

CLASSIFICATION SYSTEM DRIVES DISAGREEMENT AMONG BRAZILIAN VEGETATION MAPS AT A SAMPLE AREA OF THE SEMIARID CAATINGA

E. Bontempo^{*1,2}, M. C. Demirel³, C. Corsini¹, F. Martins^{1,2}, D. Valeriano²

¹ FUNCATE, Sao Jose dos Campos, Brazil. (edgard.silva, christianne.corsini, flora.martins)@funcate.org.br

² INPE, Sao Jose dos Campos, Brazil. (edgard.bontempo, dalton.valeriano)@inpe.br

³ Department of Civil Engineering, Istanbul Technical University, 34469 Maslak, Istanbul, Turkey. demirelmc@itu.edu.tr

KEY WORDS: Land Cover Classification, LCC, Classification Semantics, Caatinga, Vegetation Mapping

ABSTRACT:

The mapping of vegetation and Land Cover (LC) is important for research and for public policy planning but, in Brazil, although diverse maps exist there are few studies comparing them. The semiarid region of the Caatinga, in northeastern Brazil is an area long neglected by scientific research and its vegetation is diverse and relatively rich despite years of human occupation and very little preservation effort. In this study we make a comparison between the main maps made for the Caatinga from four different sources: IBGE (Brazilian Institute of Geography and Statistics), TCN (Third National Communication), ProBio (Project for Conservation and Sustainable Use of Biological Biodiversity) and MapBiomias. We also test these maps against well-known Land Cover maps from ESA and NASA: ESA's GlobCover and Climate Change Initiative (CCI) Land Cover, and NASA's MODIS MCD12Q1. This was done on a sample area where many of the Caatinga's vegetation physiognomies can be found, using well-established Difference metrics and the new SPAtial Efficiency (SPAEF) algorithm as they present complementary viewpoints to test the correspondence of mapped classes as well as that of their spatial patterns. Our results show considerable disagreement between the maps tested and their class semantics, with IBGE's and ProBio's being the most similar among all national maps and MapBiomias' the most closely related to global LC maps. The nature of the observed disagreement between these maps shows they diverge not only in the application of their classification systems, but also in their mapped spatial pattern, signaling the need for a better classification system and a better map of vegetation and land cover for the region.

1. INTRODUCTION

The classification of vegetation is a complex undertaking and, although man has studied it at least since Theophrastus of Athens wrote his notes entitled *Historia Plantarum*, sometime between 350 to 287 B.C. (Hort, 1917), there is not yet an accepted universal classification system or a "definitive" mapping of Earth's vegetation. Whichever one's objectives may be in classifying currently established vegetation, a system must be established to differentiate the observed plant communities and, a method must be chosen and utilized to effect said differentiation.

In the case of Brazil, the government (in separate initiatives) and Non-Governmental Organizations (NGOs) have produced different maps of vegetation and land use for diverse purposes and no scientific research has been done to compare or qualify these efforts. From the different regions of the country, the semiarid Caatinga has been particularly neglected by scientific research (Santos et al., 2011) and consequently, disadvantaged by most previous vegetation mapping attempts. This neglect is not in regard to the number of maps that have been produced but actually to the semantic choices made on the preparation of these maps and the fact that they can be perceived as insufficient to describe the phytogeographies of the Caatinga.

The official maps of Caatinga physiognomies, made by IBGE (IBGE, 2018), classify the local vegetation as a mosaic of foreign typologies (*i.e.*, not originated at the Caatinga but from other neighboring ecoregions like the Cerrado and the Atlantic Forests) interspersed with one main Caatinga-specific type denominated by IBGE as steppic-savanna which is subdivided into four sub-types: Forested Steppic-Savanna,

Arboreous Steppic-Savanna, Parkland Steppic-Savanna and Woody-Grassy Steppic-Savanna. The problem with this classification concerns mainly the validity of the steppic-savanna label. As we demonstrate in another work (Bontempo et al., 2020), the combination of both savanna and steppe into a compound name - originally suggested by Trochain (Schnell, 1971) but later abandoned by that same author (Trochain, 1957), is erroneous since the two classes, savanna and steppe, are incompatible by definition (Bontempo et al., 2020). Such a term is also misleading because the main characteristic defining a putative "steppic-savanna" would be a lack of clear dominance between savannic and steppic characteristics which is vague and hard to substantiate (Bontempo et al., 2020). Furthermore, the hierarchical subdivisions originally proposed for the region are a repetition of what was accepted for the savannic Cerrado: a density- and height-related organization with physiognomies containing a dense canopy at one end, the forested steppic-savanna, and another with trees absent and a mix of shrubs and seasonal forbs at the other end, the woody-grassy steppic-savanna. This is an oversimplification considering the more than eleven physiognomies that compose the best classification attempt (Andrade-Lima, 1981, Prado, 2003) and, although it can be construed as a reasonable simplification forced by the limitations of remote sensing since some of the characteristics used for distinguishing between the different physiognomies are not easily observable with current remote sensing technology, it still ignores the differences between similarly tall but essentially different physiognomies (Bontempo et al., 2020).

There are at present four main sources of vegetation maps for the Caatinga ecoregion: IBGE's own map of the Caatinga's ve-

vegetation and land use (IBGE, 2018), the map from the Third National Communication (TCN map - from the Portuguese acronym) from Brazil to the United Nations Framework Convention on Climate Change (UNFCCC); the ProBio map (from the Project for Conservation and Sustainable Use of Biological Biodiversity) (MMA, 2007); and, the MapBiomass national map of vegetation and land use from the Brazilian Annual Land Use and Land Cover Mapping Project (MapBiomass, 2019).

Considering the importance of vegetation and land cover maps for scientific research and public policy planning, we aimed to compare these four maps agreement and also, to test the coherence of their mapped classes to actual local vegetation in a small sample of the Caatinga region. Because no reference map can be found that represents a consensus, we have decided to compare them to other widely-used land cover and land use maps: the GlobCover 2009 product from ESA and Climate Change Initiative (CCI) Land Cover Classification product, also from ESA.

1.1 Objectives

Our objectives with this work are twofold:

- First, to demonstrate the differences and similarities between the existing land cover maps for the region of the Caatinga;
- Secondly, to compare the existing maps with three neutral references of LCC data produced with diverse data.

To achieve our objectives, maps were compared using the cross-tabulation based Disagreement or Difference metrics, previously suggested as a replacement for the inadequate, but widely used, Kappa index of agreement (Olofsson et al., 2014, Pontius Jr., Millones, 2011, Pontius Jr., Santacruz, 2014). Complementary to that well-established land-cover and land-use (LCLU) mapping comparison technique, the SPAtial Efficiency metric (SPAEF) was also used since it is an unbiased algorithm for the comparison of spatial patterns based on three metrics: correlation, coefficient of variation and histogram overlap (Demirel et al., 2017, Koch et al., 2018, Demirel et al., 2018). Although Pontius Jr.'s quantity, exchange and shift components of difference are robust they are still subject to eventual bias stemming from the classification systems employed in the different maps being tested (Pontius Jr., Santacruz, 2014). While SPAEF can be used to compare spatial patterns independently of their semantic relationships, testing rather their mathematical differences.

2. MATERIALS AND METHODS

2.1 Study Area

The Caatinga is a semiarid region of northeast Brazil with nearly 850,000 Km² harboring over 28 million inhabitants (Silva et al., 2017). The region is bounded by Atlantic evergreen forests to the east and by Cerrado savannas to the west and south. Caatinga vegetation is frequently armed with thorns and also commonly deciduous, it includes a range of woody and succulent species and a few ephemeral herbs and grasses distributed in a complex mosaic of physiognomies (Andrade-Lima, 1981, Pennington et al., 2000, Silva et al., 2017). Most of its area is not subject to frequent fires, as is common to savanna vegetation and, the local physiognomies include mostly

sparse and dense shrublands, savannas, Seasonally Dry Tropical Forests (STDF) and also enclaves of humid tropical forests usually isolated in landscape features favoured by orographic precipitation (Pennington et al., 2018, Silva et al., 2017). The region has been scantily studied (Santos et al., 2011, Pennington et al., 2018) and its relatively high levels of endemism (Pennington et al., 2000, Silva et al., 2017) points to an urgent need for further study and protection. Unfortunately, there are relatively few protected areas and these cover only 7.4% of its area while extensive human occupation since the 16th century has altered the local landscape mainly by the introduction of goat farming and by small-scale agriculture (Silva et al., 2017).

As a study concerning the total area of the Caatinga would be too complex and extensive to discuss in this format, we have chosen to subset the Caatinga area for this preliminary study. Therefore, we have selected an area containing several representative types of Caatinga vegetation (Fig. 1) which we have found to be often confused and misclassified, namely: seasonally dry tropical forests (STDF), savannas, steppic-savannas and Carrasco (a dense and tangled semi-deciduous vegetation composed mainly of shrubs and small trees permeated by thorny vines).

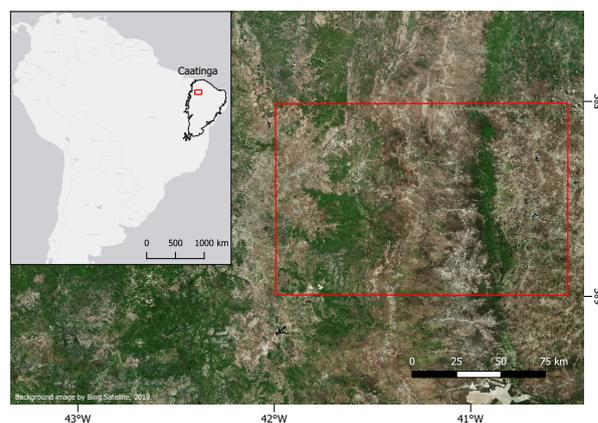


Figure 1. Caatinga and study area

2.2 Brazilian Maps

The Brazilian Institute for Geography and Statistics (IBGE - acronym from the Portuguese) has produced an official map from data gathered in the Radam Brasil Project (also spelled as RADAMBRASIL), a pioneering endeavor from the Brazilian Ministry of Energy which started in 1975, with the latest iteration dating from 2018 (IBGE, 2018). This map uses IBGE's own classification system developed in-house by a team of specialists using a morphological approach largely based on Ellemergh and Muller-Dombois' system from 1966 (IBGE, 2012).

At the request of the Brazilian Ministry of Environment (Environmental Ministry) and within the Project for Conservation and Sustainable Use of Biological Biodiversity (ProBio, also known as ProBio I) Landsat images were used to create a new map of Brazilian vegetation (MMA, 2007).

In the context of Brazil's national REDD+ framework, another map of vegetation was produced to be incorporated into the Third National Report of our country to the UNFCCC (MCTI, 2016) using a re-analysis of both the IBGE and the ProBio Map, along with more Landsat imagery. This map is locally known

as the TCN Map (from the portuguese acronym for the Third National Communication to the UNFCCC).

The IBGE, the ProBio and the TCN maps all follow IBGE's classification system and were all made at the scale of 1 : 250,000.

Completing the list of Brazilian-made vegetation maps is the one produced by the MapBiomias project (MapBiomias, 2019), a group established as a result of a seminary organized by the Greenhouse Gas Emission and Removal Estimating System (SEEG, from the Portuguese acronym). This map is the most modern of the three, made using Landsat and diverse data inputs processed through a complex chain using a Random Forest algorithm operating within Google Earth Engine. The MapBiomias map is the only national map included here not made using IBGE's classification system but, employing instead a custom "legend" as stated by the authors (MapBiomias, 2019). Due to its pixel-based mapping method and its use of Landsat imagery, it is the national map of highest spatial resolution: approximately 0.00029° by 0.00029°.

2.3 Global Maps

As previously stated, we have chosen to include three datasets representing global LCLU classification efforts, here referred to as the GlobCover map and the CCI map, from the European Space Agency (ESA) and, the MODIS map from the United States' National Aeronautics and Space Administration (NASA) agency. These were chosen due to their widespread use in the scientific community and also due to their spatial resolutions being comparable, according to Tobler's rule (Tobler, 1988), to that of the national maps described above, with the exception of the MapBiomias map.

The GlobCover map used here is from GlobCover 2009 and it was created through automatic classification of ENVISAT's Medium Resolution Imaging Spectrometer (MERIS) surface reflectance data from that year (Arino et al., 2012). Data is made available in raster format and it is gridded at approximately 0.0027° by 0.0027°.

The MODIS map used here is the MCD12Q1 Land Cover Product with in its IGBP classification layer (Friedl, Sulla-Menashe, 2015). This data is produced at approximately 0.0049° by 0.0049° and also made available in raster format.

Finally, the CCI map is from ESA's Climate Change Initiative (CCI) Land Cover product (ESA, 2018). It was made with PROBA-V data and it was included as its classification system is an improvement from GlobCover and it includes more classes that have been added after careful analysis of MERIS-based classification efforts.

2.4 Classification System Compatibility

To allow for the comparison of all chosen maps, we have made a custom system based on the current version of IBGE's vegetation classification system (IBGE, 2012) with composite classes included. Since no definitive maps exist and no adequate field classification or validation can be found for most of the Caatinga area, our choices here have been tailored from the beginning to improve compatibility whenever we had a subjective decision to make, for example: MapBiomias considers most of the local vegetation as the general class "savanna" but, since IBGE, Probio and TCN maps use the "steppic-savanna" class

to describe the majority of the area's vegetation, we converted MapBiomias' "savanna" into "steppic-savanna".

Our custom classification system is presented in the following table (Tab. 1) but the complete correlation between the systems of each used map and our custom one cannot be presented here as the space is insufficient for the multiple tables that would be necessary.

Number	Class Description
1	Water
2	Urban
3	Agriculture and pasture
4	Barren
5	Savanna / Steppic-savanna
6	Savanna / Steppic-savanna / Deciduous Forest
7	Savanna / Deciduous Forest
11	Forested Savanna
12	Woody Savanna
13	Park Savanna
14	Woody-Grassy Savanna
21	Forested Steppic-savanna
22	Woody Steppic-savanna
23	Park Steppic-savanna
24	Woody-Grassy Steppic-savanna
31	Deciduous Forest
44	Grassland

Table 1. Custom Land Cover Classification System - Classes are translations from IBGE's classification system adapted to include only types found in our study area, regardless of the dataset. Combined classes 5, 6 and 7 represent areas of mixed vegetation or ecotones.

2.5 Statistics

All maps were tested in two groups, one comparing the local Brazilian maps and another testing them against ESA's and NASA's maps. As previously stated, we use the Difference Metrics of Map Disagreement, quantity and allocation (the sum of exchange and shift) (Pontius Jr., Millones, 2011, Pontius Jr., Santacruz, 2014), and also, the SPAtial EFFiciency metric (SPAEF) defined by the following formula:

$$SPAEF = 1 - \sqrt{(\alpha - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (1)$$

where α is the Pearson correlation coefficient between map A and map B, β is the fraction of coefficient of variations representing spatial variability, and γ is the percentage of histogram intersection. For more information and detailed equations of each term on SPAEF please see (Demirel et al., 2017, Koch et al., 2018, Demirel et al., 2018).

All data processing was carried out using Qgis (QGIS Development Team, 2019), python (van Rossum, 1995) and R 3.6.1 (R Core Team, 2019), with the following R packages: diffeR (Pontius Jr., Santacruz, 2019), ggplot2 (Wickham, 2016), raster (Hijmans, 2019) and rgdal (Bivand et al., 2019).

3. RESULTS AND DISCUSSION

3.1 Brazilian Maps Comparison

Concerning the pairwise comparison of national maps, ProBio and IBGE maps show the highest spatial correlation and SPAEF, indicating the similarity between their spatial patterns

(Fig. 2). Although TCN and ProBio maps show a similar amount of correlation as can be observed between TCN and IBGE maps, the later pair has the lowest SPAEF value observed, due to their mutual coefficient of variation and divergence in the histograms of both maps (results not shown). This in turn indicates that the mapped categories are very different although the overall pattern (as indicated by spatial correlation) is not so dissimilar. Results also show that MapBiomias, although using a different classification system and having higher resolution than other maps is not markedly dissimilar from the others in its spatial pattern (Fig. 2).

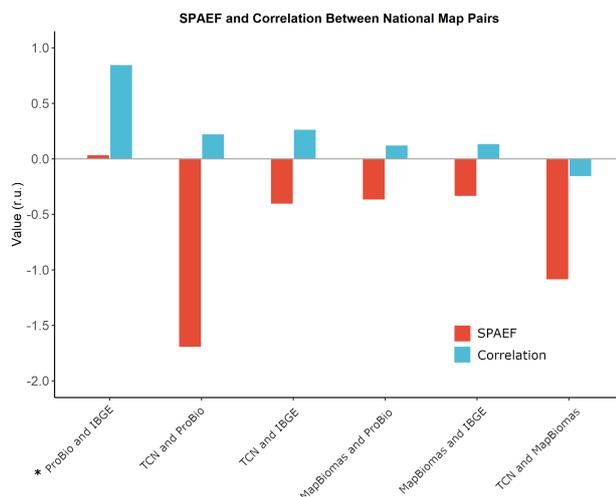


Figure 2. SPAEF Metrics Between National Map Pairs: SPAEF is the Spatial Efficiency index. A SPAEF value of 1 would indicate identical datasets. Correlation refers to Pearson Product Moment Correlation and is one of the factors used in calculating SPAEF, the others being coefficient of variation and histogram overlap.

Difference metrics' results show large disagreement between national maps (Fig. 3) averaging 64%. Although the IBGE, the ProBio and the TCN maps were all made using the same classification system we still observe overall differences (the sum of the three components: quantity, exchange and shift) at or above 50% between them.

As the majority of observed variation on differences is in the quantity component, our results show that these discrepancies pertain mostly to an imperfect match in the amount of classes in each map (Pontius Jr., Santacruz, 2014). As can also be seen, only between ProBio and IBGE there is significant exchange of mapped classes, and these are the maps found to be most similar among all compared data (overall difference at 23%). Our detailed cross-tabulation analysis of category differences shows that most confusion occurs between savanna and steppic-savanna (results not shown). As these are the categories occupying the largest proportion of the mapped area, results suggest that these maps are unreliable in differentiating between these categories. This is understandable because, as it is argued in another study (Bontempo et al., 2020) these classes are not well-defined and therefore, easily confused.

3.2 Local and Global Maps

Concerning the spatial patterns of national and global maps, our results show that MapBiomias is the most similar to all global

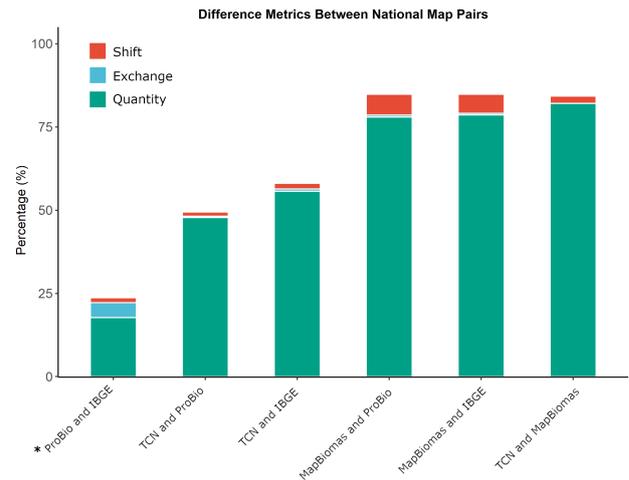


Figure 3. Difference Metrics Between National Map Pairs: Quantity is the percentage of disagreement between maps that derives from an imperfect match in the amount of the categories, Exchange and Shift are parts of the allocation component of differences and refer to confusion between the categories or classes of both maps. Exchange is the percentage of pixels reciprocally classified as category A in the first map and as category B in the second map while Shift is the percentage of non-reciprocal confusion (involving more categories).

maps (Fig. 4). The other map comparisons show mixed results but generally low similarity in general, with TCN and ProBio maps showing the overall lowest SPAEF values.

Comparisons through difference metrics shows better results for IBGE, ProBio and TCN against GlobCover data while MapBiomias was considerably similar to MODIS MCD12Q1 but, not so to the other two global maps (Fig. 5). Regardless, maps with IBGE-based classification system were still more similar among themselves than when compared to global maps (Fig. 3). MapBiomias in the other hand showed better results when compared to MODIS than with any other map and mixed results in all other comparisons (Figs. 3 and 5). National maps overall differences to international maps averaged over 77% (results not shown).

Category-based difference results against global maps corroborate our observation with national maps that most confusion was related to the steppic-savanna classes (not shown). Results of the comparison of national and global maps support the conclusion that the classification system used in IBGE, ProBio and TCN maps makes them incompatible to global data (Figs. 4 and 5). Furthermore, MapBiomias' map is also not ideal for northeastern Brazil vegetation as it labels generically as "savanna" a whole range of different vegetation types previously described at this region (Andrade-Lima, 1981, Prado, 2003).

Considering that the tested global maps were produced with classification systems based on the Plant Functional Type (PFT) concept (Arino et al., 2012, ESA, 2018, Friedl, Sulla-Menashe, 2015), and that PFTs have been shown to relate closely to biologically meaningful units and have been successfully used to improve large scale ecosystem models (Duckworth et al., 2000, Wullschleger et al., 2014), the discrepancies observed between these global maps and IBGE, ProBio and TCN maps raises

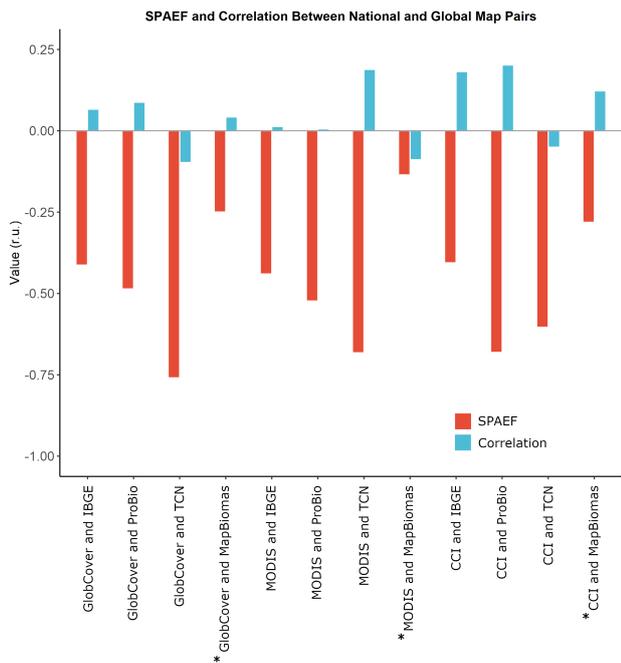


Figure 4. SPAEF Metrics Between National and Global Map Pairs: SPAEF is the Spatial Efficiency index. A SPAEF value of 1 would indicate identical datasets. Correlation refers to Pearson Product Moment Correlation and is one of the factors used in calculating SPAEF, the others being coefficient of variation and histogram overlap (not shown here).

doubts concerning their usability for scientific research. As our results suggest, the classification system used for the production of those maps hampers their credibility as realistic representations of the local plant communities and consequently, limits their application. Therefore, we suggest that a new classification system and a new map are needed to support scientific research in northeastern Brazil.

Ideally, such a system should harmonize the known complexity of Caatinga vegetation with the current technological limitations of remote sensing. For example, some of the vegetation units identified by previous research depend on the presence or absence of certain succulent species for their precise characterization (Prado, 2003) and, therefore, are not likely to constitute realistic units for mapping. Despite such limitations, we suggest that much can be done to improve the correspondence of mapped classes and basic PFTs, like for example: a more careful identification of the different scrub-like formations (e.g., sparse shrubland, dense shrubland and the vine-permeated Carasco shrublands) and the separation of *Orbignia spp.* (Babaçu) and *Copernicia sp.* (Carnaúba) -based palm formations from general savannic classes (currently mapped by IBGE as Savanna Parkland or Steppic-savanna parkland). These examples are feasible and their implementation would greatly improve the biological meaningfulness of mapped physiognomies.

Before our concluding remarks we must reiterate that the area studied here is small considering the expanse of the Caatinga region. Therefore, although we have confidence that our choice of study site was fortuitous in including diverse vegetation types, our results may change as the study progresses and more areas

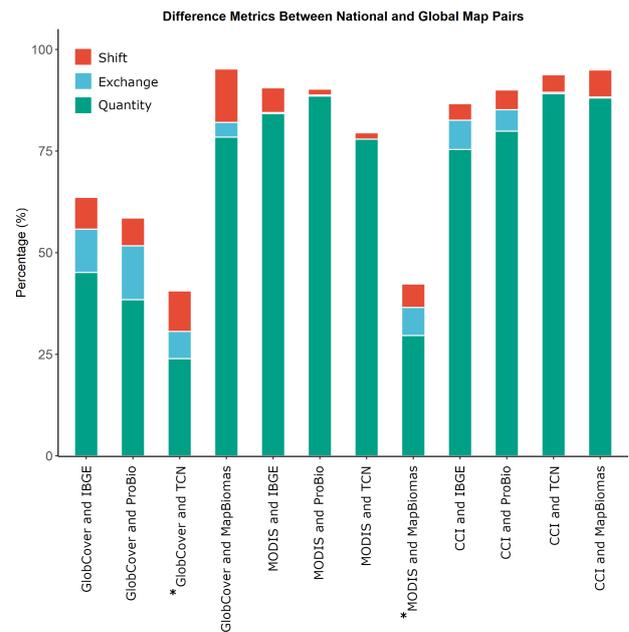


Figure 5. Difference Metrics Between National Map Pairs: Quantity is the percentage of disagreement between maps that derives from an imperfect match in the amount of the categories, Exchange and Shift are parts of the allocation component of differences and refer to confusion between the categories or classes of both maps. Exchange is the percentage of pixels reciprocally classified as category A in the first map and as category B in the second map while Shift is the percentage of non-reciprocal confusion (involving more categories).

are tested. Nevertheless, the category-related confusion we have observed seems to result from the classification system used for the whole region and, therefore, it is reasonable to assume that our conclusions will apply to most of the Caatinga.

4. CONCLUSIONS

Results show that national IBGE, ProBio, TCN and MapBiomas maps disagree considerably among themselves and also when compared to the widely used land products: Globcover 2009, MODIS MCD12Q1 and ESA's Climate Change Initiative Land Cover product. Furthermore, our results show that most class confusion was related to Steppic-savanna classes. Therefore, we suggest that a new classification system should be developed for remote mapping of Caatinga vegetation, as a new map, with more biologically accurate classes will improve support for research and public policy planning on this large and important semiarid region of South America.

ACKNOWLEDGEMENTS

Authors would like to thank Robert Gilmore Pontius Jr. for his comments and suggestions on our ideas for this study. This research was funded by the Amazon Fund (Fundo Amazônia - BNDES) through the project "Monitoramento Ambiental dos Biomas Brasileiros" - <http://www.fundoamazonia.gov.br/pt/projeto/Monitoramento-Ambiental-dos-Biomas-Brasileiros/>.

REFERENCES

- Andrade-Lima, D., 1981. The Caatingas Domain. *Revista Brasileira de Botânica*, 4, 149–162.
- Arino, O., Ramos Perez, J. J., Kalogirou, V., Bontemps, S., Defourny, P., Van Bogaert, E., 2012. Global Land Cover Map for 2009 (GlobCover 2009).
- Bivand, R., Keitt, T., Rowlingson, B., 2019. rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R package version 1.4-4.
- Bontempo, E., Corsini, C., Martins, F., Valeriano, D., 2020. The steppic-savanna fallacy: A revision into the semantics and applicability of the term to Caatinga vegetation. *In preparation*.
- Demirel, M. C., Koch, J., Stisen, S., 2017. SPAEF: SPATial Efficiency. GitHub Res., doi:10.5281/ZENODO.1158890.
- Demirel, M. C., Mai, J., Mendiguren, G., Koch, J., Samaniego, L., Stisen, S., 2018. Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model. *Hydrology and Earth System Sciences Discussions*, 2, 1–22.
- Duckworth, J. C., Kent, M., Ramsay, P. M., 2000. Plant functional types: an alternative to taxonomic plant community description in biogeography? *Progress in Physical Geography: Earth and Environment*, 24(4), 515–542.
- ESA, 2018. CCI Land Cover Product - <https://www.esa-landcover-cci.org/>.
- Friedl, M., Sulla-Menashe, D., 2015. MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 0.05Deg CMG V006 [Data set].
- Hijmans, R. J., 2019. raster: Geographic Data Analysis and Modeling. R package version 2.9-23.
- Hort, A., 1917. Theophrastus: Enquiry into Plants, and Minor Works on Odours and Weather Signs.
- IBGE, 2012. *Manual técnico da vegetação brasileira*. 2nd edn, Instituto Brasileiro de Geografia e Estatística.
- IBGE, 2018. Map of vegetation of Brazil 1 : 250, 000 scale - in Portuguese.
- Koch, J., Demirel, M. C., Stisen, S., 2018. The SPATial Efficiency metric (SPAEF): Multiple-component evaluation of spatial patterns for optimization of hydrological models. *Geoscientific Model Development*, 11(5), 1873–1886.
- MapBiomass, P., 2019. Collection 4 of Brazilian Land Cover and Use Map Series, accessed on Oct. 1, 2019, through the link: <https://mapbiomas.org/en/>.
- MCTI, 2016. Terceira Comunicação Nacional do Brasil à Convenção-Quadro das Nações Unidas sobre Mudança do Clima.
- MMA, 2007. Map of plant cover of Brazil, Caatinga - in Portuguese.
- Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., Wulder, M. A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, 42–57.
- Pennington, R., Lehmann, C. E. R., Rowland, L. M., 2018. Tropical savannas and dry forests. *Current Biology*, 28, 542–545.
- Pennington, R., Prado, D. E., Pendry, C., 2000. Neotropical Seasonally Dry Forests and Quaternary Vegetation Changes. *Journal of Biogeography*, 27(2), 261–273.
- Pontius Jr., R. G., Millones, M., 2011. Death to Kappa: Birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15), 4407–4429.
- Pontius Jr., R. G., Santacruz, A., 2014. Quantity, exchange, and shift components of difference in a square contingency table. *International Journal of Remote Sensing*, 35(21), 7543–7554.
- Pontius Jr., R. G., Santacruz, A., 2019. diffeR: Metrics of Difference for Comparing Pairs of Maps or Pairs of Variables. R package version 0.0-6.
- Prado, D., 2003. As caatingas da América do Sul. I. Leal, M. Tabarelli, J. Silva (eds), *Ecologia e Conservação das Caatingas*, Ed. Universitaria UFPE, Recife, 822.
- QGIS Development Team, 2019. QGIS Geographic Information System. Open Source Geospatial Foundation.
- R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Santos, J. C., Leal, I. R., Almeida-Cortez, J. S., Fernandes, G. W., Tabarelli, M., 2011. Caatinga: The Scientific Negligence Experienced by a Dry Tropical Forest. *Tropical Conservation Science*, 4(3), 276–286.
- Schnell, R., 1971. *Introduction a la phytogéographie des pays tropicaux, les problemes generaux*. Gauthier-Villars, Paris.
- Silva, J. M. C., Leal, I. R., Tabarelli, M. (eds), 2017. *Caatinga - The largest tropical dry forest region in South America*. Springer International Publishing.
- Tobler, W., 1988. Resolution, Resampling, and All That. H. Mounsey, R. Tomlinson (eds), *Building Data Bases for Global Science*, Taylor and Francis, London, 129–137.
- Trochain, J.-L., 1957. Accord interafricain sur la définition des types de végétation de l'Afrique Tropicale. *Bulletin de l'Institut d'Etudes centrafricaines, nouvelle série*, 13-14, 55–93.
- van Rossum, G., 1995. Python tutorial. Technical Report CS-R9526, Centrum voor Wiskunde en Informatica (CWI), Amsterdam.
- Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wulschleger, S. D., Epstein, H. E., Box, E. O., Euskirchen, E. S., Goswami, S., Iversen, C. M., Kattge, J., Norby, R. J., Van Bodegom, P. M., Xu, X., 2014. Plant functional types in Earth system models: Past experiences and future directions for application of dynamic vegetation models in high-latitude ecosystems. *Annals of Botany*, 114(1), 1–16.