FMASK algorithm applied to Landsat images: a study case in the region of Santarém, Pará, Brazil

Elton Vicente Escobar-Silva[1*], Anielli Rosane de Souza[1*], Maria Lucia Ferreira Barbosa[1*], Cláudia Maria de Almeida[1], Liana Oighenstein Anderson[2], Gilberto Ribeiro de Queiroz[1] and Thales Sehn Körting[1]

[1]Earth Observation and Geoinformatics Division, National Institute for Space Research (INPE), São José dos Campos, São Paulo, Brazil

{elton.silva, anielli.souza, maria.lucia, claudia.almeida, gilberto.queiroz, thales.korting}@inpe.br

[2]National Center for Monitoring and Early Warning of Natural Disasters (Cemaden), São José dos Campos, São Paulo, Brazil

liana.anderson@cemaden.gov.br

*These authors contributed equally to the work.

Abstract: Clouds and cloud shadows are recurrent problems in optical remote sensing images. Algorithms have been developed for automatic detection of these targets such as the Function of Mask (Fmask) model. This work presents an automatic mapping of cloud areas with Fmask in Landsat images between 2017 and 2018 in northern Santarém, Pará, Brazil. The results show that the algorithm presents a good performance (85%) in the study area and is a powerful tool in clouds and cloud shadows detection. However, for further works, it is suggested to use data at a larger time scale and the evaluation of more pixels within the images, when comparing Fmask with other algorithms.

1. Introduction

Optical remote sensing images have a variety of applications that allow for new research areas in scientific research. However, considering land use and land cover (LULC) studies, even with some advances and upgrading, optical sensors are affected by clouds and cloud shadows, a problem in the visible and infrared spectral regions. These image effects limit the monitoring of LULC by obstructing targets, reducing the useful area of analysis, and interfere with the image quality itself (Jensen, 2009; Rees, 2013).

Automatic methods for mapping clouds and cloud shadows have been developed taking into account different approaches. Some studies rely on radiometric and color attributes to detect clouds and cloud shadows (Polidorio et al., 2006). Others are based on thresholds (Cucu-Dumitrescu, 2013) and on object analysis, such as the Function of mask (Fmask) method (Zhu and Woodcock, 2012), which is widely used both to eliminate unwanted noise in the pre-processing stage and to study the objects resulting from the data output (Pletsch et al., 2018; Foga et al., 2017; Braaten et al., 2015).

Fmask automatically extracts cloud, cloud shadows, snow, and water information from satellite images. The algorithm was developed by Zhu and Woodcock (2012) and its first version was developed only for Landsat-4 to 7 series images.

However, in 2015, Zhu et al. (2015) released a new version with some improvements and the algorithm was expanded to Landsat-8 and Sentinel-2 images. Fmask can be not only applied in the pre-processing of images, to minimize the inclusion of undesired targets, but also in the result analysis (Pletsch et al., 2018), which makes it a powerful tool for image analysis in areas with high recurrence of clouds.

The Amazon, located in the intertropical convergence zone (ITCZ), is characterized by a dense presence of clouds. Thus, the presence of clouds can be a major limiting factor in the use of remote sensing images in this region. Considering the great importance of the Amazon in the context of climate change and increasing pressures to change land cover, the development and validation of a tool aimed for cloud mask is essential for improvements in the use of optical images in this region, since the oversight of these areas are impacted by the clouds. In this context, the objective of the present study is (i) to apply the Fmask algorithm to detect clouds in Landsat images in the northern portion of Santarém, Pará; (ii) to analyze the number of times a pixel is classified as cloud; and (iii) to present the classification variation of the same pixel over time.

2. Materials and Methods

2.1 Study area

The area under analysis encompasses the 227/62 path/row (Figure 1) of Landsat 8. The region is located in the *'Baixo Amazonas'* mesoregion, northern Santarém in the Legal Amazon, and it includes part of the municipalities of Belterra, Mojuí dos Campos, Monte Alegre, Prainha, Santarém, and Uruará. Moreover, the area has different land regimes such as Protected Areas, INCRA (National Institute for Colonization and Agrarian Reform) settlement projects, indigenous lands, and agro-extractive settlement projects (IBGE, 2014; Souza et al., 2017).

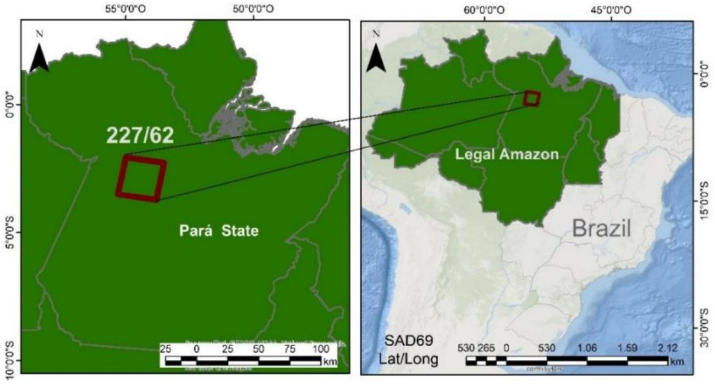

Figure 1: Geographic location of the study area.

2.2 Materials

Landsat has a totally free of charge database available on the United States Geological Survey (USGS) Earth Explorer website. On this website, images were acquired from the Operational Landsat Imager (OLI) sensor and from the Thermal Infrared Sensor (TIRS) Level-1 with radiometric and geometric corrections. The time series imagery extends over 2017 and 2018, totaling 24 images. Aiming to keep homogeneous weather

conditions throughout the time series, we selected images acquired around the same day in each month along the two analyzed years.

2.3 Methods

The methodology is structured into two major steps. The first step (Section 2.3.1) comprises the execution of Fmask and the second step (Section 2.3.2) describes the methodological approach for analyzing the algorithm results.

2.3.1 Fmask

In this study, the Python-Fmask version 0.5.4 package was used, executed in the Python language version 3.8.2. The Python code can be downloaded here. To map the clouds and cloud shadows, digital number (DN) Landsat-8 images were converted into top of the atmosphere (TOA) reflectance images by the package itself. It is worth noting that the most recent version of the Fmask package accepts both TOA and DN images as input. If one chooses to use the DN images, it is important to verify if the file containing the images metadata ('MTL.txt') is located in the same directory of the images themselves (Flood and Gillingham, 2015). Table 1 describes the spectral bands of the employed images.

Table 1. Spectral bands of Landsat-8 images used in this work.

OLI (µ m)	TIRS (µ m)
Band 1 (0.43–0.45)	Band 10 (10.60–11.19)
Band 2 (0.45–0.51)	Band 11 (11.50–12.51)
Band 3 (0.53–0.59)	
Band 4 (0.64–0.67)	
Band 5 (0.85–0.88)	
Band 6 (1.57–1.65)	
Band 7 (2.11–2.29)	
Band 8 (0.50–0.68)	
Band 9 (1.36–1.38)	

The main purpose of Fmask is mapping clouds, cloud shadows, and snow. However, with the algorithm improvements, the resulting data is a raster file containing six classes: (1) null; (2) no-cloud; (3) cloud; (4) cloud shadow; (5) snow; and (6) water. In this study, null values were removed, and the images were reclassified again from 1 to 5, starting with the class no-cloud and ending with water.

Lastly, this algorithm identifies pixels with a high probability of being clouds according to the physical properties of a cloud, separating them from other targets on the Earth's surface. The Python-Fmask package used in this work is structured in a high-level Python programming language. A concise description of the package can be found in Flood and Gillingham (2015). In the command terminal, the Fmask package was executed for all 24 images in the analyzed period.

2.3.2 Classification analysis

Initially, the frequency at which a pixel was classified as cloud in the study period was assessed. This analysis was divided into two stages. In the first one, a qualitative

analysis was carried out through visual interpretation, where images from the Landsat series were compared to the classification result, in order to observe the correspondence between areas mapped as cloud and no-cloud areas in the image. A vertical bar graph was also created using the matplotlib library, considering the total pixels mapped in each class for July 2018 as an experimental period. Furthermore, 100 random points were generated in QGIS and visually checked if they were correctly classified. In the second stage, a temporal analysis at a pixel level was performed, where it was possible to observe the class changes within the analyzed period. The raster values were extracted over time from a random pixel centroid corresponding to the Fmask classes. For visualization purposes, bar graphs were plotted using the matplotlib.

3. Results

Fmask presented a satisfactory performance in the detection of clouds in the study area. By visual comparison, it was observed that the algorithm correctly classified the resulting classes (no-cloud, cloud shadow, snow, and water). 85% of randomly generated points were classified correctly. Most of the error is associated with the classification of the pixel with cloud or shadow when in fact the pixel is neither related to these two classes. According to Qiu et al. (2019), there is no optimal global threshold for cloud detection, so local thresholds should be applied by users according to a given region´s peculiarities. Furthermore, accurately detecting cloud shadows is a major challenge as many dark surfaces (e.g., wetlands, burnt areas and terrain shadows) are easily mistaken for cloud shadows (Zhu et al. 2015; Qiu et al. 2019).

Considering the Fmask result for July 2018, the dry season, it was observed that the classification generated accurate results when compared to the Landsat image (Figure 2). In July 2018, the water class represented 11.80% of the total pixels of the classified image (Figure 3). This class is associated with the Tapajós, Arapius, and Amazon rivers, which were correctly identified by the algorithm. The cloud shadow class was registered as 7.80% of the total pixels. The cloud and cloudless classes, the largest classes in the image, represented 14.24% and 66.14% of the pixels, respectively. Finally, no pixels belonging to Class 4 (snow) were identified, indicating a good accuracy of the algorithm since this class does not occur in the study area.

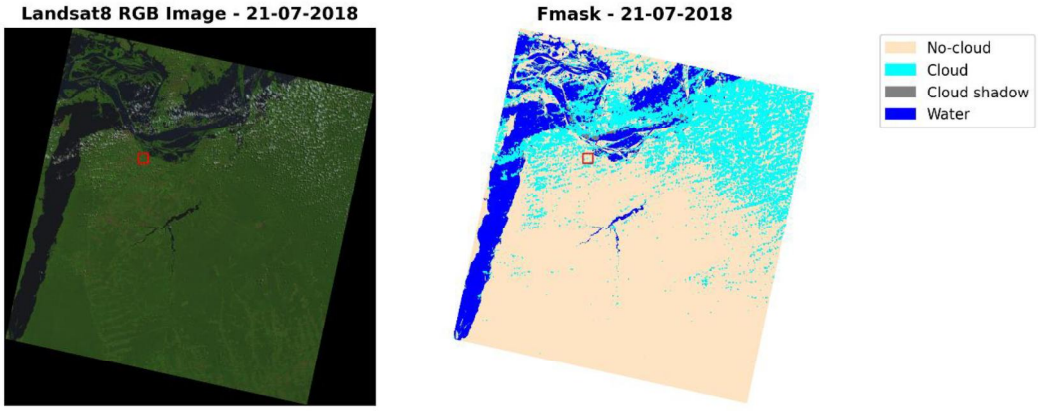

Figure 2: RGB Landsat 8 image (left) and Fmask (right) for the same scene on 17 Jul 2018. The red box on both images represents the centroid location of the analyzed pixel in this work.

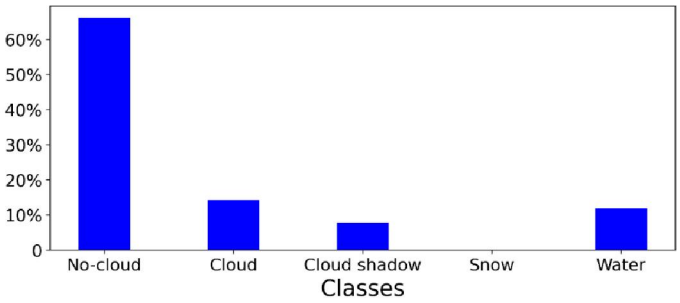

Figure 3: Percentage of mapped pixels per class in July 2018.

The class change corresponding to the analysis of a pixel over the 24 images (2017 and 2018) is presented in Figure 4. In this case, the pixel was identified within only two classes, with cloud and no-cloud, Classes 1 and 2, respectively. It is possible to identify that the pixel was identified as cloud in most months, even during the dry season in the two analyzed years. This pattern could be explained by the low representation of a pixel when compared to the full scene.

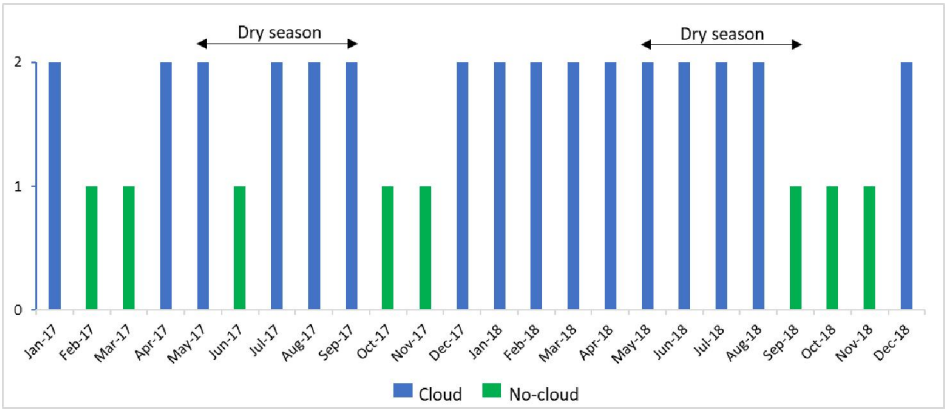

Figure 4: The class variation between 2017 and 2018 of the selected pixels. See pixel location in Figure 2.

In 2017, the analyzed pixel was identified as cloud in 58.3% of the months and as no-cloud area in 41.7%. In 2018, the cloud proportion raised to 75%, while the share of no-cloud classification dropped to 25%. When observing the two years altogether, in 66.7% of the time series, the pixel was registered as cloud, and in 33.3% of it as no-cloud. Lastly, as expected, the algorithm identified a greater presence of clouds in the images referring to the rainy season (October to April) and a greater presence of cloudless areas in the dry season (May to September).

4. Final considerations

The Fmask algorithm showed a good performance in the study area (85% of accuracy). The snow class, which does not occur in the Amazon, was not computed by the algorithm, indicating a reliable result. However, the algorithm reduces the image size by removing some columns in the image edges, which can cause a small loss of image information. As directions for future work, it is recommended to compare the performance of Fmask with other cloud detection algorithms in the same study area.

5. Acknowledgments

E.V.E.S. thanks the São Paulo Research Foundation (FAPESP, Grant 2021/11435-4). E.V.E.S., A.R.S. and M.L.F.B. thank the National Council for Scientific and Technological Development (CNPq, Funding Code 001). C.M.A. thanks FAPESP (Grant 2020/09215-3) and CNPq (Grant 311324/2021-5).

6. References

BRAATEN, Justin D.; COHEN, Warren B.; YANG, Zhiqiang. 2015. Automated cloud and cloud shadow identification in Landsat MSS imagery for temperate ecosystems. *Remote Sensing of Environment*, 169, pp.128-138.

CUCU-DUMITRESCU, Catalin. 2013. A simple method of determining cloud-masks and cloud-shadow-masks from satellite imagery. *IEEE Geoscience and Remote Sensing Letters*, 11(1), pp.10-13.

FLOOD, N.; GILLINGHAM, S. 2015. *Python Fmask documentation*. Disponível em <http://www.pythonfmask.org/en/latest/index.html> Acessado em Jun 06 2022.

FOGA, Steve; SCARAMUZZA, Pat L.; GUO, Song; ZHU, Zhe; DILLEY JR, Ronald D.; BECKMANN, Tim; SCHMIDT, Gail L.; DWYER, John L.; HUGHES, M. Joseph; LAUE, Brady. 2017. Cloud detection algorithm comparison and validation for operational Landsat data products. *Remote sensing of environment*, 194, pp.379-390.

IBGE - Instituto Brasileiro de Geografia e Estatística. *Redes e Fluxos do Território:* Gestão do Território. 2014.

JENSEN, John R., 2009. *Remote sensing of the environment:* An earth resource perspective. 2nd ed. Pearson Education India.

PLETSCH, Mikkaela A. J. S.; PENHA, T. V.; SILVA JUNIOR, C. H. L.; KÖRTING, Thales S.; ARAGÃO, Luiz E. O. e C.; ANDERSON, Liana O. 2018. Integração do algoritmo fmask ao modelo linear de mistura espectral como subsídio à detecção de áreas queimadas na Amazônia brasileira. *Revista Brasileira de Cartografia*, vol. 70, pp. 696–724.

POLIDORIO, Airton M.; IMAI, Nilton N.; TOMMASELLI, Antonio M. G.; FLORES, Franklin C. 2006. Detection and Discrimination of Shades, Clouds and Waterbodies in Remote Sensing Images. *Revista Brasileira de Cartografia*, vol. 58, pp. 223-232.

QIU, Shi; ZHU, Zhe; HE, Binbin. (2019). Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sensing of Environment*, 231, p.111205.

REES, William G. 2013. *Physical Principles of Remote Sensing*. 4th ed. Cambridge: Cambridge University Press.

SOUZA, Anielli R.; ESCADA, Maria I. S.; VIEIRA, Antônio M. V. 2017. Padrão da paisagem associado ao uso e cobertura da terra em comunidades ribeirinhas e de terra firme situadas no sudoeste do Pará. *Geografia*, vol. 42, no. 2, pp. 135–164.

ZHU, Zhe; WANG, Shixiong; WOODCOCK, Curtis E. 2015. Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote sensing of Environment*, 159, pp.269-277.

ZHU, Zhe; WOODCOCK, Curtis E. 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote sensing of environment*, 118, pp.83-94.