# LAND USE AND LAND COVER MAPPING AND SPATIOTEMPORAL SEGMENTATION BASED ON IMAGE TIME SERIES CLUSTERING

Baggio Luiz de Castro e Silva

Master's Dissertation of the Graduate Course in Applied Computing, guided by Drs. Karine Reis Ferreira Gomes, and Gilberto Ribeiro de Queiroz, approved in January 30, 2023.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/48DF32P>

INPE

São José dos Campos

2023

sid.inpe.br/mtc-m21d/2023/01.21.03.35-TDI

# LAND USE AND LAND COVER MAPPING AND SPATIOTEMPORAL SEGMENTATION BASED ON IMAGE TIME SERIES CLUSTERING

Baggio Luiz de Castro e Silva

Master's Dissertation of the Graduate Course in Applied Computing, guided by Drs. Karine Reis Ferreira Gomes, and Gilberto Ribeiro de Queiroz, approved in January 30, 2023.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34T/48DF32P>

INPE

São José dos Campos

2023

**INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS**

**DEFESA FINAL DE DISSERTAÇÃO BAGGIO LUIZ DE CASTRO E SILVA**
**BANCA Nº 006/2022, REG. 309710/2021**

No dia 30 de janeiro de 2023, as 09h, por teleconferência, o(a) aluno(a) mencionado(a) acima defendeu seu trabalho final (apresentação oral seguida de arguição) perante uma Banca Examinadora, cujos membros estão listados abaixo. O(A) aluno(a) foi APROVADO(A) pela Banca Examinadora, por unanimidade, em cumprimento ao requisito exigido para obtenção do Título de Mestre em Computação Aplicada, com a exigência de que o trabalho final a ser publicado deverá incorporar as correções sugeridas pela Banca Examinadora, com revisão pelo(s) orientador(es).

**Novo título: "Land use and land cover mapping and spatiotemporal segmentation based on image time series clustering".**

**Membros da Banca:**

Dr. Thales Sehn Körting – Presidente – INPE
Dra. Karine Reis Ferreira Gomes – Orientador(a) – INPE
Dr. Gilberto Ribeiro de Queiroz – Orientador(a) – INPE
Dr. Rafael Duarte Coelho dos Santos – Membro Interno – INPE
Dr. Alana Kasahara Neves – Membro Externo – CEF/ISA/ULisboa

Documento assinado eletronicamente por **Alana kasahara NEVES (E)**, **Usuário Externo**, em 03/02/2023, às 18:26 (horário oficial de Brasília), com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.mcti.gov.br/verifica.html, informando o código verificador **10782304** e o código CRC **329C01A3**.

---

**Referência:** Processo nº 01340.000615/2023-42                                                                              SEI nº 10782304

"Feliz aquele que transfere o que sabe e aprende o que ensina".

# ACKNOWLEDGEMENTS

# ABSTRACT

Land use and land cover mapping is fundamental in various applications such as environmental monitoring, public policy support, territorial management, and many others. The quality of this mapping directly affects its applications, reflecting the quality of the product and the methods used. Many images have been produced with different spatial, temporal, and spectral technologies through technological advances in remote sensing. In addition to producing these resources, open access policies favored the mapping and dissemination of various national and international land use and land cover products. One of the oldest and most widely used techniques in this mapping process is segmentation, which aims to find homogeneous regions in the image based on some attribute. However, this technique only considers information from a single image, which can be affected depending on the cloud cover. One such technique is spatiotemporal segmentation, which aims to find homogeneous regions in space and time. Therefore, this dissertation proposes a spatiotemporal segmentation method and a land use and land cover classification method using clustering methods with a *"time-first, space-later"* approach. These methods are based on unsupervised algorithms: Self-Organizing Map (SOM) and the Hierarchical Clustering Algorithm (HCA), where the input data are time series extracted from the satellite images to be mapped. Finally, a method for interpolating clouds in the time series of satellite images using Deep Learning was proposed since this is a crucial step in clustering methods that use distance as a similarity requirement. The three proposed approaches were examined, adhering to visual and quantitative assessment principles. In the spatiotemporal segmentation, the methodology was applied to a heterogeneous region southeast of Mato Grosso state. For land use and land cover classification, six distinct regions with varied characteristics of inland waters were selected, which served as a training base and were later applied in four other regions for mapping inland waters. Regarding cloud interpolation in time series, time series distributed throughout Brazil were chosen, using the Sentinel-2/MSI cloud mask (Scene Classification Layer - SCL) as the basis for reference interpolation.

Keywords: Satellite image time series. LULC Classification. Spatiotemporal segmentation. Inland water bodies. Clustering methods.

# MAPEAMENTO DO USO E COBERTURA DO SOLO COM BASE EM SEGMENTAÇÃO E AGRUPAMENTO DE SÉRIES TEMPORAIS

## RESUMO

O mapeamento de uso e cobertura da terra tem papel fundamental em diversas aplicações como no monitoramento ambiental, no suporte à políticas públicas, na gestão territorial, entre outras. A qualidade desse mapeamento impacta diretamente nas suas aplicações, sendo este o reflexo da qualidade do produto e das metodologias utilizadas. Graças ao avanço tecnológico na área de sensoriamento remoto nos últimos anos, grandes volumes de imagens vêm sendo produzidos com diferentes resoluções espaciais, temporais e espectrais. Aliadas à produção destes recursos, as políticas de acesso aberto favoreceram o mapeamento e a disseminação de diferentes produtos de uso e cobertura da terra nacionais e internacionais. Uma das técnicas mais antigas e difundidas neste processo de mapeamento é a segmentação, que visa encontrar regiões homogêneas na imagem com base em algum atributo. Entretanto, esta técnica considera apenas informação de uma única imagem, que pode ser prejudicada dependendo da cobertura de nuvem. Uma variante desta técnica é a Segmentação espaçotemporal, que tem como objetivo não só encontrar regiões homogêneas no espaço mas também no tempo. Desta maneira, esta dissertação propõe um método segmentação espaçotemporal e um método de classificação de uso e cobertura da terra, baseados em métodos de clusterização utilizando uma abordagem *"time-first, space-later"*. Estes métodos têm como base os algoritmos não supervisionados: Self-Organizing Map (SOM) e o Hierarchical Clustering Algorithm (HCA), em que os dados de entrada são séries temporais extraídas das próprias imagens de satélite a serem mapeadas. Por fim foi proposto um método de interpolação de nuvens em séries temporais de imagens de satélite utilizando deep learning visto que esse é um passo crucial em metodologias de clustering que usam a distância como critério de similaridade. As três abordagens propostas foram examinadas, aderindo aos princípios de avaliação visual e quantitativa. Na segmentação espaçotemporal, aplicou-se a metodologia em uma região heterogênea localizada no sudeste do Estado de Mato Grosso. Para a classificação de uso e cobertura, selecionaram-se seis regiões distintas com características variadas de águas interiores, as quais serviram como base de treinamento e foram posteriormente aplicadas em quatro outras regiões para o mapeamento de águas interiores. No que concerne à interpolação de nuvens em séries temporais, foram escolhidas séries temporais distribuídas por todo o Brasil, utilizando-se a máscara de nuvem Sentinel-2/MSI (Scene Classification Layer - SCL) como base para interpolação de referência.

Keywords: Séries temporais de imagem de satélite. Classificação de uso e cobertura. Segmentação espaço temporal. Corpos d'água interiores. Métodos de agrupamento.

# LIST OF FIGURES

# LIST OF TABLES

# CONTENTS

# 1 INTRODUCTION

Land use and land cover (LULC) maps are important indicators of the human footprint and are essential to support the development of efficient and effective environmental and territorial policies. These maps are crucial data sources for measuring global indicators such as the United Nations' Sustainable Development Goals (SDGs), including land degradation, freshwater ecosystems, or urban development (GIULIANI et al., 2020). LULC maps are mainly produced from Earth Observation (EO) satellite imagery (SAAH et al., 2020).

Nowadays, different missions are launching EO satellites and producing big open data consistently calibrated over time and space. Open data policies adopted by many space agencies worldwide make accessible petabytes of EO satellite images of different spatial, spectral, and temporal resolutions (SOILLE et al., 2018).

To produce LULC maps from these big data sets of EO satellite images, image time series analysis and machine learning methods are being widely used with promising results (GÓMEZ et al., 2016; BELGIU; CSILLIK, 2018; PICOLI et al., 2018; SIMOES et al., 2020; SIMOES et al., 2021). Distinct methods have been applied to different kinds of satellite image time series to produce LULC maps, including Time-Weighted Dynamic Time Warping (TWDTW), Random Forest (RF) and Support Vector Machine (SVM). These methods are based on supervised learning, that require training samples labeled *a priori*.

To mainly support image time series analysis, big data sets of EO satellite images have been modeled as EO data cubes (APPEL; PEBESMA, 2019; KOPP et al., 2020). The concept of a data cube refers to structuring a multidimensional data set to organize it, simplify manipulation, and improve search and analysis (KOPP et al., 2020). EO Data Cubes are multidimensional arrays with space and time dimensions and spectral derived properties created from EO satellite imagery (APPEL; PEBESMA, 2019). These arrays improve data manipulation and make them available for analysis and consequently support studies in the field that is increasingly concerned with models instead of processing the data correctly.

Recent initiatives to create EO data cubes for specific countries or regions include the Australian Data Cube (LEWIS et al., 2017), the Swiss Data Cube (GIULIANI et al., 2017), the Africa Regional Data Cube (KILLOUGH, 2019) and the Brazil Data Cube (FERREIRA et al., 2020b). Meanwhile, novel software platforms have been developed to access, prepare, index, and handle big amounts of remote sensing images

and EO data cubes, such as Open Data Cube (ODC), Processing and Analysis for Land Monitoring (SEPAL) and Google Earth Engine (GEE) (GOMES et al., 2020). These initiatives and platforms are paving the way for broadening the use of EO data to larger communities of users, supporting decision-makers with information converted into meaningful geophysical variables, and ultimately unlocking the information power of EO data (GIULIANI et al., 2019).

Brazil Data Cube is a project that is being developed by INPE (National Institute for Space Research) and that aims to create multidimensional data cubes of spatial medium-resolution remote sensing images for all Brazilian territory and to produce Land use and land cover information from these data cubes using machine learning and satellite image time series analysis.

## 1.1   Motivation and objectives

Time series extracted from EO satellite images over the same areas, modeled as data cubes, improve our understanding of environmental patterns and processes. Instead of selecting individual images from specific dates and comparing them, researchers can track changes continuously. Satellite image time series analysis captures subtle changes in ecosystems and improves the quality of land classification (SIMOES et al., 2021).

Considering a dense sequence of EO satellite images, Camara et al. (2016) define the term *space-first, time-later* to refer to approaches which apply methods separately to the entire image of this sequence and later join the results over time; and the term *time-first, space-later* to approaches which apply methods to the time series associated to individual pixels of this sequence. They argue that *time-first, space-later* approaches use the full potential of remote sensing image time series, and that temporal autocorrelation is stronger than spatial autocorrelation for change detection. According to Pasquarella et al. (2016) time series derived from remote sensing images allow detecting complex underlying processes that would be difficult to identify using bi-temporal or other traditional change detection approaches.

The main objective of this dissertation is to explore image time series clustering methods for spatiotemporal segmentation and Land use and land cover mapping, following the *time-first, space-later* approach. In this work, Self-Organizing Maps (SOM) neural network combined with Hierarchical method is used to produce clusters of image time series. SOM maps the input data (high dimensionality) onto a bi-dimensional grid (low dimensionality), keeping the topology of the neighbor-

hood; that is, similar time series tends to be close in 2D space (KOHONEN, 1990). Due to these properties, SOM has been widely used for image time series clustering (ZURITA-MILLA et al., 2013) (SANTOS et al., 2021).

Also, in the context of this dissertation, deep learning approaches were evaluated to interpolate cloudy pixels in Sentinel-2 image time series. Optical remote sensors are extremely susceptible to clouds. Clouds and their shadows affect remote sensing image processing methods to automatically identify and classify Land use and land cover types. In image time series analysis, interpolation techniques are used to produce valid values when pixels are covered by clouds or shadows. In this dissertation, we evaluated deep learning approaches to interpolate cloudy pixels.

## 1.2   Dissertation contributions

The main contributions of this dissertation are:

- A methodology for spatiotemporal segmentation of satellite image time series using Self-Organizing Maps (SOM). This methodology is described in Chapter 2, based on the article published by Silva et al. (2022);

- A methodology for Land use and land cover mapping based on image time series clustering. This methodology was applied to classify inland water bodies, that is a specific type of land cover. It is described in Chapter 3;

- An evaluation of deep learning approaches to interpolate cloudy pixels in Sentinel-2 image time series. This study is described in Chapter 4, based on the article accepted to be published in the XX Brazilian Remote Sensing Symposium (SBSR 2023) proceeding by Silva et al. (2023).

The following academic and technological contributions are part of this dissertation:

a) Authorship or co-authorship in scientific articles:

- *Spatiotemporal segmentation of satellite image time series using self-organizing map* (SILVA et al., 2022);

- *Deep Learning as tool to interpolate cloudy pixels in sentinel-2 time series* (SILVA et al., 2023);

- *Uso de séries temporais para classificações de uso e cobertura da terra em Petrolina, Pernambuco* (BRITO et al., 2023).

b) Free and open source code:

- Spatiotemporal Segmentation (GitHub: https://github.com/brazil-data-cube/st-segmentation);

- LULC Classification (GitHub: https://github.com/BaggioCastro/dissertation-content).

# 2 SPATIOTEMPORAL SEGMENTATION OF SATELLITE IMAGE TIME SERIES USING SELF-ORGANIZING MAP[1]

## 2.1 Introduction

Remote sensing imagery is essential for monitoring environmental changes, natural disasters, population growth, and many other applications. Nowadays, a large number of Earth observation satellites have generated big volumes of images with different spatial and temporal resolutions every day. In 2019, the volume of open data produced by Landsat-7 and Landsat-8, MODIS (Terra and Aqua units), and the three first Sentinel satellites (Sentinel-1, -2 and -3) were around 5 PB(SOILLE et al., 2018). This scenario poses a great challenge for storing, processing, and analyzing this large amount of data and also promotes the use of satellite image time series for spatiotemporal analysis (SIMOES et al., 2021).

Spatiotemporal segmentation consists in finding homogeneous regions in space and time from remote sensing images based on spectral features (PETITJEAN et al., 2012). Remote sensing image segmentation is widely used in object-based approaches to produce homogeneous geographical objects (XI et al., 2019). Despite being an important step in remote sensing image processing, there are few methods proposed in the literature that consider the space and time dimensions to produce homogeneous segments (PETITJEAN et al., 2012; COSTA et al., 2018b; XI et al., 2019; GARNOT; LANDRIEU, 2021)

Petitjean et al. (2012) proposed an approach that segments each image of the series and then combines these segments to provide a classification map. Costa et al. (2018b) propose a method for segmentation applied to image time series, adapting the traditional region growing algorithm and using the Dynamic Time Warping (DTW) similarity measure to detect homogeneous regions in space and time. Xi et al. (2019) describe a multiresolution segmentation to generate multiscale spatiotemporal cubes. Garnot e Landrieu (2021) present a method that processes a sequence of images in parallel by a shared convolutional encoder. At the lowest resolution, a temporal encoder produces a set of temporal attention masks for each pixel, then spatially interpolated at all resolutions.

---

[1]This chapter is an adaptation of SILVA, B.L.C.; SOUZA, F.C.; FERREIRA, K.R.; QUEIROZ, G.R.; SANTOS, L.A. Spatiotemporal segmentation of satellite image time series using self-organizing map. ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., V-3-2022, 255–261, 2022

Different from the existing methods, this document presents a spatiotemporal segmentation of satellite image time series using a self-organizing map (SOM) and hierarchical clustering algorithm. SOM is an unsupervised artificial neural network based on competitive learning that reduces a high dimensional feature input space onto a lower-dimensional feature output space (KOHONEN, 1998). It also preserves the neighborhood topology; that is, similar input data are mapped to the same neuron or a nearby one. SOM has been widely used for image time series clustering and spatiotemporal patterns discovering (ZURITA-MILLA et al., 2013) and (SANTOS et al., 2021).

The proposed approach is unsupervised and applied to satellite image time series; that is, time series associated with pixels of a satellite image sequence ordered in time. Based on a given threshold or a number of distinct clusters, it computes the similarity among all image time series to produce homogeneous regions in space and time. The proposed method was implemented and tested in a region in the Mato Grosso state, Brazil. The code and data sets used in this work are available on Github to promote reproducibility[2].

Satellite image segmentation is an important step for various remote sensing tasks, such as analysis, data exploration, and classification. Currently, there are some methods that use only one image in time for segmentation. Methods that explore time and space can use different approaches. Basically, there are methods that use snapshots that compare individually segmented images in time, and there are methods that use time series for integration of regions.

This proposed approach uses time series to integrate regions. The main difference with methods that use only one image in time for segmentation is the ability to distinguish segments based on their spatiotemporal dynamics. Compared to the model of Costa et al. (2018b), which uses the same basis, i.e. homogeneous regions in space and time using time series, the proposed method uses the clustering approach, where it is not necessary to define seeds, defining the parameters for clustering methods, such as the number of neurons for the Self Organizing Map (SOM). In the proposed approach, an unsupervised classification map is also created, which shows the temporal profiles of each segment and can be used in other applications.

An approach to finding homogeneous regions in space and time using satellite image time series is very useful for identifying some types of land use and land cover

---

[2]`https://github.com/brazil-data-cube/st-segmentation`

classes. Some classes, such as one-cycle and two-cycle agriculture types, are correctly identified when we consider the temporal variation of satellite image spectral values. Regions that are used for one-cycle and two-cycle crops are homogeneous during an cropping year. Thus, it is important to consider the spatial and temporal dimensions to extract these classes from satellite image sequences ordered in time.

## 2.2   Clustering methods

The method is based on two unsupervised clustering algorithms: SOM and agglomerative hierarchical clustering. The following describes each method in detail.

### 2.2.1   Self Organizing Map (SOM)

SOM is an unsupervised learning method based on competitive learning that reduces a high-dimensional feature input space to a low-dimensional feature output space. A dataset can be mapped and represented by a set of neurons by using weight vectors. Preserve neighborhood topology is a key feature of SOM so that similar input data is mapped to the same or a neighboring neuron.

Each output space neuron $j$ has a weight vector (codebook vector) associated $w_j = [w_{j1}, \ldots, w_{jn}]$ that contains the same dimension $n$ of the input data $x(t) = [x(t)_1, \ldots, x(t)_n]$. The algorithm consists of two main steps. First, the distances $D_j$ between a sample and each neuron in the SOM grid are calculated. The neuron with the smallest distance value $d_b$ is selected as the best matching unit (BMU) for that sample. The equations for calculating distance and BMU are

$$D_j = \sqrt{\sum_{i=1}^{n}(x(t) - w_{ji})^2} \tag{2.1}$$

$$d_b = min(D_1, ..., D_j) \tag{2.2}$$

The second step is to adjust the weight vectors of the BMU and its neighbors so that the neurons have similar properties. The equation for the adjustment is

$$w_{ji} = w_{ji} + \alpha \times h_{b,j}[x(t)_i - w_{ji}]. \tag{2.3}$$

Where $\alpha$ is the learning rate and $h_{b,j}$ is the neighborhood function.

These steps are repeated $T$ times to ensure that the neurons organize themselves into a similar neighborhood. Then, one neuron is assigned for each sample in the dataset (KOHONEN, 1998).

The grid size defines the number of neurons in the SOM, which consequently represents the number of clusters to be formed and the distribution of the input data (KöRTING et al., 2011).

### 2.2.2 Hierarchical clustering

Hierarchical clustering is a method that partitions data sequentially, creating a hierarchy of clusters (EVERITT et al., 2011). This representation makes it easier to visualize each step where level similarity occurs. Hierarchical clustering is classified into two types: agglomerative and divisive. The divisive algorithm begins with the entire dataset in one cluster and then divides it into two more similar clusters. The agglomerative method starts with each weight vector in its own cluster, computes a similarity matrix, and identifies the two most similar groups. The clusters are merged at each step, and the hierarchy is built using linkage criteria (KAUFMAN, 1990).

Hierarchical algorithms construct a binary tree called a dendrogram. This structure represents the order in which the clusters were merged. A dendrogram divides the data into groups that are homogeneous within themselves. The variability of the data can be visualized using the tree hierarchy. The objective of the dendrogram is to explore and define the appropriate number of clusters based on the level of analysis.

### 2.3 Material and methods

Figure 2.1 shows the proposed methodology for spatiotemporal segmentation using SOM and agglomerative hierarchical clustering. The input data are a time series of satellite images obtained over a data cube, consisting of one or more spectral bands. After applying the SOM method, the algorithm's output provides weight vectors, also called codebook vectors, which have the same size as the input time series. The number of codebook vectors depends on the grid size defined for the SOM, where each vector represents a set of time series. In the agglomerative hierarchical clustering step, the codebook vectors are the input data, in which the data are clustered into levels based on a similarity score. After the final number of clusters is determined, the unsupervised map is created in the spatiotemporal segmentation

**Figure 2.1** - Methodology for spatiotemporal segmentation of satellite image time series.
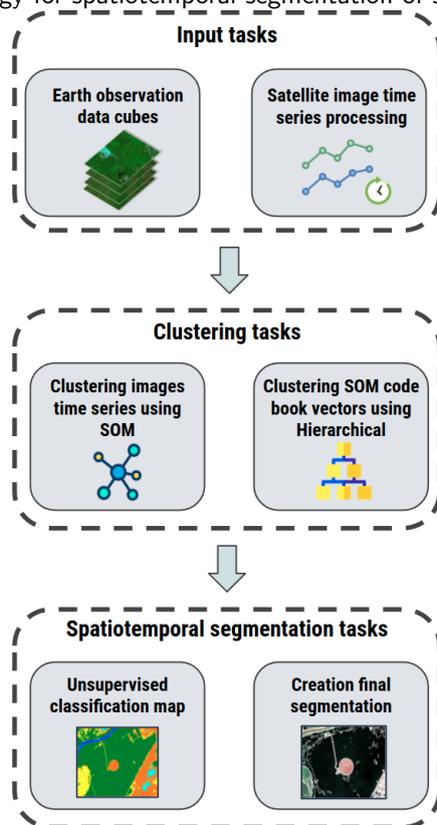


Source: From author.

tasks. However, smoothing is performed on this map, which is used as the basis for segmentation.

### 2.3.1 Earth observation data cubes

Earth observation (EO) data cubes can be defined as multidimensional arrays with space and time dimensions and spectral derived properties created from remote sensing images mainly to support image time series analysis (APPEL; PEBESMA, 2019). The Brazil Data Cube (BDC) project is producing EO data cubes from analysis-ready data of CBERS-4, Sentinel-2, and Landsat-8 satellite images for the entire Brazilian territory (FERREIRA et al., 2020a).

This document used the EO data cubes of Sentinel-2 images with a spatial resolution of 10-meters and temporal aggregation of 16 days using the best-pixel approach produced by the BDC project. For the experiments, we used these EO data cubes for the study area considering one agricultural year, 2019-08-29 to 2020-08-12, corresponding to 23 images.

To be as general as possible, only the Normalized Difference Vegetation Index (NDVI) was used for this study. However, if the goal is to segment a specific land cover, different bands can be used to segment the target better.

### 2.3.2 Satellite image time series processing

In the step of time series processing, we apply a cloud-free filter using the classes "No data", "Saturated or defective", "Dark areas", "Cloud shadow", "Cloud medium probability", "Cloud high probability", "Thin cirrus" and "Snow" contained in the SCL (Scene Classification Layer) mask, and interpolate the missing values, using using linear interpolation. Figure 2.2 shows a time series of NDVI extracted from the EO data cube of Sentinel-2 images and associated with a spatial location of the study area.

After interpolation, the image time series were formatted to serve as input to the SOM. Based on the size of the area to be studied, the scene will have $N \times M$ pixels where $N$ is the width of the image and $M$ is the height. The input format of the model is an array $[N \times M, T \times B]$, where the first component is the total size image's pixels, and the second depends on the temporal resolution $T$ and spectral bands $B$.

**Figure 2.2** - Example of an interpolated NDVI time series.
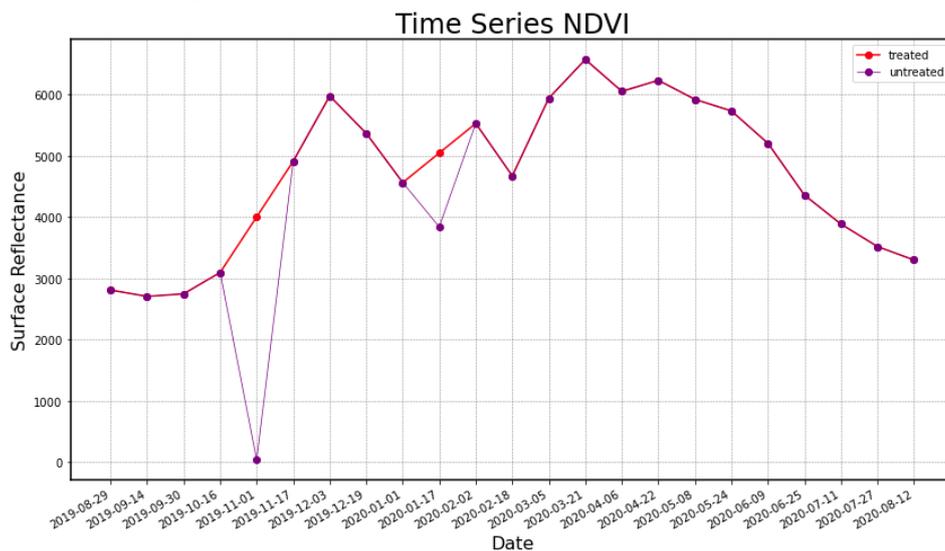


Source: From author.

### 2.3.3 Clustering and segmentation

Each time series represents one image pixel in the time domain and serves as input for the SOM algorithm. After the clustering step, each time series is assigned to a neuron, which in turn has an associated codebook vector, i.e., the codebook vector represents all-time series associated with the respective neuron, that characterizes the dimensionality reduction since an image with $N \times M$ pixels is represented by the number of neurons that make up the model.

Note that the number of patterns found in the image depends on the grid size chosen for SOM. How they are grouped also depends on the selected distance, the shape of the grid, and the neighborhood function.

After clustering with the SOM, the codebook vectors resulting from the first step are used as input for agglomerative hierarchical clustering. This is a clustering of the entire image represented by the codebook vectors, which would be impractical if all pixels were used, depending on the size of the image. After the new clustering, it is possible to create a visualization of the different levels of the clusters formed using a dendrogram. As in the previous step, the clusters are formed depending on some parameters such as the distance to be computed, the linkage criterion, and the number of the new clusters.

After creating the unsupervised classification map, which serves as the basis for creating the segments, spatial smoothing was applied to reduce noise. Because depending on the spatial resolution, there may be various disturbances due to the quality of the images, the presence of clouds, and spectral mixing models, since a pixel may contain more than one ground information that affects its reflectance value. Since it is a pixel-based approach, a pixel or a small set of noisy pixels may interfere with forming polygons more representative of the actual area (KöRTING et al., 2011).

Finally, a very simple but effective method was implemented to improve the final segments. Based on the unsupervised classification map result, a method was used to calculate the mode of a pixel neighborhood in which the value is replaced. Depending on the experiment, different window sizes such as $3 \times 3$, $5 \times 5$, or $7 \times 7$ produce different results with greater or lesser smoothing.

## 2.4 Results and discussion

### 2.4.1 Study area

This document used the EO data cubes of Sentinel-2 images with a spatial resolution of 10-meters and temporal aggregation of 16 days using the best-pixel approach produced by the BDC project. For the experiments, we used these EO data cubes for the study area considering one agricultural year, 2019-08-29 to 2020-08-12, corresponding to 23 images.

To be general as possible, only the Normalized Difference Vegetation Index (NDVI) was used for this study. However, if the goal is to segment a specific land cover, different bands can be used to segment the target better.

The study area is in the Mato Grosso state, located in the mid-west of Brazil. Mato Grosso is one of the largest states in Brazil; its territory is composed of three biomes: Cerrado, Amazon, and Pantanal. Therefore, it presents a high spatial and temporal heterogeneity. Figure 2.3 shows the study area that is located in Rondonópolis, a city in the Mato Grosso state. The image on the right side corresponds to the satellite/sensor Sentinel-2/MSI from an agricultural composition (Swir16, Nir, and Blue) with a 10-meter spatial resolution.

The selected study region has several interesting land use and land cover elements for the segmentation task. For example, an extensive urban center, the watershed of the Rio Vermelho that extends throughout the state, native vegetation, agricultural areas, and exposed soil regions.

### 2.4.2 Configuration

Based on the study area section, the images used in this work have $1954 \times 2188$ pixels. The input format of the model is an array$[4275352, 23]$, where the first component is the total size image's pixels and the second one depends on the temporal resolution and spectral bands. For this experiment, 23 time series images with only one band were used. Another configuration, for example, if we had a temporal resolution of 25 images and three spectral bands, would be the array $[4275352, 75]$, since multiplying 25 by 3 gives 75, which corresponds to the concatenation of spectral bands.

For this experiment with the following results, only two parameters were selected for the SOM, the grid size and the number of epochs, and two parameters were also selected for the hierarchical algorithm, the final number of clusters and the linkage,

**Figure 2.3** - The study area where the experiments were conducted. The image on the right was taken from Sentinel-2/MSI with a 10-meter spatial resolution. The composition (Swir16, Nir, and Blue) was used.



South America ■ Mato Grosso State - Brazil ■ Study Area
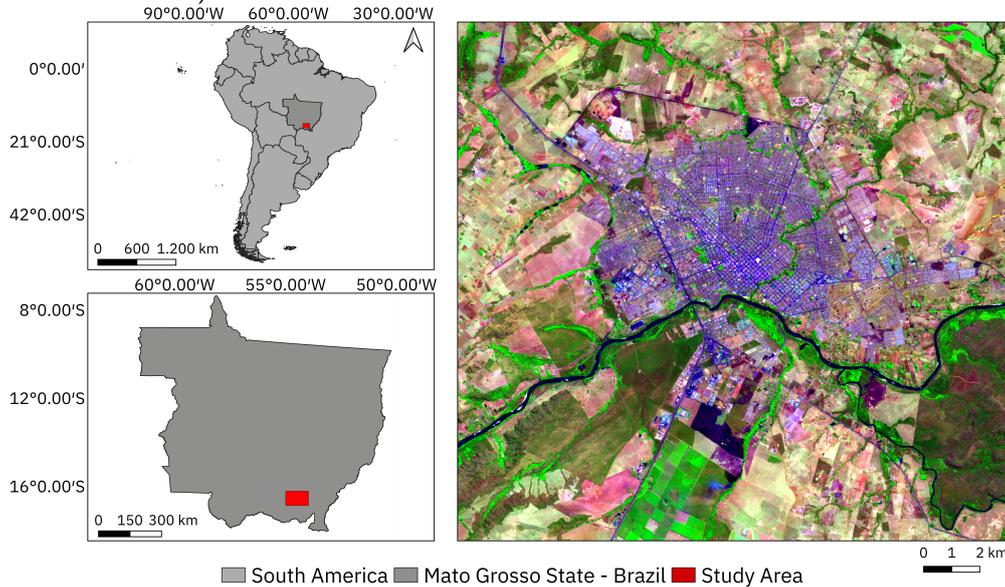
Source: From author.

the other parameters were set by default for `XPySom` (MANCINI et al., 2020) and `sklearn` (PEDREGOSA et al., 2011) Agglomerative Clustering, respectively. The size of the SOM grid was determined based on the temporal and spectral variability of the image, and empirically some grid configurations were tested. Visually satisfactory results were obtained with size $12 \times 12$, totaling 144 neurons, and 20 epochs for training. For the hierarchical algorithm, 4 and 12 final cluster numbers with the average linkage were defined to check different spatiotemporal segments level.

The results were evaluated using qualitative and quantitative approaches. In the qualitative approach, a visual analysis of the satellite image of a Sentinel-2 image and a comparison with the reference provided by the TerraClass Cerrado project for the same period. In the quantitative approach, supervised and geometric metrics were used to analyze the quality of the produced segments. The following describes the qualitative and quantitative results, respectively.

### 2.4.3 Qualitative evaluation

At the first moment, a qualitative assessment compares the segments with a Sentinel-2 image and the TerraClass Cerrado map based on visual interpretation. TerraClass Cerrado project produces land use and land cover maps for the Cerrado biome in Brazil with spatial resolution between 20 and 30 meters (ALMEIDA et al., 2016). This

13

project's methodology for producing the land use and land cover map is based on several steps, mainly visual analysis by remote sensing experts. The experts use MODIS time series and high-resolution imagery to distinguish the different agricultural cycles in each Brazilian biome. These data sets are available through the TerraClass portal[3].

The unsupervised result for the entire scene is shown in Figure 2.4. In this Figure, we present the thematic map of the TerraClass project for the 2020 year and the unsupervised maps with 4 and 12 clusters. Note that each map has its legend, and the colors of the unsupervised maps do not correspond to the TerraClass map classes.

**Figure 2.4** - Results obtained with the settings with 4 and 12 groups.
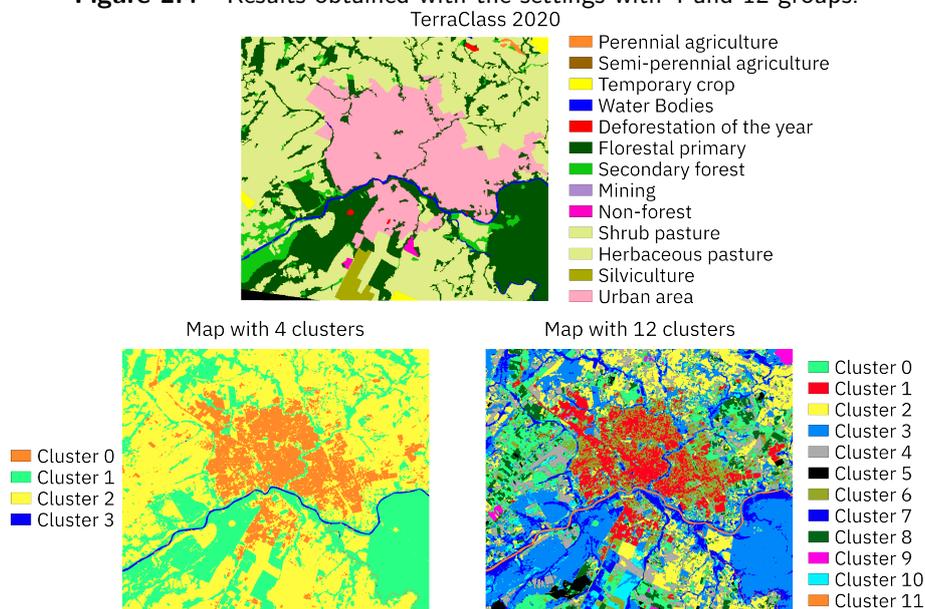


Source: From author.

In general, the two configurations were able to capture the main features of the region, namely: urban area, native vegetation, and water bodies. In the configuration with 4 groups, it was possible to identify urban areas and waters more homogeneously. However, regions with different vegetation types, cropland, and pasture land obtained a greater integration but a good separation from the other classes.

To support visual analysis and understand the dynamics of the model with 4 groups, Figure 2.5 shows the dendrogram created by the hierarchical clustering method, where 144 (12x12 SOM grid) codebooks vectors were integrated into 4 final clusters

---

[3]https://www.terraclass.gov.br/

defined only by the time series distance. Figure 2.6 shows the NDVI profiles for each codebook vector with the respective colors assigned to each final cluster.

**Figure 2.5** - Dendrogram Results.



Neurons CodeBook Number
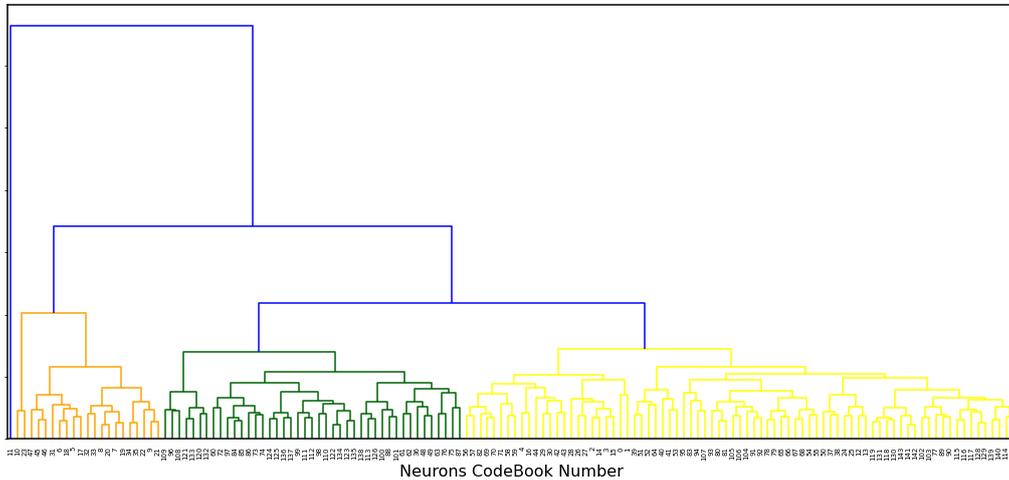
Source: From author.

**Figure 2.6** - NDVI Codebooks vector of SOM.



Source: From author.

In the model with 12 groups, it was possible to detect more nuances in land use and land cover classes. For example, the model captured different types of roofs, containers, and roads in the urban region. Another example is the detected vegetation types, where it was possible to distinguish between dense and shallow vegetation.

15

Figure 2.7 shows an average of the patterns integrated into 4 clusters in the hierarchical clustering step, which can be used as representatives of each cluster for the unsupervised classification map with 4 clusters in Figure 2.4. In this way, it is possible to add more information to each segment, which can be used for other purposes such as classification, data mining, or even to determine the number of final clusters for segmentation.



**Figure 2.7** - Time series representing each cluster.
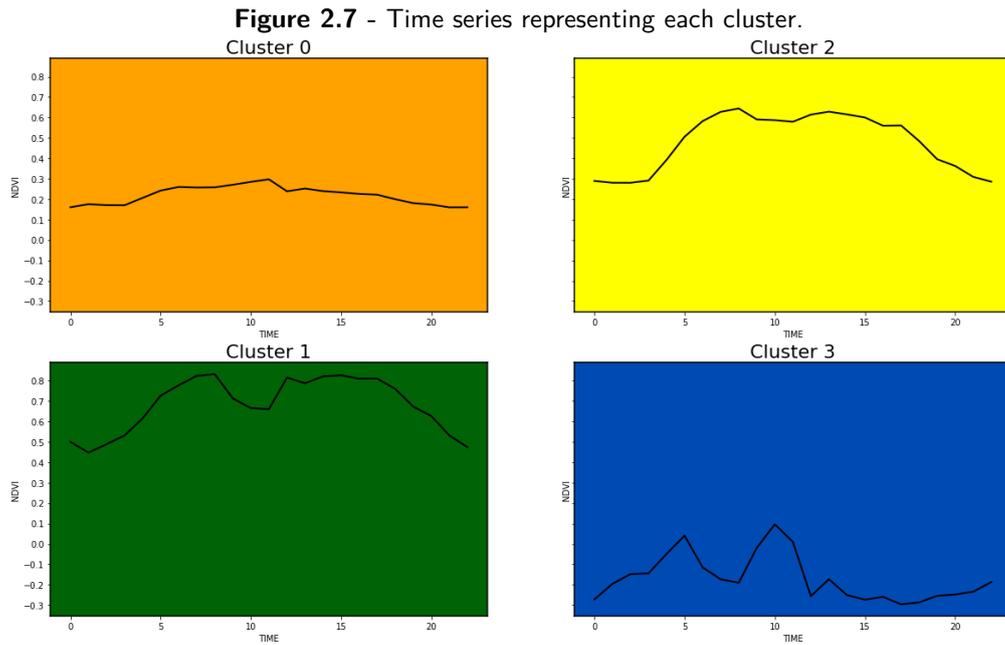
Source: From author.

Analyzing some regions of the study area in more detail, we can observe in Figure 2.8a - Figure 2.8b a heterogeneous region, sharing the space with urban areas, shallow vegetation, dense vegetation and water bodies. Note that Figure 2.8a, which corresponds to an image dated 2019-12-19 image, does not have a local pivot point, unlike Figure 2.8b dated 2020-04-06, which has a pivot point in the center of the region.

If we look at the unsupervised maps generated with 4 groups in Figure 2.8c without smoothing and Figure 2.8d with smoothing, we can notice some points. The first one is the loss of details such as trails and the smoothing of edges in some regions, especially in urban areas and along the river. Despite the great heterogeneity of the region, it was possible to make a macro delineation of the region by dividing it into four major groups, such as water bodies, dense vegetation, roofs with exposed

**Figure 2.8** - Example of a scene with images with two different dates and segmentations with and without smoothing with different number of clusters. In (a) we have a scene in the RGB composite of 2019-12-19 (b) we have a scene in the RGB composite of 2020-04-06 (c) segmentation of the scene with 4 clusters without smoothing (d) segmentation of the scene with 4 clusters with smoothing (e) segmentation of the scene with 12 clusters without smoothing (f) segmentation of the scene with 12 clusters with smoothing.
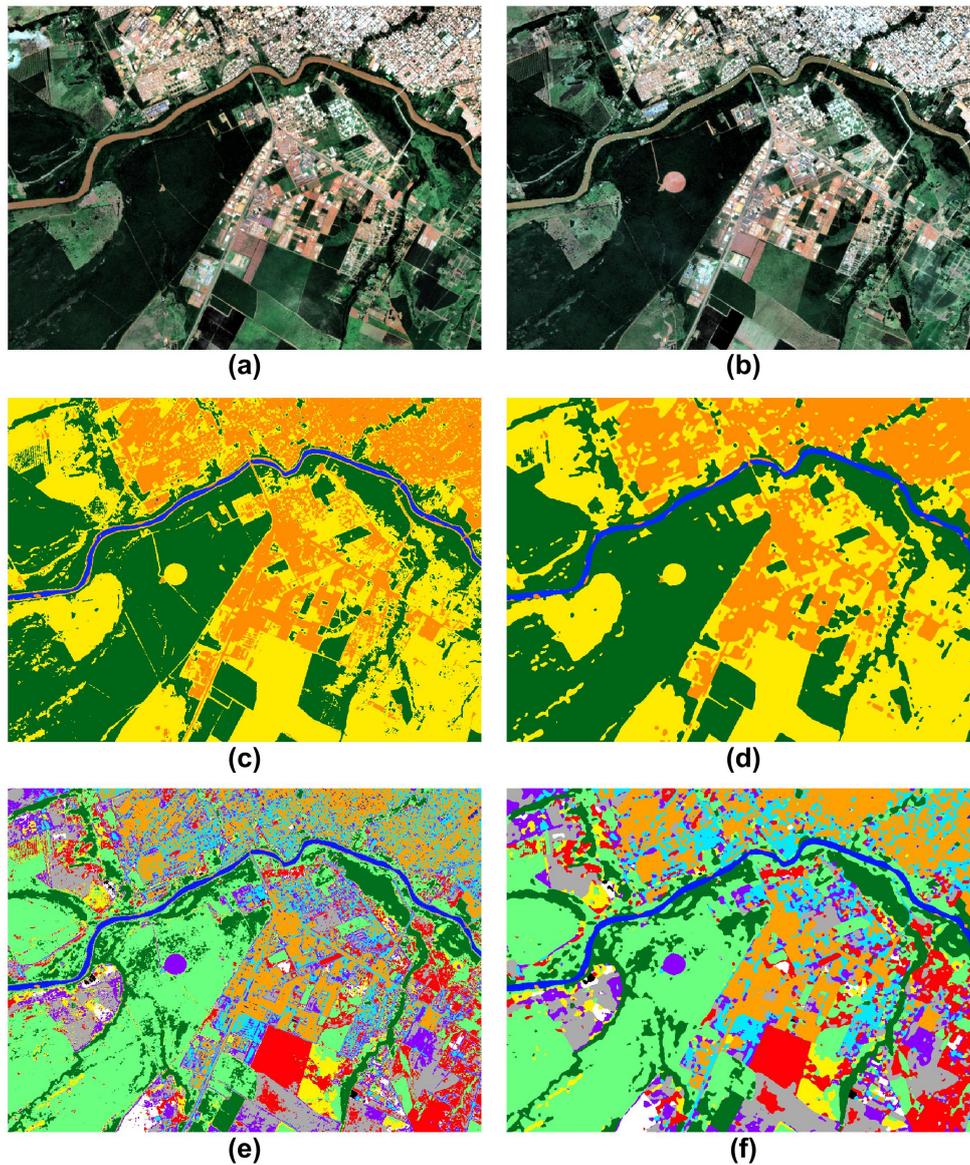
soil, and the last group that integrated regions with temporal variations, such as the region of the pivot. In Figure 2.8e and 2.8f, we see the same loss of detail of trails and edges. With 12 clusters, it was able to see greater detail than with 4 clusters, where regions were subdivided into two or more segments and produced more segments as a result. Each segment has a distinct interpretation depending on the number

of clusters selected. Each segment has a distinct interpretation depending on the number of clusters selected.

Figures 2.9a and 2.9b were selected to observe the temporal change of some land use and land cover areas and to see how the presence of the cloud in the river region did not affect the final segmentation. Note that there is still a loss of detail in Figure 2.9c and Figure 2.9d due to smoothing. Even after smoothing in Figure 2.9d, it is possible to see some orange segments in the middle of the blue river that would be sandbars in this region. In Figure 2.9e and 2.9f, there is a black region that may be an agricultural region, meaning that each segment has contextual information associated with its dynamics in time.

Figure 2.10 shows another example that has a great heterogeneity drawing attention by the good separability between vegetation, urbanized area, and the others for the map with 4 clusters. In Figure 2.10e and 2.10f, where there are 12 clusters, one can observe a greater diversification for this region with several different patterns in time, note the good delineation of the segment in the upper right corner in pink, representing temporal agriculture by TerraClass.

Temporal dynamics with characteristic patterns such as agriculture, vegetation that remains constant over time, and patterns that generally remain constant over time are easy to identify and distinguish from others.

After each step and the creation of unsupervised maps and subsequent smoothing, Figure 2.11 shows the result of the segments with 4 clusters, it is possible to see how the rivers and regions with dense vegetation are well-delineated.

### 2.4.4   Quantitative evaluation

In remote sensing image segmentation applications, determining the accuracy of segmentation undertaken quantitatively is still at an early stage of maturation (COSTA et al., 2018a). Since much of the evaluated work uses a qualitative approach to measure the quality of its results, this document used supervised metrics for quantitative analysis to determine segmentation accuracy to perform the quantitative analysis. Supervised metrics measure the similarity of generated segments to reference polygons by shape similarity. These metrics are divided into area-based and location-based. The area-based metrics evaluate the overlap region between the segmentation object and the reference polygon. The location-based metrics evaluate

**Figure 2.9** - Example of a scene with images with two different dates and segmentations with and without smoothing with different number of clusters. In (a) we have a scene in the RGB composite of 2019-12-19 (b) we have a scene in the RGB composite of 2020-04-06 (c) segmentation of the scene with 4 clusters without smoothing (d) segmentation of the scene with 4 clusters with smoothing (e) segmentation of the scene with 12 clusters without smoothing (f) segmentation of the scene with 12 clusters with smoothing.



(a)  (b)

(c)  (d)

(e)  (f)
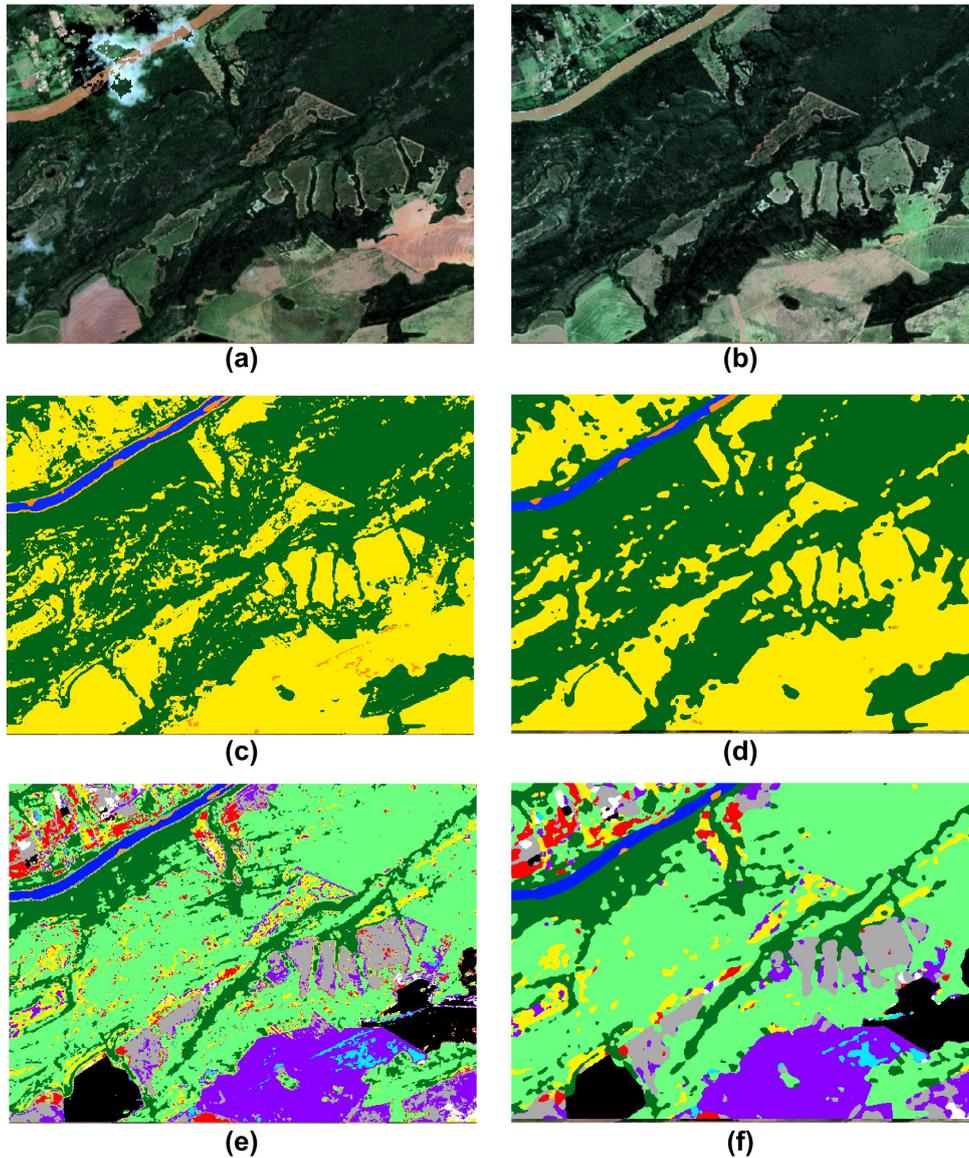
Source: From author.

the distance between the centroids of the segmentation objects and the reference polygon (CLINTON et al., 2010).

Costa et al. (2018a) recommends the use of area-based metrics for applications in which geometric representation is more important than the thematic information associated with segment objects. Therefore, in this document, we used the metrics

**Figure 2.10** - Example of a scene with images with two different dates and segmentations with and without smoothing with different number of clusters. In (a) we have a scene in the RGB composite of 2019-12-19 (b) we have a scene in the RGB composite of 2020-04-06 (c) segmentation of the scene with 4 clusters without smoothing (d) segmentation of the scene with 4 clusters with smoothing (e) segmentation of the scene with 12 clusters without smoothing (f) segmentation of the scene with 12 clusters with smoothing.
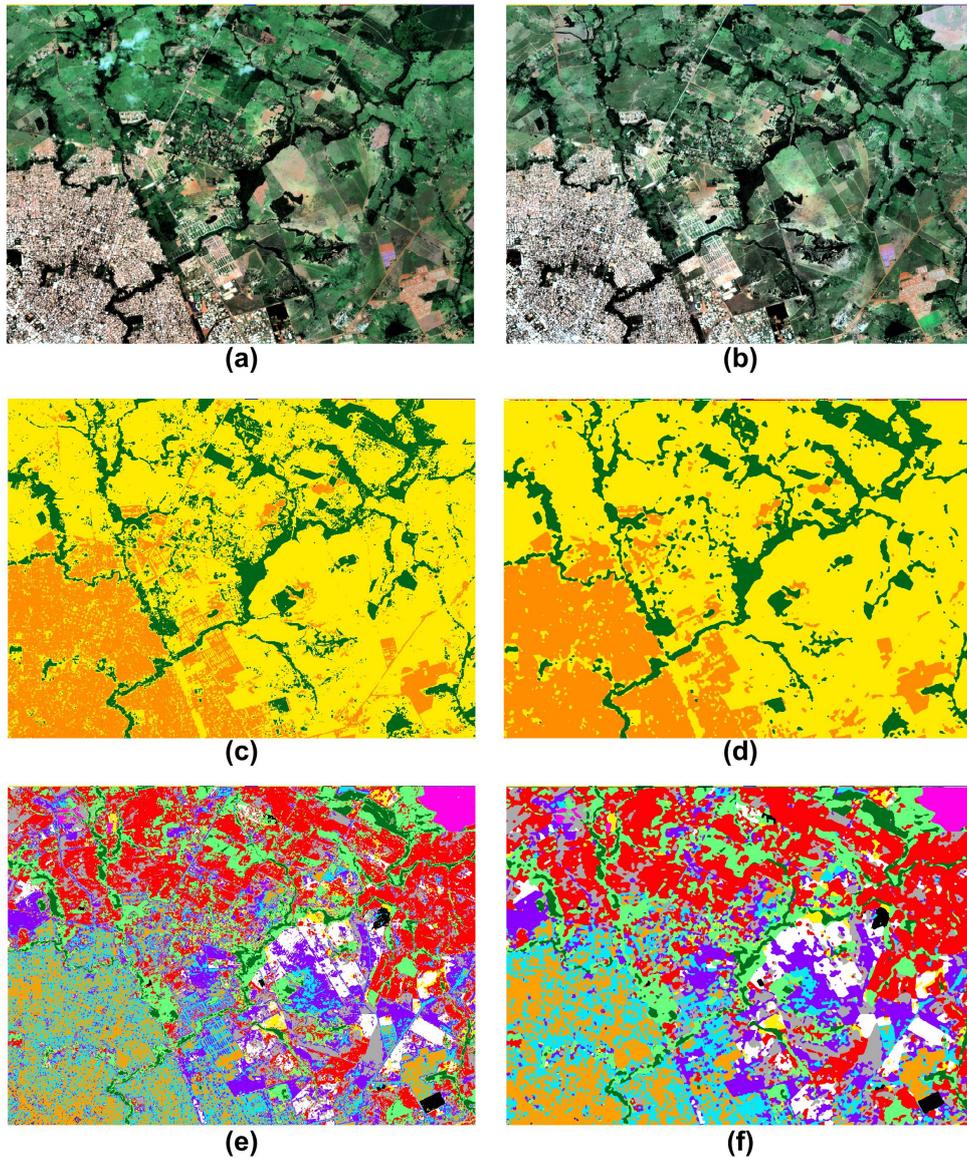


Source: From author.

*Precision* (RIJSBERGEN, 1979; ZHANG, 1996), *Recall* (RIJSBERGEN, 1979; ZHANG, 1996), *F_measure* (RIJSBERGEN, 1979; ZHANG, 1996), and *Match* (*M*) (JANSSEN; MOLENAAR, 1995; FEITOSA et al., 2010). The *Precision* metric measures the precision of the segment object, with the reference polygon being sensitive to under-segmentation. The *Recall* metric measures how similar the reference polygon is to the

20

**Figure 2.11** - Spatiotemporal segmentation results (red outlines).

Source: From author.

segment object and is more sensitive to over-segmentation. The $M$ metric measures the average match between polygons and segments, measuring under- and over-segmentation. Finally, $F\_measure$ is a combined metric representing the balance between *Recall* and *Precision*, which also measures under- and over-segmentation. These metrics range from 0 and 1, with the optimal value being 1.

The vegetation polygons were extracted based on the TerraClass thematic map and compared with the 4 clusters map to perform the quantitative assessments. Figure 2.12 shows the reference polygons extracted from the TerraClass Cerrado project and used as reference polygons. The segmentation results compared to the reference polygons can be seen in the same Figure, where the red area represents the reference polygons, and the white lines represent the segment's object. To calculate the metrics, we used the R package `segmetric` (SIMOES et al., 2021).

The results of the metrics evaluated to determine the accuracy of the generated segmentation are shown in Table 2.1. It can be noted that the *Recall* metric presented the best result (0.87), indicating a small over-segmentation error. The $M$ metric obtained the worst result since this metric calculates the average between the match of the two geometries. This value is due to some shapes that obtained a low value. The

**Figure 2.12** - Reference Polygons with fill-in red and the segments with black borders produced by the model used for quantitative evaluation with a Sentinel-2 composition (Swir16, Nir, and Blue)..
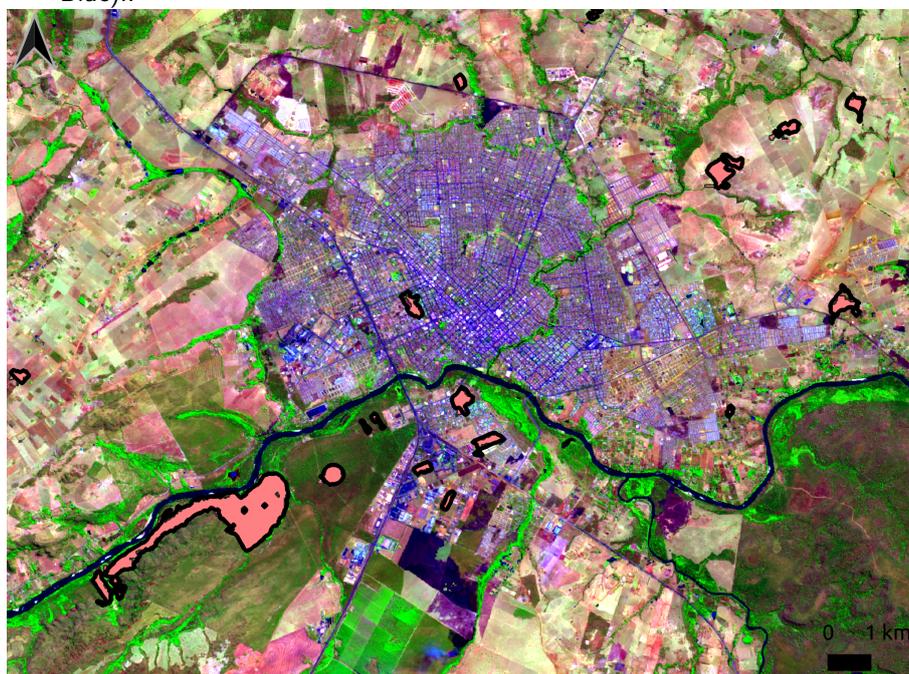


Source: From author.

**Table 2.1** - Quantitative results of the spatiotemporal segmentation.

| Metric | Value |
|---|---|
| *Recall* | 0.87 |
| *Precision* | 0.83 |
| *F_measure* | 0.85 |
| *M* | 0.64 |

Source: From author.

results obtained were satisfactory and showed a good agreement with the reference polygons.

### 2.4.5 Comparison of approaches

Figure 2.13a shows the Rondonopolis region segmented using the region-growing algorithm (ADAMS; BISCHOF, 1994) for a 2020-04-06 NDVI image. The adjustments were made empirically in the Terraview [4] application developed by the National

---

[4]http://www.dpi.inpe.br/terralib5/

Institute for Space Research (INPE) image processing division, with thresholds defined so that the number of segments was similar to the result of the model proposed in this document for comparison purposes.

Note that in Figure 2.13a, the colors of the segments contain no information other than what distinguishes them from the other segments, while in Figure 2.13b, the colors contain the temporal information of each segment.

In addition to these differences, the segments themselves may have different shapes. Remember that in Figure 2.13a we are only looking at surface reflectance for a single image in time, so it is not possible to extract features that do not change over time. For example, if there are temporal dynamics within the chosen temporal resolution that may depend on the number of clusters chosen, it is possible for a polygon to represent these dynamics over time. In a single image of the same region, this depends only on the change in reflectance.

Figure 2.13 - Study area segmented by two different techniques. In (a) we have the segmentation of an single image using Region growing (b) we have our proposal of spatiotemporal segmentation.



(a)                                    (b)
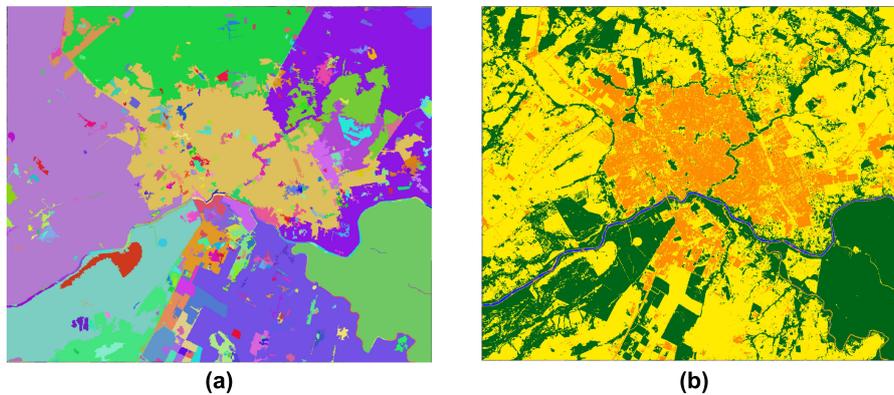
Source: From author.

## 2.5 Conclusion

This document presents a spatiotemporal segmentation based on two clustering algorithms. Two configurations were used to present the model, and qualitative and quantitative evaluations were made to analyze the segments generated by the algorithm.

Note that due to the nature of spatiotemporal segmentation, some polygons represent land use and land cover classes that are often difficult to evaluate with a single

image. On the other hand, land use and land cover types that have characteristic temporal profiles, such as agriculture, deforestation, and forest fires, can be represented by a segment. Thus, with the knowledge of experts, it is possible to generate segments of semantic information analyzing these temporal profiles.

As successive integrations of clusters are performed, the final segments may have a large extent that can interfere with processing. However, creating rules to divide these segments into subsegments is possible for better preprocessing.

The model was run on a Google Collab virtual machine with the following specifications: Intel(R) Xeon(R) CPU @ 2.20GHz with 2 cores, 12GB ram and 25GB disk space, Nvidia K80 / T4 GPU with 12GB/16GB GPU Memory, which obtained the times of 15 to 20 seconds for training with 20 epochs and approximately 5 seconds for prediction for the SOM algorithm which had the longest execution time. Further study will be done to take into account neighboring pixels for contextual information and explore different spectral bands, temporal resolutions, and other spatial resolutions for different land use and land cover applications.

# 3 LAND USE AND LAND COVER MAPPING BASED ON IMAGE TIME SERIES CLUSTERING

## 3.1 Introduction

Land use and land cover (LULC) map is a product, mainly derived from remote sensing images, with several critical applications such as urban planning, natural resource management, deforestation monitoring and food security analysis (BROWN, 2016) (SAAH et al., 2020). The quality of LULC mapping directly impacts the success of its applications. Therefore, new prospects for LULC mapping have emerged with the technological development in remote sensing and the increase in computational power (SAAH et al., 2020).

With the availability of a large amount of remote sensing data with different spatial resolutions, temporal and spectral models, and artificial intelligence models, LULC mapping is becoming increasingly robust (SOILLE et al., 2018). Data structuring, storage, and distribution technologies have enabled access to large volumes of data, such as satellite image time series, which have been widely used for land use, cover analysis, and classification (SIMOES et al., 2021). Due to their temporal characteristics, these time series are crucial for understanding dynamics such as deforestation, wildfires, flooded areas, and agriculture (PASQUARELLA et al., 2016).

To produce LULC maps from big data sets of remote sensing images, image time series analysis and machine learning methods are being widely used with promising results (GÓMEZ et al., 2016; BELGIU; CSILLIK, 2018; PICOLI et al., 2018; SIMOES et al., 2020; SIMOES et al., 2021). Satellite image time series analysis captures subtle changes in ecosystems and improves the quality of land classification (SIMOES et al., 2021). Approaches that use satellite image time series analysis are defined as *time-first, space-later* by Camara et al. (2016).

Some methods in the literature use clustering methods for classification, for example. Gonçalves et al. (2008), which also uses SOM and HCA for classification but does not use temporal information for classification and only uses an image in time and samples. Bagan et al. (2005) uses SOM for classification using time series and land use and land cover samples. However, the classification of each neuron is based on the majority of the samples contained in each neuron.

Amitrano et al. (2017) proposes a feature extraction method from SAR images that uses SOM to create segments based on the color homogeneity of a single image and

uses the concept of Geographic Object-Based Image Analysis (GEOBIA) to extract features from the segments and create a classification.

Most of the machine learning methods used to classify image time series and to produce LULC maps are supervised and thus they require a training phase using a significant number of samples labeled *a priori*. The quality of training samples is crucial in the classification process. Good samples lead to classification results with better accuracy (SANTOS et al., 2021).

This document evaluates image time series clustering methods to produce training samples and to classify land use and land cover types. The main idea it to produce these samples from the image time series and their clusters, mainly to identify land use and land cover types that have characteristic spectral and temporal behavior such as inland water bodies.

We propose a methodology for land use and land cover mapping using image time series clustering. This document presents an experiment using this methodology to map inland water bodies, a specific type of land cover, using six study areas as training grounds for the Random Forest model.

Random Forest is a robust model that predicts using many decision trees. It works by randomly selecting subsets of characteristics and training samples for each tree, reducing variation and increasing accuracy. It aggregates predictions from each tree into a final prediction, further enhancing model accuracy. Furthermore, essential variables in the categorization can be recognized, aiding in interpreting the results. It has been frequently utilized and proven in LULC classification based on time series(SIMOES et al., 2020; BELGIU; DRăGU, 2016; PELLETIER et al., 2016).

The goal is to apply this model to four other regions to create a map of water bodies. This mapping has several important applications, such as urban planning identifying areas vulnerable to flooding where the temporal premise of the model facilitates the identification of use and cover, as these areas are strongly characterized by their temporal behavior. The experiment relies on two evaluation methods: qualitative and quantitative.

The main water inland map available for Brazilian territory are those mapped by the National Water Agency (ANA) (AGÊNCIA NACIONAL DE ÁGUAS (ANA), 2020; AGÊNCIA NACIONAL DE ÁGUAS (ANA), 2022) and by the Brazilian Institute for Geography and Statistics (IBGE) (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍS-

TICA (IBGE), 2011; INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE), 2022), which can be described as official water map. However, both are made by a mixture of different compositions of satellite images, field camps, historical mapping, and supervised delimitation. On the other hand, we can find products of land use and land cover from which we can extract water bodies, as those provided by MapBiomas (SOUZA et al., 2020; MAPBIOMAS, 2022) with 30m spatial resolution, and ESA WorldCover (ZANAGA et al., 2021) with 10m spatial resolution, or even global datasets which we can filter for Brazilian territory as the case of the JRC Global Surface Water Mapping (PEKEL et al., 2016).

Despite the main water products using a set of annual images in a classification, those procedures reprocess the images to get a better pixel, or reduce time information on indexes, or even classify each time step and then include a new processing tool to reduce it to a single image.

Both procedures lose information on the process or do not use all the time series potential. According to Fu (2011) (FU, 2011) a time series "is a collection of observations made chronologically" and is always considered as a whole instead of an individual numerical field. This chronological nature gives the time series components such as trends, seasonality, and cycles (MORETTIN; TOLOI, 2008), providing more data about the event.

## 3.2 Material and methods

Figure 3.1 shows the proposed methodology in the document, which can be described in four stages: The first stage is the preparation of the data, where the time series of images are extracted, and their interpolation to serve as input data for clustering, which consists of two clustering methods, Self Organized Map (SOM) and Hierarchical Clustering Algorithm (HCA).

The clustering step is performed on time series from the data cube of the area of interest and results in an unsupervised classification map and the information of each cluster formed by the SOM grid and the dendrogram. The third stage consists of the training phase for land use and land cover classification. For each data cube to be classified, an unsupervised classification map and a SOM grid consisting of neurons with associated weight vectors are generated. These are commonly called codebook vectors, which are considered a time series. Each codebook vector represents a set of time series that characterize the dimensionality reduction. Each neuron is interpreted as a different cluster with its associated time series. For each time series associated

**Figure 3.1** - Metodology for Proposed Method.

1. **Data preparation**
2. **Time series clustering tasks**
3. **LUC classification - training task**
4. **LUC classification - predict task**
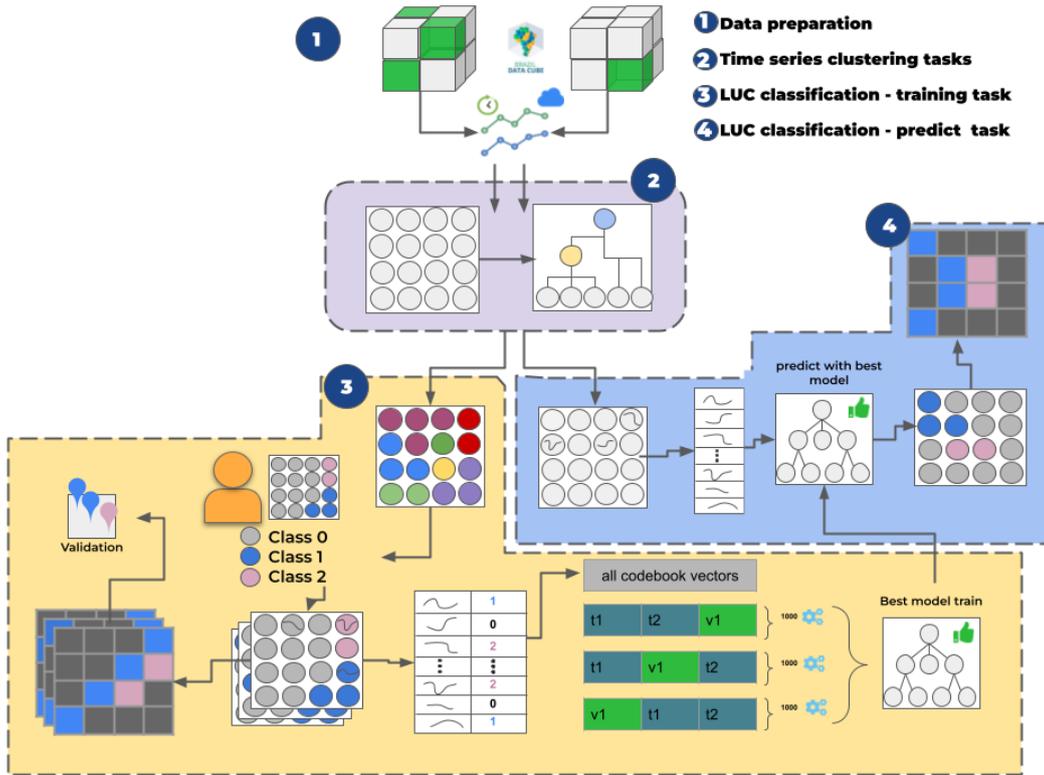
Source: From author.

with a neuron, this means that it has the shortest distance to the codebook vector compared to all others.

Each neuron is associated with a segment/cluster in the unsupervised classification map. For an expert, it is possible to assign a semantic meaning to each segment by labeling each neuron. Depending on the size of the study area and the number of neurons, it can be very costly to label each neuron. For this, the dendrogram can be used to adjust the number of final groupings, which is not necessarily the same as the number of land use and land cover classes.

The specialist determines the number of study areas for the training phase, and the neurons for each area are labeled with the classes that make up the map. As mentioned earlier, each neuron is associated with a segment on the map; therefore, after labeling all neurons, the classified land use and land cover map are created. And then, a validation process of these maps is done using Samples of land use land cover time series. After validation, and if the results are satisfactory, a supervised classification model is trained with the labeled neurons as input data, with the task

of labeling new neurons not seen by the model so that an expert is no longer needed to classify new maps.

In the fourth step, new regions are selected from the training set, which has also gone through the data preparation process and the clustering phase. The result is unsupervised maps and the SOM grid with associated neurons. Unlike the previous step, the new neurons are labeled by the trained model, where the model labels the neurons, and then the classified land use and land cover map is generated.

### 3.2.1 Data preparation

It is necessary to define some parameters for extracting the time series, such as the sensor to be used, the period to be mapped, whether data cubes with or without aggregation should be used, and which spectral bands should be used.

After defining these parameters, the time series is extracted from the data cube and subjected to a treatment to remove noise that could interfere with clustering. Based on a cloud mask, the classes to be removed are selected, and then a linear interpolation is performed using the last valid observation.

This process is of great importance in the clustering task, considering that only the distances between the time series are used as clustering criteria, i.e., the influence of a cloud in a period of the time series can be deterministic in clustering quality. It can be that two time series that are spatially close to each other remain in different clusters even though they have similar temporal and spectral dynamics because of these noises, causing salt and pepper effects in the final map.

### 3.2.2 Time series clustering task

Two clustering models are used for this process: the Self Organized Maps (SOM) and the Hierarchical Clustering, where depending on the computational power, it is possible that the input of the SOM is all the time series of the image, that is, an array [N × M, T × B ], where N is the width of the image and M is the height of the image and T is the temporal resolution and B is the spectral resolution, if it is not possible to use all the time series, a strategy can be used that does not harm the final result as long as it does not change the distribution of the input time series.

Considering the nature of the extracted time series from the image, neighboring pixels behave very similarly, so it is possible to make a stride skipping a few pixels and thus reduce the number of time series in the model.

It is worth noting that the more time series of the image used in the model input, the more the codebooks will represent their time series sets, mitigating outlier effects and making it easier to interpret the neurons in later processes. Remember also that after the training phase of the SOM to generate the map, it will be necessary to apply the model to all the time series of the image.

The SOM is an elastic network, i.e., the weight vectors are updated as a function of the input data. If a non-ideal number of input time series is selected, the outliers can shift this network by moving away from the center of the clusters. As a result, a codebook vector is not faithful to the set it represents, affecting later interpretation of the time series.

Determining the ideal size for the SOM grid can be challenging, given each region's different spectral and temporal dynamics, which can characterize different distributions. Again, depending on computational power, different grid sizes can be specified. One argument that can be used to set the upper limit of neurons at 2700 is that MapBiomas (MAPBIOMAS, 2022) has 27 land use and land cover classes, and Terra-Class[1] has 18 classes. Setting 100 different dynamics for each class seems reasonable, considering that not all regions have a maximum of 27 classes.

The number of neurons may be much larger than necessary. In the Hierarchical Clustering step, two strategies can be used. The first is defining a distance threshold for integrating neurons that removes the excessive number of neurons smaller than the defined threshold. Depending on the region, multiple neurons may not be activated, and groups formed by the hierarchical clustering may also contain null time series, so a filter can be created to remove these clusters.

A second strategy is to define a fixed number of final clusters into which several similar neurons are integrated. It should be remembered that this step is entirely changeable depending on the expert's interpretation for a better classification. After selecting the best approach, the unsupervised classification map and the final clustering of neurons are created.

---

[1]https://www.terraclass.gov.br/

### 3.2.3 LULC classification

#### 3.2.3.1 Training task

In the classification training phase, a remote sensing specialist analyzes and labels the neurons. The results of the previous stage, such as the unsupervised classification map and the clustering SOM map, are used. Information such as the average of the clustering weight vectors considered like a time series and the shape of the concerned segments on the map are some information that the specialist can take into account in the classification, in addition to the support from satellite images and other resources.

After labeling each final cluster or neuron, the unsupervised map is also labeled to create the land use and land cover map, which in turn undergoes a validation process determined by the specialist to verify the quality of the map created. This step is repeated for each map selected for the training step.

Now that the specialist has labeled the neurons, they are divided into different training and validation sets. The weight vectors assigned to the neurons are the input of a supervised classification model (Random Forest) and are trained in different configurations of hyperparameters. The best model is selected and later used in the prediction step to classify new neurons to reproduce the results of the labeling done by the expert.

#### 3.2.3.2 Predicting task

With the best model trained, new regions go through the clustering process, generating SOM maps with their respective weight vectors. The model labels each neuron, generating land use and land cover maps.

### 3.3 Results and discussion

#### 3.3.1 Water inland study area

We choose six places (Figure 3.2) that capture the different facet in the inland water mapping.

- The first place is located in the Parque Estadual do Rio Negro in Mato Grosso do Sul, Pantanal biome, local of a vast variation in the water surface, including wetland zones and seasonal floods;

31

**Figure 3.2** - Location of the six areas tested in the study with tile number in brackets.



[0830865] Lago Grande Curuai - PA
[094090] Açude Castanhão - CE
[082090] PN Jamanxim - PA
[088104] Serra da Canastra - MG
[082104] PE do Rio Negro - MS
[088105] Serra da Canastra - MG

N

0  500  1,000 km

Study Area

Biome
Amazonia
Caatinga
Savannah
Atlantic Forest
Pampa
Pantanal
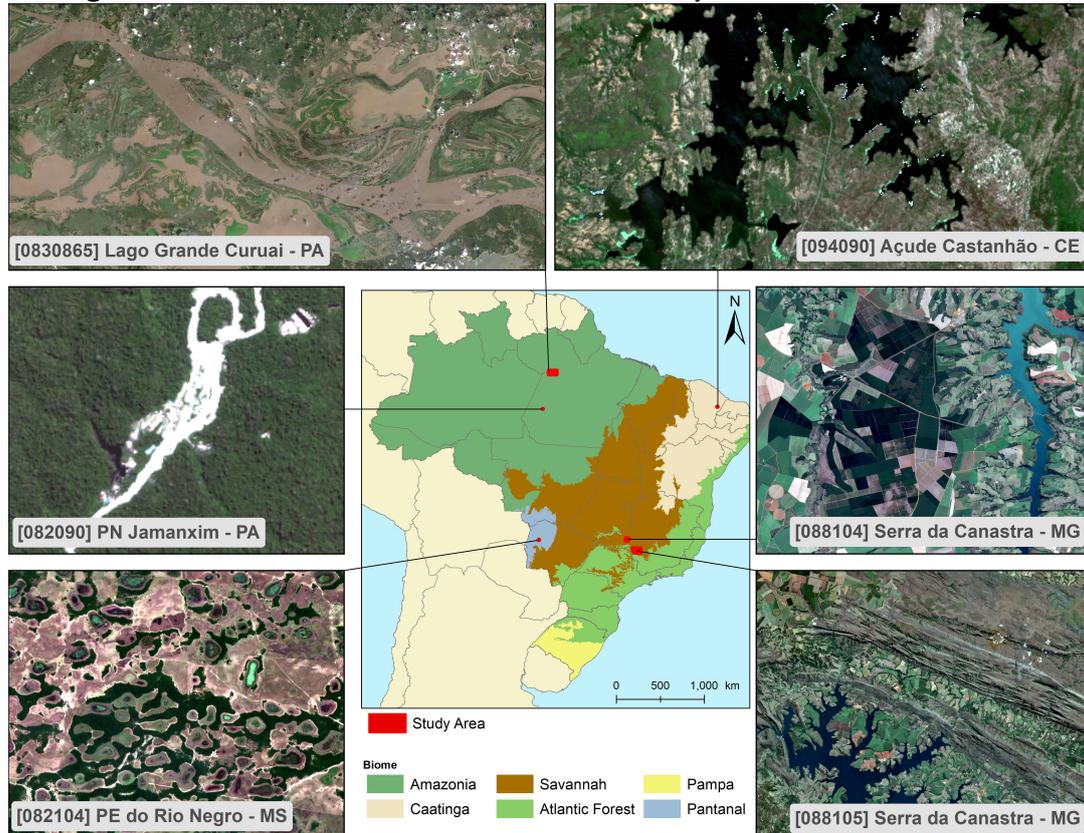
Source: From author.

- The second place is located inside the Floresta Nacional Jamanxim in Pará state. The place contains several mining spots, which dump large amounts of particulate matter into the river, causing confusion with soil during classification;

- The third place is Curaí lake in Pará state. This lake has a significant variation in its borders and affluent, changing its shape continuously;

- The fourth place is the Castanhão dam, located in the Caatinga biome, more specifically in Ceará state. The place was chosen by its variation in shape and the max flood area;

- The fifth place is located in the Serra da Canastra, Minas Gerais, inside the Cerrado biome. The place concentrates a variety of agriculture, and constant use of water can saturate the soil causing a misclassification of inland water. Was also found some glint in the major lake;

- Last place is also located in the Serra da Canastra, Minas Gerais, and was chosen by the proximity to high mountains and cliffs that makes shadows in the terrain, causing misclassification;

**Figure 3.3** - The figure shows the four regions for model prediction. Image A corresponds to the RGB image of tile 80086, B to tile 94096, C to 89101, and D to tile 84108.
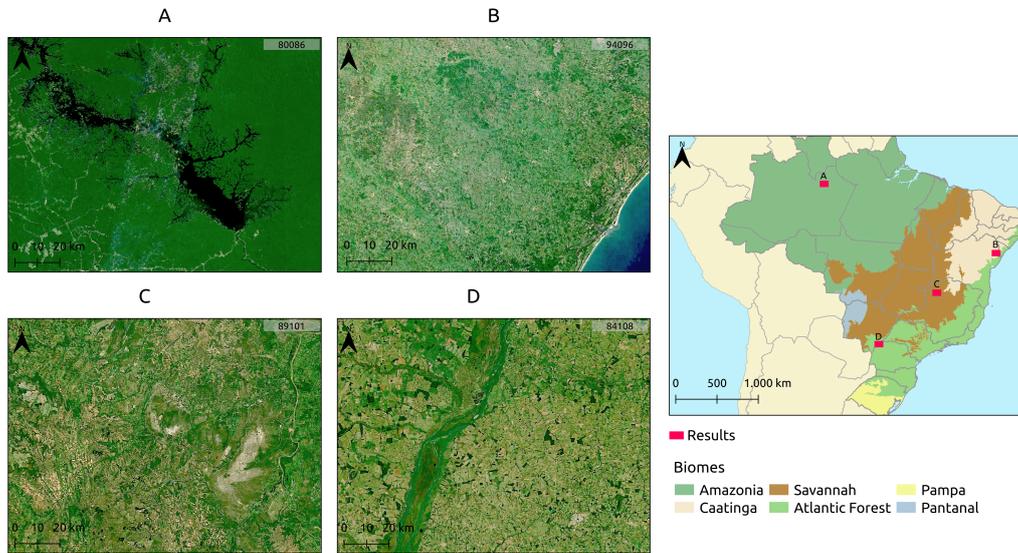


Source: From author.

After the training phase the four regions were selected for prediction by the model Figure 3.3:

- A: tile 80086 is located in the north of the country in the Amazon region, its main body of water is the Balbina reservoir;
- B: tile 94096 is located in the northeast of Brazil between the cities of Salvador and Aracaju and explores the coastal region;
- C: Tile 89101 is located in southeastern Brazil in the Cerrado biome. The main river that flows through the region is the São Francisco River;
- D: Tile 84108 is located on the border between Mato Grosso do Sul and Paraná and has the Paraná River as its main river.

### 3.3.2 Configuration

Each study area has images with the format $10986 \times 16806$ pixels, corresponding to a BDC grid tile. For the input format of the model, it is an array [92315358,69], where the first component has half the number of pixels in the image; since a stride of one pixel was made due to computational constraints, that will not harm the development of the model. The second component represents three spectral bands of the Sentinel-2 sensor "B04" (red), "B08" (nir), and "B11" (swir16), which were resampled to 10m and concatenated for 23 observations per band during the period

from 2020-01-01 to 2020-12 -31. These three spectral bands were selected for this objective because they extract different components from the water, allowing for better characterization.

The correlations between the spectral bands of water samples extracted from ESA map were analyzed, finding three large blocks with different correlations and selecting one band from each block. The first block consists of bands B01 to B05, the second block of band B06 to B09, and the third is composed of bands B11 and B12, which can be seen in Figure 3.4. With this configuration, we have good spectral and temporal characterization.

The methodology proposed by ESA is found in (ZANAGA et al., 2021), which essentially uses samples from Sentinel-1 and 2 time series, and extraction of various features, performs supervised classification, and finally applies expert rules. Because the central part of the methodology works exclusively with time series distances, it is critical to have spectral enrichment, either using indices or more than one band, because two pixels can have a very similar response in one spectral band but a more significant difference in another, which can be crucial to stay in two different clusters, with the addition of temporal features increases furthermore differences.

For this experiment, the size of the SOM grid was defined as $20 \times 20$ with 400 neurons and 40 training epochs. For the hierarchical algorithm, 14 final cluster numbers with the average linkage were used, and the Euclidean distance was defined, e the other hyperparameters were defined by default as in the first experiment.

The land cover legend was based mainly on the temporal attribute of the feature. Thus, the three main classes used to describe the inland water map were:

- Water - Include natural or anthropogenic water bodies, water courses, and water drainage channels. Composed of any features with more than 70% of water persistence during the year. Including lakes, rivers, reservoirs, ponds, lagoons, and estuaries;
- Flooded area - Perennial or periodic water. Composed of features with less than 30% of water persistence during the year includes wetlands, reservoirs, ponds;
- Not Water - Any other land use or land cover. Including, forest, agriculture, pasture, and urban areas.

**Figure 3.4** - Correlation Map for band selection.
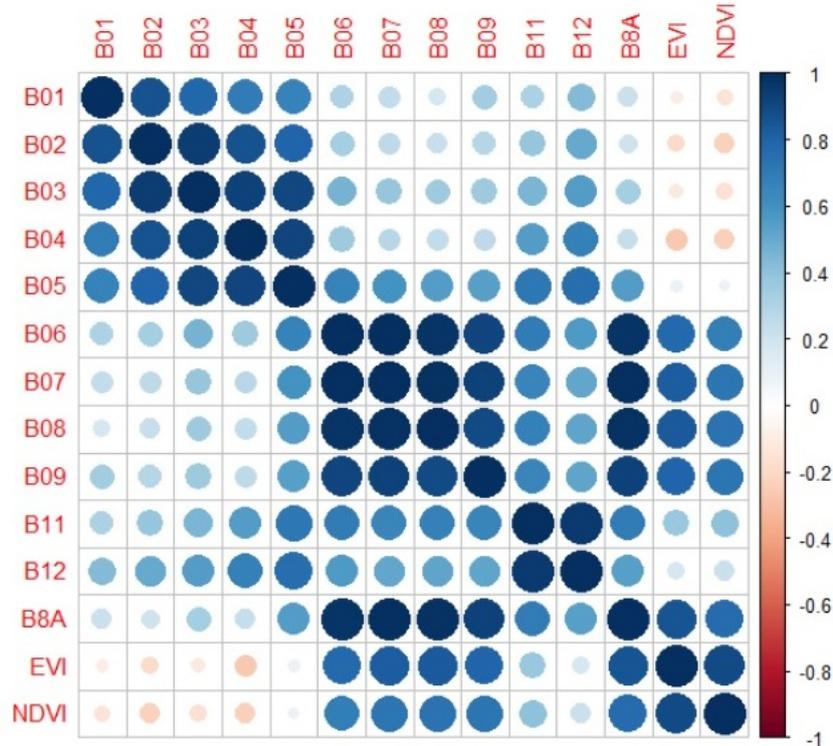
Source: From author.

The expert analyzed the final clusters resulting from the clustering step, classifying each cluster as "water," "flooded area," or "not water." After this step, a land use and land cover classification map are created based on these classes.

The validation process was based on two schemes. First, we use the class water from the ESA WorldCover v100 product, to get the area of intersection and concordance. This product is based on Sentinel-1 and 2 data, available at 10 m spatial resolution for 2020 (ZANAGA et al., 2021), and its global accuracy for South America is around 76% (EUROPEAN SPACE AGENCY (ESA), 2020).

Secondly, we use a pointed random sample, to estimate the accuracy of the classification map. In this scenario, a total of 250 points by each tile were randomly generated considering the area of each class as a stratum.

Those random points were uploaded in the BDC-TerraCollect (BRAZIL DATA CUBE (BDC), 2022b), to identify the class of each one. This platform is an online system that provides tools to check spatially and temporally the values for each point using the same data cube that was used in the classification process.

Finally, the validation points were compared to the classified map, generating the confusion matrix and its indexes, such as global accuracy, user accuracy, and product accuracy. To understand the quality of the map the same set of points was also used to check the accuracy of the ESA product.

After the classification and validation of the six areas of study, the already labeled neurons will be input to the Random Forest model, which through a tunning with *RandomizedSearchCV* with a slipt of 6 folders and 1000 interactions for each fold, used a total of 6000 different trained models and as default score, the estimator itself. The best hyperparameters (Table 3.1) were found with a final score of 97.4 percent.

Table 3.1 - Values used for each hyperparameter during the tunning process.

| Hyperparameter | Value |
|---|---|
| n_estimators | 1200 |
| min_samples_split | 12 |
| min_samples_leaf | 2 |
| max_features | log2 |
| max_depth | 150 |
| bootstrap | True |

Source: From author.

### 3.3.3 Visual analysis

In this section, a visual analysis will be made of two results, the first of the maps generated by the specialist and the second of the maps used for the automatic classification.

In Figure 3.5, there is a crop of the results of the approach proposed by this document in the second line, in the first line are their respective Sentinel-2 RGB composition images, and in the third line, the result of the ESA product of 10m spatial resolution. Note that under each result is the proportion of each class present in each tile.

A visual comparison will be made between the ESA product and the proposed method results for a small region within the six tiles of the training study area, pointing out the main differences. Differences in crop region will be evaluated, and differences that occurred in the tiles in general, which are not necessarily present in the image, considering that each tile has a length of about 18000km$^2$.

On tile 082104 Figure 3.5, located in the Pantanal biome, one of the most heterogeneous biomes in Brazil (ALHO, 2008), with a significant variation of flooded areas

**Figure 3.5** - The first line consists of a crop of the Sentinel-2 image in RGB of the six regions defined for training, the second line of the corresponding result for Proposed Method, and the third line for the ESA 10m product. Below the images, the proportion of classes for each tile.
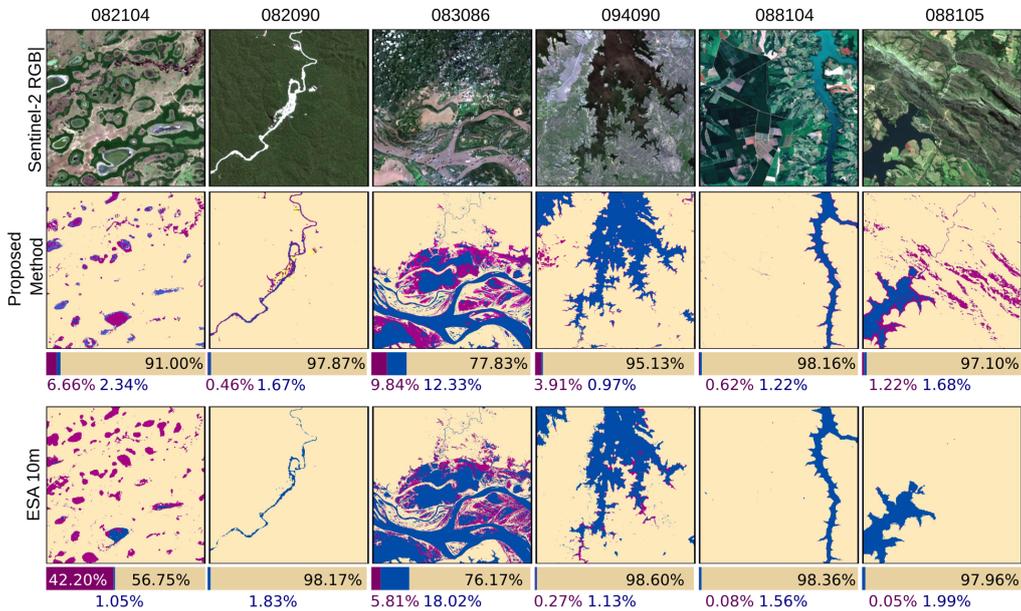
Source: From author.

at different times of the year, and therefore there is great difficulty in mapping this area. We can see a considerable difference in the flooded area between the two maps, the main difference being the legends of these maps for each class. The ESA considers natural herbaceous vegetation (cover of 10% or more) that is permanently or regularly flooded by fresh that are predominant in this tile.

An important thing to observe is that rivers of small widths are more captured and usually grouped with some regions of flooded areas, so they received this classification. A possible cause is the pixels that contemplate vegetation area and water in its mixing model, causing the Influence on the pixel reflectance, where this behavior can also be seen on the edges of rivers.

In tile 082090 Figure 3.5, rivers with a small width were captured in the proposed method result. Another significant difference was that, unlike the ESA product, the result of the proposed method covered mining regions and classified them as flooded areas.

In Tile 83086, which has rivers with a large extension, we notice a big difference between the two classifications of this area: the flooded area and permanent water.

37

However, we can notice in the proposed method result a detailing of permanent rivers in the middle of the flooded region.

In Tile 94090 of Figure 3.5, the proposed method result was better at capturing flooded areas for this region, especially in the river branches. However, these areas had a little confusion with irrigated crops, remembering that the hierarchical step clusterized these areas clustering with 14 clusters. One way to improve this separability is to increase the number of neurons or final clusters. The vast majority of regions are easily separable with 400 neurons, but similar behaviors with different classes may look at the same neuron in this process.

Tiles 88014 and 88105 in Figure 3.5 also maintained the characteristic of capturing more rivers with a small width. In tile 88105, for the proposed method, he mixed the flooded area with mountain shadows, which have a spectral behavior very close to flooded regions since these shadows have a similar response to water. With the seasonal variation in the satellite's passage, the shadows vary in size and location due to the variation in solar incidence, and the behavior resembles flooded areas.

Figure 3.7 shows more detail a behavior that occurred in all tiles where rivers with a small width were mapped by the proposed method and ESAs were not.

**Figure 3.6** - The figure consists of four components. The first is the Sentinel-2 RGB image of the region; the second one is the unsupervised map resulting from the SOM step; the third is the unsupervised map resulting from the Hierarchical Clustering step; and finally, the SOM map with 400 neurons and Clustering by hierarchical Clustering with 14 clusters corresponding to the colors of the map.



Source: From author.

38
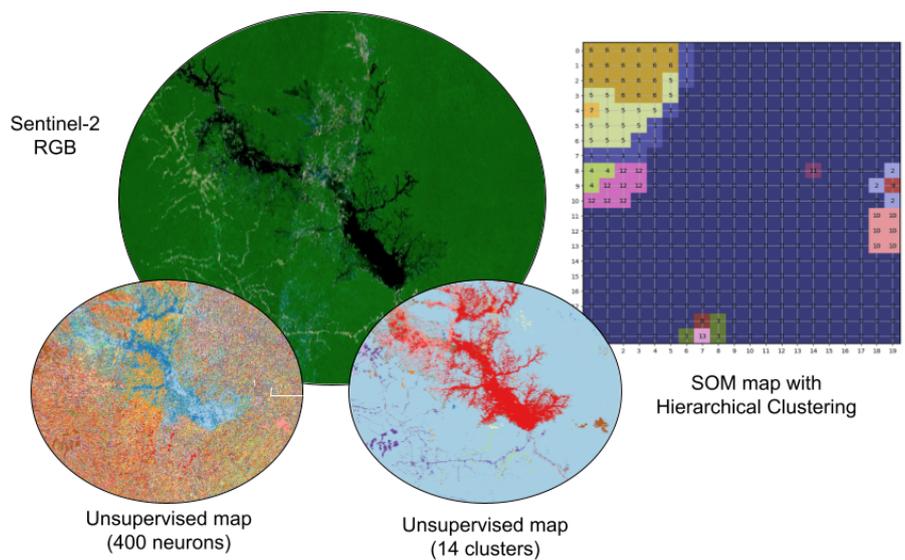
**Figure 3.7** - The Figure compares the proposed method and ESA map from tile 80086 of two zoomed regions with the Sentinel-2 RGB image corresponding to the center region.



Zoomed Proposed Method map — Zoomed Sentinel-2 RGB — Zoomed ESA 10m map

☐ Not Water ■ Water ■ Flooded area
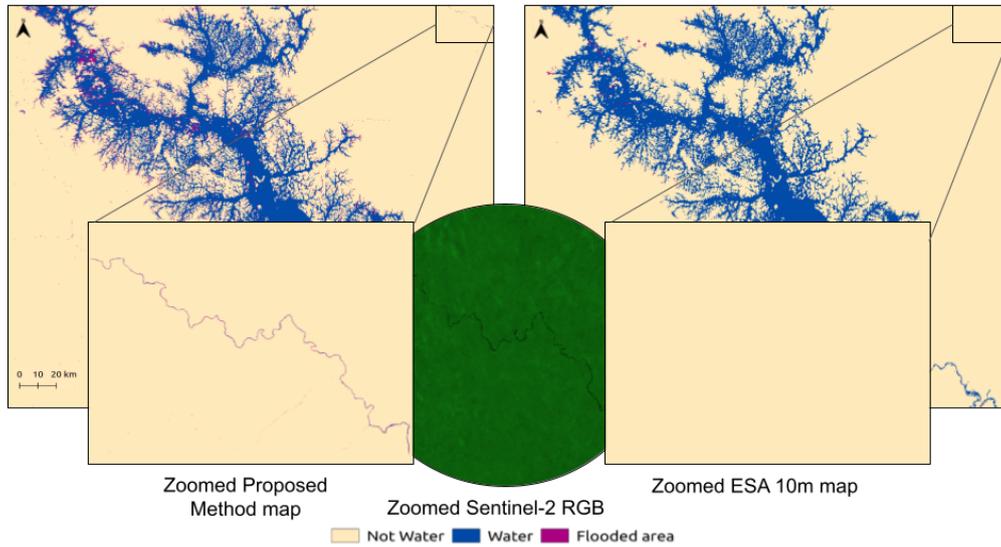
Source: From author.

After training and selecting the best model, four new region was selected, Figure 3.3, including the Balbina Dam in the state of Amazonas . There, the time series were extracted, and interpolation was performed for SOM input, resulting in a SOM map with unlabeled neurons, which in turn generates the unsupervised map with 400 neurons like as in Figure 3.6 applied for tile 80086, along with the map with clustering of neurons by HCA, which generates the map with 14 clusters also find in Figure 3.6.

The unsupervised map (400 neurons) in Figure 3.6 consists of SOM clustering, i.e., the image has 400 distinct temporal and spectral patterns. A new clustering is performed over the neurons represented by the codebook vectors. Figure 3.9 shows the SOM map with Hierarchical clustering, where colors represent the clustering of neurons. The 400 patterns represented by neurons have been reduced to 14, and it is now possible again to create an unsupervised map with (14 clusters).

Figure 3.8 shows the SOM map labeled by the trained model that labels the neurons individually based on their vector codebook. Note that the classification of neurons for water and flooded area is concentrated in one part of the SOM map due to the topology conservation property. In Figure 3.8, it is also possible to see the classification of each vector codebook with visually different standards for Nir. Neurons

39

classified as water remained below 0.2 of reflectance, those classified as flooded had midyear fluctuation, and not water remained near or above 0.2.

Note that for not water patterns, some codebook vectors for a given date exhibited some noise and reached reflectance values close to 1. This occurrence is due to the cloud mask not being 100% and not filtering the time series perfectly, so the map SOM captures these noisy time series in a given neuron.

Figure 3.9 shows the map with the proposed method result for the whole tile for four regions defined for training.

The tile 89101 in Figure 3.10 captured well the regions of flooded areas on the banks of the main river in the region. At the same time, tile 84108 showed the most significant differences in mapping small regions with irrigated agriculture as a flooded area and also captured smaller-diameter rivers.

On tile 94096, a better resolution by proposed method prevails, especially for the water tanks in the south of Figure 3.10 and for part of the river branches that were better delineated. Another difference was that ESA products near the rivers mapped more flooded areas.

**Figure 3.8** - On the left is the SOM map with the classification of each neuron; on the right is the classified vector codebook corresponding to each neuron.
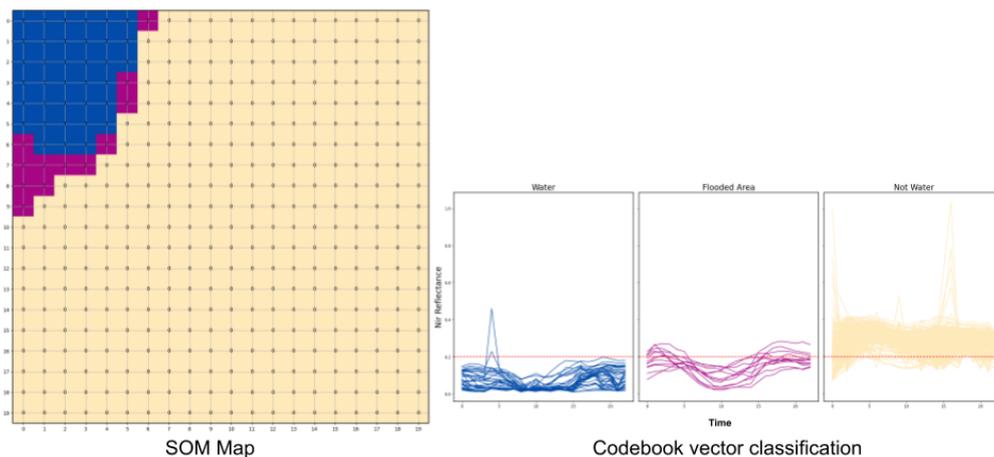


Source: From author.

**Figure 3.9** - The figure presents the Proposed Method results for the four regions selected for prediction, the tiles 80086, 94096, 89101, and 84108.
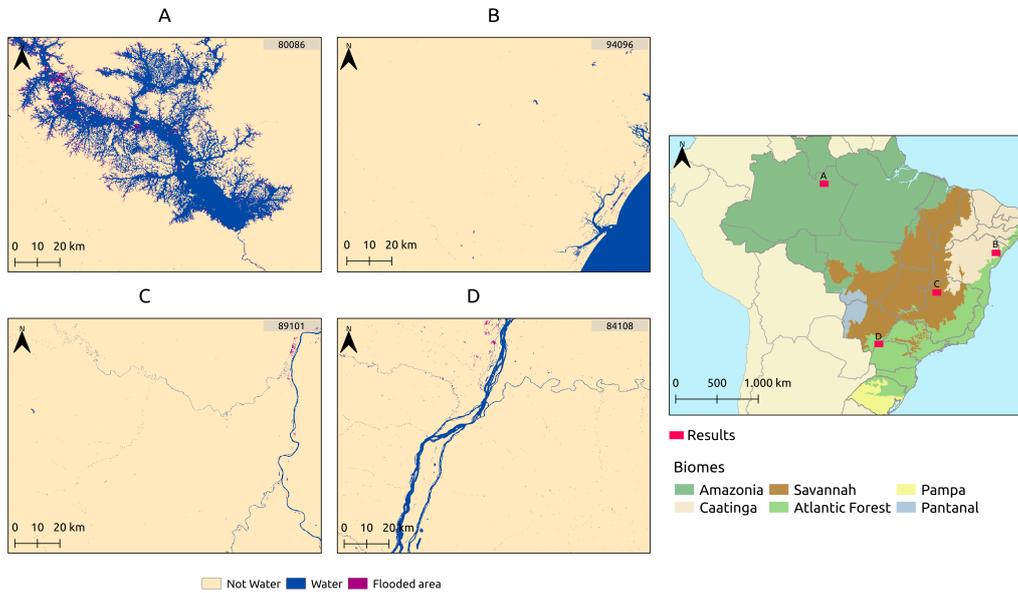


Source: From author.

It is noteworthy that all the areas mapped in Figure 3.10 were not worked on by the specialist. The trained model labeled the neurons with the respective class later used to create the map of use and coverage.

**Figure 3.10** - Visual comparison between the proposed method and the ESA product for the tile 94090, 82104 and 83086.
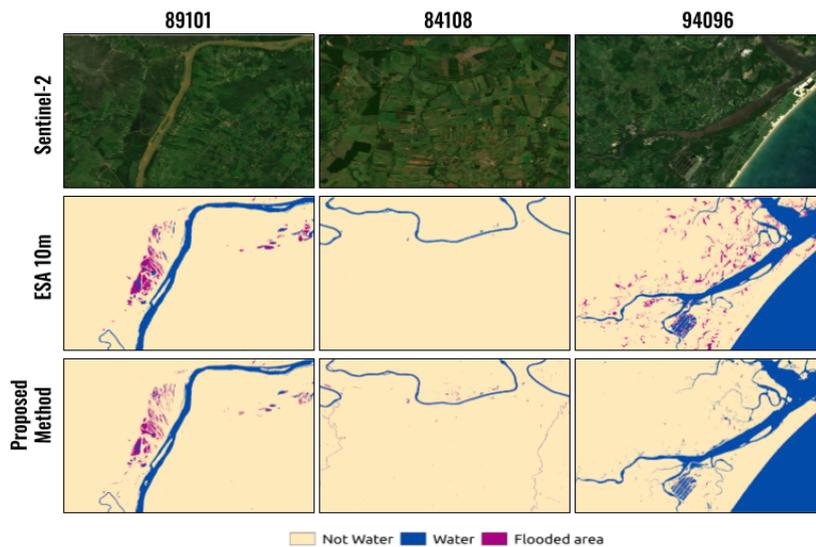


Source: From author.

### 3.3.4 Quantitative analysis

For each region with water bodies, maps of use and coverage were produced with three classes: "water", "non-water" and "wetland area" to check the quality of each map, and later it was quantitatively compared with the map of the corresponding ESA classes. The validation process was conducted by an expert who randomly generated 250 samples for each land use and land cover class based on the ESA map. The samples were selected with a minimum distance between each other. Each sample was classified blindly using only time series and satellite images.

The quantitative result of the validation can be seen in Figure 3.11, where tile 82090 for Proposed Method achieved 77.2% and ESA product 82.8% and tile 88105 Proposed Method achieved 74% with 78.8% of ESA, with ESA products having an advantage and in the other tiles Proposed Method had an advantage over ESA with tile 82104 Proposed Method achieving 85.2% compared to 72.4% for ESA, the other tiles 83086 and 94090 achieved 80% and 91.2% for Proposed Method with 73.2% and 83.3% for ESA, respectively; overall, the proposed approach achieved on average 9% more accuracy for the tiles where it was better and the ESA product achieved 5.2% more in the tiles where it was better.

**Figure 3.11** - Global Accuracy for each tile based in the time-series approach (SOM) and the ESA land use and land cover product (ESA).
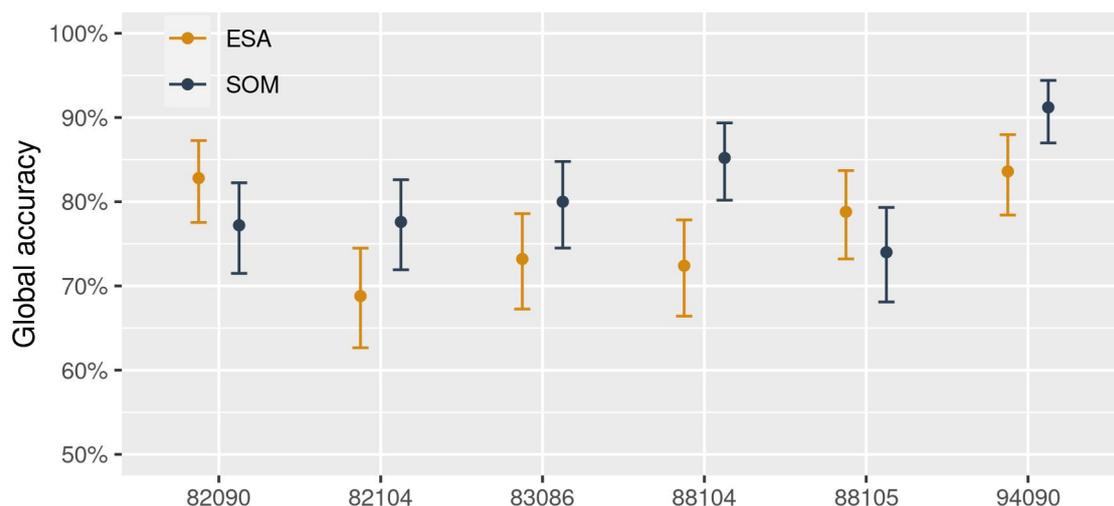


Source: From author.

Table 3.2 shows the concordance matrix of the reference samples and the Proposed Method equivalent class. As with the ESA product, the class with the lowest accuracy is the flooded class that caused the most confusion, as seen in the table.

To compare the mapped area, Table 3.3 shows a confusion matrix in km² for each tile between Proposed Method and ESA. Tile 82104 has a 6610 km² flooded area between the two mappings, expressing the slight difference between the definition of the two classes in the two areas mapped, where this region of Pantanal has a large area of natural herbaceous vegetation, whether flooded or not. The tile 83086, which has the highest water cover, obtained a difference of 1062.8 km² between the maps with a greater coverage for the ESA product; in contrast, the flooded area had a greater mapping area for the Proposed Method with 738.19 km² difference, in the end, approximately the same proportion remained in the classes that contain water.

**Table 3.2** - Confusion matrix for each tile considering the SOM result.

| Tile | Predicted Proposed Method | Reference | | | |
| | | Not water | Water | Flooded | Total |
|---|---|---|---|---|---|
| 082090 | Not water | 106 | 0 | 1 | 107 |
| | Water | 0 | 72 | 1 | 73 |
| | Flooded | 30 | 25 | 15 | 70 |
| | Total | 136 | 97 | 17 | 250 |
| 082104 | Not water | 121 | 3 | 34 | 158 |
| | Water | 0 | 53 | 8 | 61 |
| | Flooded | 1 | 10 | 20 | 31 |
| | Total | 122 | 66 | 62 | 250 |
| 083086 | Not water | 96 | 5 | 14 | 115 |
| | Water | 0 | 70 | 10 | 80 |
| | Flooded | 4 | 17 | 34 | 55 |
| | Total | 100 | 92 | 58 | 250 |
| 088104 | Not water | 137 | 5 | 0 | 142 |
| | Water | 0 | 26 | 0 | 26 |
| | Flooded | 6 | 26 | 50 | 82 |
| | Total | 143 | 57 | 50 | 250 |
| 088105 | Not water | 124 | 14 | 0 | 138 |
| | Water | 0 | 35 | 1 | 36 |
| | Flooded | 25 | 25 | 26 | 76 |
| | Total | 149 | 74 | 27 | 250 |
| 094090 | Not water | 101 | 1 | 4 | 106 |
| | Water | 0 | 88 | 3 | 91 |
| | Flooded | 10 | 4 | 39 | 53 |
| | Total | 111 | 93 | 46 | 250 |

Source: From author.

## 3.4 Conclusions

A method for classifying land use and land cover using clustering methods were presented. The method can be divided into two stages, the first to the creation of

**Table 3.3** - Cross validation between the ESA product and the SOM mapped approach.

| TILE | Proposed Method | ESA | | | Total Proposed Method |
|---|---|---|---|---|---|
| | | Not Water | Water | Wetland | |
| 082090 | Not Water | 18059.53 | 10.12 | 0.03 | 18069.68 |
| | Water | 1.64 | 306.42 | 0.00 | 308.07 |
| | Wetland | 62.63 | 22.60 | 0.09 | 85.32 |
| | Total ESA | 18123.80 | 339.15 | 0.12 | |
| 082104 | Not Water | 10191.95 | 9.79 | 6599.25 | 16800.99 |
| | Water | 28.34 | 142.60 | 261.34 | 432.28 |
| | Wetland | 208.00 | 42.84 | 978.47 | 1229.31 |
| | Total ESA | 10428.28 | 195.23 | 7839.06 | |
| 083086 | Not Water | 13923.03 | 48.28 | 398.32 | 14369.63 |
| | Water | 19.22 | 2224.39 | 33.46 | 2277.07 |
| | Wetland | 102.78 | 1067.20 | 646.40 | 1816.37 |
| | Total ESA | 14045.03 | 3339.87 | 1078.18 | |
| 088104 | Not Water | 18108.96 | 10.47 | 4.22 | 18123.66 |
| | Water | 1.52 | 223.32 | 0.04 | 224.87 |
| | Wetland | 34.44 | 68.92 | 11.18 | 114.54 |
| | Total ESA | 18144.92 | 302.71 | 15.45 | |
| 088105 | Not Water | 17910.75 | 13.49 | 3.02 | 17927.26 |
| | Water | 0.07 | 309.80 | 0.00 | 309.87 |
| | Wetland | 159.87 | 59.10 | 6.96 | 225.94 |
| | Total ESA | 18070.69 | 382.39 | 9.99 | |
| 094090 | Not Water | 17535.85 | 23.12 | 4.28 | 17563.24 |
| | Water | 18.31 | 147.98 | 12.23 | 178.52 |
| | Wetland | 643.73 | 40.34 | 37.24 | 721.31 |
| | Total ESA | 18197.88 | 211.44 | 53.74 | |

Source: From author.

the land use and land cover map by the specialist, which was used as training for a supervised model (Random Forest), and the second part, where the specialist's results were reproduced for other regions.

There is a growth towards using deep learning models for land use and land cover classification, which give excellent results. On the other hand, these models are highly dependent on training data, which must be large and of high quality. Apart from that, the first process of the proposed model could also be used to create maps for deep learning models, considering that large maps can be easily created by specialists, which have better accuracy than fully automated methods.

On the other hand, the proposed model also obtained good results for the classification of inland water bodies in the prediction process, even if only neurons are classified, which also depends on the quality of the segments generated by spatiotemporal segmentation.

A limitation of the model is the classification of land use and land cover classes that involves context, such as shrubland pasture, where a pixel-by-pixel approach would

not solve this problem. For this reason, information from neighboring pixels can be added in segmentation to delete and reduce effects such as salt and pepper and allow the classification of classes that take context into account. In LULC classification, some classes that require context could even be classified considering the geographic features of the input segments. However, it would only be possible to reproduce by extracting features for other regions and use as input to Random Forest.

In a general evaluation, the proposed method performed well in classifying inland water bodies. However, problems such as miss classification in regions such as mountain shadows could be solved with an elevation filter. For problems such as classifying rivers with a small width, using spectral indices for water and adding contextual information could also be evaluated to improve the results. The specialist could consider the shapes of the segments and disassociate the classification to only the neuron, solving type this problem, including mountain shadows.

As for processing, each tile requires a large amount of RAM and VRAM since preprocessing is done with RAM and model training is done with VRAM. A large structure is required for this extensive processing.

The LNCC/SDumont supercomputer was used to create the unsupervised map. The image download phase took an average of 1,100 s per tile, with the most significant overhead of the model occurring in the preprocessing portion, which takes an average of 2,250 s to complete. With the GPU, the execution of the SOM is high-speed, using $92,315,358$ of time series for training, which took an average of 300 s to complete. For the other processes, an average of 500 s was spent, so the entire pipeline for each tile takes an average of 1h and 20m. The SDumont queue sequana_gpu_shared was used for this processing.

The second part of the method was performed on a conventional computer with an Intel(R) Core(TM) i7-10510U CPU @ 1.80GHz with 8 cores, 16GB Ram, and 250GB hard disk space. After the Random Forest was trained, it was speedy in prediction, and then the new pixel labeling was done based on the new classification.

# 4  DEEP LEARNING AS A TOOL TO INTERPOLATE CLOUDY PIXELS IN SENTINEL-2 TIME SERIES[1]

## 4.1  Introduction

Passive sensors, such as the Multi-Spectral Instrument (MSI) onboard Sentinel-2, are extremely susceptible to clouds. Even with this highly temporal recurrence, around five days considering both satellites (EUROPEAN SPACE AGENCY (ESA), 2022), clouds affect the ability to identify land covers, decreasing the accuracy of surface parameters as they interfere in the solar and terrestrial radiation (JENSEN; EPIPHANIO, 2011; COLUZZI et al., 2018).

In Brazil, due to its continental proportion, the cloud cover follows an irregular temporal and spatial distribution. For instance, the Pampa biome presents a constant number of clear imagery regions throughout the year. At the same time, the Amazon suffers from a lack of clear images, more frequently available during the dry season but not too successfully in the Northwest of the biome (MAS et al., 2020).

Most optical imagery already uses cloud detection algorithms to correct the radiance, indicate cloudy areas, and indirectly help users work properly with remote sensing imagery. In this sense, we can point out the Fmask(QIU et al., 2019) used in the Landsat collection; Sen2Cor(LOUIS et al., 2016) used by the European Space Agency (ESA) in their Sentinel-2 satellite products, and; s2cloudless(SENTINEL HUB, 2022) used by the Sentinel Hub's in its derivation products from known satellites.

However, those techniques can overestimate cloud areas harming temporal analyzes. A study performed by Sanchez et al. (2020)(SANCHEZ et al., 2020) in the Amazon tropical forest shows that using the FMask 4 in Sentinel-2 images results in an overall accuracy greater (90%) than the regular Sen2Cor (79%). While Coluzzi et al. (2018) (COLUZZI et al., 2018) shows that globally the Sentinel-2 cloud mask underdetect systematically, reaching a maximum difference (70%) in the Amazon basin.

In analyses that use temporal information, such as deforestation, forest degradation, and phenology, the lack of valid data due to cloud periods is crucial, making the interpolation technique required. There are many available methods to interpolate

---

[1]This chapter is an adaptation of SILVA, B.L.C.; CAMPANHARO, W.A.; SOUZA, F.C.; FERREIRA, K.R.; QUEIROZ, G.R. Deep learning as a tool to interpolate cloudy pixels in Sentinel-2 time series. In: SBSR, 20° Simpósio Brasileiro de Sensoriamento Remoto, 2023.

time series, since more simple such as nearest-neighbor interpolation, until stochastic methods, such as machine learning methods (LEPOT et al., 2017).

In this sense, this document aims to verify if the use of deep learning is capable to improve visually and temporally the values of an image without the use of a cloud mask, and adopting all-time series as input.

## 4.2 Material and methods

We use the Sentinel-2 images from the Brazil Data Cube (BDC) project. The project generates multidimensional analysis-ready data cubes. We use `S2-SEN2COR_10_16D_STK-1` product that contains MSI surface reflectance at full spatial resolution (10m) and 16 days Temporal Compositing considering the SCL cloud mask. The temporal composite is generated from the images with less clouds over the 16 days (FERREIRA et al., 2020c; BRAZIL DATA CUBE (BDC), 2022a).

The Multilayer perceptrons (MLP) were chosen to test a new interpolate approach. This method is a feedforward neural network that indicates that the information flows through layers always forward until it approximates to the determined function and finally reaches the output (GOODFELLOW et al., 2016). The MLP implementation used is from the `scikit-learn MLPRegressor` library (PEDREGOSA et al., 2011).

The methodology follows five main steps, as shown in Figure 4.1. First, we collected 102 random patches across Brazilian territory, selecting areas that encompass each class of clouds following the Sentinel cloud mask (Scene Classification Layer - SCL). Each patch has a 500 x 500 pixels size, encompassing 250,000 time series, which will represent a total of 25,500,000 samples.

Secondly, we extract the time series of one year (01-01-2019 to 12-31-2019) from the Sentinel-2 data cube, including three bands (B04-Red, B08-Nir, B11-Swir). Next, those time series were interpolated using the `interp` function by the `numpy` library (HARRIS et al., 2020), based on the classes "No data", "Saturated or defective", "Dark areas", "Cloud shadow", "Cloud medium probability", "Cloud high probability", "Thin cirrus" and "Snow" contained in the SCL mask. Later, those time series were used to train the MLP model, and then the model was applied in different areas with different cloud cover structures and land use and land cover patterns.

The strategy of this document is to interpolate the time series without using a cloud mask. In this sense, the model input consisted of two-dimensional arrays. The first

**Figure 4.1** - The methodological scheme used in this study, with the main tasks: 1-random points generation; 2-extraction of the time series; 3-linear interpolation; 4-training the MLP model; 5-predicting the cloud-free values for a year of images.
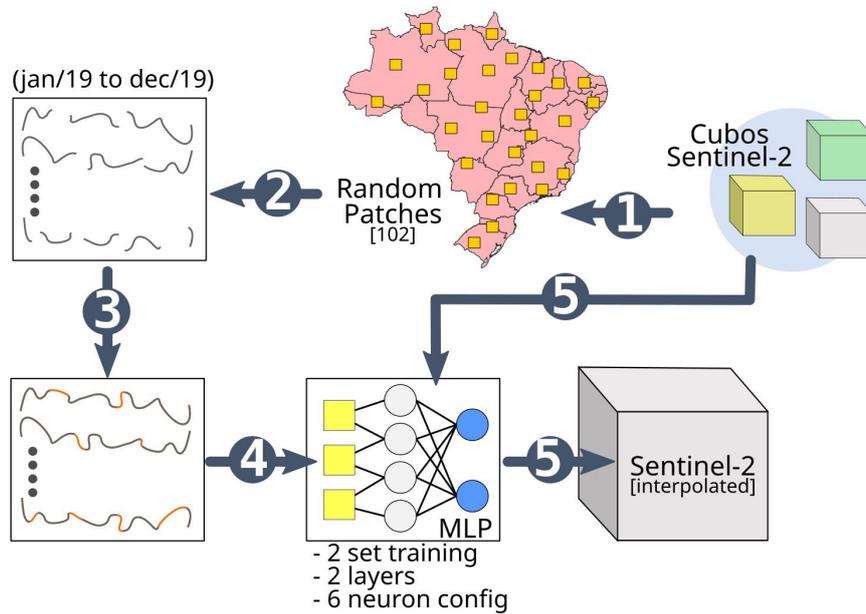
Source: From author.

dimension is the number of samples, and the second is the time resolution of the time series, resulting in an output with the same time dimension as the input.

Two hidden layers were defined with six variations in the number of neurons: [8,4], [16,8], [32,16], [64,32], [128,64], and [256,128], where the first value is the number of neurons in the first layer and the second value is the number of neurons in the second layer.

Another configuration used is related to the total training samples. From the 25,500,000 random time series extracted for the whole Brazilian territory, we set two different networks, one with a step size of 10 pixels and another with a step size of 100 pixels, summarizing 2,550,000 and 255,000 time series, respectively. The activation function used was `relu` with the solver `adam`, the `alpha` was set to equal to 0.001, `batch` size to 1000, `learning rate init` to 0.001, validation fraction of 20%, and the other parameters as default.

We use two methods to validate the deep-learning approach. First, a visual procedure was applied, this method consist in to indicate if the patterns, texture, and features were modified, and all those features were indicated by a specialist.

The second procedure was to check if the physical value (superficial reflectance) changes at the point to invalidate the intrinsic characteristic of the object. We collect random points and compare our approach with linear interpolation. The root mean squared error (RMSE), the mean squared error (MSE), and the mean absolute error (MAE) were used to check the divergences between both approaches.

## 4.3  Results

Figure 4.2 shows the application of the 12 models for one pixel of the NIR band along with the raw time series and the linear interpolated time series based on the cloud mask. The gray highlighted background represents the pixels identified as clouds and needs interpolation.

Table 4.1 shows the reference values for each model with their respective MSE, RMSE, and MAE values calculated using linear interpolation as the baseline. An improvement between the set with 255,000 and 2,550,000 samples is seen in the Table 4.1. On the other hand, in the model with fewer neurons, such as 8_4, 16_8, and 32_16, the time series becomes flatter and loses all the variation. In contrast, the model with more neurons, such as 64_32, 128_64, and 256_128, is more similar to linear interpolation, as can seen in Figure  4.2.

**Figure 4.2** - Result for the six models of each training set, comparing it to the linear interpolation and the raw data. The bars represent the cloud class extract from the SCL band.
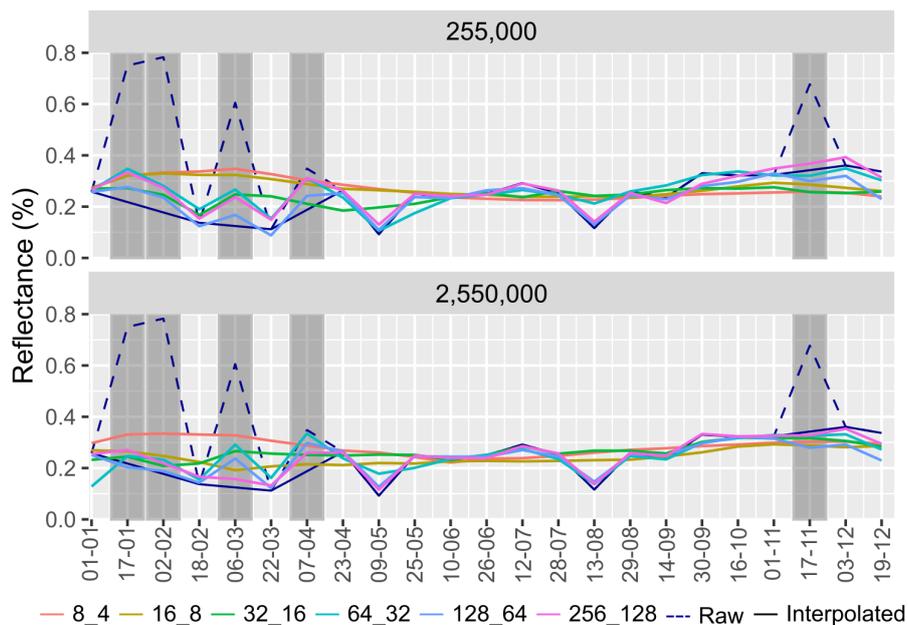


Source: From author.

50

Table 4.1 shows the reference values for each model with their respective MSE, RMSE, and MAE values calculated using linear interpolation as a baseline.

**Table 4.1** - Metrics comparing the six models of each training set with linear interpolation.

| Model configuration | S255,000 | | | S2,550,000 | | |
|---|---|---|---|---|---|---|
| | MSE | RMSE | MAE | MSE | RMSE | MAE |
| 8_4 | 0.0121 | 0.1100 | 0.0866 | 0.0100 | 0.1000 | 0.074 |
| 16_8 | 0.0100 | 0.1000 | 0.0758 | 0.0038 | 0.0616 | 0.0526 |
| 32_16 | 0.0049 | 0.0701 | 0.0577 | 0.0480 | 0.0694 | 0.0461 |
| 64_32 | 0.0038 | 0.062 | 0.0431 | 0.004 | 0.0629 | 0.0426 |
| 128_64 | 0.0014 | 0.0374 | 0.0281 | 0.0022 | 0.0466 | 0.0299 |
| 256_128 | 0.0026 | 0.0510 | 0.0335 | 0.0007 | 0.0258 | 0.0177 |

Source: From author.

Figure 4.3 shows two scenarios of cloud cover selected. In the first, there is the presence of Cirrus clouds and large shadows, indicated by the red and blue polygons. While in the second, there are dense clouds and well-defined shadows. In the baseline section, the raw image and the linear interpolated method result are displaced.

To track the visual efficiency generated by the MLP method, we distribute the images by the number of samples (column) and the number of neurons (lines).

The image reconstructions show a higher variance between the 12 results from the MLP model. Visually the low number of neurons highlighted the features more than the linear interpolations, increasing the contrast and the visual acuity.

However, the cloud's shadow is a persistent artifact in the MLP model, mainly in the model set up with more neurons. From 32_16 the cloud's shadows start to appear as an artifact of the MLP, presents only in the 2,550,000 sample set. After this point, the presence of cloud artifacts was more present, including the insertion of bare soil in the middle of an agricultural area.

## 4.4 Discussion

Table 4.1 shows a small error, with some configurations performing better quantitatively. Looking at Figure 4.3, we can observe some characteristics of each model. The models with fewer neurons were smoother and removed most of the temporal breaks. This characteristic is also evident when we reconstruct the image, as in Figure 4.3. The configurations with fewer neurons removed both the cloud and shadow influence. However, as a result, the temporal variability of use and coverage was lost.

**Figure 4.3** - Visual comparison between the original image (Raw), the linear interpolated, and the MLP results. The column shows the results with the different numbers of samples, and the lines show the variation of neurons in each layer. Composite Color Image: Red-B04, Nir-B08 e Swir-B11.
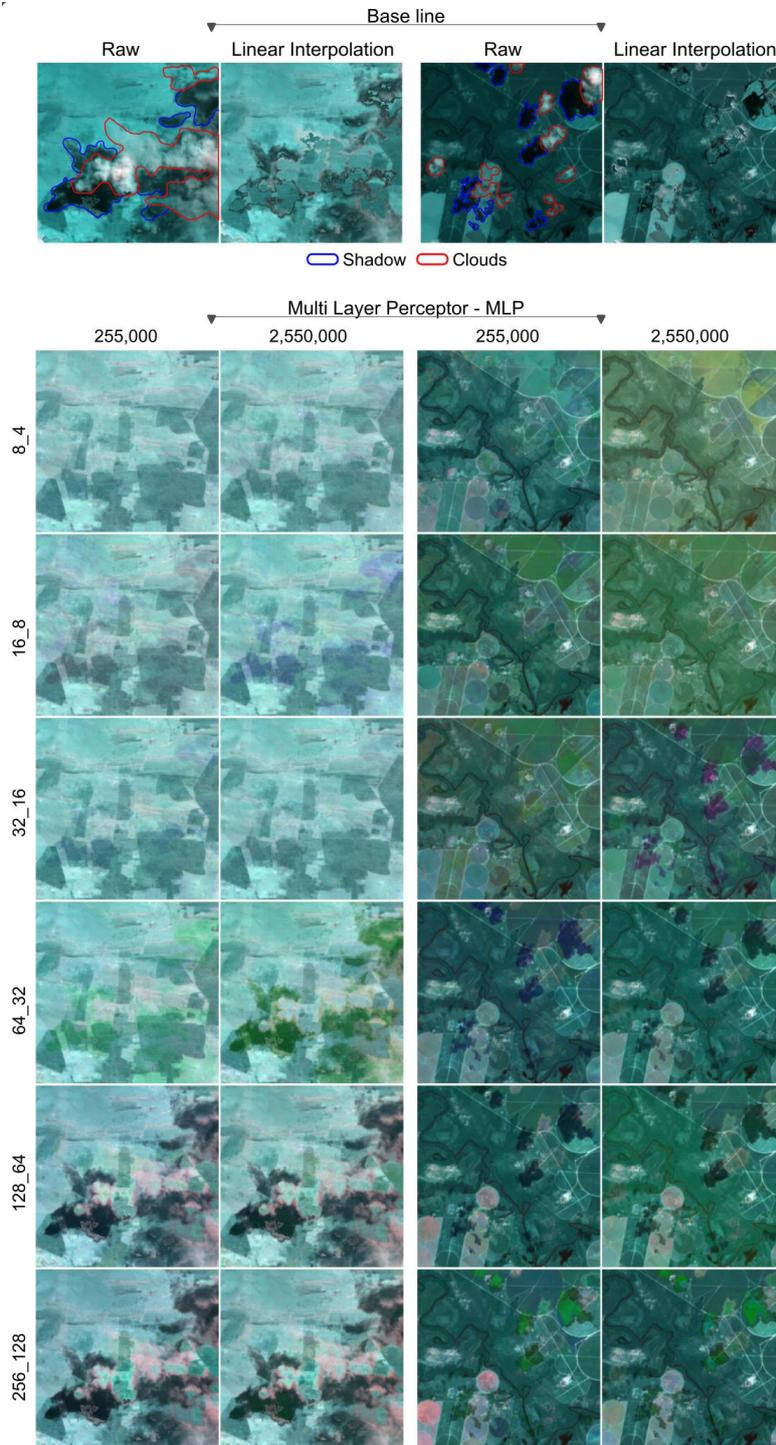


Source: From author.

At the same time, the other images showed little variability, tending to remove, for example, the crop rotation.

The models with more neurons behaved more like linear interpolation. As a result, where the cloud mask has more errors, the model also "learned" the same errors, making it very difficult to interpolate cloud shadows. Although, it can even interpolate some types of clouds well, with the advantage of preserving some temporal variability in the other images.

The Linear Interpolation (LI) uses the SCL cloud mask to orientate where the algorithm should be interpolated. In this sense, in places where the mask made a misclassification, the pixel still shows the influence of the cloud. In those areas, the imagery presents a very characteristic cloud noise.

As pointed out by some authors (COLUZZI et al., 2018; SANCHEZ et al., 2020), the SCL mask under-detects systematically clouds. Thus, even with the best method of filling a gap, the fact of needing the mask as a guide still makes remains noisy in the images.

There are some deep-learning approaches to dealing with clouds. However, they are more interested in indicating where they are and passing the users the process to fill the gap. This is the case of the S2cloudless (SENTINEL HUB, 2022). However, even this algorithm uses a cloud mask as training data.

Depending on the number of neurons used in each layer, the MLP model resembles linear interpolation. However, in those scenarios, the model shows difficulties in dealing with cloud shadow, not smoothing the valley as visible in the models with fewer neurons.

The size and quality of the samples are very important in machine learning applications (MAXWELL et al., 2018; SENTINEL HUB, 2022). The two sets did not disagree too much visually and physically in our approach.

## 4.5   Conclusions

Optical remote sensors are extremely susceptible to clouds, leaving it up to the user to remove them and deal with possible failures. The use of deep-learning models brings an advantage because, simultaneously, the presence of clouds is removed, and the data are already interpolated and smoothed, two essential steps usually performed when working with time series.

Our study shows that the Multilayer perceptron (MLP) has great potential to perform the basics process leaving the image ready to use in time series as pixel data or imagery composition. However, the choice of the number of neurons must be considered since, when reducing this parameter to remove more cloud artifacts, the final interpolation excludes significant variations of features, such as wetlands, crop rotation, and vegetation growth.

While configurations with fewer neurons have been shown to improve the reduction of clouds and cloud shadows, they may also result in a loss of temporal variability at other times, which is not desirable for most applications. Therefore, configurations with more neurons tend to be more useful, even if they do not improve cloud-shadow interpolation, because they preserve the temporal structure of the images. Consequently, it is evident that the training set must have the smallest possible error for the model to learn more effectively.

This study was limited to comparing, at the first moment, the MLP model as an alternative to the linear interpolation technique. Further studies need to be developed to check the factual accuracy of the interpolation, the depth of the model more suitable considering all Brazilian territory, and its potential use to rebuild a cloud mask.

In future work, we intend to use visual interpretation, not the SCL cloud mask, to identify clouds and shadows in the random time series used to train the MLP method. Thus, we will be able to evaluate the deep learning approach regardless of the SCL cloud mask errors.

# 5 CONCLUDING REMARKS

This document is organized as a collection of articles. Chapter 2 is based on the article published at the Proceedings of the The International Society for Photogrammetry and Remote Sensing (ISPRS) congress Silva et al. (2022). Chapter 3 is an article that will still be submitted to some journal. Chapter 4 is an article accepted to be published in the XX Brazilian Remote Sensing Symposium (SBSR 2023) proceeding by Silva et al. (2023). These articles contains the main contributions of this dissertation.

Chapter 2 presents a spatiotemporal segmentation based on two clustering algorithms. Two configurations were used to present the model, and qualitative and quantitative evaluations were made to analyze the segments generated by the proposed methodology. Note that due to the nature of spatiotemporal segmentation, some polygons represent land use and cover classes that are often difficult to evaluate with a single image. On the other hand, land use and cover types that have characteristic temporal profiles, such as agriculture, deforestation, and forest fires, can be represented by segments. Thus, with the knowledge of experts, it is possible to generate segments of semantic information analyzing these temporal profiles.

The methodology and experiments presented in Chapter 2 run on a Google Collab virtual machine with the following specifications: Intel(R) Xeon(R) CPU @ 2.20GHz with 2 cores, 12GB ram and 25GB disk space, Nvidia K80 / T4 GPU with 12GB/16GB GPU Memory, which obtained the times of 15 to 20 seconds for training with 20 epochs and approximately 5 seconds for prediction for the SOM algorithm which had the longest execution time. For the development of the work, neighboring pixels are considered for contextual information and different spectral bands, temporal resolutions, and other spatial resolutions are investigated for different land use and cover applications.

Chapter 3 presents a method for classifying inland water bodies based on clustering satellite image time series. Some maps produced by specialist achieved an accuracy of more than 80%, even 90%, even with the difficulty of mapping flooded areas. A visual interpretation was made comparing the proposed method and the ESA product in the prediction maps. Several comments were made, especially in capturing low-width rivers by the proposed method.

The incorporation of methods such as GEOBIA will be explored for future works, using the segmentation map as the basis for the specialist to extract various features

of the segments, including the segment's temporal information. Including contextual information in the clustering, phase can help classify classes that need contextual information.

Chapter 4 discussed a methodology for interpolating clouds, cloud shadows, and other noises using Deep Learning, even if still in an early stage. However, time series interpolation is a critical issue for the quality results of the proposed models, which are based on clustering methods that use time series distance as a clustering criterion. The experiments showed a large variability of results concerning the parameters used in the model. Most of the errors were passed on due to the errors contained in the training data itself.

Deep learning models require a large number of training data, in part due to the nature of the problem, where satellite image time series exhibit large variability. The ideal would be to have more accurate training data to evaluate the model results better, and even with low-accuracy data, great potential has been identified. New parameters with more accurate training data will be evaluated for future work, and a better extract of samples will be proposed.

# REFERENCES

ADAMS, R.; BISCHOF, L. Seeded region growing. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 16, n. 6, p. 641–647, 1994. ISSN 01628828. 22

AGÊNCIA NACIONAL DE ÁGUAS (ANA). **Nota Técnica n52/2020 - atualização e complementação da base de dados nacional de referência de massas d'água**. Brasília, DF: [s.n.], 2020. 27 p. 26

_____. **Massas d'água**. 2022. Disponível em: <https://metadados.snirh.gov.br/geonetwork/srv/por/catalog.search#/metadata/7d054e5a-8cc9-403c-9f1a-085fd933610c>. 26

ALHO, C. J. Biodiversity of the Pantanal: response to seasonal flooding regime and to environmental degradation. **Brazilian Journal of Biology**, v. 68, n. 4 SUPPL., p. 957–966, 2008. ISSN 15196984. 36

ALMEIDA, C. A.; COUTINHO, A. C.; ESQUERDO, J. C. d. M.; ADAMI, M.; VENTURIERI, A.; DINIZ, C. G.; DESSAY, N.; DURIEUX, L.; GOMES, A. R. High spatial resolution land use and land cover mapping of the Brazilian legal Amazon in 2008 using Landsat-5/TM and MODIS data. **Acta Amazonica**, v. 46, n. 3, p. 291–302, 2016. ISSN 00445967. 13

AMITRANO, D.; CECINATI, F.; MARTINO, G. D.; IODICE, A.; MATHIEU, P.-P.; RICCIO, D.; RUELLO, G. Using geobia for feature extraction from multitemporal sar images: preliminary results. In: IEEE. **2017 IEEE 3rd International Forum on Research and Technologies for Society and Industry (RTSI)**. [S.l.], 2017. p. 1–4. 25

APPEL, M.; PEBESMA, E. On-demand processing of data cubes from satellite image collections with the gdalcubes library. **Data**, v. 4, n. 3, 2019. ISSN 23065729. 1, 9

BAGAN, H.; WANG, Q.; WATANABE, M.; YANG, Y.; MA, J. Land cover classification from MODIS EVI times-series data using SOM neural network. **International Journal of Remote Sensing**, v. 26, n. 22, p. 4999–5012, 2005. ISSN 13665901. 25

BELGIU, M.; CSILLIK, O. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. **Remote Sensing of Environment**, v. 204, p. 509–523, jan. 2018. ISSN 0034-4257. 1, 25

BELGIU, M.; DRăGU, L. Random forest in remote sensing: a review of applications and future directions. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 114, p. 24–31, 2016. ISSN 09242716. 26

BRAZIL DATA CUBE (BDC). **Brazil Data Cube - Cubo de dados**. 2022. Disponível em: <http://brazildatacube.org/cubo-de-dados/>. 48

_____. **TerraCollect**. 2022. Disponível em: <https://brazildatacube.dpi.inpe.br/terracollect/explore>. 35

BRITO, P. V. da S.; CHAVES, M. E. D.; CARVALHO, H. F. de S.; SOUZA, F. C. de; SILVA, B. L. de Castro e; FERREIRA, K. R.; SANTOS, R. D. C. dos; QUEIROZ, G. R. de. Uso de séries temporais para classificações de uso e cobertura da terra em petrolina, pernambuco. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 20., 2023, Florianópolis, SC. **Anais...** São José dos Campos: INPE, 2023. 3

BROWN, M. E. Remote sensing technology and land use analysis in food security assessment. **J Land Use Sci**, v. 11, n. 6, p. 623–641, 2016. 25

CAMARA, G.; ASSIS, L. F.; RIBEIRO, G.; FERREIRA, K. R.; LLAPA, E.; VINHAS, L. Big earth observation data analytics: matching requirements to system architectures. In: **Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data - BigSpatial '16**. Burlingame, California: ACM Press, 2016. p. 1–6. ISBN 978-1-4503-4581-1. Disponível em: <http://dl.acm.org/citation.cfm?doid=3006386.3006393>. 2, 25

CLINTON, N.; HOLT, A.; SCARBOROUGH, J.; YAN, L. I.; GONG, P. Accuracy assessment measures for object-based image segmentation goodness. **Photogrammetric Engineering and Remote Sensing**, v. 76, n. 3, p. 289–299, 2010. ISSN 00991112. 19

COLUZZI, R.; IMBRENDA, V.; LANFREDI, M.; SIMONIELLO, T. A first assessment of the sentinel-2 level 1-c cloud mask product to support informed surface analyses. **Remote Sensing of Environment**, v. 217, p. 426–443, 2018. ISSN 0034-4257. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0034425718303742>. 47, 53

COSTA, H.; FOODY, G. M.; BOYD, D. S. Supervised methods of image segmentation accuracy assessment in land cover mapping. **Remote Sensing of**

**Environment**, v. 205, n. December 2016, p. 338–351, 2018. ISSN 00344257. Disponível em: <https://doi.org/10.1016/j.rse.2017.11.024>. 18, 19

COSTA, W. S.; FONSECA, L. M. G.; KORTING, T. S.; BENDINI, H. D. N.; SOUZA, R. C. M. D. Spatiooral segmentation applied to optical remote sensing image time series. **IEEE Geoscience and Remote Sensing Letters**, v. 15, n. 8, p. 1299–1303, 2018. ISSN 15580571. 5, 6

EUROPEAN SPACE AGENCY (ESA). **Product validation report (WorldCover_PVR_v1.1)**. [s.n.], out. 2020. 38 p. Disponível em: <https://esa-worldcover.s3.eu-central-1.amazonaws.com/v100/2020/docs/WorldCover_PVR_V1.1.pdf>. 35

_____. **Sentinel-2**. 2022. Disponível em: <https://sentinel.esa.int/web/sentinel/missions/sentinel-2>. 47

EVERITT, B.; LANDAU, S.; LEESE, M. S. **D. Cluster analysis**. [S.l.]: 5.ed. Hoboken, NJ: Wiley, 2011. 8

FEITOSA, R.; FERREIRA, R.; ALMEIDA, C.; CAMARGO, F.; COSTA, G. Similarity metrics for genetic adaptation of segmentation parameters. In: GEOGRAPHIC OBJECT-BASED IMAGE ANALYSIS CONFERENCE, 2010. **Proceedings...** [S.l.], 2010. 20

FERREIRA, K. R.; QUEIROZ, G. R.; CAMARA, G.; SOUZA, R. C.; VINHAS, L.; MARUJO, R. F.; SIMOES, R. E.; NORONHA, C. A.; COSTA, R. W.; ARCANJO, J. S.; GOMES, V. C.; ZAGLIA, M. C. Using remote sensing images and cloud services on aws to improve land use and cover monitoring. In: IEEE LATIN AMERICAN GRSS AND ISPRS REMOTE SENSING CONFERENCE. **Proceedings...** Istanbul, 2020. p. 558–562. 9

FERREIRA, K. R.; QUEIROZ, G. R.; VINHAS, L.; MARUJO, R. F.; SIMOES, R. E.; PICOLI, M. C.; CAMARA, G.; CARTAXO, R.; GOMES, V. C.; SANTOS, L. A.; SANCHEZ, A. H.; ARCANJO, J. S.; FRONZA, J. G.; NORONHA, C. A.; COSTA, R. W.; ZAGLIA, M. C.; ZIOTI, F.; KORTING, T. S.; SOARES, A. R.; CHAVES, M. E.; FONSECA, L. M. Earth observation data cubes for Brazil: requirements, methodology and products. **Remote Sensing**, v. 12, n. 24, p. 1–19, 2020. ISSN 20724292. 1

FERREIRA, K. R.; QUEIROZ, G. R.; VINHAS, L.; MARUJO, R. F. B.; SIMOES, R. E. O.; PICOLI, M. C. A.; CAMARA, G.; CARTAXO, R.; GOMES,

V. C. F.; SANTOS, L. A.; SANCHEZ, A. H.; ARCANJO, J. S.; FRONZA, J. G.; NORONHA, C. A.; COSTA, R. W.; ZAGLIA, M. C.; ZIOTI, F.; KORTING, T. S.; SOARES, A. R.; CHAVES, M. E. D.; FONSECA, L. M. G. Earth observation data cubes for brazil: Requirements, methodology and products. **Remote Sensing**, v. 12, n. 24, 2020. ISSN 2072-4292. Disponível em: <https://www.mdpi.com/2072-4292/12/24/4033>. 48

FU, T. A review on time series data mining. **Engineering Applications of Artificial Intelligence**, v. 24, n. 1, p. 164–181, 2011. ISSN 0952-1976. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0952197610001727>. 27

GARNOT, V. S. F.; LANDRIEU, L. Panoptic segmentation of satellite image time series with convolutional temporal attention networks. In: INTERNATIONAL CONFERENCE ON COMPUTER VISION, 2021. **Proceedings...** [S.l.]: IEEE, 2021. p. 4852–4861. 5

GIULIANI, G.; CAMARA, G.; KILLOUGH, B.; MINCHIN, S. Earth observation open science: enhancing reproducible science using data cubes. **Data**, v. 4, n. 4, p. 4–9, 2019. ISSN 23065729. 2

GIULIANI, G.; CHATENOUX, B.; BONO, A. D.; RODILA, D.; RICHARD, J. P.; ALLENBACH, K.; DAO, H.; PEDUZZI, P. Building an Earth observations data cube: lessons learned from the Swiss Data Cube (SDC) on generating Analysis Ready Data (ARD). **Big Earth Data**, v. 1, n. 1-2, p. 100–117, 2017. ISSN 25745417. Disponível em: <http://doi.org/10.1080/20964471.2017.1398903>. 1

GIULIANI, G.; MAZZETTI, P.; SANTORO, M.; NATIVI, S.; BEMMELEN, J. V.; COLANGELI, G.; LEHMANN, A. Knowledge generation using satellite earth observations to support sustainable development goals (sdg): a use case on land degradation. **International Journal of Applied Earth Observation**, v. 88, p. 102068, 2020. 1

GOMES, V. C.; QUEIROZ, G. R.; FERREIRA, K. R. An overview of platforms for big earth observation data management and analysis. **Remote Sensing**, v. 12, n. 8, p. 1–25, 2020. ISSN 20724292. 2

GÓMEZ, C.; WHITE, J. C.; WULDER, M. A. Optical remotely sensed time series data for land cover classification: a review. **ISPRS Journal Photogramm**, v. 116, p. 55–72, 2016. 1, 25

GONÇALVES, M. L.; NETTO, M. L.; COSTA, J. A.; JUNIOR, J. Z. An unsupervised method of classifying remotely sensed images using Kohonen self-organizing maps and agglomerative hierarchical clustering methods. **International Journal of Remote Sensing**, v. 29, p. 3171–3207, 2008. 25

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.]: MIT Press, 2016. `http://www.deeplearningbook.org`. 48

HARRIS, C. R.; MILLMAN, K. J.; WALT, S. J. van der; GOMMERS, R.; VIRTANEN, P.; COURNAPEAU, D.; WIESER, E.; TAYLOR, J.; BERG, S.; SMITH, N. J.; KERN, R.; PICUS, M.; HOYER, S.; KERKWIJK, M. H. van; BRETT, M.; HALDANE, A.; RÍO, J. F. del; WIEBE, M.; PETERSON, P.; GÉRARD-MARCHANT, P.; SHEPPARD, K.; REDDY, T.; WECKESSER, W.; ABBASI, H.; GOHLKE, C.; OLIPHANT, T. E. Array programming with NumPy. **Nature**, v. 585, n. 7825, p. 357–362, set. 2020. Disponível em: <`https://doi.org/10.1038/s41586-020-2649-2`>. 48

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Manual de procedimentos técnicos para fiscalização, controle de qualidade e validação da base cartográfica contínua na escala 1:250.000**. Rio de Janeiro - RJ: [s.n.], 2011. 31 p. 27

_____. **Bases cartográficas contínuas - Brasil**. 2022. Disponível em: <`https://www.ibge.gov.br/geociencias/cartas-e-mapas/bases-cartograficas-continuas/15759-brasil.html?=&t=acesso-ao-produto`>. 27

JANSSEN, L. L.; MOLENAAR, M. Terrain objects, their fynamics and their monitoring by the integration of GIS and remote sensing. **IEEE Transactions on Geoscience and Remote Sensing**, v. 33, n. 3, p. 749–758, 1995. ISSN 15580644. 20

JENSEN, J. R.; EPIPHANIO, J. C. N. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. São José dos Campos, SP: Parêntese, 2011. ISBN 978-85-60507-06-1. 47

KAUFMAN, P. L. **Finding groups in data: an introduction to cluster analysis**. 9. ed. [S.l.]: Wiley: Hoboken, NJ, USA, 1990. 8

KILLOUGH, B. The Impact of Analysis Ready Data in the Africa Regional Data Cube. In: INTERNATIONAL GEOSCIENCE AND REMOTE SENSING SYMPOSIUM, 2019. **Proceedings...** [S.l.]: IEEE, 2019. p. 5646–5649. 1

KOHONEN, T. The self-organizing map. **Proceedings of the IEEE**, v. 78, n. 9, p. 1464–1480, 1990. 3

_____. The self-organizing map. **Neurocomputing**, v. 21, n. 1-3, p. 1–6, 1998. ISSN 09252312. 6, 8

KOPP, S.; BECKER, P.; DOSHI, A.; WRIGHT, D. J.; ZHANG, K.; XU, H. Data cube. **Definitions**, 2020. 1

KöRTING, T.; FONSECA, L.; CâMARA, G. A geographical approach to self-organizing maps algorithm applied to image segmentation. In: **BLANC-TALON, J.; KLEIHORST, R.; PHILIPS, W.; POPESCU, D.; SCHEUNDERS, P. (Ed.). Advanced concepts for intelligent vision systems.** Berlin: Springer, 2011. 8, 11

LEPOT, M.; AUBIN, J.-B.; CLEMENS, F. H. Interpolation in time series: an introductive overview of existing methods, their performance criteria and uncertainty assessment. **Water**, v. 9, n. 10, 2017. ISSN 2073-4441. Disponível em: <https://www.mdpi.com/2073-4441/9/10/796>. 48

LEWIS, A.; OLIVER, S.; LYMBURNER, L.; EVANS, B.; WYBORN, L.; MUELLER, N.; RAEVKSI, G.; HOOKE, J.; WOODCOCK, R.; SIXSMITH, J.; WU, W.; TAN, P.; LI, F.; KILLOUGH, B.; MINCHIN, S.; ROBERTS, D.; AYERS, D.; BALA, B.; DWYER, J.; DEKKER, A.; DHU, T.; HICKS, A.; IP, A.; PURSS, M.; RICHARDS, C.; SAGAR, S.; TRENHAM, C.; WANG, P.; WANG, L. W. The Australian geoscience data cube — foundations and lessons learned. **Remote Sensing of Environment**, v. 202, p. 276–292, 2017. ISSN 00344257. Disponível em: <http://dx.doi.org/10.1016/j.rse.2017.03.015>. 1

LOUIS, J.; DEBAECKER, V.; PFLUG, B.; MAIN-KNORN, M.; BIENIARZ, J.; MUELLER-WILM, U.; CADAU, E.; GASCON, F. Sentinel-2 sen2cor: L2a processor for users. In: OUWEHAND, L. (Ed.). **ESA Living Planet Symposium 2016**. Spacebooks Online, 2016. (ESA Special Publications, SP-740), p. 1–8. Disponível em: <https://elib.dlr.de/107381/>. 47

MANCINI, R.; RITACCO, A.; LANCIANO, G.; CUCINOTTA, T. Xpysom: high-performance self-organizing maps. In: INTERNATIONAL SYMPOSIUM ON COMPUTER ARCHITECTURE AND HIGH PERFORMANCE COMPUTING, 32., 2020. **Proceedings...** [S.l.]: IEEE, 2020. p. 209–216. 13

MAPBIOMAS. **Projeto MapBiomas - Coleção 7 da série anual de mapas de cobertura e uso da terra do Brasil**. 2022. Disponível em: <https://mapbiomas.org/colecoes-mapbiomas-1>. 27, 30

MAS, J. F.; SOPCHAKI, C. H.; RABELO, F. D. B.; ARAúJO, F. S. d.; SOLóRZANO, J. V. Análise da disponibilidade de imagens landsat e sentinel para o Brasil. **Geografia Ensino & Pesquisa**, v. 24, dez. 2020. Disponível em: <https://periodicos.ufsm.br/geografia/article/view/61324>. 47

MAXWELL, A. E.; WARNER, T. A.; FANG, F. Implementation of machine-learning classification in remote sensing: an applied review. **International Journal of Remote Sensing**, v. 39, n. 9, p. 2784–2817, 2018. 53

MORETTIN, P. A.; TOLOI, C. M. d. C. **Análise de séries temporais**. São Paulo - SP: Edgard Blucher, 2008. OCLC: 319222044. ISBN 978-85-212-0389-6. 27

PASQUARELLA, V. J.; HOLDEN, C. E.; KAUFMAN, L.; WOODCOCK, C. E. From imagery to ecology: leveraging time series of all available LANDSAT observations to map and monitor ecosystem state and dynamics. **Remote Sensing in Ecology and Conservation**, v. 2, n. 3, p. 152–170, 2016. ISSN 2056-3485. 2, 25

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011. 13, 48

PEKEL, J.-F.; COTTAM, A.; GORELICK, N.; BELWARD, A. S. High-resolution mapping of global surface water and its long-term changes. **Nature**, v. 540, n. 7633, p. 418–422, dez. 2016. ISSN 0028-0836, 1476-4687. Disponível em: <http://www.nature.com/articles/nature20584>. 27

PELLETIER, C.; VALERO, S.; INGLADA, J.; CHAMPION, N.; DEDIEU, G. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. **Remote Sensing of Environment**, v. 187, p. 156–168, 2016. ISSN 00344257. Disponível em: <http://dx.doi.org/10.1016/j.rse.2016.10.010>. 26

PETITJEAN, F.; KURTZ, C.; PASSAT, N.; GANÇARSKI, P. Spatio-temporal reasoning for the classification of satellite image time series. **Pattern Recognition Letters**, v. 33, n. 13, p. 1805–1815, 2012. ISSN 01678655. 5

PICOLI, M. C. A.; CAMARA, G.; SANCHES, I.; SIMÕES, R.; CARVALHO, A.; MACIEL, A.; COUTINHO, A.; ESQUERDO, J.; ANTUNES, J.; BEGOTTI, R. A.; ARVOR, D.; ALMEIDA, C. Big earth observation time series analysis for monitoring Brazilian agriculture. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 145, p. 328–339, 2018. ISSN 09242716. Disponível em: <https://doi.org/10.1016/j.isprsjprs.2018.08.007>. 1, 25

QIU, S.; ZHU, Z.; HE, B. Fmask 4.0: improved cloud and cloud shadow detection in landsats 4–8 and sentinel-2 imagery. **Remote Sensing of Environment**, v. 231, p. 111205, 2019. ISSN 0034-4257. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0034425719302172>. 47

RIJSBERGEN, C. van. **Information retrieval**. 2. ed. London: [s.n.], 1979. 20

SAAH, D. et al. Primitives as building blocks for constructing land cover maps. **International Journal of Applied Earth Observation**, v. 85, p. 101979, 2020. 1, 25

SANCHEZ, A. H.; PICOLI, M. C. A.; CAMARA, G.; ANDRADE, P. R.; CHAVES, M. E. D.; LECHLER, S.; SOARES, A. R.; MARUJO, R. F. B.; SIMõES, R. E. O.; FERREIRA, K. R.; QUEIROZ, G. R. Comparison of cloud cover detection algorithms on sentinel–2 images of the amazon tropical forest. **Remote Sensing**, v. 12, n. 8, p. 1284, Apr 2020. ISSN 2072-4292. Disponível em: <http://dx.doi.org/10.3390/rs12081284>. 47, 53

SANTOS, L. A.; FERREIRA, K.; PICOLI, M.; CAMARA, G.; ZURITA-MILLA, R.; AUGUSTIJN, E.-W. Identifying spatiotemporal patterns in land use and cover samples from Satellite Image Time Series. **Remote Sensing**, v. 13, n. 5, 2021. ISSN 2072-4292. Disponível em: <https://www.mdpi.com/2072-4292/13/5/974>. 3, 6, 26

SENTINEL HUB. **Cloud masks and cloud probabilities**. 2022. Disponível em: <https://docs.sentinel-hub.com/api/latest/user-guides/cloud-masks/>. 47, 53

SILVA, B. L. C.; CAMPANHARO, W. A.; SOUZA, F.; FERREIRA, K.; QUEIROZ, G. R. de. Deep learning as tool to interpolate cloudy pixels in

sentinel-2 time series. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO
REMOTO, 20., 2023, Florianópolis, SC. **Anais...** São José dos Campos: INPE,
2023. 3, 55

SILVA, B. L. C.; SOUZA, F. C.; FERREIRA, K. R.; QUEIROZ, G. R.; SANTOS,
L. A. Spatiotemporal segmentation of satellilte image time series using
self-organizaing map. **ISPRS Annals of the Photogrammetry, Remote
Sensing and Spatial Information Sciences**, v. 3, p. 255–261, maio 2022. ISSN
2194-9050. Disponível em:
<https://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/
V-3-2022/255/2022/>. 3, 5, 55

SIMOES, R.; CAMARA, G.; QUEIROZ, G.; SOUZA, F.; ANDRADE, P. R.;
SANTOS, L.; CARVALHO, A.; FERREIRA, K. Satellite image time series
analysis for big earth observation data. **Remote Sensing**, v. 13, n. 13, 2021. ISSN
2072-4292. Disponível em: <https://www.mdpi.com/2072-4292/13/13/2428>. 1,
2, 5, 25

SIMOES, R.; PICOLI, M. C.; CAMARA, G.; MACIEL, A.; SANTOS, L.;
ANDRADE, P. R.; SÁNCHEZ, A.; FERREIRA, K.; CARVALHO, A. Land use
and cover maps for Mato Grosso State in Brazil from 2001 to 2017. **Scientific
Data**, v. 7, n. 1, p. 1–10, 2020. ISSN 20524463. 1, 25, 26

SIMOES, R.; SANCHEZ, A.; PICOLI, M. **segmetric: Metrics for Assessing
Segmentation Accuracy for Geospatial Data**. Sao Jose dos Campos, Brazil,
2021. Disponível em: <https://github.com/michellepicoli/segmetric>. 21

SOILLE, P.; BURGER, A.; MARCHI, D. D.; KEMPENEERS, P.; RODRIGUEZ,
D.; SYRRIS, V.; VASILEV, V. A versatile data-intensive computing platform for
information retrieval from big geospatial data. **Future Generation Computer
Systems**, v. 81, p. 30–40, 2018. ISSN 0167739X. Disponível em:
<https://doi.org/10.1016/j.future.2017.11.007>. 1, 5, 25

SOUZA, C. M.; SHIMBO, J. Z.; ROSA, M. R.; PARENTE, L. L.; ALENCAR, A.
A.; RUDORFF, B. F. T.; HASENACK, H.; MATSUMOTO, M.; FERREIRA, L.
G.; SOUZA-FILHO, P. W. M.; OLIVEIRA, S. W. de; 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. 17, 2020. ISSN 2072-4292. Disponível em: <https://www.mdpi.com/2072-4292/12/17/2735>. 27

XI, W.; DU, S.; WANG, Y. C.; ZHANG, X. A spatiotemporal cube model for analyzing satellite image time series: application to land-cover mapping and change detection. **Remote Sensing of Environment**, v. 231, n. May, p. 111212, 2019. ISSN 00344257. Disponível em: <https://doi.org/10.1016/j.rse.2019.111212>. 5

ZANAGA, D.; KERCHOVE, R. V. D.; KEERSMAECKER, W. D.; SOUVERIJNS, N.; BROCKMANN, C.; QUAST, R.; WEVERS, J.; GROSU, A.; PACCINI, A.; VERGNAUD, S.; CARTUS, O.; SANTORO, M.; FRITZ, S.; GEORGIEVA, I.; LESIV, M.; CARTER, S.; HEROLD, M.; LI, L.; TSENDBAZAR, N.-E.; RAMOINO, F.; ARINO, O. **ESA WorldCover 10 m 2020 v100**. Zenodo, out. 2021. Disponível em: <https://doi.org/10.5281/zenodo.5571936>. 27, 34, 35

ZHANG, Y. J. A survey on evaluation methods for image segmentation. **Pattern Recognition**, v. 29, n. 8, p. 1335–1346, 1996. ISSN 00313203. 20

ZURITA-MILLA, R.; GIJSEL, J. A. V.; HAMM, N. A.; AUGUSTIJN, P. W.; VRIELING, A. Exploring spatiotemporal phenological patterns and trajectories using self-organizing maps. **IEEE Transactions on Geoscience and Remote Sensing**, v. 51, n. 4, p. 1914–1921, 2013. ISSN 01962892. 3, 6