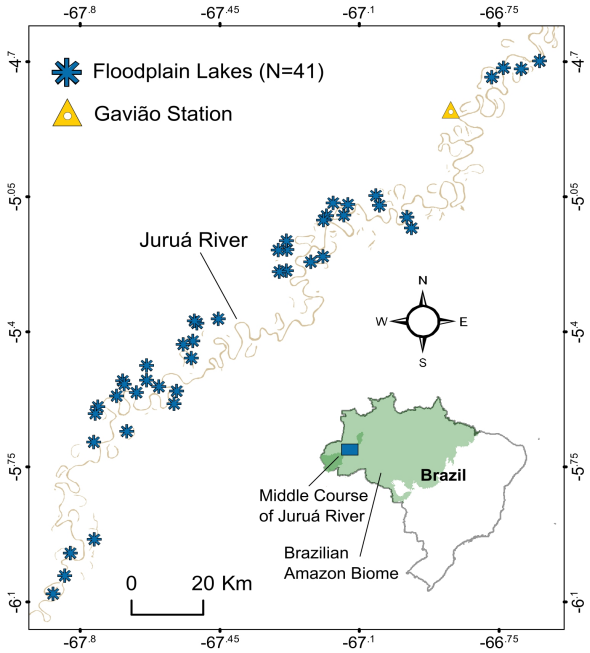


Figure 1. Study Area.

2.1. Dataset

A total of 323 scenes of the NICFI PlanetScope Monthly Mosaics were used from 2020 (September) to 2022 (March). These mosaics are composites of the best pixels observed along one month. In this selection, pixels with cloud or degraded by any spurious effects are disregarded [7]. The mosaics are freely available for the tropics in surface reflectance level, for 4 spectral bands (blue, green, red, and nir) and with high spatial resolution (around 5 meters). The data can be downloaded from https://www.planet.com/nicfi/. In this research, the NDWI index [8] was used to create a water mask (threshold based on zero value).

Hydrological connectivity data were empirically collected *in-situ*. Floodplain residents were interviewed about the monthly connectivity of 41 floodplain lakes located near the Juruá River considering the 2020 year (Figure 1). Note that the *in-situ* connectivity data were not acquired concurrent with the Planet mosaics. In this case, we considered that the lake connectivity pattern observed *in-situ* in 2020 was similar in the following years (2021 and 2022). Moreover, historical average of water level data at the Juruá River (from June 1997 to December 2016) recovered from the Gavião station (Figure 1) was also used.

2.2. Hydrological Connectivity Estimation

The HC-model was based on the machine learning *Random Forest* classifier. The model development consisted of two steps: (a) feature selection, and (b) training and validation of HC-model (Figure 2).

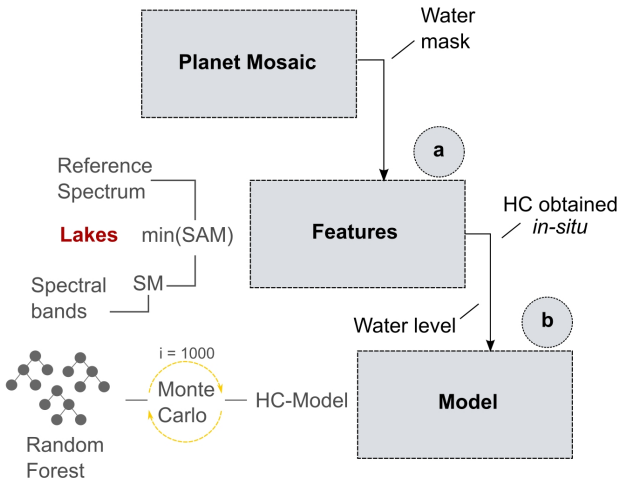


Figure 2. Flowchart of the process.

2.2.1 Feature Selection

The HC-model had seven spectral features: Planet mosaic surface reflectance (four spectral bands), two spectral similarity measures (Spectral Angle Mapper-SAM and Spectral Magnitude-SM), and the monthly average of water level at the Juruá River. The SAM and SM compute the shape and area differences below the reflectance spectra from the floodplain lakes and their main river, respectively. Samples (N=50) were collected on the scenes along the Juruá River to composite the average reference spectrum needed to estimate the SAM [9]. The minimum SAM value for each lake was the criterion used to select the geographical coordinate for which the remaining spectral features were obtained.

2.2.2 HC-model structure

The features (Section 2.2.1) are the input to the HC-model. The outputs are two classes: connected and non-connected. *In-situ* hydrological connectivity (Section 2.1) and 7 features were used to train and validate our model. The dataset (N=359) was iteratively split into training (70%) and validation (30%) using a Monte Carlo simulation (1000 iterations). For the training, the Random Under Sample process was applied to increase the minority class representation (class non-connected) in our dataset. This process balances the sample number per class by randomly selecting a data subset of similar size (majority class) to that of the minority class. A total of 1000 models were generated from the iterations, and we selected the median HC-model based on the accuracy metric value resultant from the validation process.

3. RESULTS AND DISCUSSION

3.1. Distribution of the dataset

The HC samples employed in the approach were unbalanced in the temporal dimension (Figure 3). The high-water season had 12.5% more samples than for the low-water season. Months such as March and December alone occupied about 24% of the total dataset. Otherwise, some falling-stage months (e.g., June) were represented by only 3%. This is partially explained by the fact that the Planet Mosaics inserted a larger number of scenes acquired in the rainy season. Moreover, cloud cover and factors that prevent the spectral water response retrieval (e.g., adjacency effects) limit the continuous observation of small water bodies in satellite-imagery [10]. This difference among the number of samples along the months can influence the HC-model response, since the floodplain lake hydrological connectivity is well-correlated with the basin hydrology features such as the rainfall and the river stage [11]. Consequently, the available sample number per class was also affected. The 'connected' class was better represented in the dataset. Its occurrence was up to twice as high as that of the non-connected class. Overall, the dataset distribution made it difficult to obtain a large and representative sample number for both classes, indicating, in this case, the need of balancing techniques during the training stage of the machine learning algorithms, as suggested in Section 2.2.2.

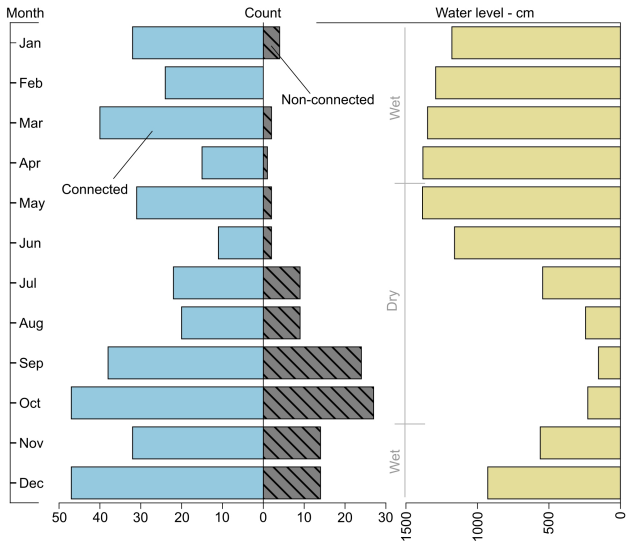


Figure 3. Dataset (*in-situ* + image-data) (N=359) split by class.

3.2. Performance of the HC-model

The HC-model performed satisfactorily as indicated by the statistical metrics computed with the validation set (total accuracy of 82%) (Figure 4). HC-model slightly overestimated the 'connected' class (recall: 89%), suggesting that it is difficult to estimate non-connected lakes' occurrence (recall: 66%). The overestimation may be an artifact of the lack of a broader range of spectra representing the 'non-connected' class in the training set. Furthermore, the disagreement between the spectral response collected from Planet mosaics and the *in-situ* hydrological connectivity may have contributed to the high level of confusion of the HC-model in the 'non-connected' class. Note that although rainfall and river stage vary from year to year, this study assumed that lake river connectivity was similar from one year to the next (Section 2.2.2).

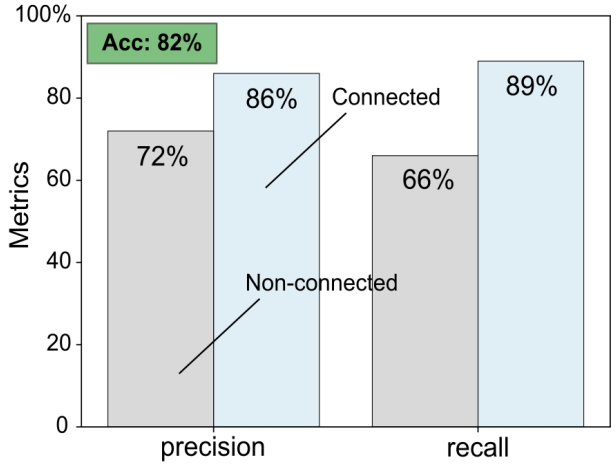


Figure 4. Validation metrics of the HC-model.

Though machine learning algorithms are of a 'black box' nature, the HC-model was based on the hydrological connectivity of small floodplain lakes and the water color and water level features. The rank analysis of the 7 features used as input helps to interpret the model outputs and the features' contributions to the HC prediction. This analysis showed that river stage level explained the HC better than the water color features (Figure 5). The importance of the water level was much larger in comparison to the other features. However, the spectral similarity measures (SAM and SM) and the red spectral band also showed ability to explain the hydrological connectivity in floodplain lakes. The red-band reflectance is highly related to the suspended sediment load in water [12] and the exchange of bio-optical water components between the Juruá River and its floodplain lakes during the wet season may have led to reasonable importance of the water color features for the model.

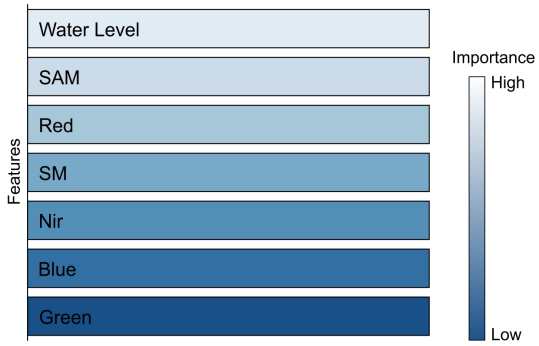


Figure 5. Selection of the feature importance.

3.2. Hydrological Connectivity of floodplain lakes

The HC model was used to estimate the monthly hydrological connectivity in 41 floodplain lakes studied for 2021 (Figure 6). Lakes observed less than six months per year had on average a three-month connection to the Juruá River (max: 6 and min: 1). On the other hand, lakes with a more significant number of observations (more than six months) showed average connectivity of seven months (max: 12 and min: 2). In some cases, there were lakes always connected to the Juruá River, especially for those lakes closer to the river channel. The month of May had 100% of lakes connected to the river (31 of 31 lakes observed), while September had only 34% (12 of 35 lakes observed) (Figure 6c). This difference can be attributed to the dynamics of river stage level during the hydrological year. The floodplain lakes are connected to the Juruá River while the water level rises and are isolated while the water level is lower, as observed in Figure 3. Unfortunately, the small number of observations prevented a broader analysis of several lakes. Otherwise, with an even monthly distribution of images, the HC model could have provided a better estimate of the connectivity pattern of the Juruá River floodplain.

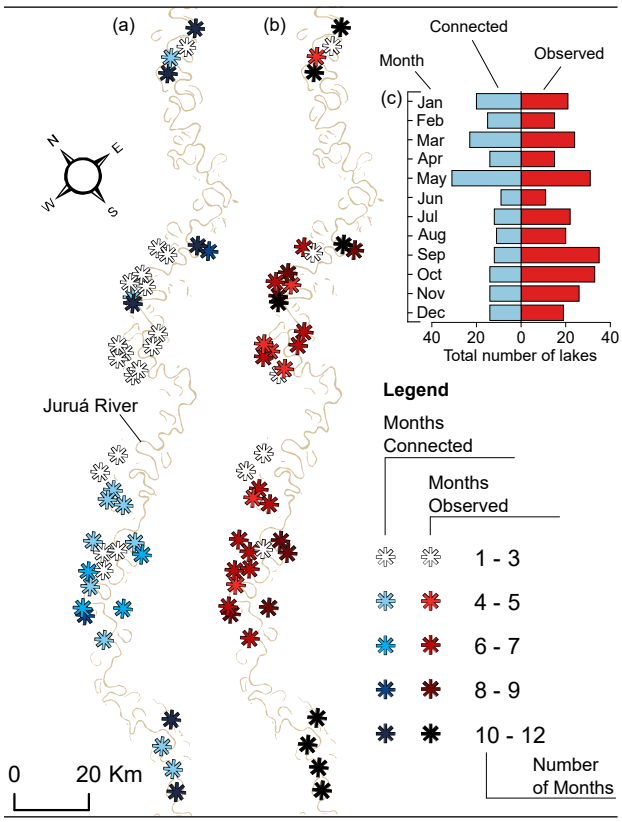


Figure 6. Evaluation of HC retrieved along the 2021 year.

3.3. Thoughts regarding developing HC products based on Planet mosaics

In this study, the Planet mosaics showed two limitations: (i) inaccurate spectral response and (ii) disagreement with *in-situ* data. The normalization process carried out on Planet mosaics can compromise the original spectral response of lake and river waters. Moreover, the limited spectral band number prevented a good estimate of the similarity features since they are based on the shape and intensity of aquatic system spectra so as to measure the difference between the lake and river water color. Noise from atmospheric and adjacency effects, and cirrus cloud cover also were observed on scenes. In our approach, these factors reduced the number of observations over several lakes, undermining the theoretical advantage provided by the Planet mosaics. Another important factor was the disagreement between the Planet mosaics and the HC data collected *in-situ*. The floodplain lakes are dynamic systems and dramatically sensitive to seasonal variations. Thus, our generalization of HC *in-situ* information for other years can not represent a true ground-truth.

4. CONCLUSIONS

This paper was the first attempt to apply machine leaning techniques and use water color retrieved from Planet monthly mosaics to estimate the hydrological connectivity in small floodplain lakes in the Amazon region. The results have demonstrated that the accurate modelling of HC based on water color requires more accurate water spectra than those provided by the Planet mosaics. In this sense, we have developed a similar approach from 3 m PlanetScope images, with better spectral response, and a large dataset based on visual interpretation of satellite imagery. Therewith, we expect to generate a model more accurate for predicting the hydrological connectivity in future applications.

5. ACKNOWLEDGMENTS

This research is funded through the 2017–2018 Belmont Forum and BiodivERsA joint call for research proposals, under the BiodivScen ERA-Net COFUND program, and with the funding organizations French National Research Agency (ANR), São Paulo Research Foundation (FAPESP-Grant 2018/12083-1), National Science Foundation (NSF), the Research Council of Norway and the German Federal Ministry of Education and Research (BMBF). The study was also funded by the SABERES “Sustaining Amazon floodplain biodiversity and fisheries under climate change” funded by the BNP Paribas foundation within its call Climate & Biodiversity Initiative 2019. Authors are in debt to Carolina Tavares de Freitas for help with *in-situ* measurements in Juruá River floodplain lakes.

6. REFERENCES

[1] Pringle C.M. Hydrologic connectivity and the management of biological reserves: a global perspective. *Ecol. Appl.*, v. 11, pp. 981–998, 2001.
[2] Saha, S. et al. Hydrological connectivity, surface water quality and distribution of fish species within sub-locations of an urban oxbow lake, East India. *Watershed Ecology and the Environment*, v. 4, pp. 44-58, 2022.
[3] Zhang, Y. et al. The concept, approach, and future research of hydrological connectivity and its assessment at multiscales. *Environmental Science and Pollution Research*, v. 28, pp. 52724-52743, 2021.
[4] Camporeale, C. et al. Significance of cutoff in meandering river dynamics. *Journal of Geophysical Research*, v. 113, 2008.
[5] da Silva, E.F.F. et al. Optical water types found in Brazilian waters. *Limnology*, v. 22, pp. 57–68, 2021.
[6] Affonso, A. G. et al. Abiotic variability among different aquatic systems of the central Amazon floodplain during drought and flood events. *Braz. J. Biol.* v. 75, pp. 60–69, 2015.
[7] Planet Team. NICFI DATA Program User Guide. Available: <https://assets.planet.com/docs/NICFI_UserGuidesFAQ.pdf>. Access in: 8 Oct 2022.
[8] McFeeters, S. K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Inter, Jour. of Rem. Sens.*, v.17, n.7, p.1425-1432, 1996.
[9] Kruse, F. A. et al. The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, v. 44, pp. 145–163, 1993.
[10] Paulino, R. S. et al. Assessment of Adjacency Correction over Inland Waters Using Sentinel-2 MSI Images. *Remote Sens.*, v. *14*, pp. 1829, 2022.
[11] Liu, X. et al. Satellite image-based investigation of the seasonal variations in the hydrological connectivity of a large floodplain (Poyang Lake, China). *Journal of Hydrology*, v. 585, pp. 24810, 2020.
[12] Fassoni-Andrade, A. C. et al. Mapping spatial-temporal sediment dynamics of river-floodplains in the Amazon. *Remote Sensing of Environment*, v. 221, pp. 94-107, 2019.