Revista Brasileira de Cartografia
ISSN 1808-0936 | https://doi.org/10.14393/revbrascartogr
Sociedade Brasileira de Cartografia, Geodésia, Fotogrametria e Sensoriamento Remoto

Evaluation of the Integrated Use of Nanosatellite Images and Classifiers based on Machine Learning for Studies of Hydrological Dynamics in the Nhecolândia Region (Pantanal)

Avaliação do Uso Integrado de Imagens de Nanossatélites e Classificadores baseados em Aprendizado de Máquina para Estudos da Dinâmica Hidrológica na Região da Nhecolândia (Pantanal)

Mariana Dias Ramos [1], Eder Renato Merino [2], Célia Regina Montes [3] and Adolpho José Melfi [4]

1 São Paulo State University, Rio Claro, Brazil. mariana.d.ramos@unesp.br
ORCID: https://orcid.org/0000-0001-8205-4624
2 University of Brasília, Brasília, Brazil. eder.merino@unb.br
ORCID: https://orcid.org/0000-0003-2155-8620
3 University of São Paulo, São Paulo, Brazil. crmlauar@usp.br
ORCID: https://orcid.org/0000-0002-5173-1909
4 University of São Paulo, São Paulo, Brazil. ajmelfi@usp.br
ORCID: https://orcid.org/0000-0001-5960-937X

Received: 11.2022 | Accepted: 03.2023

Abstract: The Lower Nhecolândia region is one of the most iconic landscapes in the Pantanal Basin. Its unique morphology comprises more than 10,000 lakes with saline-alkaline water and fresh water that coexist in an area of approximately 12,000 km². This region is subject to seasonal flooding that acts on runoff; however, little is known about its flooding dynamics. Recent advances in the area of geoprocessing have helped expand our knowledge about lacustrine environments. This work evaluates the performance of two supervised classifiers based on machine learning (Support Vector Machine and Random Forest), for characterizing the hydrological dynamics of the Nhecolândia region. The classifiers were applied to nanosatellite images (PlanetScope) using the Google Earth Engine cloud computing platform. The results showed satisfactory and similar performance of these two classifiers.
Keywords: Nanosatellites. Google Earth Engine. Supervised Classifiers. Lakes.

Resumo: A região da Baixa Nhecolândia é uma das paisagens mais icônicas da Bacia do Pantanal. Sua morfologia única é composta por mais de 10.000 lagoas com águas salino-alcalinas e água doce que coexistem em uma área aproximada de 12.000 km². Essa região está sujeita a alagamentos sazonais que atuam no escoamento superficial, porém, pouco se conhece sobre sua dinâmica de inundação. Avanços recentes na área do geoprocessamento têm ajudado a ampliar nosso conhecimento sobre ambientes lacustres. Este trabalho teve como objetivo avaliar o desempenho de dois classificadores supervisionados baseados em aprendizado de máquina (*Support Vector Machine* e *Random Forest*), para a caracterização da dinâmica hidrológica da região da Nhecolândia. Os classificadores foram aplicados em imagens de nanossatélites (PlanetScope) por meio da plataforma de computação em nuvem Google Earth Engine. Os resultados evidenciaram o desempenho satisfatório e semelhante dos dois classificadores.
Palavras-chave: Nanossatélites. Google Earth Engine. Classificadores Supervisionados. Lagoas.

1 INTRODUCTION

Remote sensing stands out due to its relevance in studies of continental aquatic resources, especially for identifying and monitoring large flooded areas, namely, wetlands (OZESMI; BAUER, 2002). The use of orbital images, obtained by radar and multispectral sensors, is largely responsible for the systematic mapping of these areas, as well as for studying their characteristics and uses (FINLAYSON; MILTON; PRENTICE, 2018). The Pantanal Basin is considered one of the largest continuous wetland regions in the world, with an approximate area of 150,000 km², occupying part of the Brazilian states Mato Grosso and Mato Grosso do Sul and also covering part of Bolivia and Paraguay (ASSINE et al., 2015; KEDDY et al., 2009).

The water dynamics involving the flood pulse and different hydrosedimentological patterns along the Pantanal Basin result in distinct and complex ecosystems, which can guide the division of this wetland into subregions of unique landscapes that configure the valuable biodiversity of the Pantanal (JUNK; BAYLEY; SPARKS, 1989; JUNK et al., 2006). The region known as Nhecolândia stands out because it has, in its southern portion, more than 10,000 lakes distributed in an approximate area of 12,000 km².

The lakes of Nhecolândia differ in their hydrobiogeochemical characteristics, ranging from the occurrence of saline-alkaline water to fresh water, and also present different types of aquatic vegetation (ALMEIDA et al., 2003). In addition to the unique concentration of lakes, this lacustrine environment has a dense network of shallow channels sporadically flooded and of variable widths, locally known as *vazantes* (MERINO; ASSINE, 2020). The processes that involve the hydrological dynamics of the Nhecolândia region are highly seasonal, such as the flooding of lakes and the action of *vazantes* on surface runoff. Although Nhecolândia presents unique characteristics and is the subject of much research, there are no studies analyzing the temporal and spatial detail of the runoff and flooding in the region (COSTA; TELMER, 2006).

To study flood flow or monitor lacustrine environments such as Nhecolândia, the main challenges of remote sensing are related to the limitations of orbital systems, such as low temporal resolution, and the fact that some flood dynamics have a cycle or a high seasonality not compatible with the revisit time of a satellite (CARVALHO JÚNIOR, 2018; PEREIRA; ABREU; MAILLARD, 2017; POURSANIDIS et al., 2019). Spatial resolution is also a limitation to be considered: many rivers have their bed with a smaller extension than the spatial resolution of the sensor system (CARVALHO JÚNIOR, 2018; POURSANIDIS et al., 2019). One can mention the series of Landsat 8 satellites, which have a resolution of 30 m, making it unviable to apply collections of Landsat images for monitoring or studying water bodies that have a smaller extension, as well as ephemeral events of less than 16 days duration.

Currently, the incorporation and use of nanosatellites in water studies is gaining ground in the debate on technological innovations and their contribution to scientific development (COOLEY et al., 2019). Nanosatellites are characterized by having reduced dimensions and weight compared to traditional conventional satellites due to the reduction of their sensors and electronic components. They are designed to orbit in constellations composed of dozens of satellites with the intention of overcoming gaps in temporal resolution, achieving imaging of the entire globe with daily revisits, as well as spatial resolution, with sensors capable of capturing information of submetric dimensions (CARVALHO JUNIOR, 2018). The largest constellation of nanosatellites today is from Planet Labs, with more than 200 PlanetScope sensor systems in orbit (STRICK et al., 2019; PEREIRA et al., 2019). This constellation manages to capture daily images of the entire globe with an average spatial resolution of three meters (PLANET LABS, 2016).

PlanetScope sensors generate multispectral products in the red (R), green (G), blue (B) and near infrared (NIR) bands, and the use of these nanosatellites has been explored by the academic community considering the information gain obtained by high spatial and temporal resolutions. Regarding the study and monitoring of water resources, Planet Labs images have already been used in research on quantification of sediment load and flow (STRICK et al., 2019), water level monitoring (EHRET et al., 2021), quality of water after extreme events (NIROUMAND-JADIDI et al., 2020), concentration of suspended sediments in intermittent rivers (PEREIRA et al., 2019), bathymetric studies in shallow waters (POURSANIDIS et al., 2019) and monitoring of lake flooding dynamics (COOLEY et al., 2019).

Despite highlighting some difficulties generated by the low spectral resolutions and different radiometric calibrations between the constellation of sensors, the use of PlanetScope images for developing

the aforementioned research showed positive and relevant results for the understanding of the dynamics of the water resources, mainly of the events of high seasonality. As alternatives for maximizing space-time data from nanosatellites, there are image processing techniques that use artificial intelligence and cloud computing (BARBOSA; NOVO; MARTINS, 2019; CARVALHO JÚNIOR, 2018).

The use of artificial intelligence, such as the supervised classification process based on machine learning, has been explored in several applications of remote sensing, including improving mapping capacity and quality (ADUGNA; XU; FAN, 2022; SOUZA et al., 2020) guidance on strategies for preventing disturbances (LUZ et al., 2022) and quality monitoring models for preserving fauna and flora (ANANIAS et al., 2022). Cloud computing also represents a great novelty for geoprocessing due to the possibility of manipulating data on a large scale; the Google Earth Engine (GEE) platform is one of the main ones today using cloud computing of spatial data and orbital images (WANG et al., 2022; SOUZA et al., 2020; TASSI; VIZZARI, 2020; GORELICK et al., 2017).

Considering the gain of spatial and temporal information generated by nanosatellites, as well as the ability to process these data on a large scale with cloud computing, the present work aimed to evaluate the performance of different supervised classifiers, Support Vector Machine (SVM) and Random Forest (RF), available on the GEE platform, with the aim of establishing a workflow that can be replicated in the time series of nanosatellite images for the study of hydrological dynamics in the Nhecolândia region on scales of great spatial and temporal detail.

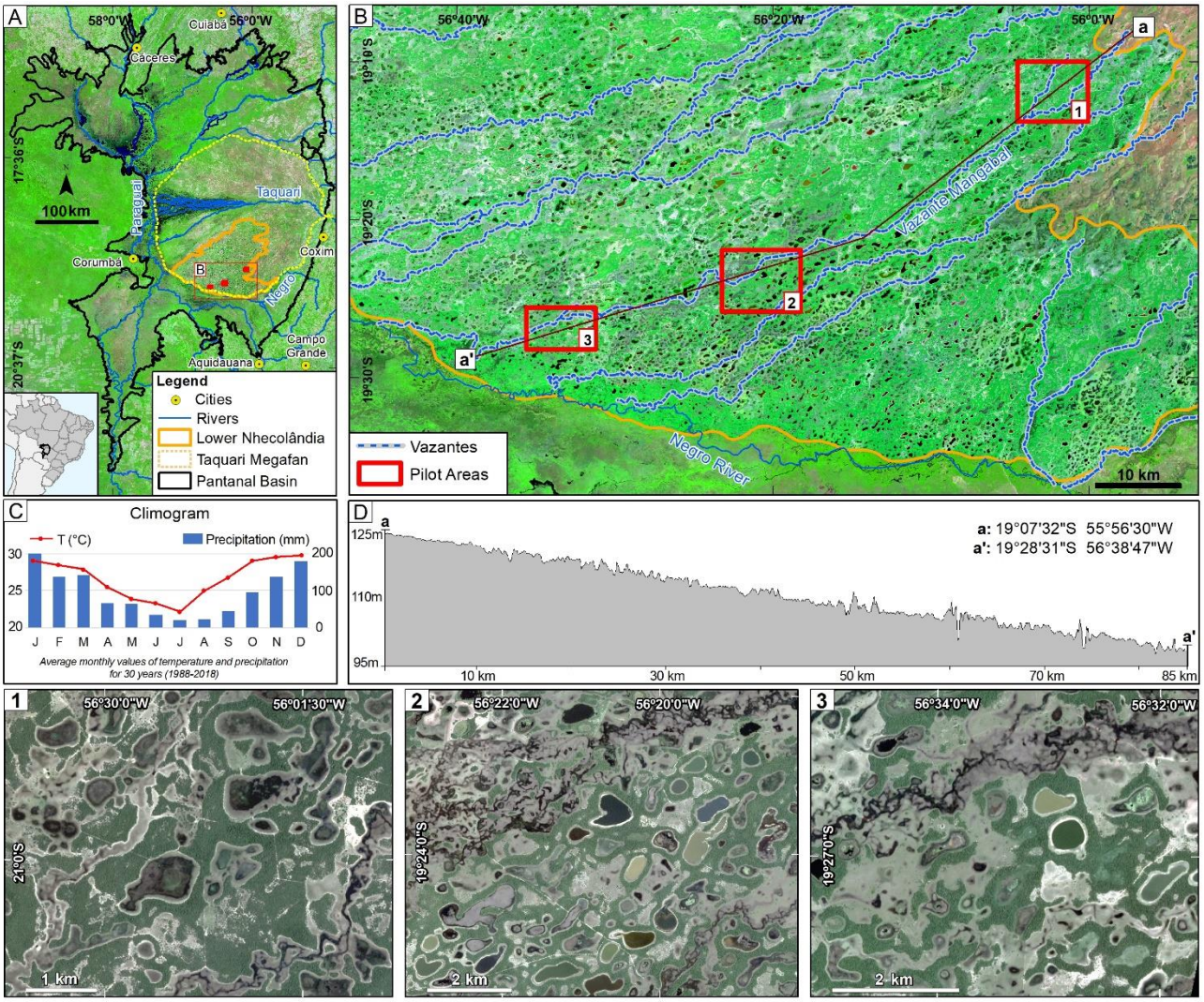
2 MATERIALS AND METHODS

The present work was based on the integrated use of data from remote sensing and processing on a cloud computing platform. Images from PlanetScope nanosatellites referring to a dry period and a flood period in the Nhecolândia region were used, considering the hydrological year from October 2017 to September 2018. The image processing was carried out in Google Earth Engine where the classifiers based on machine learning Support Vector Machine and Random Forest were applied.

2.1 Study Area and Observation Period

Three pilot areas were delimited and named Areas 1, 2, and 3, representing different areas of Nhecolândia. Each area has approximately 50 km² and is distributed along the shallow waterway named Vazante do Mangabal. This *vazante* crosses the Nhecolândia region in a northeast-southwest direction, formed in the overflow process of the Taquari River (to the north) and responsible for the surface runoff towards the Negro River (to the south). Area 1 (furthest north) and Area 3 (furthest south) are at a distance of 85 km, in this extension there is a low topographic gradient, with an altitude difference of only 25 m (Figure 1).

Figure 1 – Location of the study area: (A) Pantanal Basin with the Taquari Megafan and Baixa Nhecolândia (Landsat Mosaic 8). (B) Location of Areas 1, 2, and 3 (Landsat Mosaic 8). (C) Climogram (CHIRPS and MODIS). (D) Topographic profile (MERIT-DEM). (1.2.3) Morphology of pilot areas (PlanetScope).



Elaboration: The authors (2022).

2.2 Planet Images

Through the research incentive program of Planet Labs, a partnership was established for developing the work in which it was possible to have free access to the Planet Labs platform and the images from the PlanetScope sensor. The constellation of these nanosatellites is now composed of more than 200 sensor systems measuring $10 \times 10 \times 30$ cm that have been launched into orbit since 2016. The reduction of components and production cost allows the maintenance of the numerous constellations that capture images with an average revisit time of 30 hours (PLANET LABS, 2016; ROY et al., 2021). The PlanetScope sensor absorbs reflectance in the visible light spectrum, the blue band (B) operates in a range from 455 to 515 nm, the green band (G) from 500 to 590 nm and the red band (R) from 590 to 670 nm. The sensor also captures information in the near infrared (NIR) wavelength of 780 to 860 nm, resulting in 16-bit images of radiometric information (PLANET LABS, 2016). Images acquired on the company's platform correspond to preprocessing level 3B. These are scenes of 7×25 km, with the passing time between 9:30 and 11:30 UTC, already orthorectified, with atmospheric correction and spatial resolution of three meters (PLANET LABS, 2016).

With access to the platform, the ID's of 130 PlanetScope images covering the pilot areas were manually selected, following the visual analysis criteria of absence of clouds and biweekly frequency, to form the time series of data referring to the study period from October 2017 to September 2018.

After selecting the ID's of all the images to be used, mosaic processing and clipping was performed

using an API for integrating Planet platforms with Google Cloud and GEE. In this step, the number of images was reduced to 76 images (Area 1: 24 images, Area 2: 25 images, Area 3: 27 images) of biweekly average frequency and cropped according to the dimension of their respective pilot area, the images were also inserted via API within the GEE, forming collections of multitemporal images of Areas 1, 2, and 3.

All processing on GEE's cloud computing platform is saved as scripts composed of input data, command lines and organization of output data. The input data can be inserted from an external library as done with the Planet Labs images but can also come from the online library made available on the platform, such as those used for the preparation of the climogram of the studied region (Figure 1C), obtained through the CHIRPS (Climate Hazards Group InfraRed Precipitation with Station) model (FUNK; PETERSON; LANDSFELD, 2015), which uses historical series of pluviometric stations and MODIS (Moderate-Resolution Imaging Spectroradiometer) images with temperature data.

Based on an evaluation of the data (pluviometric and temperature) of the CHIRPS model for the period of analysis, and visual inspection of the PlanetScope image collections within the cloud computing platform, images were selected that represented scenes of periods of drought and of flooding in the Nhecolândia region for the test of supervised classifiers. That is, images corresponding to periods of low and high rainfall rates, as well as images representing the typical landscape of these periods, such as the absence or presence of water along the *vazantes*. The images from October 5, 2017 represent the peak of the dry period in the area after a sequence of months with low precipitation values, data from the National Institute of Meteorology-INMET (2022) show that for the second half of 2017, rainfall above 5 mm was observed only from October 21. The images from April 24, 2018 represent a typical flood period in the region.

2.3 SVM and RF classifiers

Supervised classifiers based on machine learning use computational methods to organize existing knowledge about an object, using artificial intelligence to label large amounts of data according to predefined parameters (MITCHELL, 1997). In the classification of orbital images, a label is assigned to each pixel according to its spatial and spectral characteristics (NOVO, 2008); this label is defined by 'inductive learning by examples'; this process can be summarized in four steps: sampling, training of samples, classification and validation (BATISTA, 2003).

Some supervised classifiers are among the image processing tools offered by GEE, two of the most used are the Support Vector Machine (SVM) and the Random Forest (RF). Support Vector Machines are machine learning methods for the purposes of classification, regression and other learning missions; the goal of an SVM is to find the ideal separating point, vector, or hyperplane from the input data entered as samples (BURGES, 1998; ANANIAS et al., 2022). On the platform, from image samples, the SVM known as Support Vector Classification seeks, within the constructed hyperplanes (Eq 1), to maximize the differentiation margin of the samples corresponding to the pre-established classes (y) (CHANG; LIN, 2011). This behavior is registered as training the classifier to be applied to all pixels of the reference image (x). In order for hyperplanes to be constructed and the margins between different classes to be maximized, SVMs use kernel functions to reproject the input data (CORTES; VAPNIK, 1995; THEODORIDIS; KOUTROUMBAS, 2008). As indicated by Hsu, Chang and Lin (2003) the kernel linear function presents better performance for treating large sets of data. Within the GEE, the parameters considered by the SVM were the spectral behavior of the samples for each of the 4 bands of the PlanetScope images; the bands also represent dimensions from which hyperplanes of ideal separation of the input data were built (Eq 2).

Form of representation of a linear separation hyperplane:

$$\boldsymbol{w^T x} + b = 0 \tag{1}$$

SVM classification function for n-hyperplanes:

$$y = sgn\left(\boldsymbol{w^T}\varphi(\boldsymbol{x}) - b\right) \tag{2}$$

where **w** is an n-dimensional weight vector that defines the orientation of the hyperplane, φ is a transformation function that maps the data into a higher-dimensional feature space with a linear kernel (K(xi,xj)) and b is the independent term (bias).

The RF supervised classifier, in turn, also uses machine learning to return classified data. From the input data, the classifier generates decision trees with branches that represent the different characteristics of the samples; this process is known as training. These characteristics are seen as variables to be considered to associate the input data (x) to a class (y) (LOH, 2011; ANANIAS et al., 2022). On the GEE platform, the RF classifier analyzes each pixel of the Planet image and, based on the behavior of the sample along the combination of all trees generated (h(x) →y), defines the best class to be assigned to the pixel with a decision called maximum vote (Eq 3) (ANANIAS et al., 2022; BREIMAN, 2001; SHETTY, 2019).

The output of the classifier can be defined as the highest value class of the function:

$$f(x) = 1 - \sum_{j=1}^{m} (p_j)^2 \quad (3)$$

where m is the number of classes and p*j* is the relative frequency of class j to x.

All available bands (NIR-R-G-B) were used to classify the images referring to the dry and flood periods in each pilot area, in accordance with the guidelines by Olofsson *et al.* (2014). Next, 4 classes were defined to label the PlanetScope images in the Nhecolândia region: water (Class 1), *vazante* (Class 2), exposed soil (Class 3, representing portions of exposed soils and sandbanks that are quite common in the Nhecolândia region), and *cordilheira* (Class 4, represent slightly elevated sandy barriers covered by dense arboreal vegetation).

For each class, 20 sample regions were inserted in a polygon format, which were saved as FeatureCollection and, in their metadata, carried the respective class value. After delimiting the regions of training samples, 40 validation points of each class were also manually delimited for later independent analysis of map accuracy. The training samples consisted of 500 points randomly distributed within the sample regions. Of the total samples, 70% were destined to actually train the classifier, while the remaining 30% were reserved for the automatic validation process of classification accuracy.

With the sample regions inserted, a standard script was written to train and apply the RF and SVM. The script variables were just the type of classifier (SVM or RF) and spatial variables depending on the image referring to Area 1, 2, or 3. It is noteworthy that the RF classifier used the construction of 100 decision trees to define the maximum vote, and SVM used a linear kernel. The processing was the same for each of the six PlanetScope images: after sample training, the entire image was classified according to the product of the supervised SVM or RF classifier, generating 12 different classifications, 6 for the SVM classifier and 6 for the RF classifier (the product of classifying a wet and dry period image for each of the three pilot areas).

2.4 Validation

Classification accuracy was verified with the 30% points of samples not used to train the classifier, an order of 150 validation points for each class and each image. In the process of constructing the confusion matrix these points can be called reference data and are compared with the product of the classifiers, which can be called classified data. Crossing the label assigned to the reference data x classified data for each class allows checking the global accuracy, producer accuracy, and consumer accuracy of each methodology (PONTIUS JUNIOR; MILLONES, 2011). The confusion matrix of each classified image was automatically generated within the GEE.

The map accuracy, in turn, is the validation of the product of the classifiers based on the comparison with independent data; that is, with the 40 points of each class manually inserted in each image independent of the classification. To check the accuracy of the map, also within the GEE, a second confusion matrix was generated in which the reference data is equivalent to 40 points.

 DOI: http://dx.doi.org/10.14393/rbcv75n0a-67656

In this way, it was possible to verify the performance of each classifier with the metrics corresponding to the performance of the classifier from the samples themselves (classification accuracy) and the performance of the classification in relation to its representation of the real (map accuracy).

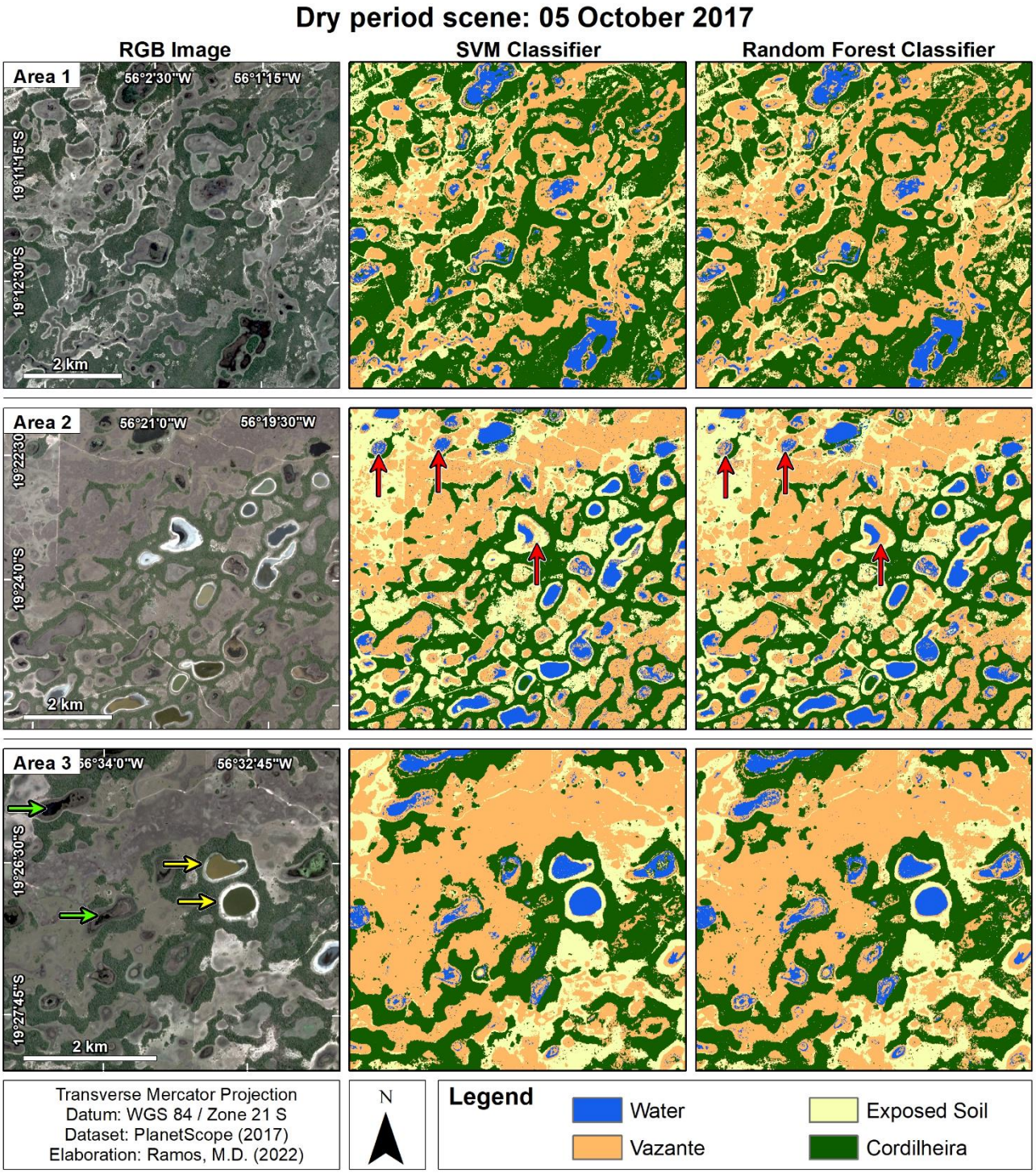
3 RESULTS AND DISCUSSIONS

3.1 Performance of Classifiers

The products of the classifiers of each PlanetScope image can be seen in Figures 2 and 3; they were divided according to the images corresponding to the dry and flood dates, respectively. The two methodologies had very similar results, with small nuances that differentiate the two classifications of the same image, such as the 3 lakes indicated in red in Figure 2.

 DOI: http://dx.doi.org/10.14393/rbcv75n0a-67656

Figure 2 – Product of the SVM (Support Vector Machine) and RF (Random Forest) Classifiers for Images of October 05, 2017. Green arrows indicate examples of 2 fresh water lakes; yellow arrows indicate 2 saline-alkaline lakes; red arrows exemplify small differences observed in the 2 classifications.

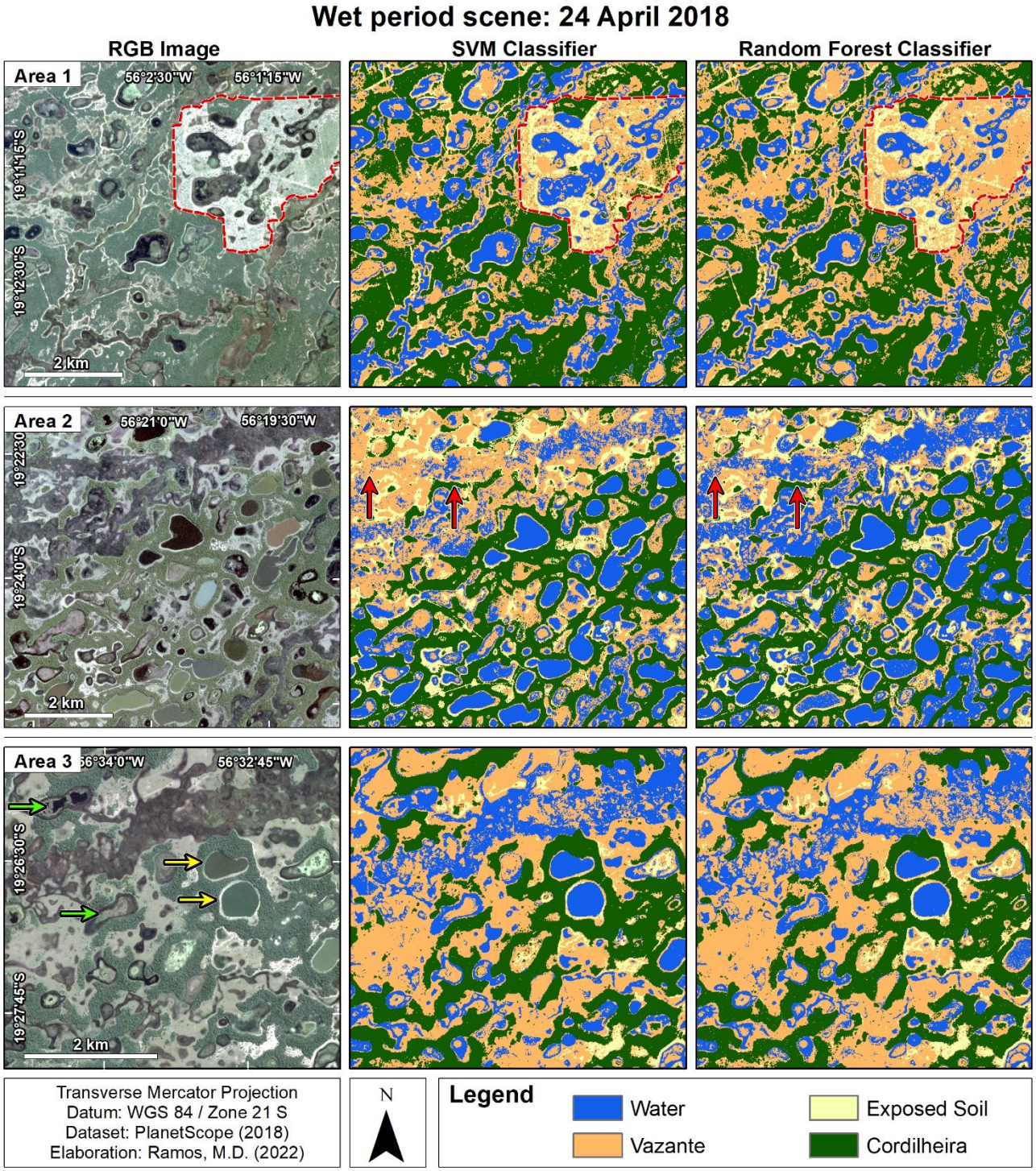


Elaboration: The authors (2022).

The performance of the classifiers exceeded expectations for the fidelity of the classification based on only 4 classes, since the PlanetScope images have spectral resolution limitations and operate in a reduced amount of bands (POURSANIDIS et al., 2019; COOLEY et al., 2019). The classifiers had to evaluate very close wavelength thresholds to define the appropriate label for each pixel, mainly due to the diversity of characteristics of the lakes found in Nhecolândia. In addition to the division between saline-alkaline and freshwater lakes, optically active constituents are very relevant aspects for the remote sensing of multispectral images such as crystalline water, colored dissolved organic matter (CDOM) and total particulate matter (BARBOSA; NOVO; MARTINS, 2019). As an example, high concentrations of cyanobacteria, which, when present, give saline-alkaline lakes a greenish color and spectral behavior similar to that of vegetation, making their classification difficult. The SVM classifier recognized and classified more homogeneously the lakes, or

saline water bodies, which have more greenish water; These are the same regions where the RF presents greater classification confusion from the spectral information of the PlanetScope sensors, especially in the driest period.

Figure 3 – Product of the SVM (Support Vector Machine) and RF (Random Forest) Classifiers for Images from April 24, 2018. Polygon dotted in red indicates an 8 km² deforested area. Green arrows indicate examples of 2 freshwater lakes; yellow arrows indicate 2 saline-alkaline lakes; red arrows exemplify differences observed in the 2 classifications.



Elaboration: The authors (2022).

The main difference between the product of the classifiers is related to the mapping of flooded and runoff areas during the flood period. In the case of the SVM classifier, the classification based on the hyperplanes responds less sensitively to the behavior of pixels referring to the water surface in runoff areas, tending to classify them as '*vazante*', while the RF classifier, after defining the maximum vote of all of its decision trees, is prone to classify these runoff areas as 'water'.

The high performance and quality of the results of the SVM and RF classifications are confirmed by

the resulting values of the confusion matrices. The global accuracy of the 6 classifications based on the SVM methodology ranged from 98.11% to 99.83%, while the values corresponding to the RF classifications ranged from 97.77% to 99.82%. The confusion matrix generated from the independent validation points also resulted in high values (≥75%) for the classifier products. The global accuracy, or here called map accuracy, of these points ranged from 93.12% to 99.37% for the SVM classifier and from 93.75% to 100% with the RF methodology. As an example, one can check the construction of the confusion matrices of global and map accuracy referring to the product of the SVM and RF classifiers for Area 1 on October 5, 2017, the dry period, in Figures 4, 5, 6, and 7.

Figure 4 – Confusion matrix of SVM classifier for Area 1 on October 5, 2017 using sample regions.

Area 1 Date: 10/05/2017		Reference Data					User Accuracy (%)
		Water	Vazante	Exposed Soil	Cordilheira	Total	
Classified Data	Water	153	0	0	0	153	100
	Vazante	0	130	1	0	131	99.24
	Exposed Soil	0	2	141	0	143	98.60
	Cordilheira	0	0	0	150	150	100
	Total	153	132	142	150	577	Overall Accuracy
Producer Accuracy (%)		100	98.48	99.30	100	Overall Accuracy	99.48

Elaboration: The authors (2022).

Figure 5 – Confusion matrix of RF classifier for Area 1 on October 5, 2017 using sample regions.

Area 1 Date: 10/05/2017		Reference Data					User Accuracy (%)
		Water	Vazante	Exposed Soil	Cordilheira	Total	
Classified Data	Water	153	0	0	0	153	100
	Vazante	0	131	1	0	131	100
	Exposed Soil	0	3	140	0	143	97.90
	Cordilheira	0	0	0	150	150	100
	Total	153	134	141	150	577	Overall Accuracy
Producer Accuracy (%)		100	97.76	99.29	100	Overall Accuracy	99.48

Elaboration: The authors (2022).

Figure 6 – Confusion matrix of SVM classifier for Area 1 on October 5, 2017 using independent samples.

Area 1 Date: 10/05/2017		Reference Data					User Accuracy (%)
		Water	Vazante	Exposed Soil	Cordilheira	Total	
Classified Data	Water	40	0	0	0	40	100
	Vazante	0	30	10	0	40	75.00
	Exposed Soil	0	0	40	0	40	100
	Cordilheira	0	0	0	40	40	100
	Total	40	30	50	40	160	Overall Accuracy
Producer Accuracy (%)		100	100	80.00	100	Overall Accuracy	93.75

Elaboration: The authors (2022).

Figure 7 – Confusion matrix of RF classifier for Area 1 on October 5, 2017 using independent samples.

Area 1 Date: 10/05/2017		Reference Data					User Accuracy (%)
		Water	Vazante	Exposed Soil	Cordilheira	Total	
Classified Data	Water	40	0	0	0	40	100
	Vazante	0	32	8	0	40	80
	Exposed Soil	0	0	40	0	40	100
	Cordilheira	0	0	0	40	40	100
	Total	40	32	48	40	160	Overall Accuracy
Producer Accuracy (%)		100	100	83.33	100	Overall Accuracy	95.00

Elaboration: The authors (2022).

The two classifiers show greater confusion in differentiating areas of runoff and wet soil, understood here as '*vazante*', from areas of non-wet soil, understood as 'exposed soil'. This difficulty can be understood as a consequence of the spectral resolution of the PlanetScope sensor, since the range of the longest recorded wavelength refers to the NIR band, which, in this sensor, represents waves in the range of 780 to 860 nm. Soil moisture monitoring, in turn, is recommended for wavelengths greater than 900 nm, as it is in the near infrared (800 to 1500 nm) and short wave (1500 to 3000 nm) ranges in which the interaction of the soil with electromagnetic energy has higher reflectance indices (NOVO, 2008).

3.2 Quantification of Water Surface

By analyzing the RGB images and the products of the classifiers, it is possible to spatially quantify the alterations of each class between the two periods. In the drought image of Area 1, for example, the count of the number of pixels classified as 'water' totaled 222,556 pixels according to RF and 228,557 according to SVM, while at the height of flooding this value was 1,100,686 and 1,289,669 pixels, respectively. As each pixel of the PlanetScope images represents 9 m² of the Earth's surface area, according to the product of the RF classifier, Area 1 had an increase of 7.89 km² in the water surface due to the flooding of the lakes and flooding of *vazantes*. The product of the SVM classifier, in turn, indicates that the increase was 9.55 km² (Table 1). The increase in the water surface of Areas 2 and 3 for each of the classifiers can also be seen in Table 1.

The difference in the presence of surface water between the driest and most flooded periods represents a significant change in the landscape of each pilot area. Considering that each of the areas has approximately 50 km², there is an average change of 20% in the soil cover only due to the dynamics of flooding in the region.

Table 1 – Calculation of water surface increase between the driest (October 05, 2017) and most flooded (April 24, 2018) images of the pilot areas.

Area	Classifier	Period	Number of 'water' pixels	Area in km²	Increase in km²
1	RF	Dry	223,556	2.01	7.89
		Wet	1,100,686	9.91	
	SVM	Dry	228,557	2.06	9.55
		Wet	1,289,669	11.61	
2	RF	Dry	493,462	4.44	17.54
		Wet	2,442,906	21.99	
	SVM	Dry	504,365	4.54	11.22
		Wet	1,751,254	15.76	
3	RF	Dry	219,337	1.97	6.45
		Wet	936,377	8.43	
	SVM	Dry	180,895	1.63	7.77
		Wet	1,044,770	9.40	

Elaboration: The authors (2022).

3.3 Classifier Choice and Processing Scalability

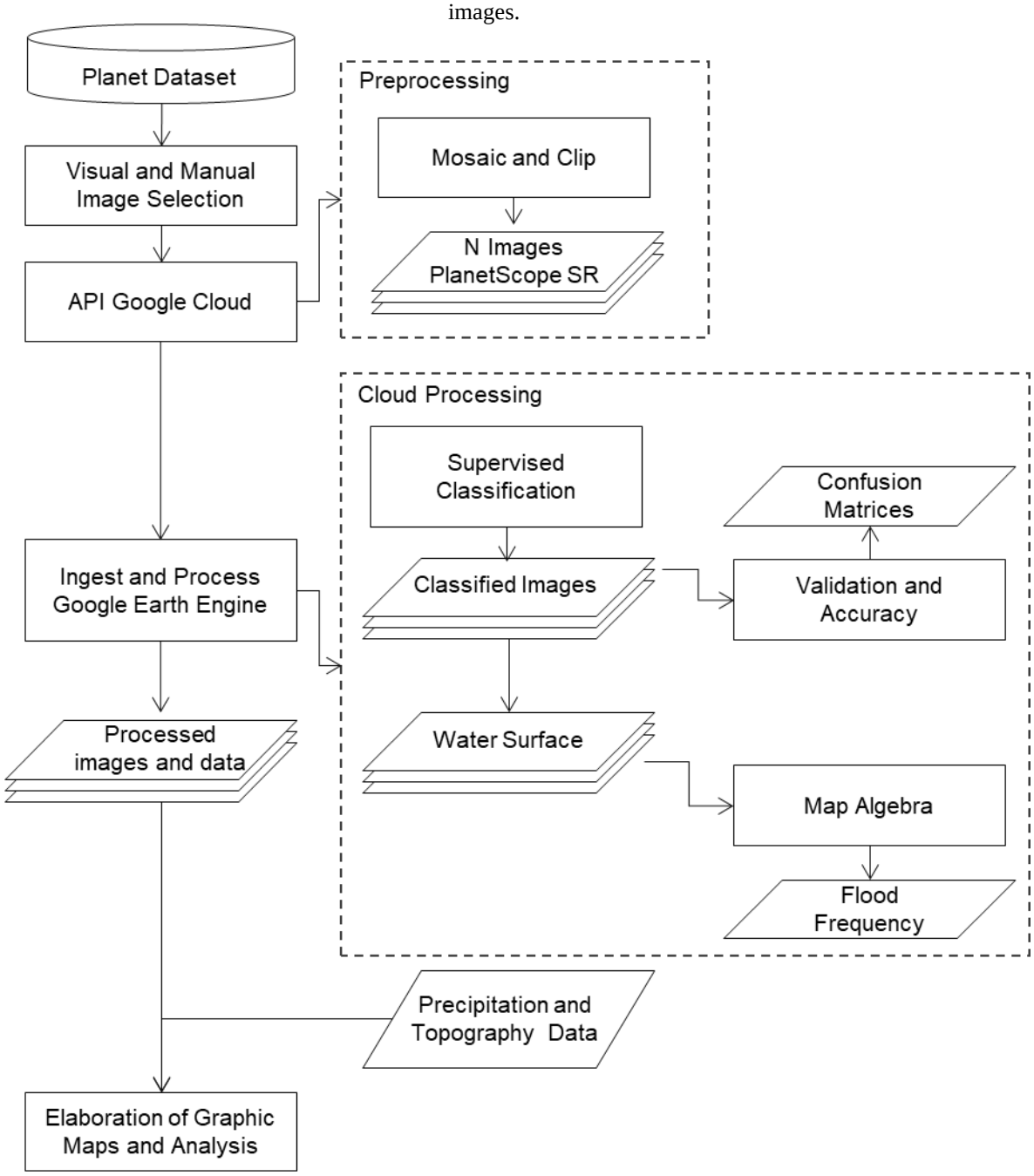
Despite the similar values of the confusion and accuracy matrices for the two classifiers, the visual analysis of the products and the quantitative analysis of integrated 'water' pixels allows identifying relevant

differences to guide the choice of a classifier in the process of studying hydrological dynamics from highly seasonal areas such as Nhecolândia. As noted, the main differences between the classifications are related to the confusion in the mapping of water and *vazante* classes during the flood period. The RF is less conservative in classifying the water surface in the runoff areas, with emphasis on the flood in Area 2, while the SVM tends to be conservative with the mapping of water in the runoff areas, diverging some pixels with the RF and classifying them as *vazante*. At the same time, the SVM classifies the water surface of saline-alkaline lakes (which have different colors) with less omission.

Considering the divergences between the two products, it can be stated that the choice of the classifier to guide a study on hydrological dynamics should start from the objectives to be met. In the Nhecolândia region, for studies of isolated water bodies such as saline-alkaline lakes, the SVM tends to map the target more homogeneously, while for surface runoff studies, the RF may be the most indicated. It is also worth noting that postclassification processing steps to improve mapping, such as the use of spatial filters, should be considered.

The analysis and choice of the classifier is recommended as a step before processing scalability. Once this decision has been made, the GEE platform allows, by means of functions and parallelization, the rapid reproduction of processing for an entire collection of images. The supervised classification of images for qualitative values opens up different possibilities to explore data with reclassification, conversion to quantitative data, performing map algebra processes and integrated analysis with different data sources. To study the hydrological dynamics of Nhecolândia, for example, it is essential to map and understand flooding over time, the flood frequency data, in turn, can be obtained from the result of the supervised classification carried out on a large scale for a time series. Still considering this study, the processing flow can be the same for both the RF and the SVM, depending on the mapping objectives as exemplified in Figure 8.

Figure 8 – Large-scale classified image processing flowchart to obtain flood frequency data from PlanetScope images.



Elaboration: The authors (2022).

4 FINAL CONSIDERATIONS

The integrated use of cloud computing tools and data from nanosatellites allows for new work possibilities in geoprocessing and studies of the Earth's surface. The SVM and RF methodologies, available on the Google Earth Engine platform, had a satisfactory performance in processing and classifying images from the PlanetScope sensor for mapping the Nhecolândia region. Changes in the landscape between the driest and flooded periods are very intense, with around 20% of the total area being converted into water surface. The significant and highly seasonal changes in the landscape observed in Nhecolândia, due to its water dynamics, highlight the need to obtain highly detailed information about this unique region. Despite the spectral and radiometric calibration limitations of this sensor, which is, for now, used mainly for commercial purposes, the Planet images processed in the GEE represent an important source of data for studying highly seasonal lacustrine environments. With this ease of processing a large amount of orbital data, it will be possible to obtain significant information about the landscape and surface runoff in the area, even allowing the elaboration of a flood frequency map in the Nhecolândia region, which would be the first high-resolution

systematization of seasonal flooding in the area.

Acknowledgments

The authors would like to thank the São Paulo Research Foundation (FAPESP) for providing funding for the scientific initiation research for this work (process 2019/1440-0), as well as for the postdoctoral research (2017/26318-8) and the Thematic Project (2016/14227-5) to which this work is related. The authors also extend their appreciation to the National Council for Scientific and Technological Development (CNPq; process no. 307024/2018 307024/2018-0) for the CRM research grant. Jim Hesson of Academic English Solutions edited the manuscript (https://www.academicenglishsolutions.com/editing-service).

Authors Contribution

M.D.R.: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Visualization, Writing - Original Draft Preparation, and Writing - Review and Editing; E.R.M.: Conceptualization, Methodology, Supervision, Validation, and Writing - Review and Editing; C.R.M.: Funding Acquisition, Resources, Supervision, Validation, and Writing - Review and Editing; A.J.M.: Funding Acquisition, Resources, Supervision, Validation, and Writing - Review and Editing.

Conflict of Interest

The authors declare no conflicts of interest.

5 REFERENCES

ADUGNA, T.; XU, W.; FAN, J. Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. **Remote Sensing**. v. 14, n. 3, p. 574, 2022. DOI: 10.3390/rs14030574.

ALMEIDA, T. I. R. de; SÍGOLO, J. B.; FERNANDES, E.; QUEIROZ NETO, J.P.; BARBIERO, L.; SAKAMOTO, A. Y. Proposta de classificação e gênese das lagoas da Baixa Nhecolândia-MS com base em sensoriamento remoto e dados de campo. **Revista Brasileira de Geociências**. v. 33, n. 2-Suplemento, p. 83-90, 2003. DOI: 10.25249/0375-7536.200333s28390.

ANANIAS, P. H. M.; NEGRI, R. G.; DIAS, M. A.; SILVA, E. A.; CASACA, W. A Fully Unsupervised Machine Learning Framework for Algal Bloom Forecasting in Inland Waters Using MODIS Time Series and Climatic Products. **Remote Sensing**. v. 14, n. 17, p. 1-22, 2022. DOI: 10.3390/rs14174283

ASSINE, M. L.; MERINO, E.; PUPIM, F.; WARREN, L.; GUERREIRO, R.; MCGLUE, M. Geology and Geomorphology of the Pantanal Basin. In: BERGIER, I.; ASSINE, M. L. (ed.) **Dynamics of the Pantanal Wetland in South America**. Springer International Publishing, (The Handbook of Environmental Chemistry), v. 37, p. 23- 50, 2015. DOI: 10.1007/698_2015_349.

BARBOSA, C.C.F.; NOVO, E.M.L.M.; MARTINS, V.S., **Introdução ao Sensoriamento Remoto de Sistemas Aquáticos: princípios e aplicações**. 1a edição. Instituto Nacional de Pesquisas Espaciais. São José dos Campos. 161p. 2019.

BATISTA, G.E.A.P.A. **Pré-processamento de dados em aprendizado de máquina supervisionado**. 2003. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2003. DOI: 10.11606/T.55.2003.tde-06102003-160219.

BREIMAN, L. Random Forests. Machine Learning, [s.l.], **Springer Science and Business Media LLC**. v. 45, n. 1, p. 5-32, 2001. DOI: 10.1023/a:1010933404324.

BURGES, C. J. A Tutorial on Support Vector Machines for Pattern Recognition. **Data Mining and Knowledge Discovery 2**, p. 121–167, 1998. DOI: 10.1023/A:1009715923555.

CARVALHO JÚNIOR, O. A., 2018. Aplicações E Perspectivas Do Sensoriamento Remoto Para O Mapeamento De Áreas Inundáveis. **Revista de Geografia (Recife)** v. 35, n. 4 (especial XII SINAGEO), 412–431p., 2018.

CHANG, C. C., C., LIN, C.J., LIBSVM: a library for support vector machines. **ACM Transactions on Intelligent Systems and Technology**, v. 2 n. 27, p. 1-27, 2011. DOI: 10.1145/1961189.1961199.

COOLEY, S. W., SMITH, L. C., Ryan, J. C., PITCHER, L. H., PAVELSKY, T. M. Arctic-Boreal Lake dynamics revealed using CubeSat imagery. **Geophysical Research Letters**, v. 46, p. 2111–2120, 2019. DOI: 10.1029/2018GL081584.

CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, v. 20, p. 273–297, 1995. DOI: 10.1007/bf00994018.

COSTA, M., TELMER, K., Utilizing SAR imagery and aquatic vegetation to map fresh and brackish lakes in the Brazilian Pantanal wetland. **Remote Sensing of the Environment,** v. 105, p. 204–213, 2006. DOI: 10.1016/j.rse.2006.06.014.

FINLAYSON C. M., MILTON G. R., PRENTICE R.C., Wetland Types and Distribution. In: FINLAYSON C., MILTON G., PRENTICE R., DAVIDSON N. (eds) **The Wetland Book. Springer**, Dordrecht, 2018. DOI: 10.1007/978-94-007-4001-3_186.

FUNK, C., PETERSON, P., LANDSFELD, M., The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. **Scientific Data 2**, n. 150066, p.1-21, 2015. DOI: 10.1038/sdata.2015.66.

GORELICK, N., HANCHER, M., DIXON, M., ILYUSHCHENKO, S., THAU, D., MOORE, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone **Remote Sensing of the Environment,** v. 202, p. 18–27, 2017. DOI: 10.1016/j.rse.2017.06.031.

HSU, C. W., CHANG, C. C., LIN, C. J., **A Practical Guide to Support Vector Classification**, Technical report, Department of Computer Science, National Taiwan University. 2003.

Instituto Nacional de Meteorologia (INMET). **Banco de Dados Meteorológicos para Ensino e Pesquisa - BDMEP**. Brasília, DF, Brasil. Disponível em: http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep. Acesso em:09 de junho de 2020.

JUNK, W.; BAYLEY, P.; SPARKS, R. The Flood Pulse Concept in River-Floodplain Systems**. Canadian Journal of Fisheries and Aquatic Sciences** 106, p. 110 – 127, 1989.

JUNK, W.; CUNHA, C. N.; WANTZEN, K. M.; PETERMANN, P.; STRÜSSMANN, C.; MARQUES, M. I.; ADIS, J. Biodiversity and its conservation in the Pantanal of Mato Grosso**,** Brazil**. Aquatic Sciences**, v. 68, p. 278–309, 2006. DOI: 10.1007/s00027-006-0851-4.

KEDDY, P. A., FRASER, L.H., SOLOMESHCH, A. I., JUNK, W. J., CAMPBELL, D.R., ARROYO, M.T.K., ALHO, C.J.R. Wet and Wonderful: The World's Largest Wetlands Are Conservation Priorities. **Bioscience**, v..59, p. 39-51. 2009. DOI: 10.1525/bio.2009.59.1.8.

LOH, W., Classification and Regression Trees. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,** v.1, p 14 – 23, 2011. DOI:10.1002/widm.8.

LUZ, A. E. O.; NEGRI, R.G.; MASSI, K. G.; COLNAGO, M.; SILVA, E.A.; CASACA, W. Mapping Fire Susceptibility in the Brazilian Amazon Forests Using Multitemporal Remote Sensing and Time-Varying Unsupervised Anomaly Detection. **Remote Sensing**, v. 14, n. 2429, p. 1-17, 2022. DOI: 10.3390/rs14102429.

MERINO, E. R.; ASSINE, M. L., Hidden in plain sight: How finding a lake in the Brazilian Pantanal improves understanding of wetland hydrogeomorphology. **Earth Surface Processes and Landforms**, v. 45, p. 440–458, 2020. DOI: 10.1002/esp.4745.

MITCHELL, T. M. **Machine Learning**, 1ª ed, McGraw-Hill, Inc., USA,1997.

NOVO, E. M. L. M. **Sensoriamento Remoto: Princípios e Aplicações.** 3ª ed, São Paulo:Edgard Blucher, 2008.

OLOFSSON, P.; FOODY, G.M.; HEROLD M.; STEHMAN, S.V.; WOODCOCK, C.E.; WULDER, M.A.

Good practices for estimating area and assessing accuracy of land change, **Remote Sensing of Environment**, v. 148, p. 42-57, 2014. DOI: 10.1016/j.rse.2014.02.015.

OZESMI, S. L.; BAUER, M. E. Satellite remote sensing of wetlands. **Wetlands Ecology and Management**, v.10, p. 381–402. 2002. DOI: 10.1023/A:1020908432489.

PEREIRA, E.; ABREU, L., MAILLARD, P., Altimetria Por Satélite Radar Aplicada A Hidrologia No Brasil**. Revista Brasileira de Cartografia,** v. 69, p. 347-360, 2017. DOI: 10.14393/rbcv69n2-44022.

PEREIRA, F. J. S., COSTA, C. A. G., FOERSTER, S., BROSINSKY, A., DE ARAÚJO, J. C., Estimation of suspended sediment concentration in an intermittent river using multi-temporal high-resolution satellite imagery. **International Journal of Applied Earth Observation and Geoinformation**, v. 79, p. 153–161, 2019. DOI: 10.1016/j.jag.2019.02.009.

PLANET LABS, **Planet Imagery Product Specification 56**, 2016.

POURSANIDIS, D.; TRAGANOS, D.; CHRYSOULAKIS, N.; REINARTZ, P., Cubesats allow high spatiotemporal estimates of satellite-derived bathymetry. **Remote Sensing**, v. 11, p. 1-12, 2019. DOI: 10.3390/rs11111299.

PONTIUS JUNIOR, R.G.; MILLONES, M., Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment, **International Journal of Remote Sensing**, v. 32, n. 15, p. 4407-4429, 2011. DOI: 10.1080/01431161.2011.552923.

SHETTY, S. **Analysis of Machine Learning Classifiers for LULC Classification on Google Earth Engine**. 2019. Tese de mestrado na Universidade de Twente. 2019.

SOUZA, C. M., JR.; Z. SHIMBO, J.; ROSA, M. R.; PARENTE, L. L.; A. ALENCAR, A.; RUDORFF, B. F. T.; HASENACK, H.; MATSUMOTO, M.; G. FERREIRA, L.; SOUZA-FILHO, P. W. M.; DE OLIVEIRA, S. W.; ROCHA, W. F.; FONSECA, A. V.; MARQUES, C. B.; DINIZ, C. G.; COSTA, D.; MONTEIRO, D.; ROSA, E. R.; VÉLEZ-MARTIN, E.; WEBER, E. J.; LENTI, F. E. B.; PATERNOST, F. F.; PAREYN, F. G. C.; SIQUEIRA, J. V.; VIERA, J. L.; NETO, L. C. F.; SARAIVA, M. M.; SALES, M. H.; SALGADO, M. P. G.; VASCONCELOS, R.; GALANO, S.; MESQUITA, V. V.; AZEVEDO, T. Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. **Remote Sensing**. v. 12, n. 2735, 2020. DOI: 10.3390/rs12172735.

STRICK, R. J. P., ASHWORTH, P. J., SAMBROOK SMITH, G. H., NICHOLAS, A. P., BEST, J. L., LANE, S. N., PARSONS, D. R., SIMPSON, C. J., UNSWORTH, C. A., DALE, J., Quantification of bedform dynamics and bedload sediment flux in sandy braided rivers from airborne and satellite imagery. **Earth Surface Processes and Landforms**, v. 44, p. 953–972, 2019. DOI: 10.1002/esp.4558.

TASSI, A.; VIZZARI, M. Object-Oriented LULC Classification in Google Earth Engine Combining SNIC, GLCM, and Machine Learning Algorithms. **Remote Sensing**, v. 12, n, 3776, 2020. DOI: 10.3390/rs12223776.

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**, 4th ed.; Academic Press: San Diego, CA, USA; p. 984, 2008.

WANG, R.; FENG, Q.; JIN, Z.; MA, K.; ZHANG, Z.; LIANG, T. Identification and Area Information Extraction of Oat Pasture Based on GEE—A Case Study in the Shandan Racecourse (China). **Remote Sensing**. v. 14, n. 4358, 2022. DOI: 10.3390/rs14174358.

Main author biography

Mariana Dias Ramos was born in Americana (SP) in 1998. She has a bachelor's degree in Geography from Unesp (São Paulo State University 'Júlio de Mesquita Filho') Rio Claro campus, with emphasis on geoprocessing and environmental analysis. She developed the scientific initiation project 'Evaluation of high resolution images (spatial and temporal) from the Planet collection of satellites for analysis of flood frequency characteristics in the Nhecolândia region' funded by FAPESP. She is currently participating in the MapBiomas project, on the team responsible for data on use, coverage and deforestation of the Pantanal and Atlantic Forest biomes.

This work is licensed under a Creative Commons Attribution 4.0 International License – CC BY. This license allows others to distribute, remix, adapt and create upon your work, even commercially, as long as they credit you for the original creation.