EVOLUÇÃO DA ACURÁCIA DE IMPORTANTES CATEGORIAS DE COBERTURA DO MAPBIOMAS PARA TODO O ESTADO DE SÃO PAULO (SP)

Nadinne da Silva Fernandes[1], *Eduardo Arraut* [1]

[1]AfiliaçãoInstituto Tecnológico de Aeronáutica (ITA), Praça Marechal Eduardo Gomes, 50 - Vila das Acácias, São José dos Campos - SP, 12228-900

RESUMO

A cobertura da terra é importante porque influencia, dentre outras coisas, clima, distribuição espacial de áreas verdes, disponibilidade hídrica e biodiversidade. O objetivo deste estudo foi avaliar a evolução da acurácia de importante categorias de cobertura do MapBiomas para todo o Estado de São Paulo (SP), considerando desde a coleção 3.1 à 7.0. Foram calculados para cada classe (Formação Florestal, Floresta Plantada, Pastagem e Cana-de-açúcar) de cada coleção as Acurácias Global, do Produtor (AP, ou 100% - erro de Omissão), e do Usuário (AU, ou 100% - Erro de Comissão). Os valores de acurácia global atingidos em todas as coleções foram maiores ou iguais a 83%, indicando boa acurácia no mapeamento desde o início do MapBiomas, tendo o menor valor sido atingido na coleção 4.1 e o maior valor na coleção 7.0.

Palavras-chave **—** cobertura da terra, acurácia, MapBiomas.

ABSTRACT

Land cover is important because it influences, among other things, climate, spatial distribution of green areas, water availability and biodiversity. The objective of this study was to evaluate the evolution of the accuracy of important MapBiomas coverage categories for the entire State of São Paulo (SP), considering from collection 3.1 to 7.0. For each class (Forest Formation, Planted Forest, Pasture and Sugar cane) of each collection, the Global Accuracy, Producer (AP, or 100% - Omission error), and User (AU, or 100% error) - Commission error). The global accuracy values, achieved in all collections were greater than or equal to 83%, indicating good mapping accuracy since the beginning of MapBiomas, with the lowest value being reached in the 4.1 collection and the highest value in the 7.0 collection.

Key words — land cover, *accuracy, MapBiomas*

1. INTRODUÇÃO

A cobertura da terra é importante porque influencia, dentre outras coisas, clima, distribuição espacial de áreas verdes, disponibilidade hídrica e biodiversidade.

Como exemplos disso, um estudo de como a paisagem se altera em uma área de desmatamento do Cerrado brasileiro compreendeu que dos anos de 2013 a 2020, o percentagem da cobertura nativa diminuiu 3,85% no bioma Cerrado no estado da Bahia (Assis, Esscada e Amaral, 2021) [1]

Utilzando o MapBiomas, também, avaliaram o desmatamento na Amazõnia e a relação com a emissão de carbono, demonstraram que até 2017 seu estoque total de carbono compensou menos de 10% das emissões resultantes da perda de florestas antigas (Smith *et al*, 2020) [2]. No Brasil, o Projeto de Mapeamento Anual do Uso e Cobertura da Terra (MapBiomas), nascido em 2015 e resultante da colaboração entre universidades, ONGs e startups de tecnologia, mapeia anualmente todo território brasileiro e disponibiliza seus mapas de maneira gratuita.

Para quantificar a incerteza do mapeamento, o que é necessário para o uso científico de um mapa (CONGALTON, 1991) [3], o MapBiomas realiza uma avaliação da acurácia baseada em um amostra de ~75.000 píxeis criada via interpretação visual de imagens Landsat, chamada de base de dados de referência. Esse tamanho da amostra foi pré-determinado por amostragem estatística. Entretanto, o tamanho e distribuição da amostra de avaliação da acurácia do MapBiomas foram dimensionados para o Brasil inteiro, e por isso não podem ser tomados como base para a acurácia do mapeamento quando se estudam áreas bem menores, como um estado ou município. Estudos que usam o MapBiomas em análises regionais ou locais são, entretanto, frequentes (MapBiomas, 2022) [4].

O objetivo deste estudo foi avaliar a evolução da acurácia das principais categorias de cobertura do MapBiomas para todo o Estado de São Paulo (SP), considerando desde a coleção 3.1 à 7.0. Espera-se com ele contribuir para a evolução da qualidade do mapeamento do MapBiomas, e prover uma base de referência da incerteza do mapeamento para estudos focados no estado de SP.

2. MATERIAL E MÉTODOS

2. 1. Caracterização da área de estudo

A área de estudo é o estado de São Paulo, localizado na região Sudeste do Brasil. Ela contém uma paisagem altamente heterogênea e fragmentada. São Paulo possui uma área, segundo o Instituto Brasileiro de Geografia e Estatística (IBGE) (2021) [5], de 248.219,485 km², e uma população estimada de 46,3 milhões de habitantes. É o estado mais populoso do Brasil, concentrando 21,9% da população total do país (IBGE, 2020) [6]. De acordo com ranqueamento feito pelo IBGE, , o estado de SP contém 3

das 17 cidades com mais de 1 milhão de habitantes, estando sua capital, São Paulo, em 1º lugar, Guarulhos em 13º e Campinas em 14º. A região metropolitana de São Paulo é a mais populosa do país, com 21,9 milhões de habitantes.

2. 2. Dados

Foram escolhidas as coleções 3.1, 4.1, 5.0, 6.0 e 7.0, todas do ano de 2017. A definição das coleções foi feita com base na disponibilidade de mapas passíveis de serem baixados na plataforma do MapBiomas, e o ano de 2017 foi escolhido por ser o ano mais atual que continha mapas de todas as coleções.

Para que a comparação entre coleções fosse possível, as classes consideradas na avaliação da acurácia foram apenas aquelas presentes em todas as coleções do MapBiomas analisadas. Assim, por exemplo, as classes Café, Citrus e Restinga Arborizada, que aparecem apenas nas coleções 6 e 7, e a classe Lavoura Perene que só aparece na coleção 5, e a classe Lavoura Temporária que só existe nas coleções 3.1 e 4.1, não foram consideradas. As classes finais analisadas foram: Formação Florestal, Floresta Plantada, Pastagem e Cana de Açúcar.

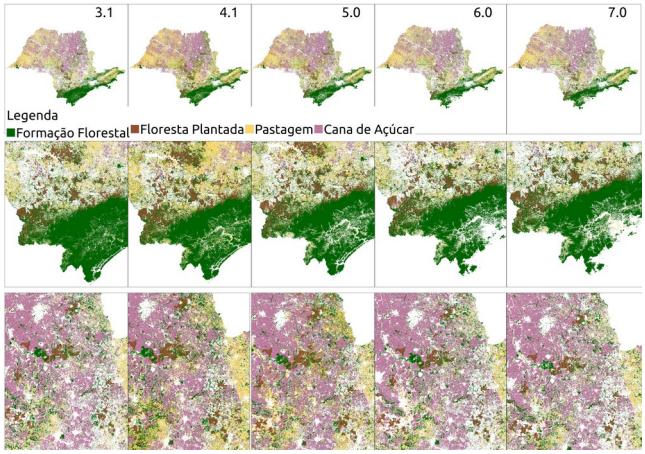


Figura 2. Evolução do MapBiomas

2. 3. Método

O software utilizado foi o QGIS 3.22 com GRASS 7.2.2 para calcular a área das classes, e o plugin AcATama foi usado para gerar os pontos aleatórios e realizar a matriz de confusão. Para calcular a área das classes, primeiro as imagens foram reprojetadas para coordenadas UTM, e em seguida foi utilizada a função *r.report* que reporta a área de cada classe conforme a unidade de medida escolhida. A unidade amostral foi o ponto vetorial, e diante da grande variabilidade na área de cada classe o desenho amostral consistiu em usar o plugin AcATama para posicionar 2890 pontos de referência usando uma amostragem estratificada. O cálculo do tamanho da amostra por classe foi feito usando a fórmula (Stehman e Foody, 2019) [7]:

$$N = (Z^2p(1-p))/d^2$$

Onde n é o número de amostras, z é o intervalo de confiança, p é a porcentagem da área da classe e d é a margem de erro.

A tabela 1 apresenta a quantidade de amostras para cada classe considerada na análise, por coleção. A quantidade de amostras calculadas foi com 90% de grau de confiança e 5% de margem de erro.

Classe	Nome	3.1	4.1	5.0	6.0	7.0
3	Formação Florestal	173	173	174	166	164
9	Floresta Plantada	35	45	42	35	36
15	Pastagem	162	211	181	162	164
20	Cana de Açúcar	188	200	201	190	188
		558	**629**	**598**	**553**	**552**
Total: 2890						

Tabela 1 - Quantitativo de amostras por coleção

O desenho resposta, que é o protocolo por meio do qual a classe de referência foi associado a cada unidade amostral, consistiu em atribuir ao ponto amostral a classe predominante dentro do buffer de 10m gerado para os pontos. A avaliação da classe de referência foi feita com base na interpretação visual de até quatro imagens ou mapas diferentes: Google Satellite, Yandex, ESRI e o projeto Canasat [8]. Ao comparar os mapas base com a função *time lapse* do Google Earth Pro (GEP) observamos que os mapas base possuem datas diferentes para visualização, sendo o Yandex com datas mais antigas, com a maioria em 2016, e o Google Satellite o mais atual. O Projeto Canasat é um mapeamento de cana de açúcar para o estado de São Paulo de 2013 realizado pelo Instituto Nacional de Pesquisa Espacial (INPE), e o uso desse dado facilitou a interpretação da cana-de-açúcar, que por ser uma cultura perene não é de fácil distinção com base apenas em imagens. Quando ainda restava dúvida sobre a classe de referência, usando o próprio plugin abria-se o GEP e interpretava-se a classe com base na série histórica de alta-resolução espacial disponível.

Foram calculados para cada classe de cada coleção as Acurácias Global, do Produtor (AP, ou 100% - erro de Omissão), e do Usuário (AU, ou 100% - Erro de Comissão).

3. RESULTADOS

As tabelas 2 a 6 mostram as matrizes de confusão para as coleções 3.1, 4.1, 5.0, 6.0 e 7.0, respectivamente. A classe 34 foi chamada de “outros”, que é quando a interpretação visual da série histórica de altíssima resolução espacial mostra que o MapBiomas está errado. Por exemplo, Mapbiomas classificou como Pastagem mas é claramente Infraestrutura Urbana.

		MapBiomas 3.1						
		3	**9**	**15**	**20**	**34**	**total**	**AU**
Mapa Base	**3**	154	1	2	0	0	157	98%
	9	8	32	0	0	0	40	80%
	15	8	1	147	7	0	163	90%
	20	0	0	10	177	0	187	95%
	34	0	1	2	4	0	7	
	total	170	35	161	188	0		
	AU	91%	91%	91%	94%	0%		84,0 %

Tabela 2 - Matriz de confusão - coleção 3.1

		MapBiomas 4.1						
		3	**9**	**15**	**20**	**34**	**total**	**AU**
Mapa Base	**3**	157	3	8	2	0	170	92,4%
	9	13	38	6	0	0	57	66,7%
	15	2	2	182	8	0	194	93,8%
	20	0	1	10	189	0	200	94,5%
	34	0	0	0	0	0	0	
	total	172	44	206	199	0		
	AP	91,3 %	86,4%	88,3 %	95,0%	0%		83,0%

Tabela 3 - Matriz de confusão - coleção 4.1

		MapBiomas 5.0						
		3	**9**	**15**	**20**	**34**	**total**	**AU**
Mapa Base	**3**	161	1	5	1	0	168	96%
	9	9	37	1	0	0	47	79%
	15	1	3	156	6	0	166	94%
	20	1	1	12	193	0		93%
	34	1	0	5	1	0		
	total	173	42	179	201	0		
	AP	93,06 %	88,10 %	87,15 %	96,02 %	0%		83,6%

Tabela 4 - Matriz de confusão - coleção 5.0

		MapBiomas 6.0						
		3	**9**	**15**	**20**	**34**	**total**	**AU**
Mapa Base	**3**	155	2	2	1	0	160	97%
	9	6	33	0	1	0	40	83%
	15	5	0	149	4	0	158	94%
	20	0	0	9	184	0	193	95%
	34	0	0	2	0	0	2	
	total	166	35	162	190	0		
	AP	93%	94%	92%	97%	0%		94,2 %

Tabela 5 - Matriz de confusão - coleção 6.0

		MapBiomas 7.0						
		3	**9**	**15**	**20**	**34**	**total**	**AU**
Mapa Base	**3**	157	3	1	0	0	161	97,5%
	9	6	32	0	0	0	38	84,2%
	15	0	0	155	2	0	157	98,7%
	20	0	0	5	185	0	190	97,4%
	34	1	1	2	1	0	5	0,0%
	total	164	36	163	188	0		
	AP	95,7 %	88,9 %	95,1 %	98,4%	0 %		96,01 %

Tabela 6 - Matriz de confusão - coleção 7.0

Com relação à acurácia global, nota-se que houve uma redução da coleção 3.1 para a 4.1, seguida de uma melhora constante até a coleção 7.

Do ponto de vista do produtor, nota-se aumento da acurácia em todas as classes da coleção 7. Pode-se observar também que as classes Floresta Plantada (9), Pastagem (15) e Cana-de-áçúcar (20) tiveram sua pior acurácia na coleção 4.1, mas voltaram a ter um aumento progressivo. A Silvicultura obteve seu pico na coleção 6.

Do ponto de vista do usuário, nota-se que a classe Formação Florestal (3) teve sua maior acurácia na coleção 3.1 (98%), pior na coleção 4.1 (92%), voltando a melhorar gradativamente até a coleção 7.0 (0,96). A classe Pastagem (15) ao longo das coleções aumentou sua acurácia. Por fim, a classe Cana-de-açúcar (20) diminuiu de acurácia até a coleção 5, voltando a melhorar a partir da coleção 6.

4. CONCLUSÃO

Os valores de acurácia global atingidos em todas as coleções foram maiores ou iguais a 83%, indicando boa

acurácia no mapeamento desde o início do MapBiomas. A menor acurácia global foi atingida na coleção 4.1 e a maior na coleção 7.0.

Ao comparar os valores calculados para todo o Brasil pelo projeto MapBiomas com os que calculamos para o estado de São Paulo, verificamos que a melhor acurácia para ambos os casos foi na coleção 7. Isso indica que, tomando ao menos a paisagem no estado de SP, o algoritmo de classificação do MapBiomas vem evoluindo.

Os resultados mostram que o produto de cobertura da terra do MapBiomas apresenta um bom grau de confiabilidade para as classes analisadas. Espera-se que os valores de acurácia obtidos aqui possam dar mais robustez a estudos que usam o MapBiomas como dado de entrada em outras análises, envolvendo por exemplo estocagem de carbono, disponibilidade de habitat para a biodiversidade, ou quantidade e qualidade da água.

8. REFERÊNCIAS

[1] Assis, T.O.; Escada, M.I.S.; Amaral, S. Effects of Deforestation over the Cerrado Landscape: A Study in the Bahia Frontier. Land 2021, 10, 352. https://doi.org/10.3390/land10040352

[2] Smith, CC , Espírito-Santo, FDB , Healey, JR , et al. As florestas secundárias compensam menos de 10% das emissões de carbono mediadas pelo desmatamento na Amazônia brasileira . Globo. Troque Biol. 2020 ; 26 : 7006-7020 _ _ . https://doi.org/10.1111/gcb.15352

[3] CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, v.49, n.12, p.1671-1678, 1991

[4] MAPBIOMAS, Análise de Acurácia. Disponível em https://mapbiomas.org/analise-de-acuracia. Acesso em: 01 jun 2022.

[5] IBGE, Instituto Brasileiro de Geografia e Estatística. Cidades e Estados. Disponível em https://www.ibge.gov.br/cidades-e-estados/sp.html. Acesso em: 07 jun 2022.

[6]IBGE, Instituto Brasileiro de Geografia e Estatística. Estatísticas Sociais. 27 ago 2020. Disponível em: https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/28668-ibge-divulga-estimativa-da-populacao-dos-municipios-para-2020. Acesso em: 13 out 2022

[7]Stephen V. Stehman, Giles M. Foody. Key issues in rigorous accuracy assessment of land cover products. Remote Sensing of Environment, v 231, 2019, 111199, ISSN 0034-4257. Disponível em: https://doi.org/10.1016/j.rse.2019.05.018. Acesso: 16 ago 2022

[8] CANASAT, Monitoramento da Cana-de-açúcar. Disponível em http://www.dsr.inpe.br/laf/canasat/. Acesso em: 01 jun 2022.