Artículo en edición

ESTIMATION OF THE FLOODED AREA OVER THE PANTANAL, A SOUTH AMERICAN FLOODPLAIN, USING MODIS DATA

Anthony Schrapffer[1,2,3], Lucía María Cappelletti[1,2,3] y Anna Sörensson[1,2,3]

[1]Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales. Buenos Aires, Argentina.
[2]CONICET – Universidad de Buenos Aires. Centro de Investigaciones del Mar y la Atmósfera (CIMA). Buenos Aires, Argentina.
[3] CNRS – IRD – CONICET – UBA. Instituto Franco-Argentino para el Estudio del Clima y sus Impactos (UMI 3351 IFAECI). Buenos Aires, Argentina

Autor correspondiente: Anthony Schrapffer, anthony.schrapffer@cima.fcen.uba.ar

Manuscrito recibido el 5 de julio de 2021, en su versión final el 22 de diciembre de 2021

ABSTRACT

Tropical floodplains, such as Pantanal in Central South America, are important features for land-atmosphere interactions. Schemes to account for floodplains should therefore be included in Earth System Models, but this requires observations of flooded area for validation. Satellite data is a possible solution to estimate the flooded area but it is important to evaluate the different flood detection algorithms available in order to use the most efficient for the region. This work explores different methods to estimate the flooded area from the MODIS MOD09A1 satellite surface reflectance product using spectral indexes (mNDWI, NDMI, NDMI-NDVI) to detect the presence of water. We include the traditional threshold-based methods but also some unsupervised classification methods such as the k-means and the Principal Component Analysis applied on the water-related spectral indexes. The calibration and validation of these methods are based on the hydrological knowledge of the region, coming from land surface models, river discharge observation and from previous satellite estimations of the flooded area. The NDMI index seems too sensible to the vegetation which leads to error in the estimation of the flooded area. The other methods were spatially and temporally consistent with previous studies over the Pantanal.

Key Words: Floodplains, flood detection, remote sensing, Pantanal, MODIS.

RESUMEN

Las llanuras de inundaciones tropicales, como el Pantanal en Suramérica Central, son importantes para las interacciones suelo-atmósfera. Por lo tanto, los esquemas que representan las llanuras de inundación tienen que ser incluidos en los Modelos del Sistema Tierra, pero eso requiere observaciones del área inundada para validación. Los datos satelitales son una posible solución para estimar la superficie inundada, pero es importante evaluar los diferentes algoritmos disponibles para utilizar el más eficiente para cada región de interés. Este trabajo explora diferentes métodos para estimar la superficie inundada con el producto de reflectancia de la superficie MODIS con el uso de índices espectrales (mNDWI, NDMI, NDMI-NDVI) para detectar la presencia de agua sobre Pantanal. Incluimos los métodos más comunes basados en el uso de umbrales y también algunos métodos de supervisión no clasificada como los k-means y el Análisis de Componentes principales aplicados a los índices espectrales relacionados con la presencia de agua. La calibración y la validación de estos métodos está basado en los conocimientos hidrológicos de la región, proviniendo de modelos de superficie, observaciones de caudal y de estimaciones de la superficie inundada por satélite realizada en trabajos anteriores. El índice NDMI parece demasiado sensible a la vegetación lo que lleva a errores en la estimación de la superficie inundada. Los otros métodos son espacial y temporalmente consistente con estudios previos sobre el Pantanal.

Palabras clave: Llanuras de inundaciones, Detección de inundaciones, Teledetección, Pantanal, MODIS.

1) INTRODUCTION

The floodplains are wetlands which are temporarily or permanently flooded and where there are strong interactions between the different terrestrial hydrological processes such as river discharge, the evapotranspiration from plants, the evaporation from open water surfaces and the vertical movement of water between the surface soil and the saturated zone. These large floodplains are places of rich biodiversity and provide important ecosystem services such as water purification, river stream regulation and carbon sequestration. The monitoring and improved comprehension of these regions are vital for their revalorization and conservation. Remote sensing products are powerful tools to monitor the spatiotemporal evolution of these extensive floodplains with a reasonable frequency. Satellite estimations of the flooded areas are also necessary to develop a correct representation of the hydrology of these regions in Land Surface Models and Earth System Models.

The periodic flooding of the floodplains related to the overflow of the river is fundamental for the local ecosystem as it is driving the lateral exchange of water and nutrients in the river floodplains system (cf. flood pulse concept, Junk et al., 1989). These exchanges are one of the reasons why the floodplains are very productive ecosystems and considered as biodiversity hotspots. However, large floodplains are also regions where the in-situ observations are not sufficient to reconstruct their full dynamics, as opposed to smaller and more homogeneous wetlands and to unvegetated regions which can be more easily monitored and where the estimation can be carried out more directly by spectral indices. Thus it is difficult to estimate the temporal variability and map the spatial variability of the floods over large floodplains.

Large tropical floodplains, such as the Pantanal in central South America, are regions of strong land-atmosphere interactions due to due to a high level of evaporation in relation with the presence of open-water surfaces and of transpiration in relation with the increased soil moisture (Schrapffer et al., 2020). This induces strong gradient of land-atmosphere fluxes and temperature between the floodplains and the neighbouring regions. This is why the floodplains processes tend to be ever more integrated in Land Surface Models (Schrapffer et al., 2020; Dadson et al., 2010; Getirana et al., 2021) because this improves the representation of the hydrological cycle and it will change the sensible and latent fluxes which may have an impact on atmospheric conditions and, thus on the regional precipitation (Taylor, 2010). These are important advances regarding the growing interest of coupled simulations to study the land-atmosphere interactions. In order to be able to calibrate and evaluate the floodplains scheme in Land Surface Models, the estimates of the temporal and the mapping of the spatial evolution of the flooded surfaces are crucial.

Remote sensing has proven to be a helpful tool to estimate large-scale land processes and may be helpful to estimate the flooded area over large tropical floodplains (Padovani, 2010; Ogilvie et al., 2015). There are two types of sensors which can be used to estimate the flooded areas: the Optical and Synthetic Aperture Radar (SAR) sensors. SAR data presents some advantages to detect the flooded area as it is not affected by clouds because it uses the microwave bands and because it can provide data during both day and night (Pereira et al., 2019). Despite this, SAR data may be affected by speckle noise (Inglada et al., 2016) and may be largely impacted by confounding effects associated with the surface conditions. Moreover, the processing of this type of data is more complex compared to optical data (Niedermeier et al., 2005). On the other hand, optical data are relatively easy to manipulate and allow to obtain both the flooded area and the presence of vegetation or other features in a relatively simple way. Therefore, in this work, we chose to employ optical data. There are two major difficulties to handle in this work: (1) the relatively large extension of the region and (2) the issue of the cloud cover over such a large region. The first point can be managed by using a satellite product with a lower resolution such as a MODIS product. For the second point, a post-processed product which uses a lower temporal resolution can be used. Some of these lower temporal resolution products are created by merging the different images available to produce images with the lowest cloudiness possible. There are two similar MODIS products which correspond to this type of post processing: MOD09A1 and MYD09A1.

Traditional methods used to estimate the extension of flooded surfaces rely on spectral indices and thresholds (Ogilvie et al., 2015). Some spectral indices may highlight the presence of water by higher values. However, some other land features may generate noise and make it difficult to directly detect the flooded area using a threshold. For example, the estimation of flooded area over regions containing lush vegetation may be confounded with the vegetation water content due to the large annual variability of water content related to the flood pulse.

This is why, although the presence of water may be overestimated by higher values in some spectral indices, some land features such as the vegetation might generate noise and make it difficult to directly detect the flooded area using a threshold. Thus more sophisticated methods may lead to an improvement of the estimate. The spectral indices considered in this study contain information about the water content and the status of the vegetation such as the modified Normalized Water Index (mNDWI), the Normalized Difference Moisture Index (NDMI) and Normalized Difference Vegetation Index (NDVI).

This study aims to compare the use of different methods based on spectral indexes to estimate the flooded area and to overcome the difficulties of estimating the flooded surface over large and complex regions such as Pantanal. This is done by comparing different traditional approaches: (1) using the classical approach of applying a threshold over spectral indexes and (2) using unsupervised classification methods such as the k-means and the Principal Component Analysis (PCA). We aim at an optimized method that is both as robust and as simple as possible. The estimates obtained are then validated by a previous satellite estimate made by Padovani (2010) and by the river height at Ladário station.

This paper is organized as follows. Section 2 contains the Methodology and Dataset used. Section 3 contains the results and the evaluation of the temporal and spatial estimation of the flooded area by the different methods considered. Section 4 contains the discussion and conclusion.

2) METHODOLOGY AND DATASETS

2.a) REGION OF INTEREST: THE PANTANAL

The Pantanal, the world's largest floodplains, has an extension of 150.000 km^2 and is located in the tropical region of southwestern Brazil (see Figure 1). The flat lands of Pantanal range between 80 and 150 m.a.s.l. of altitude while the surrounding mountain ranges of the Cerrados from its north/northeast to its southeast ranges between 200 and 1.400 m.a.s.l. (Alho, 2005). It has a regular annual cycle of flooding driven by the precipitation over the Cerrados during the rainy season (December to February). Due to the flat slopes of the Pantanal, it takes between 3 and 5 months for the water flowing from the Cerrados to cross the Pantanal. This excess of water flowing into Pantanal through the river system and slowed down by the topography generates important floods. The climatological season of floods occurs between February and May (Penatti et al., 2015).

2.b) MODIS data: MOD09A1

The Moderate Resolution Imaging Spectroradiometer (MODIS) Terra MOD09A1 and MYD09A1 products have been chosen to perform this study for various reasons. First, they have a resolution of 500m which is higher than some other surface reflectance products such as Landsat (30 m resolution) but it is sufficient and more manageable as we are dealing with an extensive region. The MOD09A1 (MYD09A1) product is constructed from an 8-day composite period and gives an estimate of the surface spectral reflectance for the 7 first bands of Terra (Aqua) MODIS with corrected atmospheric effects (gases, aerosols, Rayleigh scattering). This correction consists in (1) an adjustment to include the effect of the solar zenith angle in order to obtain the top-of-atmosphere value and (2) the correction of the error related to the atmospheric scattering and absorption due to the presence of gases and aerosols in the atmosphere and to the spherical albedo (Vermote et al., 2006). The MODIS satellites provide data for each location each 1-2 days. This permits creating a composite image, selecting for each 8-days period the highest quality data for each pixel (lower view angle, absence of clouds, clouds shadow and aerosols) to obtain the MOD09A1 and MYD09A1 products. Two tiles were considered to fully include the Pantanal: h12v10 and h12v11. Both products have been retrieved from the NASA Earth Data Search (https://search.earth data.nasa.gov).

The flooding cycle of the Pantanal is annual, thus a temporal resolution from a couple of weeks to a month is acceptable. Thus, both products can be used for this purpose. Although this product intends to avoid clouds and other inconveniences, during the rainy season the images can still be affected by the presence of clouds due to an excessive cloud coverage during the rainy season. The presence of clouds has been assessed in two steps. Firstly, the Quality Bit Flags of the MODIS products over the Pantanal were used to obtain the mask of the Pantanal which is not cloudfree nor covered by clouds shadows. For values of cloud cover fraction over the Pantanal higher than 5%, the image was discarded. After that, all the images retained were checked visually to verify that they didn't contain coarse cloud features over the Pantanal that remained undetected by the Quality Bit Flags. Between 2002 and 2021, 54% of the images available were considered cloudless over Pantanal in MOD09A1 and 35% for MYD09A1. The dates available without clouds for MOD09A1 and for MYD09A1 have been compared. It should be highlighted that the major differences between MOD09A1 and MYD09A1 are the availability of data as MOD09A1 was launched in 2000, two years before MYD09A1 (Savtchenko et al., 2004). During the period they have in common, MOD09A1 has 140 cloudless dates which are considered as cloudy in MYD09A1 while MYD09A1 only has 7 cloudless images which are considered as cloudy in MOD09A1. These 7 images represent the dry season, a period of lower cloudiness and thus of major availability of images also in MOD09A1. For these reasons, only the product MOD09A1 has been retained although the use of both products MYD09A1 may be considered to complete the data in further studies. All the MOD09A1 cloudless images have been confirmed as such by the visual check, while MYD09A1 was not checked visually since this product was not used for this study.

The different methods of flood detection presented in this study have been calibrated over the 2002-2004 period. Figure 2 represents for each month the total number of MODIS MOD09A1 images available and the quantity of exploitable images, i.e. cloudless. As expected the number of cloudless images is strongly affected by the wet season (November to March).

2.c) SPECTRAL INDEXES

The Spectral Indexes have two main objectives: (1) to isolate some specific land features signals such as signals related to the vegetation (Xue and Su, 2017), the presence of water (Acharya et al., 2018) or the soil composition (van der Meer et al., 2012); while (2) they are insensitive to other perturbing signals (Verstraete and Pinty, 1996).

The spectral indices presented here are based on normalized differences between reflectance at different wavelengths. The NDVI emphasizes the presence of vegetation while the rest of the indices try to underline the presence of water bodies. All these indexes are resumed in Table I.

The main spectral indexes are constructed based on some basic processes: the vegetation strongly reflects the Near InfraRed (NIR) and the Green but has a very low reflectance in the Red wavelength. The ShortWave InfraRed (SWIR) is very sensitive to the water content and in particular to the vegetation water content.

These indexes are shown in Figure 3 over the Pantanal region for two different dates: one during the dry season (21st August 2002) and one during the wet season (15th April 2003).

The NDMI gives a good indication over the flooded vegetation may be falsely detecting highly vegetated regions as flooded. To combine the information contained in NDMI with NDVI seems a possible solution to better distinguish between these two land covers (cf. NDMI-NDVI index in Figure 3.d and Figure 3.h).

2.d) FLOODED AREA DETECTION

Two methods are tested in this study: (1) a threshold-based method using the indices that seemed to better represent the presence of water (mNDWI; NDMI; NDMI-NDVI) and (2) using two different unsupervised classification methods using 3 indices (mNDWI, NDMI and NDVI).

The first method of unsupervised classification is the k-means (Lloyd, 1982) which is a clustering method to regroup the data into different categories. The number of categories or clusters is given by the parameter “k”. number of clusters. Each cluster is defined by its centroid and the membership of each data point to a certain cluster will be determined according to the nearest centroid. The algorithm tries to minimize the total distance between the centroids and the data. The election of the k-value depends on the problem that is being clusterized. Different values have been evaluated. For k-value under 6, the output was not stable while k-values higher than 6 added more complexity to the description of the data which wasn’t necessary adding value to the discrimination between flooded and not flooded pixels. Thus, the k-value chosen for this study is 6.

The second unsupervised classification method uses the Principal Component Analysis (PCA – Jollife and Cadima, 2016) method which finds an orthogonal projection that best fits the data and allows to reduce the number of dimensions. As the data has 3 dimensions (due to the 3 indexes considered), the maximal number of dimensions that can be considered for the PCA is 3. The number of dimensions considered in this study is set at 2. The second dimension refers to the spatial structure of the flooded area. This is not the case for the first dimension of the PCA which seems to be representing other processes such as the vegetation. Higher values in the second dimension of the PCA corresponds with areas with higher values of mNDWI, NDMI and to the spatial structure of the floods (cf. Figure 1). The value of the pixels over this axis resumed the flood related information from the 3 indexes. Then, a threshold has to be established to classify each pixel into the flooded / not flooded categories.

2.e) UNSUPERVISED CLASSIFICATION INPUT

In order to have a single model that would take into account the variability of the vegetation along the year and that would underline the flood processes, the sample input data to generate the PCA and the k-means model have been randomly selected from two images. As the Pantanal has a very marked wet and dry season, one of these images corresponds to the dry season (from June to September) and the other one at the end of the wet season (From November to March) which also corresponds to the climatological season of floods. The images chosen correspond to the dates that were used to illustrate the spectral indexes in Figure 3: the 21st of August 2002 for the dry season and the 15th of April 2003 for the wet season. A total of 10000 pixels per image has been used.

The PCA and k-means processes are quite sensible to the input. In this case, the objective is to represent the variability of the flooded area. The data is mainly composed of not flooded pixels as demonstrated by the distribution in Figure 4 whose maximum is located in the low-NDMI / low-mNDWI region. For this reason, although the 10000 pixels per image were randomly selected, pixels with higher values of mNDWI have been favored. Another filter has been applied to avoid selecting the outlier which were mainly pixels with extremely low NDMI value (cf. Figure 4).

The k-means clustering with k=6 is shown in Figure 5. In the spatial location of the clusters in the (mNDWI, NDMI) space (Figure 5.c), the cluster number 0 to 3 have a low mNDWI value, reasons why they are considered as not flooded and their NDMI index value is growing from the cluster 0 to the cluster 3. Looking at the difference between the k-means representation of the dry season image (Figure 5.b) and the wet season image (Figure 5.a) maps, we can see that they may represent different conditions of vegetation and that low vegetation regions in the dry season image become high vegetation region during the wet season. Pixels in the clusters 4 and 5 have a higher mNDWI value and can be considered as flooded. The pixels in cluster 5 include the pixels with maximal mNDWI values, thus we can consider that cluster 5 represents the open water pixels and pixel 4 the flooded vegetation.

The second dimension of the PCA is shown in Figure 6. We can deduce that higher values along this dimension represent the flooded pixels.

2.f) VALIDATION DATA

Ground-based observations of the flooded area over such a large area as the Pantanal are scarce. The validation of a flood estimate method may rely on two aspects: (1) the knowledge of the local hydrological network and the characteristics of the regions; (2) the comparison with previous satellite estimates.

Hamilton et al. (1996) is a reference for the flooded area estimate over the region. It found a relationship between the flooded area over the Pantanal estimated between 1979 and 1987 and the river gauge at the Ladário station obtained from the Brazilian National Water Agency (Agência Nacional de Águas - ANA). The flooded area has been estimated using the brightness temperature from a satellite passive microwave sensor. Hamilton et al. (2002) further extended this relationship from 1900 to 2000 to obtain an estimation of the evolution of the flooded area. Although these results are not available for the period of availability of MODIS, they point out that the river gauge data from the Ladário station (see Figure 1) can be used to assess the flooded area as these data are strongly correlated.

We will also for comparison use the estimation of (Padovani, 2010) which has been validated in comparison with Hamilton et al. (2002). Padovani (2010) applied a Linear Model of Spectral Mixture (LMSM) to MODIS MOD13Q1 images to estimate the temporal and map the spatial evolution of the flooded area over the Pantanal. The MOD13Q1 product includes vegetation description (NDVI and EVI indexes) and the corresponding Red, Near Infrared, blue and Mid-Infrared bands from MODIS. A 16-days composite image is created by selecting the highest quality data for each pixel (lower view angle, absence of clouds) and by favoring higher values of NDVI/EVI indexes. Thus, although this product is also constructed from MODIS data, it differs from MOD09A1 because of its focus on vegetation processes and because of the lower temporal resolution (images each 16 days instead of 8). The method developed by Padovani (2010) uses a single image (May 25th 2007) to calibrate by finding a linear relationship between the reflection at different wavelengths available and the soil, vegetation and water cover. By applying this relationship to the other images, it allows estimating the fraction of soil, vegetation and water cover. The flooded area is then determined by applying a threshold on the water cover fraction.

Other types of datasets have been considered for the spatial validation of the methods presented in this study such as WaterMAP (Pekel et al., 2016) and GFPLAINS250m. WaterMAP is a global dataset available between 1984 and 2015 which contains the monthly estimate of the surface water location constructed from optical sensors (Landsat 5 TM; Landsat 7 ETM+ and Landsat 8 OLI), regional datasets and from inventories. GFPLAINS250m is a 250m resolution dataset drawing the delimitations of what can be considered as floodplains based on Digital Elevation Model datasets.

The different thresholds required were calibrated with the Padovani (2010) time series for the period 2002-2004. The respective threshold values for the different methods are resumed in Table II.

The mean flood frequency from Padovani (2010) will also be used as a comparison. However, as it improves the readability of the map, the modification of the flood frequency map from Padovani (2010) presented in Fluet-Chouinard et al., (2015) can also be used to assess the spatial representation of the floodplains in the different methods used in this work.

3) RESULTS

The different satellite estimate methods that have been described in Section 2 and calibrated over the 2002-2004 period have been applied to MODIS MOD09A1 between 2002 and 2009. Their temporal evolution and spatial representation are assessed in comparison with Padovani (2010).

3.a) EVALUATION OF THE TEMPORAL EVOLUTION

The temporal evolution of the flooded area estimated over the Pantanal by the threshold-based methods are presented in Figure 7 and Figure 8 shows the evaluation of the results for the unsupervised classification methods. The comparison of the different estimates with Padovani (2010) is summarized through some basic comparative statistical indexes in Table III (correlation, root mean square error - RMSE - and percentage bias – PBIAS).

Except the NDMI index, the different methods are coherent with the study of Padovani (2010). Among them, the PCA and NDMI-NDVI have higher values of flooded area while the mNDWI index and the k-means have lower values of the flooded area.

The NDMI-based estimation is less correlated than the other methods with Padovani but this correlation increases when integrating the information from the NDVI index (Figure 7.c). This difference may be related to the influence of the vegetation in the NDMI index.

The river stage at Ladário is delayed compared to both Padovani (2010) and the methods evaluated although the amplitude of the river gauge and the estimated flooded area are similar. Following the Hamilton et al. (1996) estimation of the flooded area, the river stage at Ladário should be strongly correlated. Further analysis should be performed to understand these differences.

3.b) EVALUATION OF THE SPATIAL EVOLUTION

Figure 9 shows the comparison of flood frequency between 2002 and 2009 in the different methods presented in this study in order to compare them with the flood frequency map from Padovani (2010), WaterMAP and the floodplains delimitations from GFPLAINS250m.

As seen in the first overview of the spectral indexes, the NDMI index is strongly influenced by the vegetation which creates a bias for the detection of flooded areas. For the other estimation methods, the results are more coherent with the flood frequency map from Padovani (2010) and WaterMAP although WaterMAP seems to consider only the most flooded area of the Pantanal. Except for the NDMI-based method, the large rivers such as the Main Paraguay River at the North and South of the Pantanal, the São Lourenço river at the northeast and the Taquari river at the East of the Pantanal are clearly visible in the different flood detection methods. All the results are also coherent with the GFPLAINS250m floodplains delimitation which is based on a DEM. The only exception is the central region of the Pantanal, the Taquari Megafan, which may be related to local changes in the orography (Assine, 2005).

3.c) EXPLORATION OF A CASE STUDY

The simple flood detection methods presented previously may have a large variety of applications. This subsection aims to illustrate their potential by using the mNDWI-based flood detection method and the NDVI index to explore the evolution of the extent of the floods along the years. The floods are evaluated during the month of march which is one of the most flooded months for the Pantanal. The images chosen have a cloud cover lower than

2% following the quality flag of MODIS. Three dates have been selected to perform this study: 21/03/2004 (t0), 22/03/2007(t1), 06/03/2021 (t2). t0 (respectively t1) corresponds to the year of lower maximum (respectively higher maximum) flood extent over the 2002 and 2010 period. t2 has been chosen in order to compare the two previous dates to the actual situation which corresponds to drier conditions and with the vegetation cover affected by important wildfires during the 2020 dry season.

Figure 10 shows the NDVI index (Fig. 10.a) for t0 over the Pantanal as well as the difference of NDVI between t1 and t0 (Fig. 10.c) and between t2 and t0 (Fig. 10.c). Figure 10.d-f shows the flooded area estimated with the mNDWI based method for the three dates. Comparing the flooded area in t0 and t1, the floods in t1 are much more extended but they show similar patterns. The regions where the flood became more important in t1 are the northwest and central Pantanal. Some flooded areas also appear in the South of the Pantanal. The vegetation seems to be reduced over some of the flooded area which may be related to the floods replacing the vegetation or at least reducing the NDVI. However, the NDVI increases around the shape of the floodplains in t1 compared to t0. A larger extent of flooded area reduces locally the NDVI while the NDVI increases at its border due to the higher water availability.

In t2, the floods are at their minimal extent and are principally around the Paraguay river and over the Taquari Megafan in the central region of the Pantanal. The northwest region has almost no floods in t2 but has increased NDVI compared to t0. This means that there is water allowing for the development of the vegetation but there is not enough water so it can be considered as flooded. The NDVI is lower in t2 compared to t0 over the regions with higher values of NDVI in t0 which may be related to the wildfire. It should also be noted that there is an increase of the NDVI values compared to t0 over the NorthEast of the Pantanal. This region is not usually flooded so this may be more related to the impact of the local precipitation on the vegetation during the wet season.

4) DISCUSSION AND CONCLUSION

The estimation of the flooded area over large floodplains is a difficult task. The satellite products may be precious tools. This paper explored different methods to estimate the flooded area using the surface reflection from optical remote sensing products. The water-related information is extracted by using different spectral indexes related to the presence of water content and vegetation. Then, different methods are developed using directly the spectral indexes to determine the presence of water: threshold-based methods and unsupervised classification to use the information from different spectral indexes at the same time. The different methods evaluated were coherent with the previous works although there is some delay between the temporal evolution of the estimated flood area and the river height at Ladário. The NDMI index has an issue to represent the flooded area as it is influenced by the vegetation during the wet period. However, considering the vegetation through the NDMI-NDVI index seems to improve the representation of the flooded area. The spatial map of the flooded area represents well the known hydrological features of the Pantanal. It should be noted that the threshold based methods have lower computational costs for similar results but the unsupervised classification methods can bring extra information.

The methods of flood detection presented in this study are simple methods which are based on spectral index and do not require important preprocessing. They may be divided into two categories: the threshold based methods on the one hand and the PCA based and k-means methods on the other hand. The threshold based methods consist in applying a threshold to the spectral indexes to detect the flooded area. This threshold can be determined in comparison with other data which gives an indication either on the flooded area or on the spatial extent of the floods. Different thresholds can be determined depending on the sensibility expected for its use. The PCA and k-means methods use unsupervised classification tools applied to a combination of the spectral indexes related to the presence of water. For the PCA method, the method consists in identifying the dimension related to the presence of water to calibrate and apply a threshold to this dimension. For the k-means, it consists in identifying the clusters which correspond to the flooded area.

The advantage of all the above methods is that they can be easily applicable if the user has some observational data to establish a threshold. Then, it is possible to calculate other spectral indexes corresponding to other processes using the same optical satellite data to obtain a global panorama of the hydrological processes over a certain region quite easily with a reduced pre-processing. Nevertheless, these methods also present some disadvantages. The main disadvantage is related to the presence of cloud cover in optical satellite images which requires the filter of images containing clouds and, thus, may reduce the quantity of images available. Another disadvantage is the fact that the spectral indexes may be affected by other processes which impact the presence of water without being related to floods such as it may be the case with the presence of lush vegetation.

Different solutions can be considered in order to face the issues presented previously although this may involve more sophisticated methods. Concerning the cloud cover, the combination of optical and SAR satellite data have been proven to improve the flood detection being able to solve both cloud cover issue for the optical satellite and noise from the SAR data (Prigent et al., 2020; Niedermeier et al., 2005; Inglada et al., 2016). Concerning the interaction of other processes with the flood detection when using optical satellite data, there are other methods that can be considered. The simpler process consists of developing customized spectral indexes using a linear combination of the spectral bands in order to better differentiate the vegetation from the flooded water such as it is done in other application such as the Floating Algae Index (FAI) (Dogliotti et al., 2018) used to differentiate the presence of algae in the water. Another option is, instead of evaluating the presence of flood over each pixel individually, to consider the pixels by group of pixels such as it can be done with the Object Based Image Analysis (Blaschke et al., 2014). This may help to better determine if the pixels in an object are flooded by using (1) the distribution of the reflection of the pixels composing each group and (2) the shape of the object (Louzada et al., 2020). The flood detection can also be improved by using additional ancillary data about the local orography using Digital Elevation Models. Finally, some more advanced methods of machine learning classification can be used but they require more precise information on the pixels which are flooded in order to fit the model. Unfortunately, this type of information is not always available.

Finally, we would like to emphasize that the difficulty to detect the flooded vegetation also lies in the difficulty to define a limit to qualify whether a pixel is flooded or whether it is just a pixel representing a moist soil.

ACKNOWLEDGEMENTS: We would like to gratefully acknowledge the support of the Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT), Argentina (PICTs 2017-1406, 2018-02511); the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Argentina (PIP 11220200102141CO); the French-Argentina project ECOS-Sud 2018 co-financed by the Ministerio de Ciencia, Tecnología e Innovación (MINCyT), Argentina and the Université Sorbonne Paris Nord, Francia (ECOS-A18D04); and the French National LEFE program (Les Enveloppes Fluides et l'Environnement; LEFE 12962).

REFERENCIAS

Acharya, T.D., Subedi, A., Lee, D.H, 2018: Evaluation of Water Indices for Surface Water Extraction in a Landsat 8 Scene of Nepal. Sensors, 18, 2580. doi:10.3390/s18082580

Alho, C.J.R., 2005: The Pantanal. In L. Fraser & P. Keddy (Eds.), The World's Largest Wetlands: Ecology and Conservation, Cambridge University Press, pp. 203-271. doi:10.1017/CBO9780511542091.008

Assine, M.L., 2005: River avulsions on the Taquari megafan, Pantanal wetland, Brazil. Geomorphology, 70(3-4), pp. 357-371. doi:10.1016/j.geomorph.2005.02.013

Blaschke, T., Hay, G.J., Kelly, M., Lang, S., Hofmann, P., Addink, E., Feitosa, R.Q., van der Meer, F., van der Werff, H., van Coillie, F,Tiede, D., 2014: Geographic Object-Based Image Analysis - Towards a new paradigm. ISPRS Journal of Photogrammetry and Remote

Sensing, 87, pp. 180-191. doi:10.1016/j.isprsjprs.2013.09.014

Boschetti M., Nutini F., Manfron G., Brivio P.A., Nelson A., 2014: Comparative Analysis of Normalised Difference Spectral Indices Derived from MODIS for Detecting Surface Water in Flooded Rice Cropping Systems. PLoS ONE 9, 2. doi:10.1371/journal.pone.0088741

Dadson, S. J., Ashpole, I., Harris, P., Davies, H. N., Clark, D. B., Blyth, E., Taylor, C. M., 2010. Wetland inundation dynamics in a model of land surface climate: Evaluation in the Niger inland delta region. Journal of Geophysical Research Atmospheres, 115, D23114. doi:10.1029/2010JD014474

Dogliotti, A. I., Gossn, J. I., Vanhellemont, Q., Ruddick, K. G., 2018: Detecting and quantifying a massive invasion of floating aquatic plants in the Río de la Plata turbid waters using high spatial resolution ocean color imagery. Remote Sensing, 10, pp. 1140-1154. doi:10.3390/rs10071140

Fluet-Chouinard, E., Lehner, B., Rebelo, L. M., Papa, F., Hamilton, S. K., 2015: Development of a global inundation map at high spatial resolution from topographic downscaling of coarse-scale remote sensing data. Remote Sensing of Environment, 158, pp 348-361. doi:10.1016/j.rse.2014.10.015

Gao, B. C., 1996: NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58, 3, pp 257-266. doi: 10.1016/S0034-4257(96)00067-3

Getirana, A., Kumar, S. V., Konapala, G., Ndehedehe, C. E., 2021: Impacts of Fully Coupling Land Surface and Flood Models on the Simulation of Large Wetlands' Water Dynamics: The Case of the Inner Niger Delta. Journal of Advances in Modeling Earth Systems, 13, 3. doi:10.1029/2021MS002463

Gond, V., Bartholomé, E., Ouattara, F., Nonguierma, A., Bado, L., 2004: Surveillance et cartographie des plans d'eau et des zones humides et inondables en régions arides avec l'instrument VEGETATION embarqué sur SPOT-4. International Journal of Remote Sensing, 25, 5, pp. 987-1004. doi:10.1080/0143116031000139908

Hamilton, S. K., Sippel, S. J., Melack, J., 1996: Inundation patterns in the Pantanal Wetland of South America determined from passive microwave remote sensing. Archiv für Hydrobiologie, 137, 1, pp. 1-23. doi:10.1127/archiv-hydrobiol/137/1996/1

Hamilton, S. K., Sippel, S. J., Melack, J. M., 2002: Comparison of inundation patterns among major South American floodplains. Journal of Geophysical Research Atmospheres, 107, D20, pp. LBA 5-1-LBA 5-14. doi:10.1029/2000JD000306

Hu, C., 2009: A novel ocean color index to detect floating algae in the global oceans. Remote Sensing of Environment, 113(10), 2118–2129. doi:.1016/J.RSE.2009.05.012

Inglada, J., Vincent, A., Arias, M., Marais-Sicre, C., 2016: Improved early crop type identification by joint use of high temporal resolution sar and optical image time series. Remote Sensing, 8, pp. 362-382. doi:10.3390/rs8050362

Jollife, I. T., Cadima, J., 2016: Principal component analysis: A review and recent developments, 374, 2065. doi:10.1098/rsta.2015.0202

Junk, W. J., Bayley, P. B., Sparks, R. E., 1989: The flood pulse concept in river-floodplain systems. D. P. Dodge [ed.] Proceedings of the International Large River Symposium. Can. Spec. Publ. Fish. Aquat. Sci., 106, pp. 110–127.

Lloyd, S. P., 1982: Least Squares Quantization in PCM. IEEE Transactions on Information

Theory, 28, 2, pp. 129-137.

Louzada, R. O., Bergier, I., & Assine, M. L., 2020: Landscape changes in avulsive river systems: Case study of Taquari River on Brazilian Pantanal wetlands. Science of the Total Environment, 723, 138067. doi:10.1016/j.scitotenv.2020.138067

Nardi, F., Annis, A., Baldassarre, G. Di, Vivoni, E. R., Grimaldi, S., 2019: GFPLAIN250m, a global high-resolution dataset of earth’s floodplains. Scientific Data, 6, 180309. doi:10.1038/sdata.2018.309

Niedermeier, A., Hoja, D., Lehner, S., 2005: Topography and morphodynamics in the German Bight using SAR and optical remote sensing data. Ocean Dynamics, 55, pp. 100-109. doi:10.1007/s10236-005-0114-2

Ogilvie, A., Belaud, G., Delenne, C., Bailly, J. S., Bader, J. C., Oleksiak, A., Ferry, L., Martin, D., 2015: Decadal monitoring of the Niger Inner Delta flood dynamics using MODIS optical data. Journal of Hydrology, 523, pp. 368-383. doi:10.1016/j.jhydrol.2015.01.036

Padovani, 2010: PhD thesis "Dinâmica Espaço-Temporal das Inundações do Pantanal". Escola Superior de Agricultura Luiz de Queiroz, Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, pp. 175.
https://www.teses.usp.br/teses/disponiveis/91/91131/tde-14022011-170515/pt-br.php

Pekel, J. F., Cottam, A., Gorelick, N., Belward, A. S., 2016: High-resolution mapping of global surface water and its long-term changes. Nature, 540, 418-422. doi:10.1038/nature20584

Pereira, G. H. D. A., Júnior, C. C., Fronza, G., Deppe, F. A. C., 2019: MULTITEMPORAL ANALYSIS OF SAR IMAGES FOR DETECTION OF FLOODED AREAS IN

PANTANAL. Raega-O Espaço Geográfico em Análise, 46(3), 88-100. doi:10.5380/raega.v46i3.66988

Prigent, C., Jimenez, C., Bousquet, P. (2020). Satellite-Derived Global Surface Water Extent and Dynamics Over the Last 25 Years (GIEMS-2). Journal of Geophysical Research: Atmospheres, 125(3). doi:doi.org/10.1029/2019JD030711

Rouse, W., Haas, R. H., Deering, D. W., 1974: Monitoring vegetation systems in the Great Plains with ERTS. Third Earth Resources Technology Satellite-1 Symposium- Volume I: Technical Presentations. NASA SP-351, pp. 309.

Savtchenko, A., Ouzounov, D., Ahmad, S., Acker, J., Leptoukh, G., Koziana, J., Nickless, D., 2004: Terra and Aqua MODIS products available from NASA GES DAAC. Advances in Space Research, 34(4), 710-714. doi:10.1016/j.asr.2004.03.012

Schrapffer, A., Sörensson, A., Polcher, J., Fita, L., 2020: Benefits of representing floodplains in a Land Surface Model: Pantanal simulated with ORCHIDEE CMIP6 version. Climate Dynamics, 55, pp. 1303-1323. doi:10.1007/s00382-020-05324-0

Taylor, C. M., 2010. Feedbacks on convection from an African wetland. Geophysical Research Letters, 37, 5. doi:10.1029/2009GL041652, 2010

van der Meer, F. D., van der Werff, H. M., van Ruitenbeek, F. J., Hecker, C. A., Bakker, W. H., Noomen, M. F., van der Meijde, M., Carranza, E. J. M., de Smeth, J. B., Woldai, T., 2012: Multi- and hyperspectral geologic remote sensing: A review, 14, 1, pp. 112-128. doi:10.1016/j.jag.2011.08.002

Vermote, E. F., Saleous, N. Z., 2006: Operational atmospheric correction of modis visible to middle infrared land surface data in the case of an infinite lambertian target. Earth Science

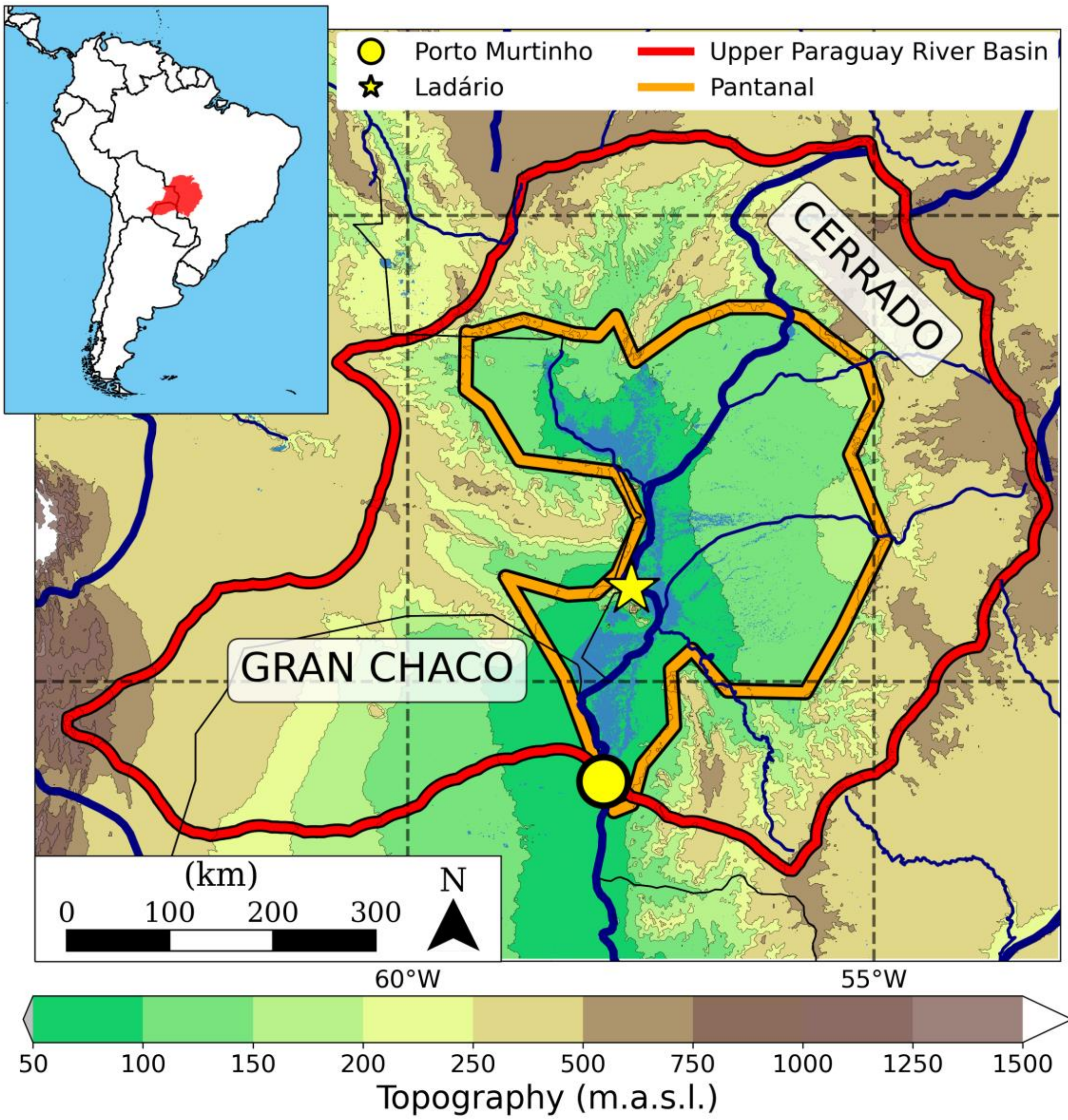
Satellite Remote Sensing: Science and Instruments. Springer, pp. 123-153. doi:10.1007/978-3-540-37293-6_8

Verstraete, M. M., Pinty, B., 1996: Designing optimal spectral indexes for remote sensing applications. IEEE Transactions on Geoscience and Remote Sensing, 34, 5, pp. 1254-1265. doi: 10.1109/36.536541

Xu, H., 2006: Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. International Journal of Remote Sensing, 27, 14, pp. 3025-3033. doi:10.1080/01431160600589179

Xue, J., Su, B., 2017: Significant remote sensing vegetation indices: A review of developments and applications. Journal of Sensors, 2017. doi:10.1155/2017/1353691

634 **Figuras y Tablas**



635

636

637 Figure 1: Localization and description of the Pantanal wetlands inside the Upper Paraguay
638 River Basin. The blue layer corresponds to the flood extent from WaterMap (Pekel et al.
639 2016; Source: EC JRC/Google).

640

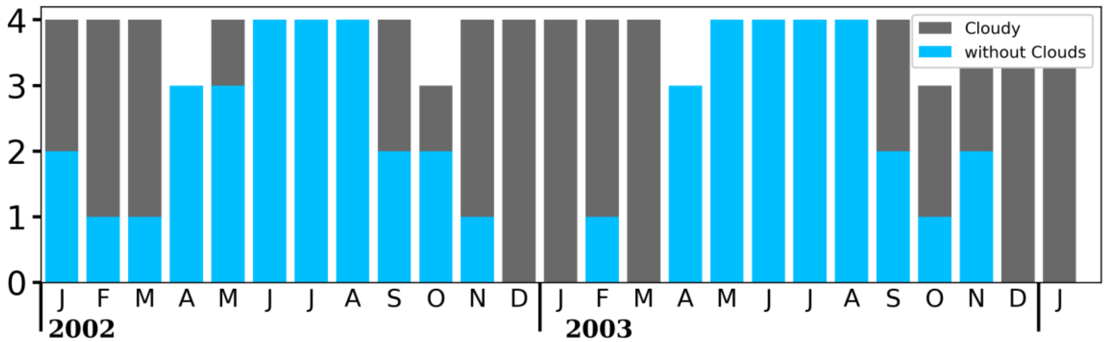


Figure 2: Number of monthly available data for this MODIS product and number of dates available without clouds between 2002 and 2004.

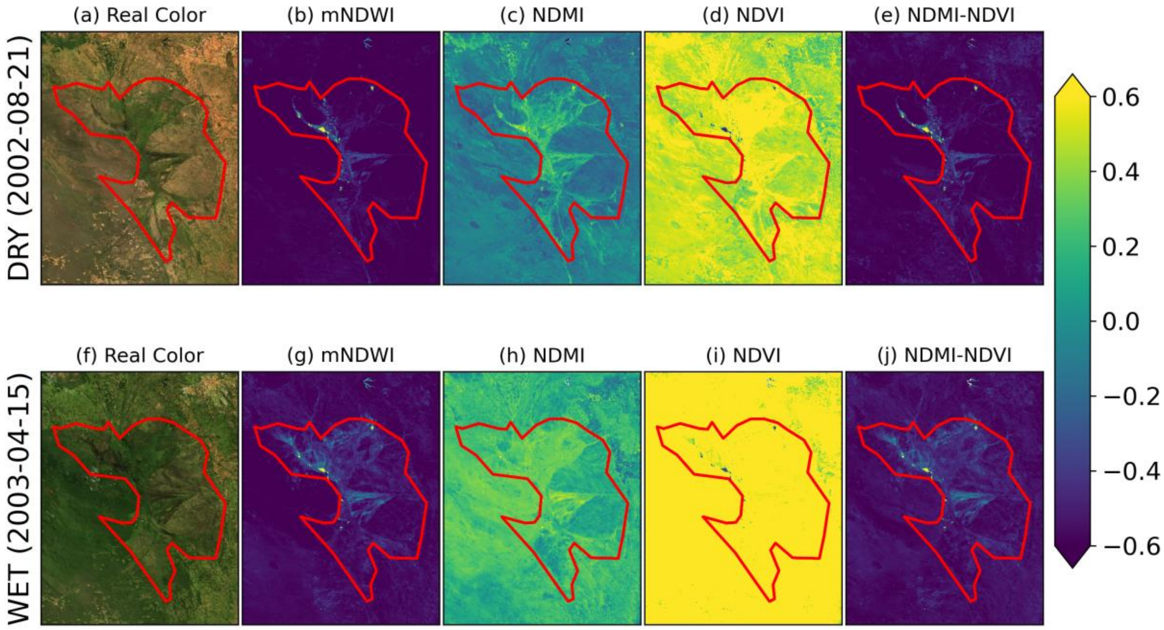


Figure 3: Results of the spectral indexes for two different dates: one during the dry period and one during the wet period.

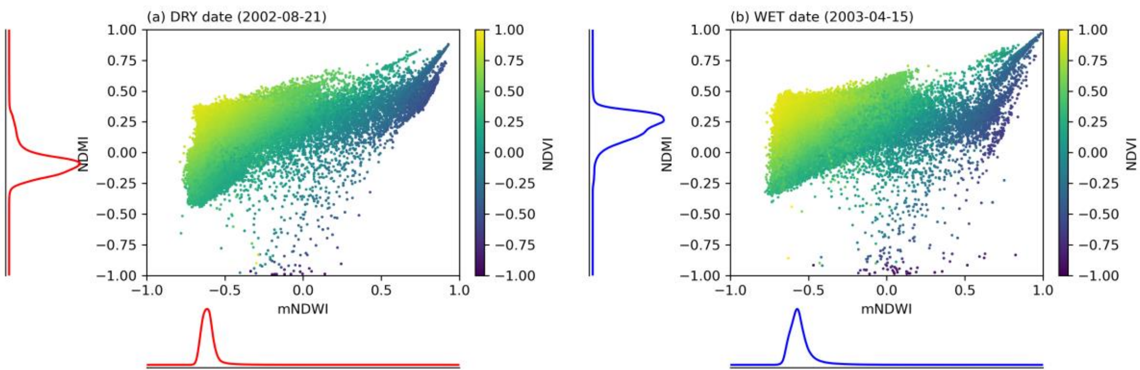


Figure 4: Distribution of the mNDWI / NDMI / NDVI values of the pixels.

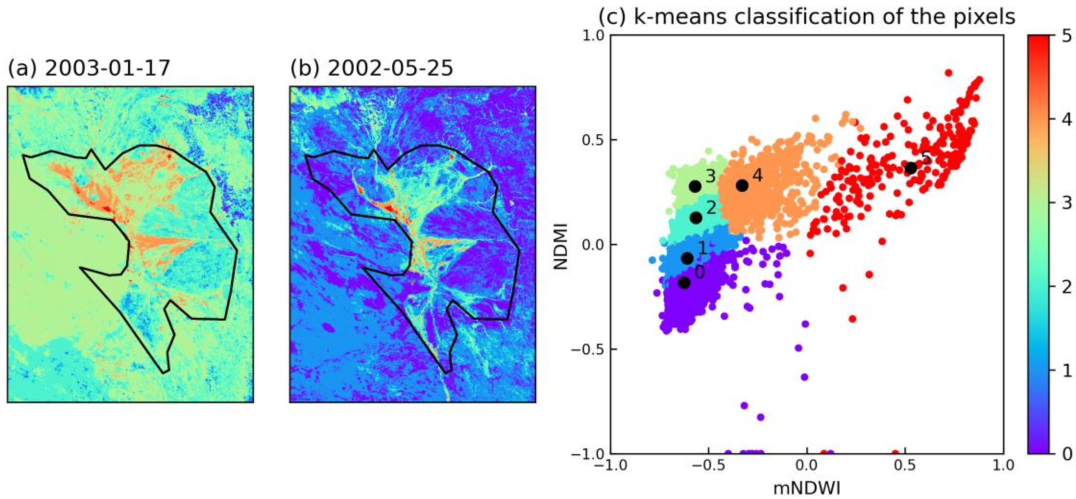


Figure 5: Illustration of k-means model output for k = 6 for (a) the wet and (b) the dry reference images and (c) distribution of the cluster in the (mNDWI / NDMI space).

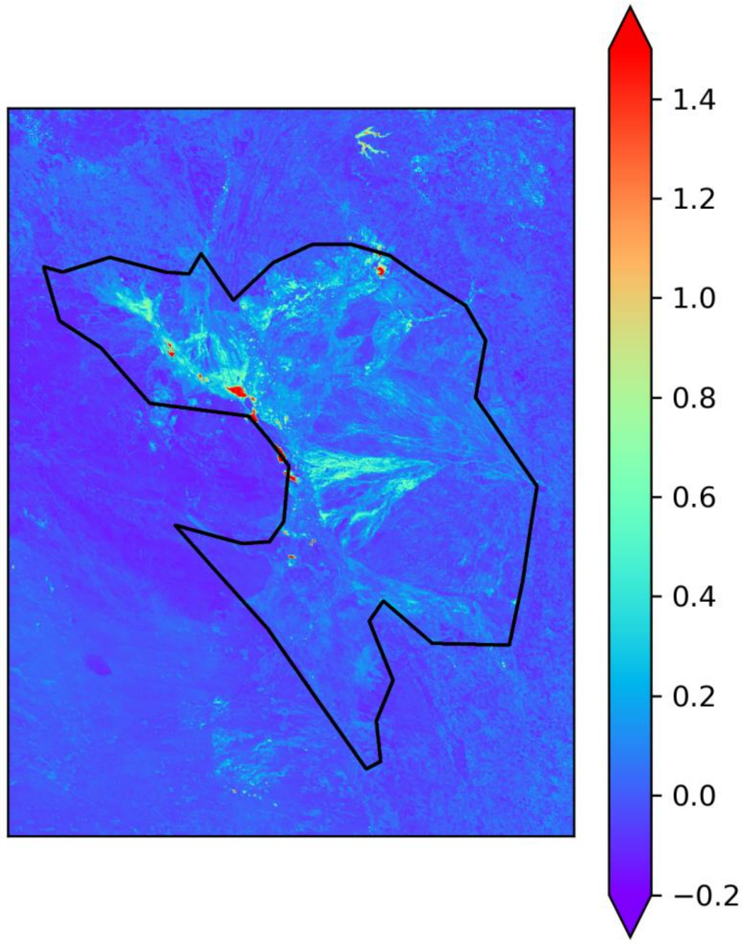


Figure 6: Values for the second dimension of the PCA for the wet reference image.

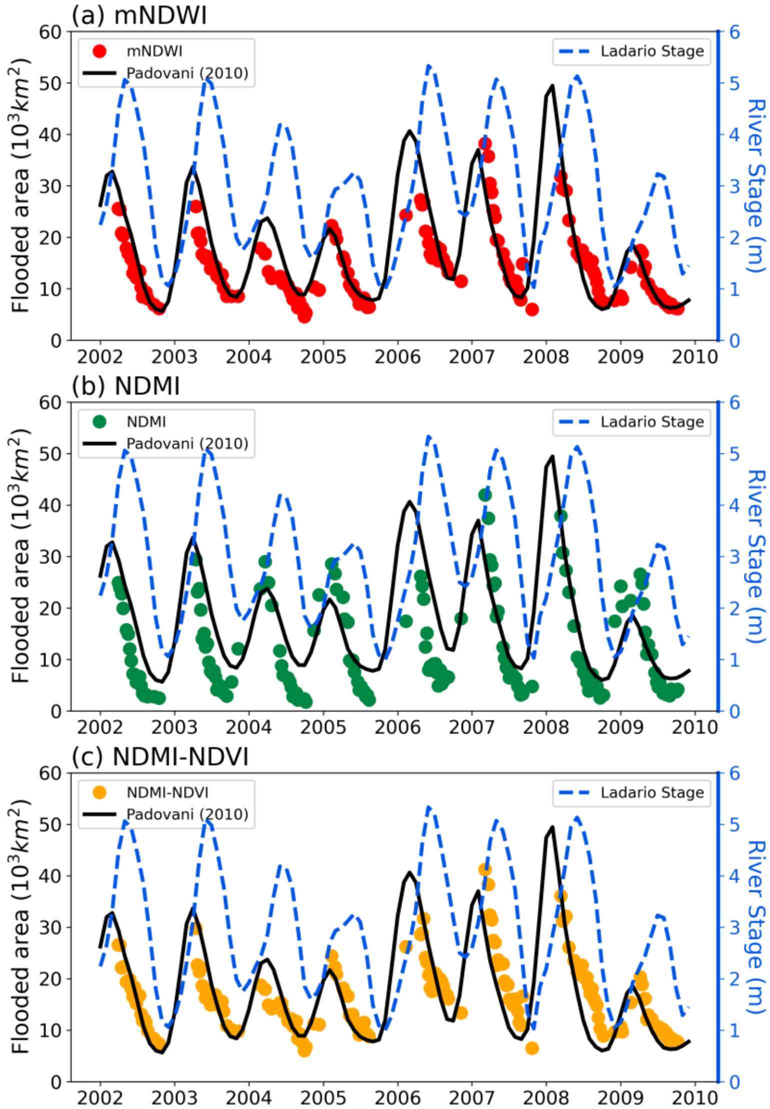


Figure 7: Time series of Padovani (2010), the river height at Ladário and of the results from the threshold-based methods using (a) mNDWI, (b) NDMI and (c) NDMI-NDVI.

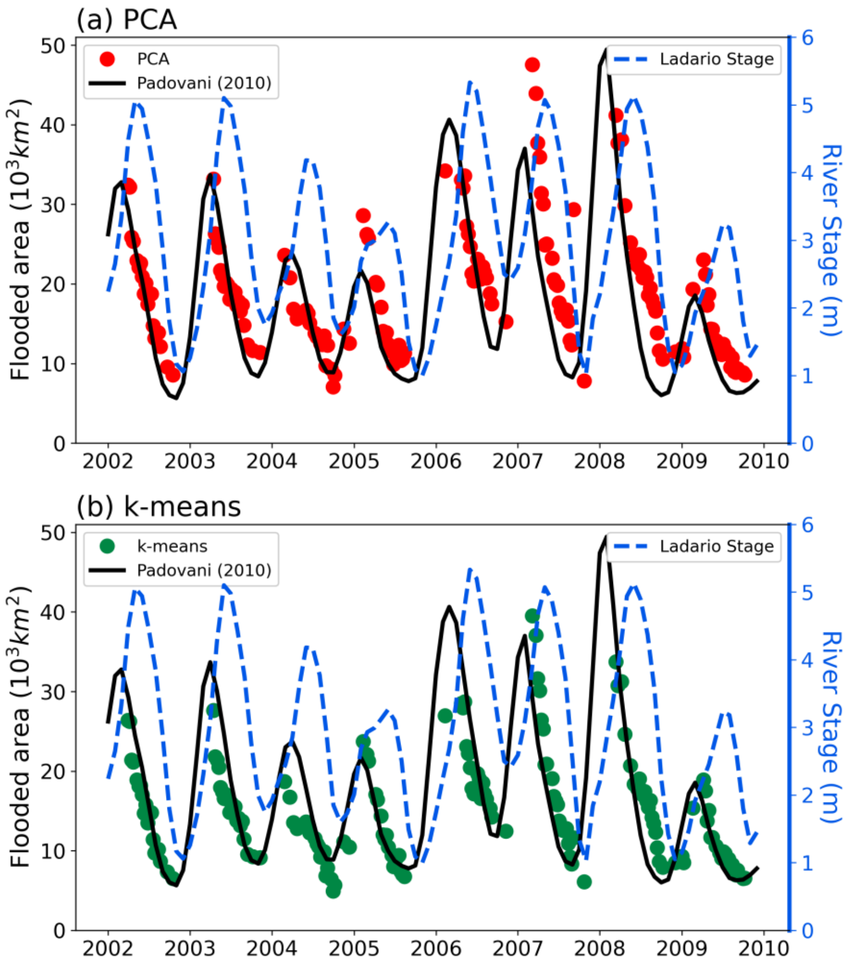


Figure 8: Time series of Padovani (2010), the river height at Ladário and of the results from the threshold-based methods using (a) the Principal Component Analysis (PCA) method and (b) the k-means algorithm with k = 6.

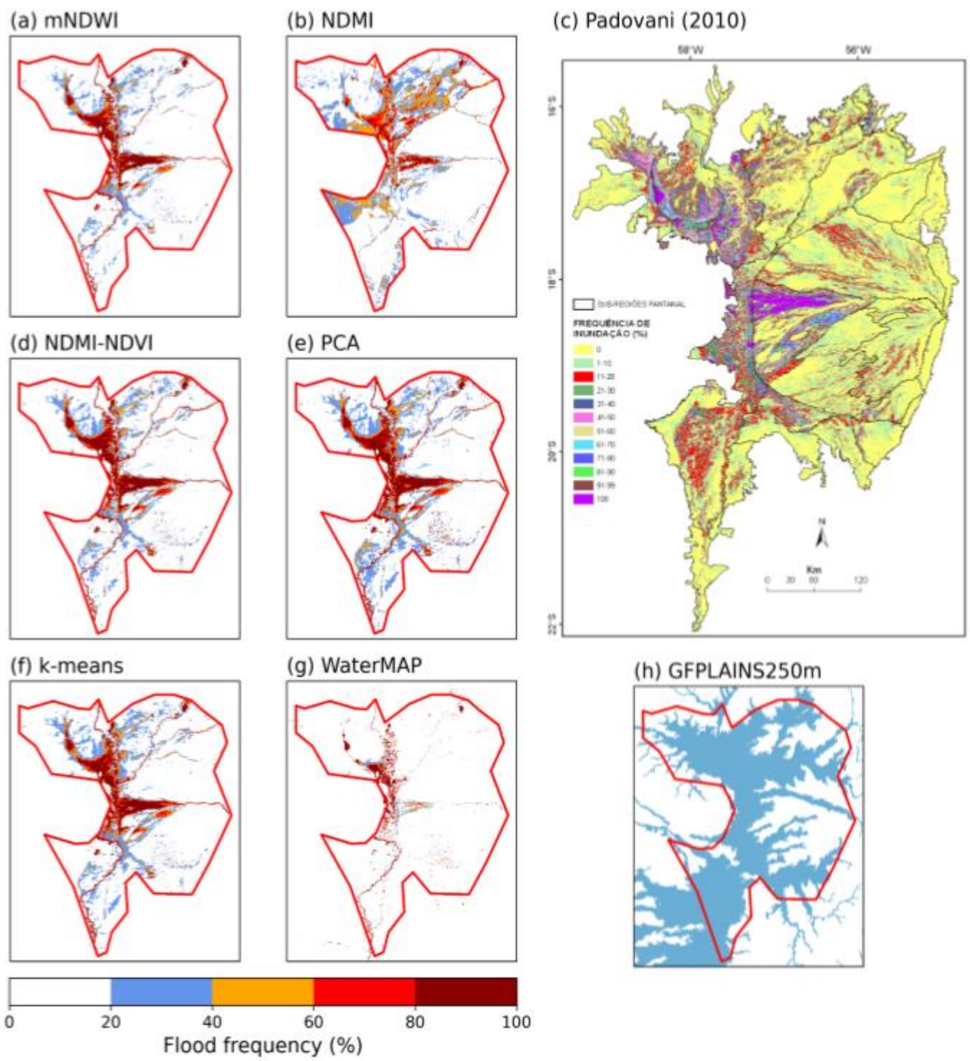


Figure 9: Flood frequency between 2002 and 2009 obtained from the different methods presented: 3 threshold-based methods using the (a) mNDWI, (b) NDMI and (d) NDMI-NDVI index and 2 unsupervised classification methods: (e) Principal Component Analysis and (f) k-means. Occurrence of flood from (c) Padovani (2010) and (g) WaterMAP (Pekel et al. 2016; Source: EC JRC/Google) between 1984 and 2015 and floodplains delimitation from GFPLAINS250m (Nardi et al., 2019).

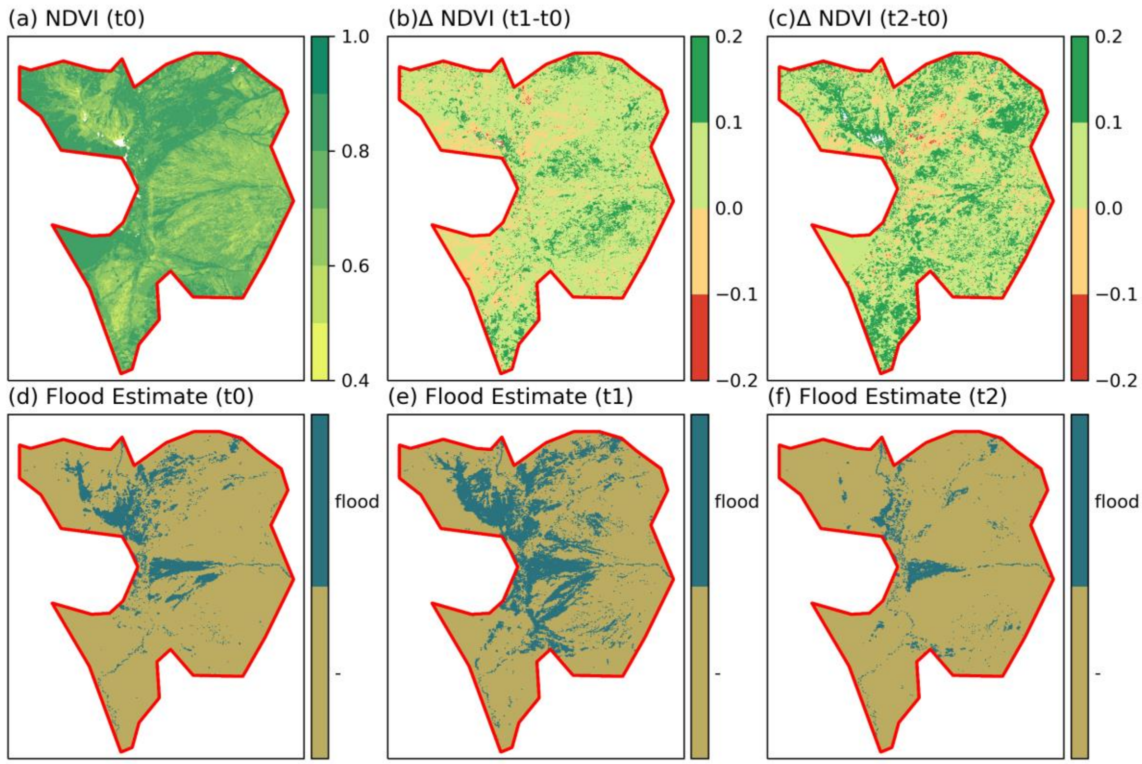


Figure 10: NDVI (a,b,c) and flood estimate (d,e,f) and for t0 (21/03/2004; a and d), t1(22/03/2007; b and e) and t2 (06/03/2021; c and f).

Spectral Indexes	References	Specificity
mNDWI = $\frac{-SWIR}{+SWIR}$	Xu (2006) Ogilvie et al. (2015)	Water detection
NDMI = $\frac{NIR-SWIR}{NIR+SWIR}$	Ogilvie et al. (2015)	Water detection
NDVI = $\frac{NIR-}{NIR+}$	Rouse et al. (1974)	Vegetation and water detection
NDMI-NDVI	Gond et al. (2004) Boschetti et al. (2014)	Rice flood mapping, water bodies and wetland

Table I: Spectral indexes considered in this study with some reference papers and the specificity of these indexes.

Method	Threshold
Threshold-based mNDWI	-0,465
Threshold-based NDMI	0,32
Threshold-based NDMI-NDVI	-0,45
K-Means	Cluster 4 and 5
PCA	0,09

Table II: Methods and their corresponding threshold values.

Method	PBIAS	RMSE	Correlation
mNDWI	-12,74	4.894	0,8
NDMI	-23,83	6.839	0,81
NDMI-NDVI	9,89	5.243	0,82
K-Means	11,46	5.119	0,77
PCA	-9,83	4.871	0,78

Table III: Resume of the statistics (Percentage bias - PBIAS, Root-Mean Square Error - RMSE, Correlation) comparing Padovani (2010) estimate with the different methods: threshold-based applied to mNDWI, NDMI and NDMI-NDVI, Principal Component Analysis (PCA) method and k-means with k = 6. The correlations are significant with a significance level of 99 %.