Satellite-Based Land Cover Classification in the Itajaí River Basin: Methods and Analysis

Guilherme Francisco Zucatelli [1,2], Jorge Antonio Silva Centeno [1], Carlos Antonio Oliveira Vieira [2]

[1] Universidade Federal do Paraná, UFPR – Curitiba, Brazil – guilherme.zucatelli@ufpr.br, centeno@ufpr.br
[2] Universidade Federal de Santa Catarina, UFSC – Florianópolis, Brazil – carlos.vieira@ufsc.br

Keywords: Land Cover Classification, Remote Sensing, Itajaí River Basin, CBERS-4A

Abstract

The cities within the Itajaí River Basin in Brazil have been experiencing continuous flooding in recent years, causing significant material damage to the population and public services. This study aims to present land use and classification techniques based on image analysis obtained from the CBERS-4A satellite, using the WFI sensor. The image was cropped to include only the Itajaí River basin. Four classification methods were employed: Maximum Likelihood (Maxver), Minimum Distance, Spectral Angle, and Random Forest. Classified maps were generated for each algorithm, and the results of each land cover class area were analyzed. Validation was performed using sample areas of each class in the original image, presenting the area of each class obtained by each algorithm and the confidence level for each.

1. Introduction

The cities within the Itajaí River Basin, Brazil, have been experiencing continuous flooding in recent years. As a consequence of this process, in 2023, the city of Rio do Sul faced seven major floods (Bertoli, 2023), causing material damage to the population and public services.

The dynamics of surface water flow are directly influenced by land use and occupation (FERREIRA, 2017). Thus, mathematical models for river flow depend on knowledge of the area and type of land cover in the hydrographic basin under study. Remote sensing emerges as a crucial tool for this research, as image processing allows for precise classification.

This study will present techniques for land use and classification based on image analysis obtained from the CBERS-4A satellite, using the WFI sensor. The image will be cropped to include only the Itajaí River basin. Four classification methods will be employed: Maxver, Minimum Distance, Spectral Angle, and Random Forest. Maps containing the classified classes will be generated, and the results of each land cover class area will be analyzed. Finally, maps will be generated for the sub-basins of the Itajaí River to identify which has the lowest vegetation index based on the data obtained.

2. Study Area

Located within the coordinates of 26° 27' to 27° 53' south latitude and 48° 38' to 50° 29' west longitude, the study area (figure 1) is demarcated by the physiographic features of the Serra Geral and Serra dos Espigões to the west, the Serras da Boa Vista, Faxinais, and Tijucas to the south, and the Serras da Moema and Jaraguá to the north, with the Atlantic Ocean bordering the east.

Encompassing an approximate area of 15,000 km², constituting 16.15% of Santa Catarina's State territory and 0.6% of Brazil's landmass, this region stands as the largest hydrographic system along the Atlantic slope in Santa Catarina. The Itajaí-Açu River extends for about 200 km from its primary source to its estuary, presenting a drainage density of 1.61 km/km². Its mean discharge is recorded at 205.0 m³/s, with minimum and maximum flow rates of 50.0 m³/s and 1,120 m³/s, respectively (Schettini, 2002).

The trajectory of the Itajaí-Açu River can be segmented into three discernible sections. The Upper Itajaí-Açu, marked by the confluence of the Itajaí do Sul and Itajaí do Oeste Rivers, exhibits a sinuous, gently sloping course spanning roughly 26 km. This micro-region features an Atlantic slope hydrological network, with the Itajaí-Açu serving as the primary river, supplemented by the Itajaí-Mirim, Benedito, and dos Cedros. The urban expanse of Blumenau is notably influenced by tributaries such as Garcia, da Velha, Itoupava, Fortaleza, and Testo (SEPLAN/Florianópolis-SC).

Over the span of 150 years, the region has witnessed 67 floods, leading to substantial losses in crops, livestock, residential structures, and industrial assets. This historical pattern has prompted a deep reflection on the intricate interplay between human activity and natural phenomena, evolving from an initial reactive stance to the recognition of the imperative for proactive measures that foster a balanced coexistence with the Itajaí-Açu River.

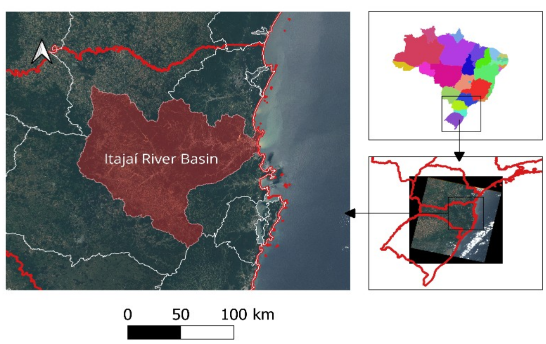

Figure 1. Location of the Itajaí River Basin and the area covered by an image from the CBERS-4A WFI sensor.

The Itajaí River Basin is characterized by a diverse land use pattern. The Upper Itajaí Valley is notable for its agricultural activities, including the cultivation of onions and tobacco, which are significant crops in the region. Small towns predominate throughout the basin, providing a rural charm while supporting agricultural economies. Additionally, forestry operations are widespread across the basin, contributing to both local economies and the regional landscape. The Itajaí National Park, a forest reserve within the basin, plays a crucial role in preserving the native Atlantic Forest and maintaining ecological balance.

Given the importance of the Itajaí River Basin's land use dynamics, remote sensing emerges as a vital tool for monitoring land use and occupation. Satellite imagery allows for the continuous observation of land cover changes, which is essential for understanding the interplay between land use and water flow dynamics. This capability is particularly crucial for regions like the Itajaí River Basin, where the balance between urbanization, agriculture, and conservation is delicate and directly impacts the hydrological behavior of the river system. The use of remote sensing in land cover classification helps inform better management practices and policy decisions, ultimately contributing to the sustainable development and resilience of the basin.

3. Data and Preprocessing

The satellite image was obtained from the website of the Brazilian National Institute for Space Research (Instituto Nacional de Pesquisas Espaciais, n.d.) through their image catalog portal. The selected sensor was the Wide Field Imager (WFI) of the CBERS-4A satellite, which has the capability to cover the entire study area in a single image. The CBERS (China-Brazil Earth Resources Satellite) program is a cooperative effort between China and Brazil, aimed at developing and operating remote sensing satellites for Earth observation. The CBERS-4A, launched in December 2019, is equipped with several sensors, including the WFI, which is designed to capture large-scale images with a spatial resolution of 64 meters and a swath width of 866 kilometers. The WFI sensor includes four spectral bands: blue (0.45 - 0.52 µm), green (0.52 - 0.59 µm), red (0.63 - 0.69 µm), and near-infrared (0.77 - 0.89 µm).

The image selected for this study is dated November 5, 2023, with a cloud cover percentage of 20%. According to the image preview available on the INPE website, the clouds were primarily located over the ocean, ensuring minimal interference with the study area. The choice of this image date and quality was critical to ensure that the analysis could be performed with the highest possible accuracy.

For image processing, QGIS 3.28 software was used, along with the "Semi-Automatic Classification Plugin" (Congedo, 2021) and the "Orfeo Toolbox" (Grizonnet et al., 2017). These tools facilitated the supervised classification process. A spectral signature training shapefile was created to define the land cover classes: urban area, water, dense vegetation, low vegetation, and exposed soil. This shapefile is essential for all the classification methods employed in this study.

3.1 Definition of Classes

Supervised classification is a common method in remote sensing where the user defines training areas for each land cover class. These areas are used to "train" the classification algorithm to recognize the spectral signatures associated with each class. A shapefile containing polygons of sample areas for spectral signatures was created, divided into five classes: urban area, water, dense vegetation, low vegetation, and exposed soil. For the urban area, the center of São José city was selected due to its highly dense region. Dense vegetation areas were chosen within the Tabuleiro Reserve Park and the Itajaí National Reserve Park. Low vegetation areas were selected from pastures and horticultural crops. Exposed soil included coastal dune areas and plowed land. Water samples were selected from maritime parts and dam lakes.

3.2 Classification Algorithms

The classification of land cover was performed using four distinct algorithms: Spectral Angle, Maximum Likelihood (Maxver), Minimum Distance, and Random Forest. These methods were chosen for their varied approaches to handling spectral data, offering a comprehensive evaluation of classification techniques suitable for remote sensing applications.

Spectral Angle Mapper classification considers the angular geometry between the spectral vectors of image pixels and the spectral signature vectors of training classes. It calculates the angle between these vectors, with the assigned class being the one with the smallest angle, indicating the greatest spectral similarity. This method is particularly effective in reducing the impact of illumination differences (Rashmi et al., 2014).

Maximum Likelihood Classification, or Maxver, is a statistical approach that assigns a class to each pixel by maximizing the statistical likelihood based on training data. This method assumes that the statistics for each class in each band are normally distributed and calculates the probability of a pixel belonging to each class (Richards, 2022). The pixel is then assigned to the class with the highest probability.

Minimum Distance classification relies on the spectral distance between image pixels and predefined training pixels for each class. It assigns each pixel to the class with the shortest spectral distance to the class's spectral signatures. This method is straightforward and computationally efficient, making it suitable for large datasets (Kruse et al., 1993).

Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training and outputting the mode of the classes (classification) or mean prediction (regression) of the individual trees. It is known for its robustness and accuracy, particularly in handling high-dimensional data and various types of variables (Breiman, 2020). This method is effective in dealing with complex interactions and variability in the dataset.

After generating the raster files with classified land cover classes for the entire satellite image, a subset of the image focusing exclusively on the Itajaí River Basin was created. This subset was created using a shapefile of the Itajaí River Basin, obtained from the Agricultural Research and Rural Extension Company of Santa Catarina (EPAGRI, n.d.)) map library portal. The raster calculator in QGIS was then employed to determine the area of each land cover class within the basin. This approach ensured that the analysis was specific to the study area, providing accurate and relevant land cover data for the Itajaí River Basin.

3.3 Validation of Results

After the classification process, it is essential to assess the accuracy of land cover classification to identify and measure map errors. Accuracy assessment is performed with the calculation of an error matrix, which is a table that compares map information with reference data for a number of sample areas (Congalton, 2001). The error matrix compares the identified land cover classes with known reference classes. The items in the major diagonal are the number of samples correctly identified, while the other items represent classification errors.

The overall accuracy (OA) is calculated as the ratio between the number of correctly classified samples (the sum of the major diagonal) and the total number of sample units. The producer’s accuracy (PA) for each class is calculated as the ratio between correct samples and the column total, while the user’s accuracy (UA) is the ratio between correct samples and the row total.

An area-based error matrix is recommended (Olofsson et al., 2014), where each element represents the estimated area proportion of each class. This allows for estimating the unbiased user’s accuracy (UAA) and producer’s accuracy (PAA), the unbiased area (UA) of classes according to reference data, and the standard error (SE) of area estimates.

The formula for the overall accuracy (OA) is:

$$OA = \frac{\sum a_{ii}}{n} \tag{1}$$

where $\sum a_{ii}$ = is the sum of the diagonal elements (correctly classified samples)
$\boldsymbol{n}$ = the total number of samples.

The formula for the user’s accuracy (UA) for each class is:

$$UA = \frac{a_{ii}}{\sum a_{ij}} \tag{2}$$

where a_{ii} = is the number of correctly classified samples for a particular class
$\sum a_{ij}$ = the total number of samples classified as that class.

The formula for the producer’s accuracy (PA) for each class is:

$$PA = \frac{a_{ii}}{\sum a_{ji}} \tag{3}$$

where a_{ii} = is the number of correctly classified samples for a particular class
$\sum a_{ji}$ = the total number of samples that actually belong to that class.

The area-based error matrix provides a more comprehensive assessment of classification accuracy by incorporating spatial information, ensuring a more accurate representation of land cover distribution. This method is particularly useful in large-scale studies where spatial accuracy is critical.

To compare the four classification algorithms it is important to note that the spectral signatures used for classification are consistent across all methods. This consistency allows for a fair comparison of the algorithms' performance. By using the same spectral signatures, any differences in classification accuracy can be attributed to the inherent strengths and weaknesses of each algorithm rather than variations in the input data. This approach ensures that the comparison focuses on the algorithms' ability to accurately classify land cover types, providing valuable insights into their relative effectiveness for the study area.

4. Results

4.1 Classified Images

Classified images for each algorithm were generated to visually represent the distribution of land cover types in the Itajaí River Basin. These images illustrate significant differences in class delineation, highlighting the strengths and weaknesses of each classification method.

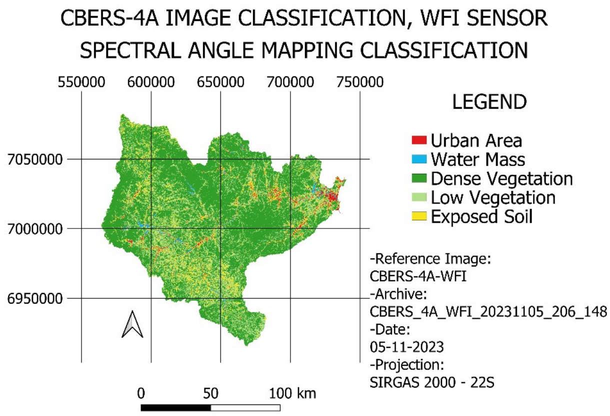

Figure 2. Resulting Image from Spectral Angle Mapping Classification

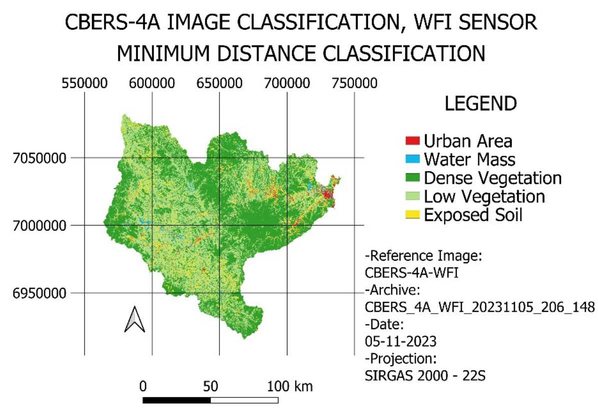

Figure 3. Minimum Distance Classification Resulting Image

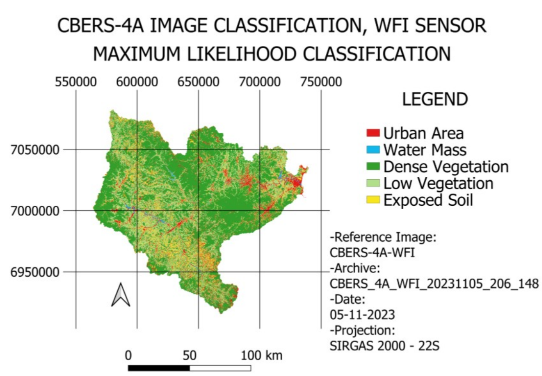

Figure 4. Maxver Classification Resulting Image

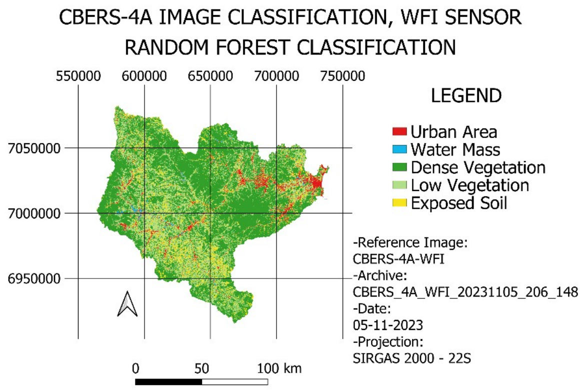

Figure 5. Random Forest Classification Resulting Image

4.2 Area Estimation

The results of satellite image classification for land cover in the Itajaí River Basin are detailed in Table 1, which presents the area values for each land cover class obtained from the four different classifiers in square kilometers.

Classifier	Urban Area	Water	Dense Veg.	Low Veg.	E. Soil
Spectral Angle	252	129	9759	3568	1219
Maximum Likelihood	1406	88	7966	3913	1557
Minimum Distance	196	143	7461	5914	1216
Random Forest	628	77	8528	4069	1624
Mean	621	109	8428	4366	1403
Standard Deviation	557	29	1076	1000	178

Table 1. Area of Land Cover Classes (in square kilometers) for Each Classification Algorithm

4.3 Classification Accuracy

The classification accuracy analysis is essential for evaluating the effectiveness of the applied algorithms. Table 2 presents the overall accuracy (OA) for each algorithm. Tables 3 to 6 provide the area-based error matrices for each algorithm. These tables detail the producer's accuracy (PA) and user's accuracy (UA) for each land cover class, offering a comprehensive understanding of each algorithm's performance.

Classification Algorithm	Overall Accuracy (OA) (%)
Random Forest	97.08
Minimum Distance	95.26
Maximum Likelihood (Maxver)	97.24
Spectral Angle	94.29

Table 2. Overall accuracy of each Classification algorithm

Class	1	2	3	4	5	Area (km²)
1	0.076	0.000	0	0.018	0.002	1355
2	0	0.029	0	0	0	405
3	0	0	0.676	0	0	9537
4	0	0	0	0.164	0	2317
5	0	0.009	0	0	0.026	496
Total	0.076	0.038	0.676	0.182	0.028	14054
Estimated area (km²)	1071	538	9537	2571	393	14054
SE (km²)	26	17	0	25	19	
PA (%)	100	75	100	90	94	
UA (%)	79	100	100	100	74	

Table 3. Area-Based Error Matrix - Random Forest

Class	1	2	3	4	5	Area (km²)
1	0.015	0	0	0	0.006	293
2	0.000	0.040	0	0	0	565
3	0	0	0.676	0.030	0	9921
4	0.000	0	0.001	0.151	0.008	2260
5	0.012	0	0	0	0.061	1013
Total	0.027	0.040	0.677	0.181	0.075	14054
Estimated area (km²)	384	559	9515	2544	1049	14054
SE (km²)	25	4	35	40	31	
PA (%)	55	100	100	84	81	
UA (%)	72	99	95	94	84	

Table 4. Area-Based Error Matrix - Spectral Angle

5. Discussion

The results from the land cover classification in the Itajaí River Basin reveal significant differences in the performance of the four algorithms: Random Forest, Minimum Distance, Maximum Likelihood (Maxver), and Spectral Angle. Each algorithm's ability to accurately classify the different land cover types varies, impacting the precision and reliability of the classification results.

The Random Forest algorithm demonstrated the highest overall accuracy (97.08%), indicating its robustness and effectiveness in handling complex datasets and variable interactions. This performance can be attributed to the ensemble approach of Random Forest, which combines multiple decision trees to improve classification accuracy. The algorithm effectively classified dense vegetation, low vegetation, and urban areas, making it suitable for detailed land cover studies in diverse regions like the Itajaí River Basin.

Class	1	2	3	4	5	Area (km²)
1	0.015	0	0	0	0.005	287
2	0.001	0.040	0	0	0	568
3	0	0	0.652	0.001	0.003	9213
4	0.000	0	0.026	0.180	0	2892
5	0.011	0	0	0.000	0.066	1092
Total	0.027	0.040	0.677	0.181	0.075	14054
Estimated area (km²)	384	559	9515	2544	1049	14054
SE (km²)	24	5	34	32	27	
PA (%)	56	100	96	99	88	
UA (%)	74	98	99	87	85	

Table 5. Area-Based Error Matrix - Minimum Distance

The Maximum Likelihood algorithm also showed high accuracy (97.24%), performing well in classifying dense vegetation and urban areas. Its statistical approach, which maximizes the likelihood of a pixel belonging to a particular class based on training data, is particularly effective in areas with distinct spectral signatures. However, it may be less effective in regions with mixed land cover types or where spectral signatures overlap.

Class	1	2	3	4	5	Area (km²)
1	0.026	0	0	0	0.0198	644
2	0	0.040	0	0	0.0011	574
3	0	0	0.676	0	0	9506
4	0	0	0	0.176	0	2484
5	0.002	0	0	0.004	0.054	844
Total	0.027	0.040	0.677	0.181	0.075	14054
Estimated area (km²)	384	560	9516	2545	1050	14054
SE (km²)	23	7	5	14	28	
PA (%)	94	100	100	98	72	
UA (%)	56	97	100	100	90	

Table 6. Area-Based Error Matrix - Maximum Likelihood

The Minimum Distance algorithm achieved an overall accuracy of 95.26%. This method, which classifies pixels based on the shortest spectral distance to class centroids, is straightforward and computationally efficient. While it performed adequately, its reliance on spectral distance can lead to misclassification in areas with similar spectral properties across different land cover types.

The Spectral Angle algorithm had the lowest overall accuracy (94.29%). Although this method is effective in reducing the impact of illumination differences by considering the angular similarity between reflectance vectors, it may struggle with subtle spectral variations within the same land cover class. This limitation is reflected in its lower accuracy compared to the other algorithms.

The choice of algorithm significantly influences the classification results, which in turn impacts decision-making in environmental management and urban planning. For instance, the high accuracy of the Random Forest algorithm makes it a preferred choice for detailed and precise land cover mapping, essential for flood management and urban development in the Itajaí River Basin. Conversely, the faster and simpler Minimum Distance algorithm may be suitable for preliminary assessments or regions with less complex land cover dynamics.

6. Conclusion

This study demonstrated the effectiveness of using CBERS-4A satellite images and machine learning algorithms for land cover classification in the Itajaí River Basin. The Random Forest and Maximum Likelihood algorithms showed the highest accuracy, making them suitable for detailed environmental and urban planning applications. The varied performance of the algorithms highlights the importance of selecting the appropriate method based on the specific requirements of the study.

The results provide a clear understanding of the land use dynamics in the Itajaí River Basin, contributing to better environmental management and policy-making. The high accuracy of the classification methods, particularly Random Forest, underscores the potential of advanced machine learning techniques in remote sensing applications.

Future research should focus on integrating additional data sources and exploring hybrid classification methods to further improve accuracy and reliability. The continuous monitoring of land cover changes using advanced remote sensing technologies will be crucial in addressing environmental challenges and supporting sustainable development in the Itajaí River Basin.

References

Bertoli, B., 2023: Rio do Sul enfrenta mais uma enchente, a sétima só em 2023. NSC Total. Available at: https://www.nsctotal.com.br/noticias/rio-do-sul-enfrenta-mais-uma-enchente-a-setima-so-em-2023

Breiman, L., 2020: Random Forests. Machine Learning, 12343 LNCS, 503–515. doi.org/10.1007/978-3-030-62008-0_35

Congalton, R.G., 2001: Accuracy assessment and validation of remotely sensed and other spatial information. International Journal of Wildland Fire, 10(3–4), 321–328. doi.org/10.1071/wf01031

Congedo, L., 2021: Semi-Automatic Classification Plugin: A Python tool for the download and processing of remote sensing images in QGIS. Journal of Open Source Software, 6(64), 3172. doi.org/10.21105/joss.03172

EPAGRI, n.d.: Mapas Digitais de Santa Catarina. Retrieved December 22, 2023, from https://ciram.epagri.sc.gov.br/mapoteca/

Ferreira, P.D.S., 2017: Modelagem hidrológica para estimativa da vazão na bacia hidrográfica do Rio Brígida e a disponibilidade hídrica frente às mudanças climáticas. Universidade Federal de Pernambuco.

Grizonnet, M., Michel, J., Poughon, V., Inglada, J., Savinaud, M., and Cresson, R., 2017: Orfeo ToolBox: open source processing of remote sensing images. Open Geospatial Data, Software and Standards, 2(1), 0–7. doi.org/10.1186/s40965-

017-0031-6

Instituto Nacional de Pesquisas Espaciais, n.d.: Catálogo de Imagens. Retrieved June 19, 2024, from http://www.dgi.inpe.br/catalogo/explore

Kruse, F.A., Lefkoff, A.B., Boardman, J.W., Heidebrecht, K.B., Shapiro, A.T., Barloon, P.J., and Goetz, A.F.H., 1993: The spectral image processing system (SIPS)—interactive visualization and analysis of imaging spectrometer data. Remote Sensing of Environment, 44(2–3), 145–163. doi.org/10.1016/0034-4257(93)90013-N

Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., and Wulder, M.A., 2014: Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment, 148, 42–57. doi.org/10.1016/j.rse.2014.02.015

Rashmi, S., Addamani, S., Venkat, and Ravikiran, S., 2014: Spectral Angle Mapper Algorithm for remote Sensing Image Classification. International Journal of Innovative Science, Engineering & Technology, 1(4), 201–205.

Richards, J.A., 2022: Remote Sensing Digital Image Analysis. Springer International Publishing. doi.org/10.1007/978-3-030-82327-6

Schettini, C.A.F., 2002: RBRH-Revista Brasileira de Recursos Hídricos Volume 7 n Caracterização Física do Estuário do Rio Itajaí-açu, SC. 7, 123–142.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-3-2024-623-2024