Coastal Zone– Appendix

Collection 6

Version 1

General coordinator
Pedro Walfir Martins e Souza Filho

Team
Alexandre Filho
Cesar Guerreiro Diniz
Luis Waldir Rodrigues Sadeck
Luiz Cortinhas Ferreira Neto
Maria Luize Silva Pinheiro

1   Overview

The Brazilian coastal zone presents a very diverse suite of coastal environments that evolved during the Quaternary in response to changes in climate and sea-level changes, showing an interaction between different sediment supplies and a geologic heritage that dates back to the breakup of South America and Africa (Dominguez, 2009). Among this diversity of coastal features, five classes are mapped in the MapBiomas Collection 5; Mangrove, Beaches and Dunes, Aquaculture, Salt Flats.

In comparison to Collection 4, the Collection 5 of the coastal zone classes presents no huge methodological shift. The major difference is the higher quantities of training samples used to training the Aquaculture U-NET classifier. This larger training dataset allowed the convolutional neural network (CNN) to better identify the aquaculture sites throughout the Brazilian Coastal Zone (BCZ). The area of aquaculture reported in Collection 5.0 is three times higher than it is in Collection 4.0. The whole classification process is described below, Figure 1.
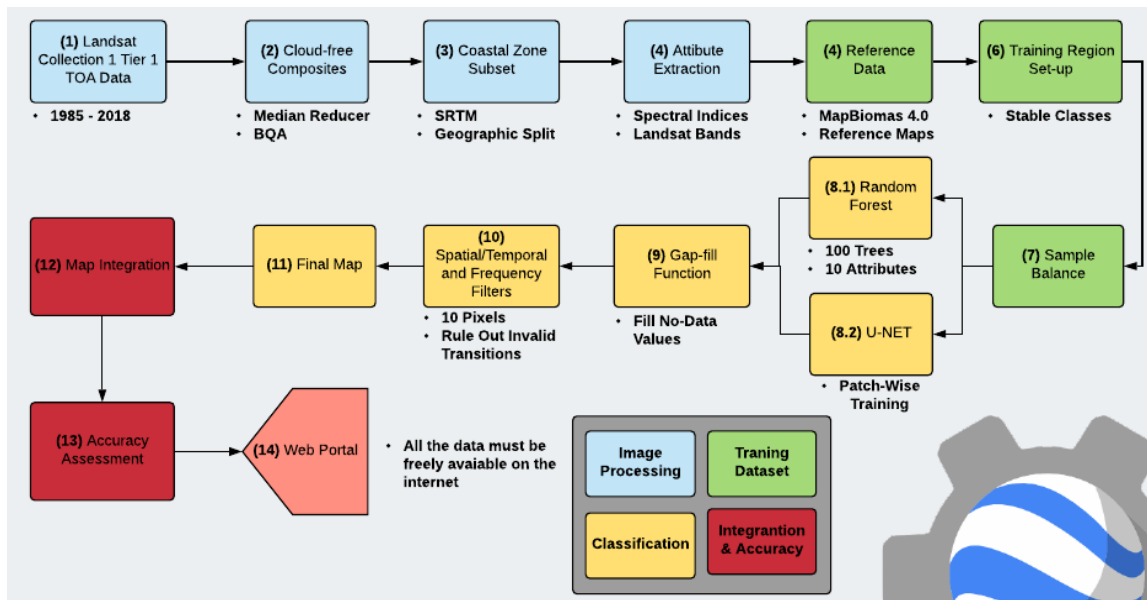
2   Landsat image mosaics

The classification of the cross-cutting themes "**Coastal Zone and Mining**" uses Landsat mosaics that differs from the mosaics used to classify the natural vegetation of the Brazilian biomes. The Coastal mosaics were defined to preserve the maximum of the coastal zone land area while capturing the smallest possible cloud cover. These Landsat mosaics are the third generation of the methodology developed specifically for these cross-cutting themes.

2.1   Definition of the temporal period

Coastal areas are severely affected by atmospheric nebulosity, a condition that is intensified by its proximity to the oceans and its tropical location. On the other hand, the

attempt to identify a time interval that covers only the driest season of the year, as alternative to reduce cloud persistency, ends-up severely reducing the number of images available to cover the entire coastal region. Thus, the annual cloud free composites are generated, ranging from 1st of January to 31st of December.

2.2 Mosaic Subsets
2.2.1 Coastal Zone

Since the Brazilian coastal zone (BCZ) is an extensive region, approximately 8,500 kilometers from Oiapoque to Chui (without counting reentrances), and affected by a variety of atmospheric systems, of lesser or greater influence of nebulosity's, the ZCB is here divided into 7 different sectors, figure 1.

Sector 1 - Amapa (AP), coastal region of the state of Amapa. Sector 2 - Marajo Island (MAR), coastal region of Marajo Island. Sector 3 - Para / Maranhao (PAMA), a coastal sector of the states of Para and Maranhao. Piaui / Bahia (PIBA), a coastal sector of the states of Piaui to Bahia. Sector 5 - Espirito Santo / Sao Paulo (ESSP), region that includes the states of Espirito and São Paulo. Sector 6 – Parana/Laguna (PRLA), a coastal region that goes from the state of Parana to the municipality of Laguna in Santa Catarina and finally Sector 7 (LARS), a region that ranges from Laguna to the state of Rio Grande do Sul.
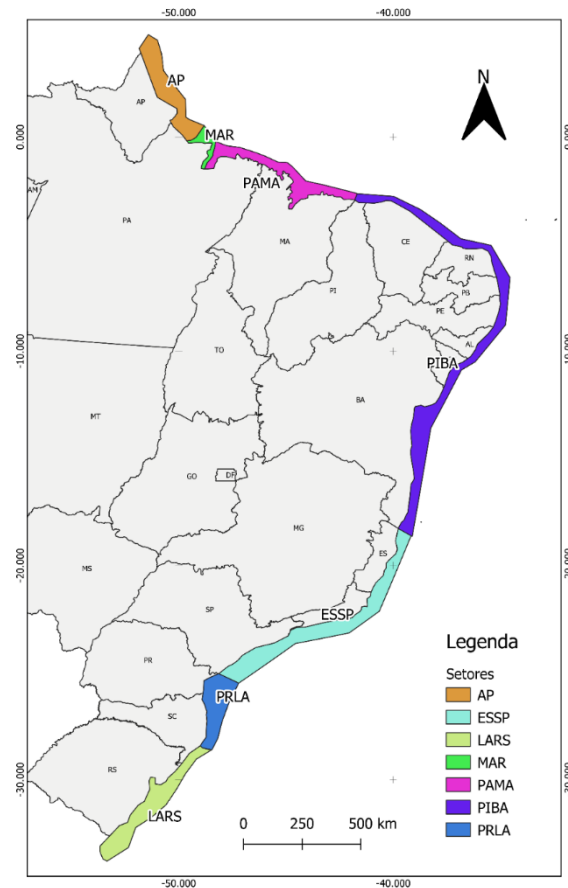
Figure 1 - The seven sectors delimited in different colors, along the Brazilian Coastal Zone (ZCB).

2.3   Image selection

The cloud/shadow removal script takes advantage from the Landsat Collection 1 Level-1 QA band and the GEE median reducer. In Collection 1 Tier 1 data, each pixel in the QA band contains unsigned integers values that represent a certain surface, atmospheric, and sensor conditions that may affect the overall usefulness of a given pixel. When effectively used, QA values can improve the data integrity by indicating which pixels might be affected by instrument artifacts or subject to cloud contamination (USGS, 2017). In conjunction to that, GEE can be instructed to pick the median pixel values in a stack of images. By doing so, GEE rejects values that are too high (e.g. clouds) or too low (e.g. shadows) and picks the median pixel value in each band, over time, Figure 3.

Figure 3 - Left, Collection 2 "cloud-free composite". Right, Collection 3 "cloud-free.

2.4 Final quality

The mosaic quality is related to Landsat cloud-free availability during the image selection period. However, from 1985 to March 1998, only the Landsat 5 satellite remained operational. In this period, for the BCZ, the average number of images per year was ~500. In the last decade between 2008 and 2018, this figure tripled to ~1500 images per year, as shown in Figure 4.
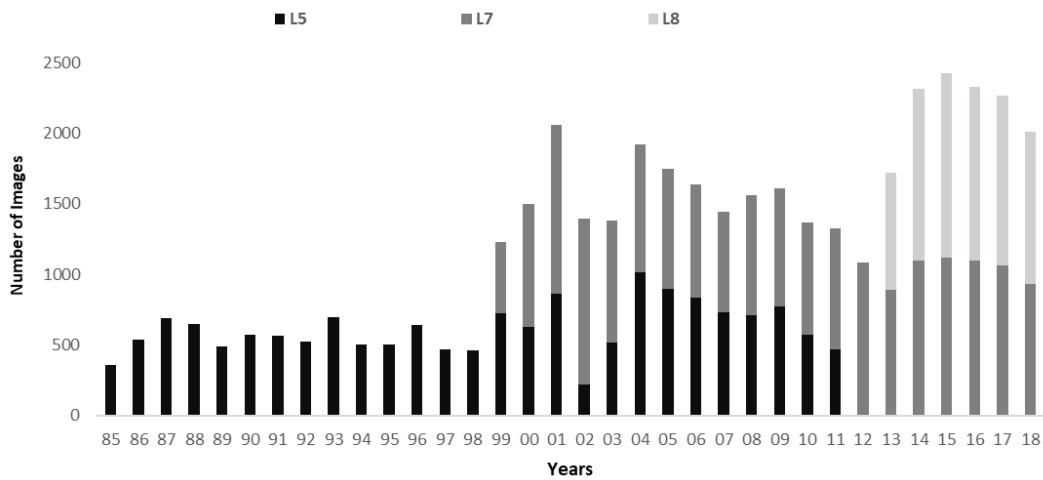

Figure 4 – Landsat image availability from 1985 to 2018. The bars show the distribution of Landsat images along the time series. L5 stands for Landsat 5, L7 refers to Landsat 7, and L8 stands for Landsat 8.

3 Classification

The automatic classification of the Landsat mosaics was completely performed on Google Earth Engine platform, based on the Random Forest classifier (Breiman, 2001).

3.1 Classification scheme

Each interested class were classified separately. In account of this, four independent classification process were performed: 1) Mangrove; 2) Beaches and dunes; 3) Apicum; 4)

Aquaculture and 5) Mining. The classification process was carried out always considering only two possible classes for each pixel, the interest class (mangrove, beaches and dunes, apicum, aquaculture and mining) or non-interest class (all that is different from the interest class).

For the supervised classification of the Landsat mosaics, we have selected training points based on reference maps and MapBiomas 4.0 collection. The details of the parameters used in the Random Forest classifier, the reference maps used for each interest class, and the feature space produced for each classification are presented in the sections to follow.

3.2 Coastal Zone Feature space

The Tables 1 and 2 shows, respectively, all spectral indices and bands used for the BCZ classification.

Table 1 – Spectral Indices used for coastal zone classification.

Index	Expression	Reducer	Reference
EVI2	2.5 * ((NIR - RED) / (NIR + 2.4*RED + 1))	Median and Standard Deviation	Liu & Huete, 1995
NDVI	(NIR - RED) / (NIR + RED)	Median and Standard Deviation	Tucker, 1979
MNDWI	(GREEN - SWIR1) / (GREEN + SWIR1)	Median and Standard Deviation	Xu, 2006
NDSI	(SWIR1 - NIR) / (SWIR1 + NIR)	Median and Standard Deviation	Rogers, 2004
MMDI	Modular Mangrove Detection Index	Median and Standard Deviation	Diniz et. al, under review

Table 2 - Table of bands used to classify Coastal Zone classes.

Variable	Description	Reducer
GREEN	Landsat Green band median value	Median and Standard Deviation
RED	Landsat Red band median value	Median and Standard Deviation
NIR	Landsat NIR band median value	Median and Standard Deviation

SWIR1	Landsat SWIR1 band median value	Median and Standard Deviation
SWIR2	Landsat SWIR2 band median value	Median and Standard Deviation

3.3 Classification algorithm, training samples and parameters

When lacking reference maps that matches the classes and/or year to be classified, reference maps of the closest possible timeframe to the median composites were used. When no reference map was available, then the classifications results of the previous year were used for the training subsequent one. This was done for each one of the years without an external reference training guide. The Table 4, shows the Random Forest parameters used to classify each one of the years.

Table 4 - Random Forest parameters used to classify each one of the years.

Parameter	Value
Number of trees	100
Number of points	100000
Number of Variables	20 (Coastal Zone)
Classes	2 (binary classification)

3.3.1 Mangroves

As in any supervised method, the Random Forest classifier needs to rely on a training dataset. For mangrove cover recognition the training data was obtained from MapBiomas Collection 5. and Dr. Giri Chandra global mangrove cover (Giri *et al.*, 2013), Figure 5. The consolidate results of the mangrove are available on (Diniz *et al.*, 2019)
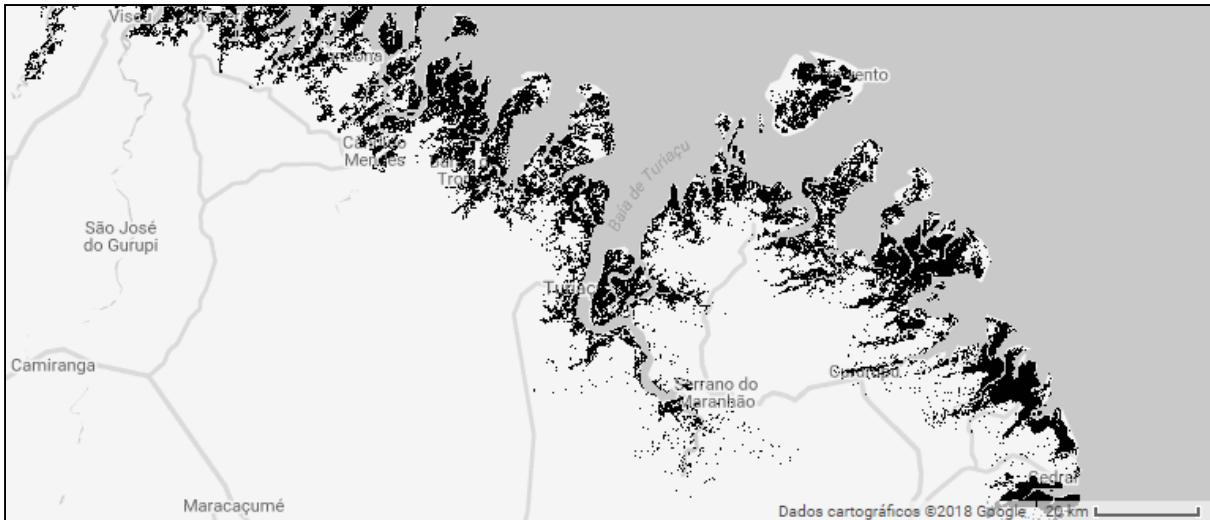
Figure 5 - Global Mangrove Cover data was used as mangrove mapping reference, from 1999 to 2002.

3.3.2 Apicum

In general, the less frequently flooded area of a mangrove swamp, in the transition to topographically elevated lands, is usually devoid of arboreal vegetation. In Brazil, this area is referred to as "*Apicum*". In the international scientific literature, this transition zone is usually called salt flat or hypersaline tidal flat. Three different reference maps were here used, the "Atlas Dos Remanescentes Florestais da Mata Atlântica" from 2016/2017, covering the Mata Atlantica coastal region and the "Carta de Sensibilidade Ambiental ao Oleo -Para-Maranhão-Barreirinhas" referent to 2017 and covering most of the Brazilian north coastal region and the data from the MapBiomas 5 Collection, Figure 6.

Figure 6 – Apicum reference maps, the "Atlas Dos Remanescentes Florestais da Mata Atlântica" from 2016/2017, covering the Mata Atlantica coastal region and the "Carta de Sensibilidade Ambiental ao Oleo -Para-Maranhão-Barreirinhas 2017", covering most of the Brazilian north coast region

3.3.3 Beaches and Dunes

Mapped without distinction between one another, in here the "Beaches and Dunes" class refers to sandy strands, bright white in color, where there is no predominance of

vegetation of any kind. The training data for this landcover was obtained from MapBiomas Collection 4.1, ranging from 2000 to 2016 and from the "Atlas Dos Remanescentes Florestais da Mata Atlântica" (SOSMA e INPE, 2018) from 2016/2017, Figure 7.
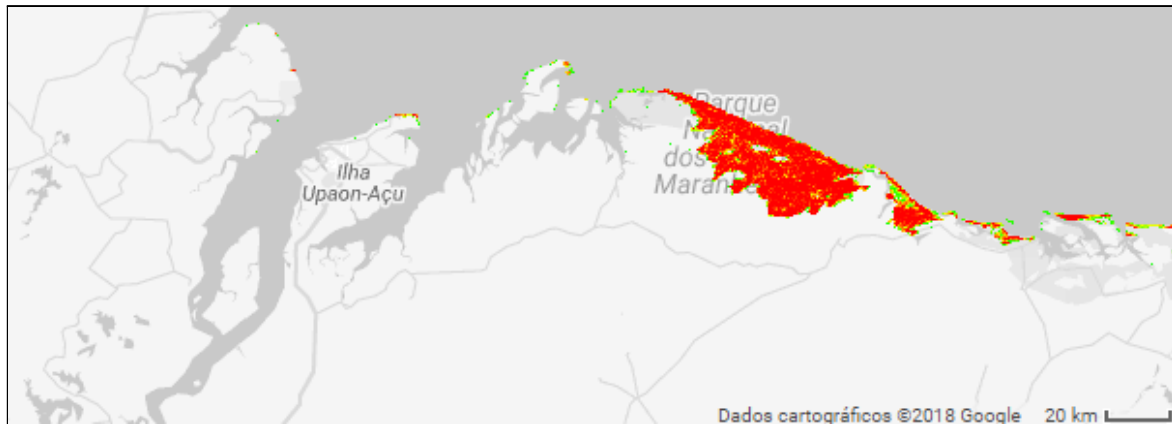

Figure 7 - The training data obtained from MapBiomas Collection 2.0, ranging from 2000 to 2016 and from the "Atlas Dos Remanescentes Florestais da Mata Atlântica 2016/2017".

3.3.4   Aquaculture/Salt culture

In comparison previous Collections, the Collection 6 aquaculture mapping consolidated the use of Deep-Learning model in replacement of the traditional Random Forest Algorithm. Its results are now published on (Diniz *et al.*, 2021). In this scenario, traditional machine learning algorithms make use spectral-temporal data to classify targets according to similarities of their spectral-temporal patterns (Breiman, 2001). Although temporal and spectral properties might be not enough to discriminate "super-similar" targets. Targets that behave similarly in both spectral and temporal domains. That is the case of most surface water targets, as aquaculture ponds, rivers, lakes and open-waters, Figure 8.

Spectrally speaking, water is water and, unless it presents a high concentration of external compounds (minerals, suspended sediments, algae, etc.), not much can be done to spectrally differentiate between numerous of surface water targets. On the other hand, the temporal domain may not present much useful discriminatory data either. In Brazil, aquaculture is a traditional and coastal-related economic activity. Thus, in 35 years of data (Barbier e Cox, 2003; Guimarães *et al.*, 2010; Queiroz *et al.*, 2013; Tenório *et al.*, 2015; Thomas *et al.*, 2017), a diverse set aquaculture frequency may exist. As a result of that, the temporal domain renounces to distinguish between well-consolidated aquaculture, main river channels and open waters, once all these features present high temporal persistence throughout the entire time-series.
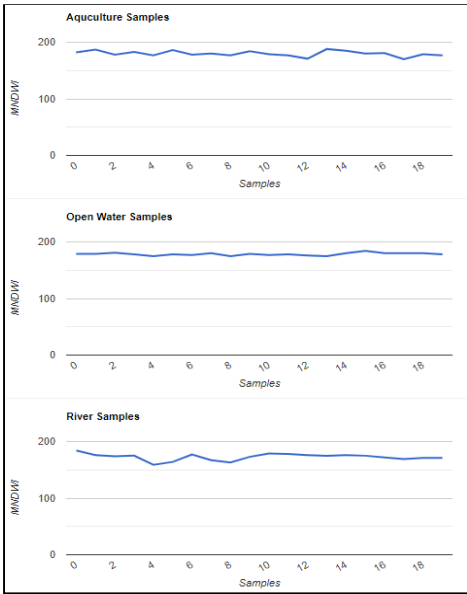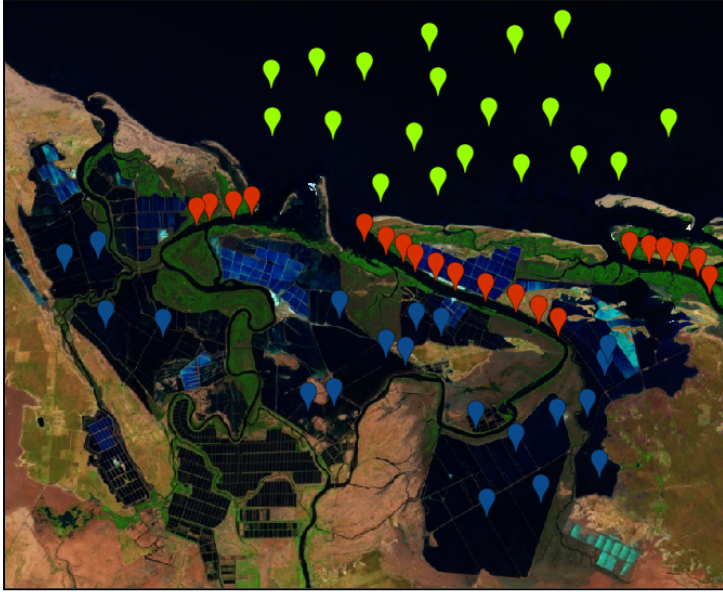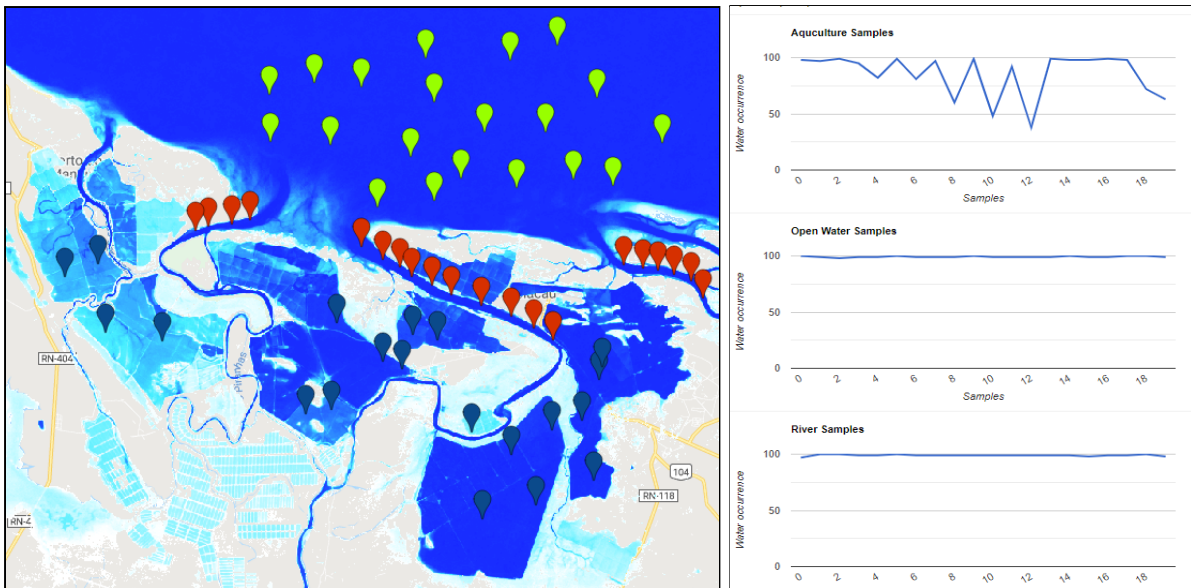
Figure 8 – Spectral and temporal patterns of the aquaculture, rivers and open waters classes. In the top-left corner, the median cloud-free composite from Macau-RN, northeast of Brazil. The markers in dark-blue, green and red, represent aquaculture, open water and river samples. In the top-right, NMDWI values for each one of the samples. In the bottom-left, JRC occurrence data. In the bottom-right, occurrence frequency to each one of the samples.

In cases like this, the "context domain", may hold the key elements to allow the distinction between rivers, aquaculture and open waters pixels. In the context analysis scenario, the U-Net: Convolutional Networks (Abadi *et al.*, 2015) has the advantage of predicting the class label of each pixel, by providing as input a local region (patches or chips) around that pixel. Such characteristic of working with "patches" or "chips", is what gives to the U-Net the ability to access the "context domain" of the image, instead of using isolated pixel. The U-Net initial training was guided by Collection 5 and 4.1 aquaculture data.

4   Post-classification

Due to the pixel-based nature of the classification method and the very long temporal series, a chain of post-classification filters was applied. The post-classification process includes the application of a gap-fill, a temporal, a spatial and a frequency filter.

4.1   Gap-Fill filter

The chain starts by filling in possible no-data values. In a long time-series of severely cloud-affected regions, such as tropical coastal zones, it is expected that no-data values may populate some of the resultant median composite pixels. In this filter, no-data values ("gaps") are theoretically not allowed and are replaced by the temporally nearest valid classification. In this procedure, if no "future" valid position is available, then the no-data value is replaced by its previous valid class. Up to three prior years can be used to fill in

persistent no-data positions. Therefore, gaps should only exist if a given pixel has been permanently classified as no-data throughout the entire temporal domain. To keep track of pixel temporal origins, a mask of years was built, as shown in Figure 9.
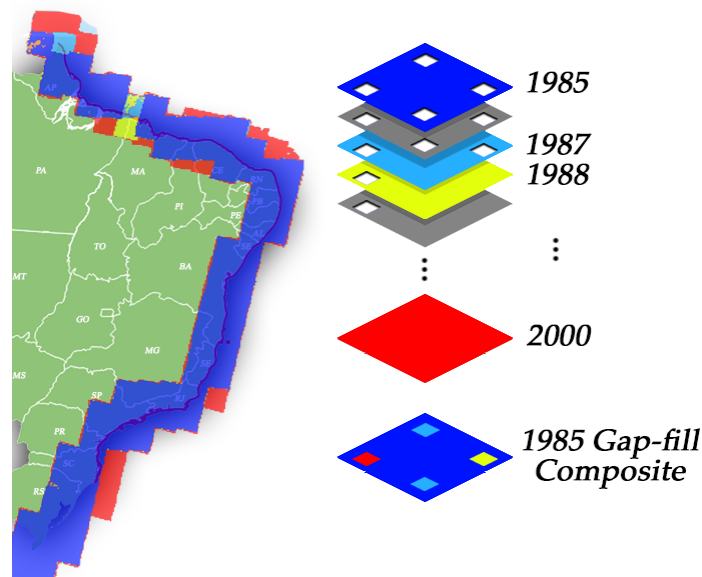

Figure 9 – Gap-filling mechanism. The next valid classification replaces existing no-data values. If no "future" valid position is available, then the no-data value is replaced by its previous valid classification, based on up to a maximum of three (3) prior years. To keep track of pixel temporal origins, a mask of years was built.

4.2 Temporal filter

After gap filling, a temporal filter was executed. The temporal filter uses sequential classifications in a 3-year unidirectional moving window to identify temporally non-permitted transitions. Based on a single generic rule (GR), the temporal filter inspects the central position of three consecutive years ("ternary"), and if the extremities of the ternary are identical but the centre position is not, then the central pixel is reclassified to match its temporal neighbour class, as shown in Table 6.

Table 6 - The temporal filter inspects the central position of three consecutive years, and in cases of identical extremities, the centre position is reclassified to match its neighbour. T1, T2 and T3 stand for positions one (1), two (2) and three (3), respectively. GR means "generic rule", while Mg and N-Mg represent mangrove and non-mangrove pixels.

Rule	Input (Year)			Output		
	T1	T2	T3	T1	T2	T3
GR	Mg	N-Mg	Mg	Mg	Mg	Mg
GR	N-Mg	Mg	N-Mg	N-Mg	N-Mg	N-Mg

4.3 Spatial filter

Next, a spatial filter was applied. To avoid unwanted modifications to the edges of the pixel groups (blobs), a spatial filter was built based on the "connectedPixelCount" function. Native to the GEE platform, this function locates connected components (neighbours) that share the same pixel value. Thus, only pixels that do not share connections to a predefined number of identical neighbours are considered isolated, as shown in Figure 10. In this filter, at least ten connected pixels are needed to reach the minimum connection value. Consequently, the minimum mapping unit is directly affected by the spatial filter applied, and it was defined as 10 pixels (~1 ha).
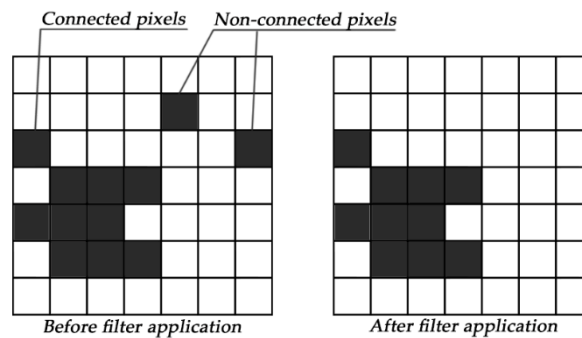

Figure 10 – The spatial filter removes pixels that do not share neighbours of identical value. The minimum connection value was 10 pixels.

4.4 Frequency filter

The last step of the filter chain is the frequency filter, as shown in Figure 11. This filter takes into consideration the occurrence frequency of given class throughout the entire time series. Thus, all class occurrences with less than 10% temporal persistence (3 years or fewer out of 33) are filtered out and incorporated into the non-class binary. This mechanism contributes to reducing the temporal oscillation of the classification signal, decreasing the number of false positives and preserving consolidated class pixels.
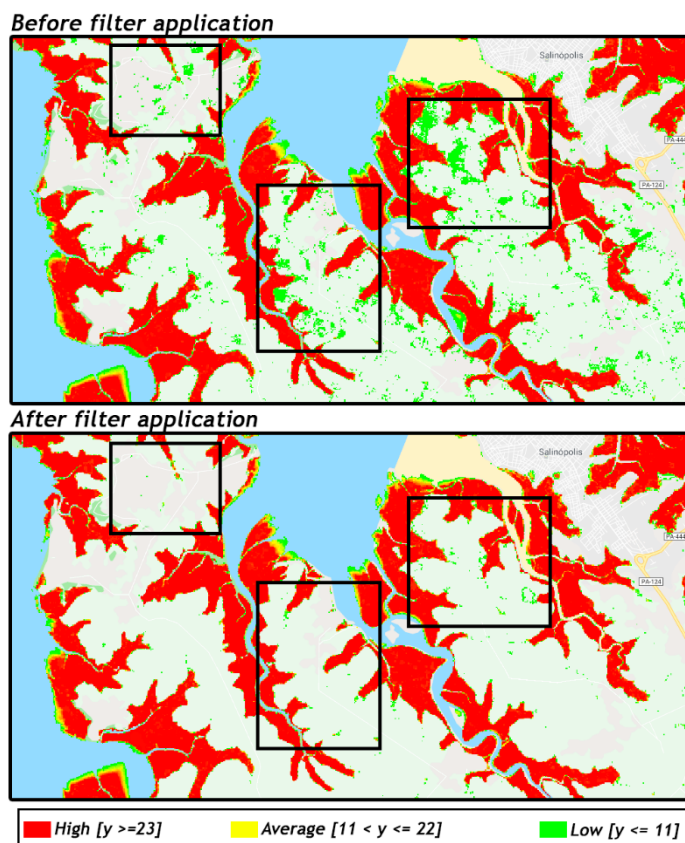
Figure 11 – Red, yellow and green represent mangrove pixels with high (23 or more years, y >=23), average (between 11 and 22 years, 11 <= y <= 22), and low (ten years or less, y < 11) occurrence frequencies, respectively. The top image shows mangrove pixels before applying the frequency filter. The bottom image shows mangrove pixels after applying the frequency filter. The black boxes are centred on areas that have been significantly affected by the filter. Note that all mangrove occurrences with less than 10% temporal persistence (3 years in 33 possible years) were filtered out.

4.5   Integration with biomes themes

After the application of the filter-chain, the cross-cutting themes and the Biomas data are integrated. This integration is guided by a set of specific hierarchical prevalence rules (Table 5). As output of this step, a final vegetation cover/land use map for each chart of the MapBiomas projet.

Table 5 - Prevalence rules for combining the output of digital classification with the crosscutting themes in Collection 4.

Collection 4	Prevalence Rule
4.1. Beaches and Dunes	1
1.1.3. Mangrove	2
5.2. Aquaculture	3
2.3. Apicum	4
5. Water	5
5.1. River, Lake and Ocean	5
1.2. Forest Plantations	6
4.4. Mining	7

References

ABADI, M. *et al.* TensorFlow: Large-scale machine learning on heterogeneous systems. **Methods in Enzymology**, 2015.

BARBIER, E. B.; COX, M. Does Economic Development Lead to Mangrove Loss? A Cross-Country Analysis. **Contemporary Economic Policy**, v. 21, n. 4, p. 418–432, 1 out. 2003.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

DINIZ, C. *et al.* Brazilian Mangrove Status: Three Decades of Satellite Data Analysis. . 2019.

___. **A Large-Scale Deep-Learning Approach for Multi-Temporal Aqua and Salt-Culture MappingRemote Sensing** , 2021.

DOMINGUEZ, J. M. L. The Coastal Zone of Brazil. *In*: **Geology and Geomorphology of Holocene Coastal Barriers of Brazil**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. p. 17–51.

GIRI, C. *et al.* **Global Mangrove Forests Distribution, 2000**Palisades, NY, NYNASA Socioeconomic Data and Applications Center (SEDAC), , 2013. Disponível em: <http://dx.doi.org/10.7927/H4J67DW8>

GUIMARÃES, A. S. *et al.* Impact of aquaculture on mangrove areas in the northern Pernambuco Coast (Brazil) using remote sensing and geographic information system. **Aquaculture Research**, v. 41, n. 6, p. 828–838, 13 maio 2010.

QUEIROZ, L. *et al.* **Shrimp aquaculture in the federal state of Ceará, 1970–2012: Trends after mangrove forest privatization in Brazil**. [s.l: s.n.]. v. 73

SOSMA; INPE. ATLAS DOS REMANESCENTES FLORESTAIS DA MATA ATLÂNTICA PERÍODO 2016-2017. **Fundação SOS Mata Atlântica e Instituto de Pesquisas Espaciais**, 2018.

TENÓRIO, G. S. *et al.* Mangrove shrimp farm mapping and productivity on the Brazilian Amazon coast: Environmental and economic reasons for coastal conservation. **Ocean & Coastal Management**, v. 104, p. 65–77, 2015.

THOMAS, N. *et al.* Distribution and drivers of global mangrove forest change, 1996–2010. **PLOS ONE**, v. 12, n. 6, p. e0179302, 8 jun. 2017.

USGS. **LANDSAT COLLECTION 1 LEVEL 1 PRODUCT DEFINITION**. [s.l.] Earth Resources Observation and Science (EROS) Center, 2017.