# Assessing Land Use and Land Cover Maps and Legends between MapBiomas and Brazil's Fourth Emission Inventory

Sabrina G. Marques<sup>1</sup>, Pedro R. Andrade<sup>1</sup>, Aline C. Soterroni<sup>2</sup>

<sup>1</sup>National Institute for Space Research (INPE) 122.279-00 – São José dos Campos – SP – Brazil

<sup>2</sup>Nature-based Solutions Initiative, Department of Biology – University of Oxford Oxford, UK

Abstract. Comparing LULC maps is essential for understanding landscape dynamics, alteration patterns, and environmental implications. This study uses an algorithm to harmonize the maps of Brazil National Inventory and MapBiomas based on the spatial distribution of LULC classes. This investigation aims to compute the agreement between two initiatives while examining the uncertainties of both. Furthermore, the results highlight the classes and areas of potential inconsistency or ambiguity, allowing to identify and correct discrepancies, proposing a harmonized legend between then. For all Brazil, we achieved a maximum concordance of 81% between the two maps; out of the 44 equivalences, the algorithm correctly identified 84% of the mappings between the classes.

#### 1. Introduction

The Earth, comprised of a complex network of ecosystems, has been a subject of study and engagement since the beginning of human civilization. The relationship between humans and their environment has significantly shaped cultural, social, and economic practices. However, in the last decades, there has been an observed reversal in this relationship. With the expansion of civilization and the advancement of technology, humanity has transitioned from being mere inhabitants to a dominant force that actively changes and modifies the environment to meet its needs [Verburg et al. 2013, Pielke Sr. et al. 2011, Ellis et al. 2013]. In the context of climate change, the Agriculture, Forestry, and Other Land Use sector emerges as a critical component. According to the 2023 IPCC report [IPCC 2023], this sector is responsible for approximately 22% of human-made greenhouse gas (GHG) emissions. Therefore, precise monitoring through Land use and land cover (LULC) maps is necessary to compile inventories of GHG emissions and removals [Shukla et al. 2019].

LULC maps represent the physical space of a chosen region through abstractions that describe the covered areas. They allow a systematic categorization of geographical regions based on specific human uses and natural characteristics. These categorizations represent the spatial distribution of human activities, serving as indicators of human-made pressures on natural ecosystems [Jansen et al. 2008]. In addition, the analytical and symbolic capabilities of LULC maps are indispensable tools in the scientific field. They not only document the current state of the environment but also, when employed for comparisons, provide a perspective for examining human-induced changes over time and their

ecological and climatic consequences. As a result, they play a critical role in forming evidence-based decision-making regarding the management and conservation of natural resources [Verburg et al. 2013].

Comparing LULC maps is a valuable resource in environmental and geographical studies. Sequentially overlaying these maps reveals environmental changes and transformations trends, providing information about deforestation rates, urban expansion, changes in water bodies, and other critical aspects. This comparative analysis is essential for evaluating the impacts of land-use policies and projecting future scenarios [Ellis et al. 2013].

In Brazil, several initiatives use open data to produce LULC maps, such as MapBiomas [MapBiomas Brasil 2021], TerraClass [INPE 2019], PRODES [INPE 2021], IBGE [IBGE 2019], and the National Communications to the United Nations Framework Convention on Climate Change (UNFCCC) [Brasil 2021]. Although each of these initiatives has different objectives, interests, and mapping standards, there are differences in the maps produced for the same area, some of which might be related to the nature of the input data or the developed methodology. This limits the compatibility and comparability of these data. Different maps might have been produced at different intervals and aggregating this information can allow for more granular time-series analyses.

Harmonization of these LULC maps is challenging due to the different methods, classification systems, and legends adopted by each project. These differences may stem from the choice of satellite imagery, classification methods, field support data, and more. Besides technical discrepancies, there are practical challenges, like differences in resolution, projection, and coordinate systems. In addition, harmonizing legends presents excellent challenges due to their nature. Differences in class naming, changes in class definitions, and the addition or deletion of classes in maps covering the same region at different times or in different initiatives create difficulties to separate actual changes over time from differences in category definitions. Thus, establishing equivalencies between classes from different maps is vital for effective comparisons.

Typically, comparing LULC maps involves constructing a key based on the semantics of each category. Frequently, categories are grouped into broader classifications to minimize discrepancies or are excluded by lacking similarity explanations. Some classification systems can also standardize keys and render maps comparable. These types of methods can be observed in the works of [Capanema et al. 2019], [Reis; et al. 2017], [Reis et al. 2018], and [Neves et al. 2020].

While traditional methods primarily start from the semantics of LULC classes, examining the spatial distribution of categories can yield additional insights. This study uses the algorithm presented in [Marques et al. 2022] to compare the LULC maps of Brazil's Fourth National Inventory and MapBiomas. This algorithm computes the highest agreement between two classifications while examining the uncertainties. We perform an analysis at the biome level and on a national scale. A mapping between their categories was created using category descriptions and the mapping derived from maximum agreement.

## 2. Methodology

In this section, we present the two maps that are subject to this study and then we describe the method to assess them.

## 2.1. MapBiomas

The Annual Land Use and Land Cover Mapping Project in Brazil, known as Map-Biomas, was created by the Greenhouse Gas Emissions Estimate System initiative of the Climate Observatory (SEEG/OC). The MapBiomas methodology consists of a pixel-by-pixel classification of Landsat satellite images, with 30 m of spatial resolution that provides LULC maps from 1985 to 2020 [MapBiomas Brasil 2022, Souza et al. 2020, MapBiomas Brasil 2021]. Figure 1 presents an overview of the data from this collection.

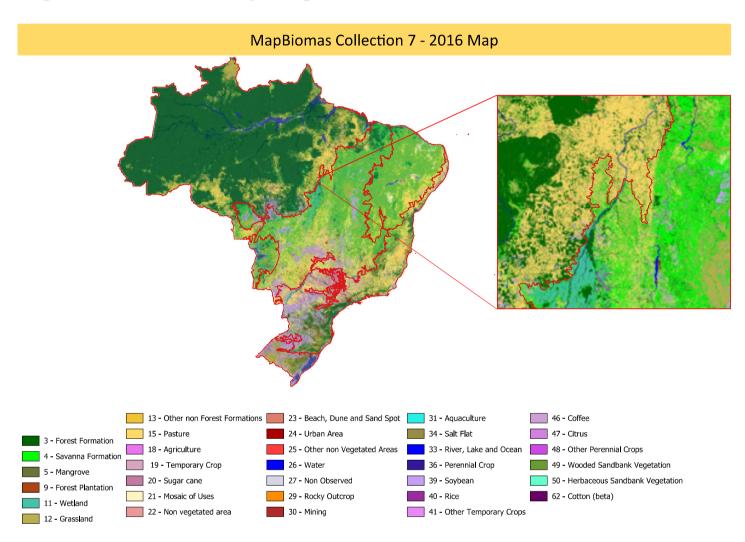


Figure 1. Map of Land Use and Land Cover of MapBiomas Collection 7.1 for the year 2016.

#### 2.2. Brazilian National Inventory

The Brazilian National Inventory, henceforth called *Inventory*, mission is part of Brazil's National Communication to the United Nations Framework Convention on Climate Change (UNFCCC). The National Communication provides anthropogenic emissions of GHGs no longer managed via the Montreal Protocol. The Ministry of Science, Technology and Innovations (MCTI) coordinates and improves the inventory. Emission estimates are primarily based on the LULC map developed by the National Inventory. This mapping uses images of the TM/OLI sensors of the Landsat-5/8 satellite and the MSI/Sentinel 2A and 2B sensor at a scale of 1:250,000, with a minimal region of 6 ha [MCTI 2021, Brasil 2021, MCTI 2020]. Figure 2 presents an overview of the produced map.

The LULC maps are vector representations, overlaying the years 1994, 2002, 2005 (only for the Amazon biome), 2010, and 2016, and are divided via means of biomes

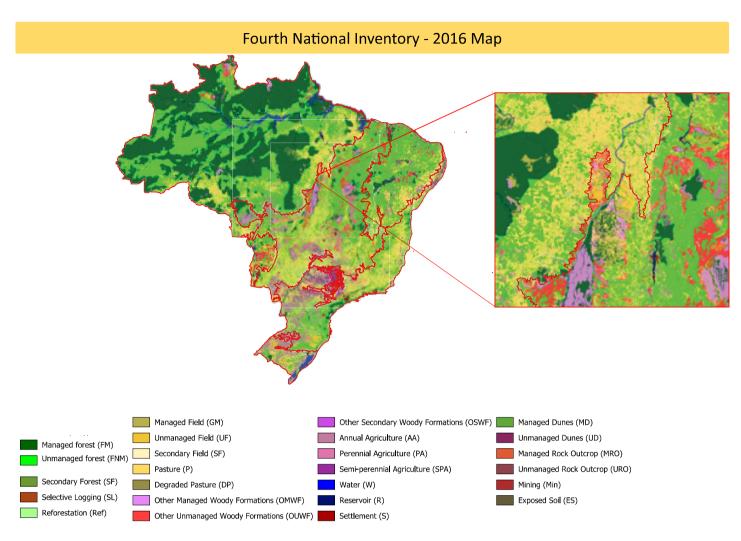


Figure 2. Map of Land Use and Land Cover of the Fourth National Inventory for 2016.

following the limits set by IBGE in 2004 [MCTI 2020]. LULC maps are available from National Emissions Registry System (SIRENE).

# 2.3. Assessment methodology

We use the legend harmonization algorithm presented by [Marques et al. 2022] to compare the two maps. This algorithm matches the legends using the maps themselves. The algorithm aims to be the first step in the harmonization process of LULC map legends, providing a proposed harmonized legend based on the spatial distribution of the classes in the maps, which delivers the highest possible accuracy between them. The algorithm has three steps. Initially, it generates a cross-tabulation matrix between the two maps using the pixel count of each class. Using this matrix, the algorithm calculates the concordances of the classes from one map to another using the maximum values of each row and each column of the matrix, creating two sets of equivalences between the maps. The union of these sets creates the harmonized legend between the maps, containing all the concordances obtained by the row and column harmonizations.

Using this procedure, given two maps, Map 1 and Map 2, the algorithm determines which classes from Map 2 are spatially equivalent to Map 1 and then repeats the process for the classes of Map 2. The grouping of these two sets of concordant classes forms the harmonized legend, which encompasses three possible cases of equivalence between the classes: (1) when there is a mapping from one class to another, both by row and column; (2) when a class is only mapped in one of these harmonization sets; and (3) when the

mapping of a class differs in the row and column harmonization. For more information about how the algorithm works, see [Marques et al. 2022].

The algorithm can capture subtle nuances in class definitions between distinct maps, reflecting unidirectional and bidirectional correspondences. Furthermore, it highlights potential inconsistencies or ambiguities, allowing users to identify and fix them.

In practical terms, the automation provided by the algorithm facilitates the integration of data from different sources, optimizing the efficiency of the process and minimizing errors that can arise from manual approaches. It is an initial step for mapping classes between maps, and it's up to the user to check if the obtained mappings are coherent or if the legend needs to be adapted. It's worth noting that since the legend produced by the algorithm provides the combination with the highest concordance between the maps, any changes will result in a lower concordance.

We compare both maps by biomes and the whole country. As the most updated map for the Inventory is for year 2016, we use it to compare with MapBiomas using the same year.

### 3. Results

Table 1 displays the maximum concordances achieved in each biome<sup>1</sup> This value is obtained if the harmonized legend produced by the algorithm was applied to both maps, considering the lowest hierarchy level of the classes. Figure 3 shows the harmonizations between the Fourth National Inventory and MapBiomas, as generated by the algorithm for the entire country.

Table 1. Maximum concordance obtained in each of the harmonizations and the area of each applied region.

	Area (km <sup>2</sup> )	Maximum
		Agreement
Amazon	4.253.027	92.39%
Caatinga	843.615	75.27%
Cerrado	1.983.655	74.33%
Atlantic	1.116.119	77.86%
Forest		
Pampa	203.965	79.32%
Pantanal	150.972	55.51%
Brazil	8.604.500	81.03%

The Amazon biome has the largest area among all the listed biomes, totaling 4,253,027 km², with the highest concordance of 92.39%. Much of this is due to the vast expanse of classes defined as forest, which favors the overlap between them and their correct identification. All forest classes of the Inventory (Managed Forest/I², Unmanaged Forest/I, Secondary Forest/I, and Selective Logging/I) were mapped to Forest

<sup>&</sup>lt;sup>1</sup>The charts and other harmonizations for the biomes can be viewed in detail on the project's GitHub page.

<sup>&</sup>lt;sup>2</sup>We use /I for classes of Inventory and /M for MapBiomas.

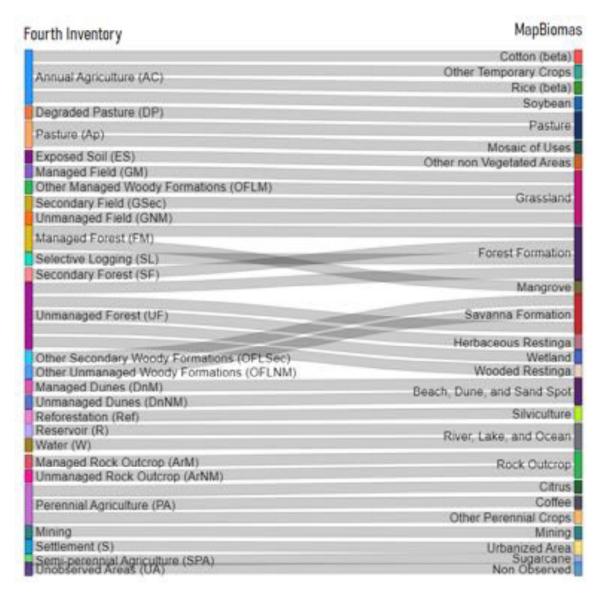


Figure 3. Harmonized subtitle produced by the algorithm for all of Brazil.

Formation/M. More granular classes, such as Beach, Dune, and Sand Spot/M, Other non-Vegetated Areas/M, Rice (beta)/M, and Unmanaged Dunes/I were incorrectly matched as Forest. In contrast, Perennial Agriculture/I was identified as Pasture/M in the harmonization. The small area of these classes in the Amazon biome leads to a low impact on the overall harmonization. Still, it raises points of attention, especially considering classes related to pasture and agriculture being identified as forest since, on a small scale, they can have implications for conservation policies or zoning.

In the Cerrado biome, the harmonization produced a concordance of 74.33%. However, it is important to highlight that the Managed Forest/I class was associated with the Aquaculture/M. Additionally, most other classes were predominantly grouped under the MapBiomas Savanna Formation, with 33% of the entire concordance area being labeled as this class, including the Secondary Field/I class that was incorrectly associated. Also, in this biome, another 32% of the concordance area was labeled as Pasture/I, with 11% of this total mapping the Mosaic of Uses/M class as Pasture/I.

For the Caatinga biome, the produced legend achieved a concordance of 75.27%. In this biome, some mappings stood out between the maps: the Mosaic of Uses/M class was incorrectly mapped as Unmanaged Forest/I, just as the classes of Unmanaged Field/I and Secondary Field/I were also incorrectly mapped as Savanna Formation/M. The classes for water and agriculture were mostly correctly mapped. From this, it can be inferred that

in this biome, the classes for forests and fields showed a lot of confusion between the maps, which might indicate that the semantic definitions of these classes may be very similar between the initiatives, especially when considering that this biome is characterized by shorter vegetation.

In the Atlantic Forest biome, where a concordance of 77.86% was observed, there was a trend to group various classes from the National Inventory into the Forest Formation category of MapBiomas. In this biome, the harmonized legend has 41 combinations of classes, due to the diversity of the biome and most of the classes do not have the same harmonization by row and column. It is worth noting that classes such as Managed Field/I and Secondary Field/I were labeled as Forest Formation/M, along with the Managed Dunes/I class. The Herbaceus Restinga from MapBiomas was identified as Pasture from the National Inventory.

The Pampa biome presents a 79.32% concordance between the two maps. Most of the classes from the National Inventory were labeled as Grassland, which might suggest that MapBiomas overestimates the field classes in this region, given that 15% of the entire biome was labeled by the pair Pasture/I and Grassland/M. This also happened with Unmanaged Forest/I, where 8% of the total area was labeled as Grassland/M.

The Pantanal showed the lowest concordance among all biomes, registering only 55.51%. This discrepancy may be attributed to the unique spatial distribution of classes in this biome. The predominance of certain classes in distinct areas might have influenced a lower concordance between the initiatives. It is noteworthy to mention that the classes related to Forest were correctly mapped, with the exception of the Secondary Forest/I class, which was identified as Grassland/M. Another highlight was the classes Other Unmanaged Woody Formations/I and Wetland/M identified as equivalents, representing 9% of the entire equivalence area of the biome.

When analyzing the harmonization obtained for Brazil, which presents a maximum agreement of 81%, there are some interesting trends and characteristics. The Amazonia biome was the only one that showed a concordance higher than Brazil's by 11%, mainly due to the size and homogeneity of the forest classes. It is evident that, on a national scale, extensive forested and agricultural areas exhibit relatively strong correspondence between the two maps. This alignment is a positive indicator for macroecological assessments and large-scale policy considerations. On the other hand, this general accuracy should not overshadow biome all particularities and the idiosyncrasies of data in more specific areas.

In the harmonizations of the Amazon, Cerrado, Atlantic Forest, Pantanal, and throughout Brazil, the Mosaic of Uses/M class was identified as Pasture/I, indicating that most of this class overlaps with the National Inventory's pasture class and could be attributed to this class in the final harmonization for the sake of accuracy. Meanwhile, for Brazil, Caatinga, Atlantic Forest, and Pampa, the classes Unmanaged Rock Outcrop/I and Forest Formation/M were associated, raising an alert given their semantic differences. Similarly, this also occurs between Unmanaged Forest/I and Rock Outcrop/M. The classes Managed Forest/I and Cotton (beta)/M were incorrectly associated in three of the harmonizations. This might occur due to the small area that encompasses the Cotton (beta)/M class, being more subject to erroneous overlaps. This can also occur with more emphasis

on transition areas between biomes, which is more difficult to classify accurately due to a more significant variability in native vegetation.

Certain relations become evident When examining all the obtained harmonizations. The Annual Agriculture/I and Soybean/M classes were correctly identified in all seven harmonizations, indicating a good match between the two maps regarding annual agricultural areas dedicated to soy. Similarly, the class Pasture in both maps was correctly associated in all cases. For Reforestation/I and Silviculture/M, both maps have a good match for reforestation or silviculture areas, correctly identifying them in all regions. The National Inventory's Reservoir and Water classes were also attributed in all harmonizations to the River, Lake and Ocean/M class. This also occurred between Settlement/I and Urbanized Areas/M, as well as Unmanaged Forest/I and Forest Formation/M.

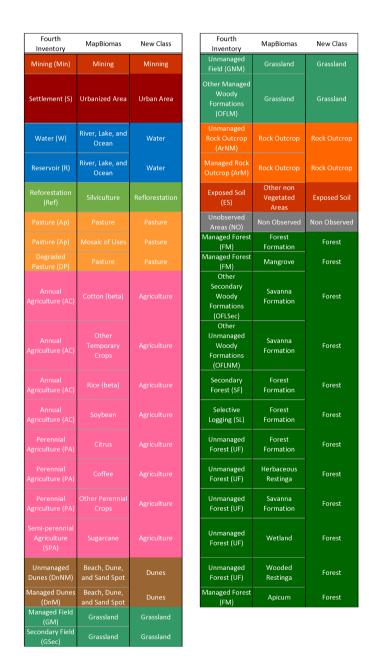


Figure 4. Harmonized legend built from the algorithm legend.

In Figure 4, we have the harmonized legend and a semantical analysis of classes between the maps from MapBiomas and the National Inventory based on the harmonization algorithm. The harmonization generated by the algorithm and the official harmonization

nization are largely aligned for most classes. However, some areas of divergence exist, particularly in the nuances of forest formations and pastures. This is mainly due to the characteristics of the classes assigned to each biome, as both initiatives define the classes of natural vegetation, especially forest and field classes, according to the characteristics of each biome. This causes discrepancies between the classes and leads to confusion between pasture and field classes, given their height and similar characteristics in some biomes. The same applies to some forest classes, which, in biomes characterized by shorter and less dense vegetation, the different classifications used by the initiatives lead to some confusion between these forests and fields, as well as between field and pasture classes.

#### 4. Conclusion

The legend harmonization algorithm provides a first automated step for the class mapping process, a frequent challenge in LULC studies. One of the main strengths of this method is its comprehensive approach, ensuring a clear equivalence for every class in every map. This approach has to be complemented by a double check, where classes are compared in rows and columns, reinforcing the accuracy of the process.

The integrity and precision of LULC maps are essential for understanding landscape dynamics, land alteration patterns, and their environmental implications. By comparing and harmonizing LULC maps from different initiatives, this study emphasized the importance of robust and comprehensive approaches, such as the presented legend harmonization algorithm.

It is important to emphasize that for the algorithm to perform well, both classifications should accurately represent reality. Otherwise, when most of the obtained maps are incorrect, the entire mapping between classes will need to be done manually based on the semantics of the classes.

The harmonization between the maps of both initiatives showed a good concordance rate with some reservations, especially when considering the Pantanal biome. It was possible to observe excellent mappings in significant classes such as forests and reforestation, urban areas, pastures, and water. When analyzing the harmonization for all of Brazil, it is possible to notice that the main class confusions that occurred in each biome diminish when aggregating all areas, in addition to reinforcing the classes that were similarly mapped in all biomes.

In biomes with a predominance of low vegetation, it was noticeable that there was an increased confusion among the field, pasture, and forest classes between the maps, especially in Pampa and Caatinga. Therefore, greater attention is needed in these cases when adapting to a coherent harmonization between the maps. The proposed legend, obtained from the algorithm's results, addresses the discrepancies between the classes identified during the initial agreement and may aid future studies.

In practical terms, the automation provided by the algorithm facilitates the integration of data from different sources, optimizing the efficiency of the process and minimizing errors that can arise from manual approaches. This optimization saves time and improves data interpretability, establishing a common standard that benefits researchers, decision-makers, and other stakeholders.

It is possible to assess changes over time and the influence of land use policies and practices by highlighting the similarities and differences. Moreover, this comparison becomes even more relevant in the absence of inventories in subsequent years. It allows for extrapolation of trends and analysis of carbon emissions by biome, ultimately providing insights for the future.

## References

- Brasil (2021). Quarta Comunicação Nacional do Brasil à Convenção Quadro das Nações Unidas sobre Mudança do Clima.
- Capanema, V. et al. (2019). Comparação entre os produtos temáticos de uso e cobertura da terra do terraclass amazônia e mapbiomas: Teste de aderência entre classes. *Anais do XIX Simposio Brasileiro de Sensoramento Remoto*, pages 724–727.
- Ellis, E. C., Kaplan, J. O., Fuller, D. Q., Vavrus, S., Goldewijk, K. K., and Verburg, P. H. (2013). Used planet: A global history. *Proceedings of the National Academy of Sciences*, 110(20):7978–7985.
- IBGE (2019). Cobertura e uso da terra. Acesso em: 24 abr. 2022.
- INPE (2019). Terraclass. Acesso em: 24 abr. 2023.
- INPE (2021). Monitoramento do desmatamento da floresta amazônica brasileira por satélite. Acesso em: 24 abr. 2023.
- IPCC (2023). Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland.
- Jansen, L. J., Groom, G., and Carrai, G. (2008). Land-cover harmonisation and semantic similarity: some methodological issues. *Journal of Land Use Science*, 3(2-3):131–160.
- MapBiomas Brasil (2021). Mapbiomas brasil. Acesso em: 24 abr. 2022.
- MapBiomas Brasil (2022). *MapBiomas General "Handbook" Algorithm Theoretical Basis Document (ATBD) Collection 6.* 1 edition.
- Marques, S. G., Andrade, P. R., and Soterroni, A. C. (2022). Um algoritmo de máxima concordância para harmonizar legendas de mapas de uso e cobertura da terra. In Santos, L. B. L. and de Arruda Pereira, M., editors, *XXIII Brazilian Symposium on Geoinformatics GEOINFO 2022, São José dos Campos, SP, Brazil, November 28 30, 2022*, pages 211–216. MCTIC/INPE.
- MCTI (2020). Quarto Inventário Nacional de Emissões e Remoções Antrópicas de Gases de Efeito Estufa Relatório de Referência Setor Uso da Terra, Mudança do Uso da Terra e Florestas. Brasília.
- MCTI (2021). Comunicações nacionais do brasil à convenção-quadro das nações unidas sobre mudança do clima. Acesso em: 24 abr. 2022.
- Neves, A. K. et al. (2020). Assessment of terraclass and mapbiomas data on legend and map agreement for the Brazilian amazon biome. *Acta Amazonica*, 50:170–182.
- Pielke Sr., R. A., Pitman, A., Niyogi, D., Mahmood, R., McAlpine, C., Hossain, F., Goldewijk, K. K., Nair, U., Betts, R., Fall, S., Reichstein, M., Kabat, P., and de Noblet, N.

- (2011). Land use/land cover changes and climate: modeling analysis and observational evidence. *WIREs Climate Change*, 2(6):828–850.
- Reis;, M. S., Escada;, M. I. S., Sant'Anna;, S. J. S., and Dutra, L. V. (2017). Harmonização de legendas formalizadas em Land Cover Meta Language-LCML. *Anais do XVIII Simposio Brasileiro de Sensoramento Remoto*, 4(1):1–23.
- Reis, M. S. et al. (2018). Towards a reproducible LULC hierarchical class legend for use in the southwest of Pará state, Brazil: Data-driven hierarchies. *Land*, 7.
- Shukla, P. R. et al. (2019). Climate change and land: an IPCC special report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems.
- Souza, C. M., Z. Shimbo, J., Rosa, M. R., Parente, L. L., A. Alencar, A., Rudorff, B. F. T., Hasenack, H., Matsumoto, M., G. Ferreira, L., Souza-Filho, P. W. M., de Oliveira, S. W., Rocha, W. F., Fonseca, A. V., Marques, C. B., Diniz, C. G., Costa, D., Monteiro, D., Rosa, E. R., Vélez-Martin, E., Weber, E. J., Lenti, F. E. B., Paternost, F. F., Pareyn, F. G. C., Siqueira, J. V., Viera, J. L., Neto, L. C. F., Saraiva, M. M., Sales, M. H., Salgado, M. P. G., Vasconcelos, R., Galano, S., Mesquita, V. V., and Azevedo, T. (2020). Reconstructing three decades of land use and land cover changes in brazilian biomes with landsat archive and earth engine. *Remote Sensing*, 12(17).
- Verburg, P. H., Erb, K.-H., Mertz, O., and Espindola, G. (2013). Land system science: between global challenges and local realities. *Current Opinion in Environmental Sustainability*, 5(5):433–437. Human settlements and industrial systems.