

Um algoritmo de máxima concordância para harmonizar legendas de mapas de uso e cobertura da terra

Sabrina G. Marques¹, Pedro R. Andrade¹, Aline C. Soterroni²

¹Instituto Nacional de Pesquisas Espaciais, São José dos Campos, SP, Brasil

²Department of Biology, University of Oxford, Oxford, UK

sabrina.marques@inpe.br, pedro.andrade@inpe.br,
aline.soterroni@biology.ox.ac.uk

Abstract. *The growing production of land use and land cover maps has made it easier to monitor land changes. However, the integration and harmonization of maps can be a challenging task, since this diversity of initiatives also implies in a diversity of legends and class definitions. This article proposes an algorithm, based on the concept of maximum agreement, to help mapping legends from land use and land cover maps. The proposed algorithm was used to compare maps produced by TerraClass and MapBiomias in a small region of Tocantins. The results indicate that the automated production of a maximum agreement legend between two maps can be a feasible initial step in the legend harmonization process, being performed even before any semantic mapping between classes.*

Resumo. *A crescente produção de mapas de uso e cobertura da terra facilitou o monitoramento das mudanças de uso e cobertura do solo. No entanto, harmonizar e integrar esses mapas pode ser uma tarefa desafiadora, uma vez que essa diversidade de iniciativas também implica em uma diversidade de legendas e classes. Esse artigo propõe um algoritmo, baseado no conceito de concordância máxima, para ajudar no mapeamento de legendas de diferentes mapas de uso e cobertura da terra. O algoritmo proposto foi usado para comparar mapas produzidos pelo TerraClass e MapBiomias em uma pequena região do Baixo Tocantins. Os resultados indicam que a produção automatizada de uma legenda de concordância máxima entre dois mapas pode ser um passo inicial viável no processo de harmonização de legendas, sendo executado antes mesmo de qualquer mapeamento semântico entre as classes.*

1. Introdução

As atividades humanas provocam impactos na superfície da terra, sendo uma das principais forças nas alterações do sistema terrestre e no clima do planeta. Em escala global, o setor de agricultura, florestas e outros usos da terra é responsável por quase um quarto das emissões de gases de efeito estufa [Shukla et al. 2019]. Nesse contexto, os mapas de Uso e Cobertura da Terra (UTC) podem ser usados para acompanhar as dinâmicas do solo, bem como para estudar as interações sociais que afetam esses processos.

Com os avanços tecnológicos e o aumento significativo na disponibilidade de dados geoespaciais, temos como resultado diversas iniciativas produzindo mapas de UCT regionais e globais. No entanto, harmonizar e integrar mapas de UCT é uma tarefa desafiadora, uma vez que essa diversidade de iniciativas e dados disponíveis também implica em uma diversidade de metodologias, produtos e legendas que são desenvolvidos de

acordo com os objetivos dos estudos e as aplicações desejadas. Com isso, nos deparamos com uma variedade de sistemas de classificação com diferentes semânticas para as classes propostas. Por melhor que sejam esses mapeamentos, essas diferenças podem indicar inconsistências ao comparar mapas de UCT de iniciativas diferentes, ou até mesmo para versões distintas de uma mesma iniciativa [Verburg et al. 2011, Köhl et al. 2000].

O processo usual para comparar mapas de UCT consiste em elaborar uma legenda harmonizada a partir das descrições semânticas de cada uma das classes apresentadas pelos dois mapas [Neves et al. 2017, Reis et al. 2018, Capanema et al. 2019]. Dessa forma, algumas classes são ignoradas por falta de concordância ou informação. Outras são agrupadas em uma classe maior e mais genérica a fim de diminuir as dissimilaridades entre os conjuntos. Sistemas de classificação também costumam ser usados para traduzir essas diferenças semânticas em uma nova legenda igual para ambos os mapas, possibilitando compará-los através de diferentes métricas. O *Land Cover Classification System* (LCCS) é um dos mais relevantes sistemas de classificação e foi proposto pela *Food and Agriculture Organization* (FAO) com intuito de fornecer uma estrutura consistente para a classificação e mapeamento hierárquico da cobertura da terra. Ele foi desenvolvido para padronizar as terminologias dos atributos usados a partir de um conjunto de critérios independentes [Gregorio 2016].

Abordagens usando dados extraídos dos próprios mapas, considerando a distribuição espacial dessas classes, não são usuais. Este trabalho assume como hipótese que realizar uma análise prévia sobre esses dados pode auxiliar no processo de harmonização de legendas, uma vez que pode fornecer informações tanto sobre classes cujo mapeamento é trivial quanto sobre dúvidas e confusões no mapeamento. Nesse sentido, o objetivo deste trabalho é propor um algoritmo de harmonização automatizada de legendas que produza a maior concordância possível entre dois mapas.

2. Metodologia

Esta seção apresenta o algoritmo de harmonização de legendas de dois mapas de UCT. A ideia do algoritmo é prover informações que auxiliem o usuário a realizar o mapeamento entre as classes de dois mapas de uso da terra. Ele computa uma concordância máxima entre os dois mapas, o que significa que qualquer mapeamento de classes diferente do escolhido pelo algoritmo irá produzir uma concordância menor ou igual à concordância resultante do algoritmo.

O algoritmo recebe como entrada dois mapas de uso da terra M_1 e M_2 e retorna ao usuário essa proposta de legenda de acordo com as concordâncias identificadas. M_1 possui as classes x_1, x_2, \dots, x_m , e M_2 possui as classes y_1, y_2, \dots, y_n . O algoritmo assume que as representações espaciais dos dois mapas são compatíveis. É necessário que ambos mapas estejam na mesma projeção, em formato matricial, possuindo o mesmo número de linhas e colunas e tenham a mesma resolução espacial. Caso não estejam, é necessária uma etapa de pré-processamento desses mapas que não é considerada pelo algoritmo.

O algoritmo possui 3 passos. O primeiro passo consiste em produzir uma tabulação cruzada entre os mapas, dada pela matriz A onde o valor da posição a_{ij} da matriz representa a contagem de *pixels* de concordância entre a classe x_i de M_1 e a classe y_j de M_2 . Ela é da forma:

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, A \in N^{m \times n}, \quad (1)$$

No segundo passo são determinadas as concordâncias por linha e por coluna a partir de A , formando dois conjuntos, C_L e C_C , a partir das funções

$$c_l(x_i) = \{y_k : n_l(a_{ik}) = \max_{p=1, \dots, n} \{n_l(a_{ip})\}\}, \text{ onde} \quad (2)$$

$$n_l(a_{ij}) = \frac{a_{ij}}{\sum_{p=1}^n a_{ip}}, \text{ e} \quad (3)$$

$$c_c(y_i) = \{x_k : n_c(a_{ki}) = \max_{p=1, \dots, m} \{n_c(a_{pi})\}\}, \text{ onde} \quad (4)$$

$$n_c(a_{ij}) = \frac{a_{ij}}{\sum_{p=1}^m a_{pj}}, \text{ respectivamente.} \quad (5)$$

O conjunto C_L contém m duplas, $(x_i, c_l(x_i))$, $i \in \{1, \dots, m\}$ formadas pelas classes de M_1 as respectivas classe de M_2 que apresentaram maiores concordâncias de acordo com a matriz A . De forma análoga, o conjunto C_C é formado por n duplas, $(c_c(y_i), y_i)$, $i \in \{1, \dots, n\}$, da concordância das classes de M_2 em relação às classes de M_1 , de acordo com A .

Após formados os dois conjuntos, o último passo do algoritmo consiste em produzir a legenda harmonizada que será dada por

$$C_L \cup C_C, \quad (6)$$

contendo todas as concordâncias obtidas pelas funções, c_l e c_c . Esse conjunto poderá ter como tamanho máximo a soma das quantidades de classes dos dois mapas e como tamanho mínimo a quantidade de classes do mapa com mais classes.

Essa união entre os conjuntos C_L e C_C engloba três possibilidades de mapeamentos entre as classes dos mapas. A primeira é a mais simples, ocorrendo quando existe um mapeamento por linha da classe x_i para a classe y_j e também existe um mapeamento por colunas entre y_j e x_i no segundo mapeamento, assim a dupla (x_i, y_j) é considerada na harmonização. O segundo caso ocorre quando o mapeamento de uma classe existe somente em uma das harmonizações. Por exemplo, quando existe um mapeamento por linha da classe x_i para a classe y_j , mas a classe x_i não é mapeada na harmonização por colunas. Nesse caso a dupla (x_i, y_j) é considerada na harmonização. No último caso, existe um mapeamento por linha da classe x_i para a classe y_j mas no mapeamento por coluna a classe y_j é mapeada com a classe x_k , $k \neq i$, indicando uma inconsistência. De qualquer forma, as duplas (x_i, y_j) e (x_k, y_j) serão consideradas na legenda harmonizada.

3. Estudo de Caso

Como estudo de caso, o algoritmo proposto foi aplicado em uma pequena região do baixo Tocantins, que compreende os municípios paraenses de Mocajuba, Baião e Cametá. Esse

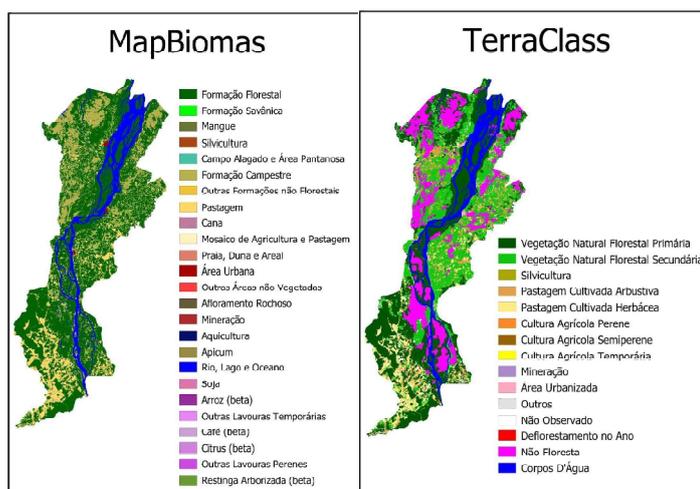


Figura 1. Mapas de Uso e Cobertura da Terra da região de interesse para 2014.

local foi escolhido com o intuito de replicar o trabalho feito por [Capanema et al. 2019], que realizou uma comparação entre o TerraClass Amazônia e o MapBiomias Coleção 2, ambos para o ano de 2014. Como os dados da Coleção 2 não se encontram mais disponíveis, neste estudo foram usados os dados do MapBiomias Coleção 6. A Figura 1 mostra os mapas de UCT de ambas as iniciativas para a área de estudo.

A Figura 2 mostra as equivalências por linha e coluna obtidas pelo algoritmo. As linhas azuis mostram as classes que tiveram a mesma equivalência nas harmonizações, ou seja, na harmonização por linha e por coluna foi formada a mesma dupla de classes. Como exemplo, temos o caso das classes de ‘Formação Florestal’ do MapBiomias e ‘Vegetação Natural Florestal Primária’ do TerraClass que foram mapeadas tanto na harmonização por linhas, quanto por colunas. As linhas laranjas representam classes que tiveram diferentes equivalências. Na harmonização por linha temos que a classe ‘Outras Lavouras Perenes’

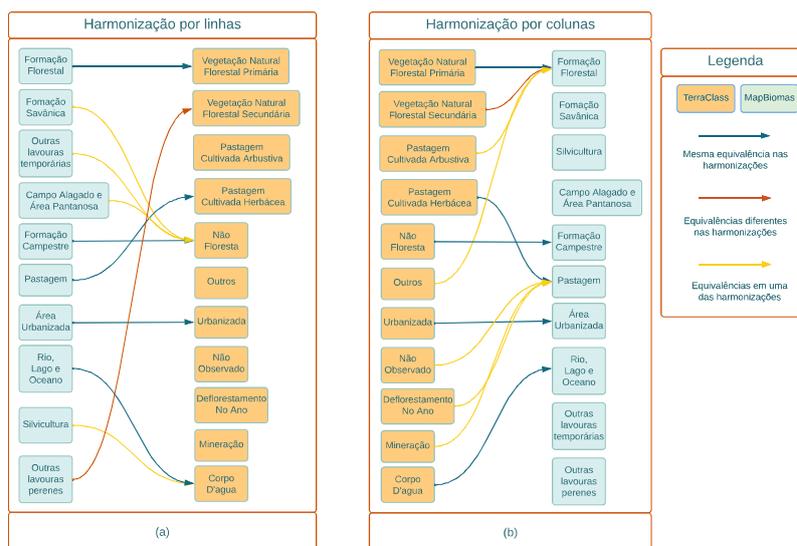


Figura 2. Equivalências por linha e coluna.

foi mapeada como ‘Vegetação Natural Florestal Secundária’, mas essa classe foi mapeada como ‘Formação Florestal’ na harmonização por colunas. Em amarelo estão as classes que foram identificadas em apenas uma das equivalências, como o caso da classe ‘Pastagem Cultivada Arbustiva’ do TerraClass que foi mapeada como ‘Formação Florestal’ na harmonização por colunas, mas na harmonização por linhas não houve mapeamento.

É importante ressaltar que nesse estudo de caso ocorreram as três possibilidades de mapeamento discutidas na metodologia. O segundo e o terceiro caso são aqueles em que o pesquisador deverá se atentar ao comparar os mapas ao fazer a harmonização das legendas, pois é onde as maiores divergências acontecem e, eventualmente, pode ser necessário desconsiderar algumas dessas classes na análise a ser realizada.

Por fim, após produzidos esses conjuntos de harmonizações, eles são agrupados formando a legenda harmonizada. A Figura 3 apresenta a proposta de legenda harmonizada, baseada nas equivalências por linha e por coluna obtidas. É possível observar que a legenda ótima produzida pode apresentar algumas inconsistências, como o mapeamento entre a classe de ‘Silvicultura’ no Mapbiomas e ‘Corpo D’água’ no TerraClass (note que ‘Corpo D’água’ no TerraClass foi mapeado como ‘Rio, Lago e Oceano’ no MapBiomas), o que fornece uma evidência das semânticas adotadas pelos diferentes mapeamentos.

A concordância entre os mapas usando o mapeamento produzido pelo algoritmo foi de 78%, enquanto que [Capanema et al. 2019] obtiveram uma concordância de 52% para a mesma região com uma versão anterior do MapBiomas. Esse aumento pode estar associada à melhor qualidade das versões mais recentes do MapBiomas. Em [Neves et al. 2017], a concordância entre essas iniciativas para o estudo do Pará foi de 84%, enquanto que em [Neves et al. 2020] a concordância foi de 87% e 92% para a região do bioma Amazônia, entre diferentes versões dessas iniciativas. Ambos os estudos apresentaram um concordância maior do que aquela apresentada pelo algoritmo, no entanto tratam-se de análises em regiões maiores, de forma que as diferenças locais podem ter impacto significativo na concordância.

Esta análise preliminar possibilita uma visão mais ampla das classes de uma le-

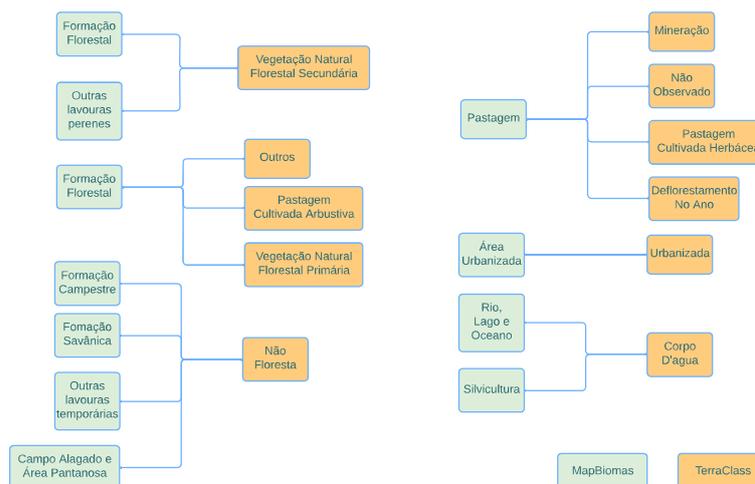


Figura 3. Legenda harmonizada obtida pelo algoritmo.

genda que não possuem equivalência direta com outra, como é o caso da classe ‘Mosaico de Agricultura e Pastagem’ do MapBiomas, que não foi identificada na área de estudo considerada. Caso fosse, seria possível analisar a qual classe ela está majoritariamente mapeada, possivelmente remapeando ao invés de desconsiderar em um estudo.

4. Conclusão

O algoritmo proposto produz uma harmonização automática entre legendas de mapas mapas de UCT que possui a maior concordância entre os dois mapas. Adicionalmente, ela poderá indicar eventuais inconsistências entre classes e fornecer uma análise das classes que não possuem uma equivalência direta com outra legenda. Esse algoritmo pode então ser um primeiro passo automatizado para o processo de mapeamento de classes. Ao invés de iniciar o processo de mapeamento das classes apenas a partir da semântica das legendas, o usuário poderá elaborar uma legenda usando como suporte esse mapeamento baseado na concordância das classes. Vale ressaltar que qualquer legenda diferente desta terá uma concordância menor ou igual à obtida pelo algoritmo proposto neste trabalho.

Os resultados obtidos no estudo de caso se mostraram promissores e indicam que a produção automatizada de uma legenda ótima pode trazer benefícios para a área de UCT. O código do experimento desse artigo está disponível em um *notebook*¹. Como trabalho futuro será desenvolvido um pacote *Python* com o algoritmo proposto neste trabalho e um *plugin* para o QGIS com intuito de facilitar o seu uso.

Referências

- 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.
- Gregorio, A. D. (2016). *Land Cover Classification System: Classification concepts*. Food and Agriculture Organization of the United Nations (FAO), Rome.
- Köhl, M. et al. (2000). Harmonisation and standardisation in multi-national environmental statistics – mission impossible? *Environmental Monitoring and Assessment*, 63:361–380.
- Neves, A. K. et al. (2017). Terraclass x mapbiomas: Comparative assessment of legend and mapping agreement analysis. *Proceedings of the Brazilian Symposium on Geoinformatics*, 2017-Decem:295–300.
- 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.
- 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.
- Verburg, P. H. et al. (2011). Challenges in using land use and land cover data for global change studies. *Global Change Biology*, 17(2):974–989.

¹<https://www.kaggle.com/code/sabrinamarques/algoritmo-de-harmoniza-o-estudo-de-caso>