

AVALIAÇÃO DA ACURÁCIA NA CLASSIFICAÇÃO DE USO E COBERTURA DA TERRA: UM ESTUDO DE CASO NA REGIÃO INTERMEDIÁRIA DE SÃO LUÍS COM O MAPBIOMAS

**EVALUATION OF ACCURACY IN LAND USE AND
COVER CLASSIFICATION: A CASE STUDY IN THE
INTERMEDIATE REGION OF SÃO LUÍS WITH MAPBIOMAS**

**EVALUACIÓN DE LA PRECISIÓN EN LA CLASIFICACIÓN
DEL USO Y COBERTURA DEL SUELO: UN ESTUDIO DE
CASO EN LA REGIÓN INTERMEDIA DE SÃO LUÍS CON
MAPBIOMAS**

DOI 10.33360/geonordeste.v36i.21680

Thomas Victor de Sousa Malheiros Rocha
Mestre em Ciência e Tecnologia Ambiental
Universidade Federal do Maranhão
thomasvictor990@gmail.com

Sérgio Souza Costa
Doutor em Computação Aplicada
Professor Associado da Universidade Federal do Maranhão (UFMA)
sergio.costa@ufma.br

Luís Fernando Cirqueira da Silva Correia
Mestre em Ciência e Tecnologia Ambiental
Universidade Federal do Tocantis (UFT)
luis.correia@discente.ufma.br

Denilson da Silva Bezerra
Doutor em Ciência do Sistema Terrestre
Professor Adjunto A da Universidade Federal do Maranhão (UFMA)
denilson.bezerra@ufma.br

RESUMO

A classificação de uso e cobertura da terra é essencial para gerenciar recursos naturais. O uso da terra inclui atividades como agricultura e urbanização, enquanto a cobertura se refere a características físicas e biológicas, como florestas e corpos d'água. Com o avanço tecnológico, a classificação automática tornou-se viável em grandes países, como o Brasil, com destaque para o trabalho anual do MapBiomas. Além da classificação, esse projeto fornece dados sobre acurácia, códigos e amostras usadas na validação, permitindo avaliar a acurácia em áreas específicas, como a região intermediária de São Luís. Adicionalmente, algumas classes de interesse, como Mangue, Apicum, Praia e Duna, ainda não têm avaliação de acurácia disponível. Deste modo, este estudo analisou a acurácia destas e outras classes dentro do agrupamento SA-23-Z, que cobre 72% da região intermediária de São Luís. Os resultados mostraram alta acurácia para Praia e Duna, mas um alto erro de comissão para Mangue e exclusão de Apicum devido à baixa amostragem. O erro de alocação foi de 20%, acima da média nacional, mas o índice Kappa indica que a classificação do MapBiomas apresenta uma ótima acurácia, apesar das limitações.

Palavras-Chave: Classificação de Imagens; Sensoriamento Remoto; Google Earth Engine.

ABSTRACT

Land use and land cover classification is essential for managing natural resources. Land use includes activities such as agriculture and urbanization, while land cover refers to physical and biological characteristics, such as forests and water bodies. With technological advances, automatic classification has



become feasible in large countries, such as Brazil, with the annual work of MapBiomass standing out. In addition to classification, this project provides data on accuracy, codes, and samples used in validation, allowing the assessment of accuracy in specific areas, such as the intermediate region of São Luís. Additionally, some classes of interest, such as Mangrove, Apicum, Beach, and Dune, do not yet have accuracy assessment available. Therefore, this study analyzed the accuracy of these and other classes within the SA-23-Z cluster, which covers 72% of the intermediate region of São Luís. The results showed high accuracy for Beach and Dune, but a high commission error for Mangrove and exclusion of Apicum due to low sampling. The allocation error was 20%, above the national average, but the Kappa index indicates that the MapBiomass classification presents excellent accuracy, despite its limitations.

Keywords: Image Classification; Remote Sensing; Google Earth Engine.

RESUMEN

La clasificación del uso y la cobertura del suelo es esencial para la gestión de los recursos naturales. El uso del suelo incluye actividades como la agricultura y la urbanización, mientras que la cobertura del suelo se refiere a características físicas y biológicas como los bosques y las masas de agua. Con los avances tecnológicos, la clasificación automática se volvió viable en países grandes, como Brasil, con énfasis en el trabajo anual de MapBiomass. Además de la clasificación, este proyecto proporciona datos sobre precisión, códigos y muestras utilizadas en la validación, permitiendo evaluar la precisión en áreas específicas, como la región intermedia de São Luís, además de algunas clases de interés, como Mangrove, Apicum, Praia y Duna, aún no disponen de una evaluación de precisión. Por lo tanto, este estudio analizó la precisión de estas y otras clases dentro del grupo SA-23-Z, que cubre el 72% de la región intermedia de São Luís. Los resultados mostraron una alta precisión para Praia y Duna, pero un alto error de comisión para Mangrove y exclusión de Apicum debido al bajo muestreo. El error de asignación fue del 20%, por encima del promedio nacional, pero el índice Kappa indica que la clasificación de MapBiomass tiene una precisión excelente, a pesar de las limitaciones.

Palabras clave: Clasificación de Imágenes; Teledetección; Motor Google Earth.

1. INTRODUÇÃO

A classificação de uso e cobertura da terra é uma ferramenta essencial para o entendimento e gestão dos recursos naturais e ambientais. Essa prática envolve a categorização da superfície terrestre em diferentes classes com base no tipo de uso humano e nas características da vegetação e outras coberturas naturais (IBGE, 2013). A precisão e a atualização constante dessas informações são cruciais para uma série de aplicações, desde o planejamento urbano até a conservação da biodiversidade.

O uso da terra refere-se à maneira como os seres humanos utilizam a terra e seus recursos, incluindo agricultura, áreas residenciais, industriais e de lazer. Já a cobertura da terra diz respeito às características físicas e biológicas presentes na superfície terrestre, como florestas, corpos d'água, áreas urbanizadas e campos agrícolas (TURNER, 1994). Compreender a dinâmica entre uso e cobertura da terra permite uma avaliação mais detalhada dos impactos ambientais, contribuindo para políticas públicas mais eficazes e sustentáveis.

Com o avanço das tecnologias de sensoriamento remoto e sistemas de informação geográfica (SIG), a obtenção e análise de dados de uso e cobertura da terra tornaram-se mais acessíveis e precisas. Nesse contexto, destaca-se o MapBiomass, que foi lançado em 2015, utiliza tecnologias avançadas de sensoriamento remoto e processamento de dados para mapear anualmente a cobertura e uso da terra no Brasil, desde 1985 até os dias atuais (MAPBIOMAS, 2023). Através de uma abordagem colaborativa, o MapBiomass integra dados de satélite com técnicas de



aprendizado de máquina para gerar mapas detalhados que cobrem todo o território brasileiro (MAPBIOMAS, 2023). Os resultados são disponibilizados gratuitamente para o público por meio de uma plataforma online, permitindo acesso fácil e transparente às informações.

O MapBiomias tem sido amplamente utilizado em diversos estudos científicos, no monitoramento e análise de transformação no uso e cobertura da terra, como nos estudos de Azevedo e Matias (2022), Souza Jr. *et al.* (2020), Barbosa de Souza *et al.* (2023) entre outros, os quais demonstram a relevância dos dados do projeto como uma ferramenta essencial para a pesquisa e monitoramento ambiental.

O sistema de classificação do MapBiomias é estruturado em diferentes níveis hierárquicos, permitindo uma análise detalhada e precisa das variações do uso e cobertura da terra. Esses níveis de classificação proporcionam uma visão abrangente e específica, adaptando-se às necessidades de diversos usuários, desde gestores ambientais a pesquisadores e formuladores de políticas públicas.

No primeiro nível, a classificação abrange categorias amplas de uso e cobertura do solo, fornecendo uma visão geral da paisagem. As principais classes incluem: Floresta, Formações Naturais Não Florestais, Agropecuária e Corpo D'Água. No segundo nível, a classificação se torna mais detalhada, subdividindo as classes gerais em subclasses específicas. Por exemplo, dentro da classe "Agropecuária", podem existir subcategorias como: Pastagem, Agricultura e Silvicultura. Essa subdivisão permite uma análise mais detalhada das variações dentro de cada categoria principal, fornecendo informações mais precisas para estudos específicos.

O terceiro nível de classificação envolve um detalhamento ainda mais fino, identificando características específicas dentro das subclasses. Por exemplo, dentro da subclasse "Agricultura", pode haver distinções entre lavoura temporária e perene. Que em um quarto nível pode ser Soja, Cana, Arroz, Café, entre outros (MAPBIOMAS, 2023). Esse nível de detalhamento é crucial para estudos específicos de uso da terra, permitindo uma compreensão detalhada das práticas agrícolas e seus impactos ambientais. Contudo, no painel de estatísticas de acurácia (MAPBIOMAS, 2023) não estavam disponíveis as acurácias de diversas classes destes níveis mais detalhados. No nível 2 não aparece a acurácia de classes como: Mangue, Apicum, Praia e Duna, Mineração e Área Urbanizada.

Este trabalho tem como interesse avaliar a acurácia de uma região litorânea, mais especificamente a região intermediária de São Luís, composta por 73 municípios, localizada na costa leste do Brasil. Segundo dados do IBGE (2022), a área total é de aproximadamente 68.393 km², abrigando uma população de aproximadamente 3 milhões de habitantes. Essa combinação de características torna a região um importante polo de pesquisa e desenvolvimento cultural (IBGE, 2023).

Nessa região, classes de cobertura como Mangue, Apicum, Praia, Duna e Areal tem uma grande importância, por possuírem extensas áreas, de grande importância econômica e social, sendo as de Mangue e de Areia (lençóis maranhenses), algumas das maiores do Brasil, segundo dados do MapBiomas (2023). Contudo, no painel de estatísticas de acurácia de diversas destas classes ainda não estavam publicadas. Conhecer essa acurácia é um passo importante para a utilização destes dados nesta região. Deste modo, este trabalho tem como objetivo avaliar a acurácia das classes de nível 2 na região geográfica intermediária de São Luís, oferecendo uma avaliação mais detalhada e específica para uma região de interesse.

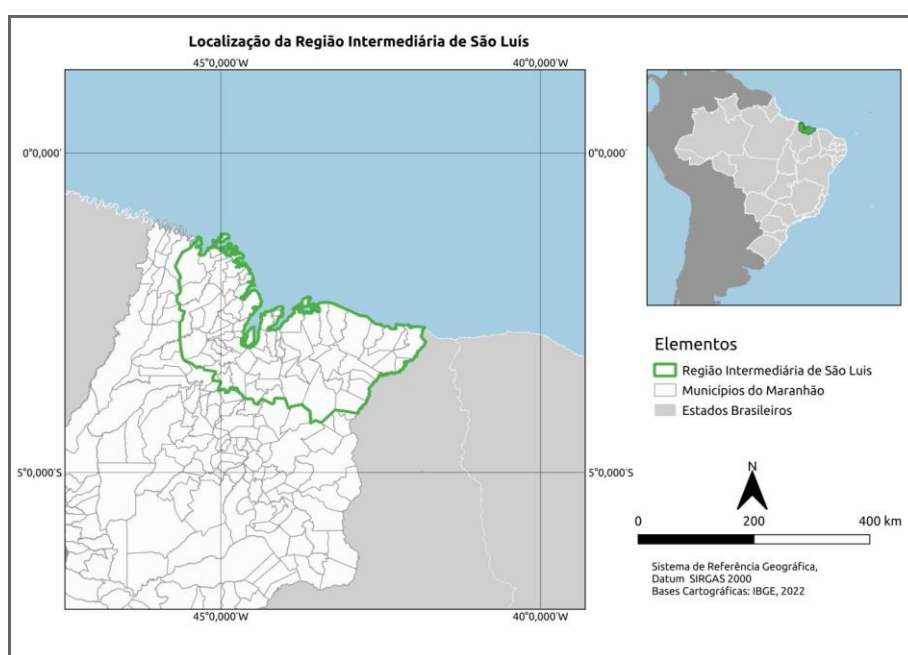
2. METODOLOGIA

2.1 ÁREA DE ESTUDO

A Região Geográfica Intermediária de São Luís (Figura 1) está entre as cinco regiões intermediárias do estado do Maranhão, no Brasil, se distingue por sua paisagem costeira diversificada, compreendendo extensas florestas de mangue, Praia e Duna, e é uma de um total de 134 destas regiões estabelecidas pelo Instituto Brasileiro de Geografia e Estatística (IBGE) em 2017. Abrangendo uma área total de cerca de 70.813 km², esta região inclui 73 municípios e apresenta ecossistemas únicos que sustentam a biodiversidade local.

Notavelmente, as praias, como o renomado Parque Nacional dos Lençóis Maranhenses, são mundialmente reconhecidas por seu fascínio natural, atraindo turistas de várias regiões do Brasil e de outros lugares. Além disso, as florestas de mangue desempenham um papel vital na proteção das costas e na manutenção da vida marinha, servindo como habitats para uma variedade de espécies (IBGE, 2020).

Figura 1 - Região Geográfica Intermediária de São Luís.



Fonte: Elaboração ou Organização: os autores, 2024.



2.2 CLASSES DE USO E COBERTURA DA TERRA

Os dados de uso e cobertura da terra utilizados neste trabalho foram obtidos a partir do MapBiomas, iniciativa que disponibiliza mapas anuais produzidos com base em imagens Landsat. A metodologia é fundamentada em algoritmos de aprendizado de máquina, em especial o classificador *random forest*, implementado na plataforma Google Earth Engine. Esse ambiente de processamento de grandes volumes de dados geoespaciais facilita a descoberta científica, ao fornecer acesso gratuito a extensos conjuntos de dados de sensoriamento remoto, além de garantir escalabilidade e eficiência computacional (Tamiminia *et al.*, 2020). A série histórica do MapBiomas tem início em 1985 e é continuamente expandida, pois novas coleções são lançadas anualmente, incorporando melhorias metodológicas e novos tipos de uso e cobertura. À época da análise, a versão mais recente disponível era a Coleção 8, que contempla 29 classes distintas de uso e cobertura da terra (MAPBIOMAS, 2024).

O sistema de classificação adotado pelo MapBiomas possui uma estrutura hierárquica, organizada em dois níveis. No primeiro nível, encontram-se categorias amplas, como Floresta, Formações Naturais Não Florestais, Agropecuária e Corpos D'água. No segundo nível, essas categorias são detalhadas em subclasses, como Pastagem, Agricultura e Silvicultura, entre outras (MAPBIOMAS, 2022).

Neste estudo, a análise considerou as classes de nível 2, apresentadas no Quadro 1, que oferecem maior detalhamento e são adequadas para investigar as transformações no uso e cobertura da terra nos diferentes biomas brasileiros.

Quadro 1 - Resumo das classes de nível 2 que serão analisadas.

Classe	Descrição
Formação Florestal	Inclui tipologias ombrófila, decidual e semidecidual, além de parte das formações pioneiras. Dossel contínuo, textura rugosa, em drenagens e encostas úmidas.
Formação Savânica	Abrange fitofisionomias do Cerrado, como Cerrado sentido restrito, Parque de Cerrado, Palmeiral e Vereda. Estrato arbóreo esparsa sobre estrato graminoso contínuo. Cor predominante verde acinzentado.
Mangue	Formação pioneira halófila e hidrófila em zonas entre marés em regiões estuarinas e lagunas costeiras tropicais e subtropicais. Porte arbóreo baixo, troncos e pneumatóforos expostos. Cor predominante verde escuro.
Floresta Alagável (beta)	São ecossistemas florestais localizados principalmente na Amazônia, que se desenvolvem ao longo de cursos d'água e são periodicamente ou permanentemente inundados. Elas incluem matas de várzea (rios de água branca) e matas de igapó (rios de água preta ou clara), com adaptações à inundações.
Restinga Arbórea	Formação florestal costeira que se desenvolve em solos arenosos ou dunas, abrangendo uma faixa que vai de Alagoas ao Rio Grande do Sul. Em 2022, 88% da área de restinga estava localizada na Mata Atlântica, enquanto 12% se encontrava no Pampa.
Silvicultura	Áreas com espécies arbóreas plantadas para fins comerciais como eucalipto, pinus e araucária. Caracterizada por árvores de porte alto e textura uniforme.



Campo Alagado	Formação pioneira com espécies herbáceas e arbustivas adaptadas a solos hidromórficos e sujeitas a inundações periódicas, ocorrendo em áreas alagadas e banhadas.
Formação Campestre	Inclui formações campestres com predominância de estrato herbáceo ou herbáceo-lenhoso e manchas com estrato arbustivo-herbáceo desenvolvido. Pode incluir áreas pastejadas nos biomas Pampa e Pantanal.
Pastagem	Áreas com gramíneas forrageiras cultivadas ou naturalizadas para pastejo. Apresenta textura lisa.
Agricultura	Áreas com cultivos anuais e, em algumas regiões, cultivos perenes. Caracteriza-se também por uma textura lisa.
Mosaico de Usos	Representa áreas com diferentes usos da terra, onde não é possível determinar um uso predominante. A textura varia conforme os usos presentes.
Praia, Duna e Areal	Depósitos arenosos no litoral, cor branca e textura aquosa. Pode apresentar formas curvas devido à ação do vento e associada às sombras.
Área Urbanizada	Superfícies não permeáveis como infraestrutura e expansão urbana ou mineração. Apresenta baixa rugosidade.
Outras Áreas não Vegetadas	Solo exposto, leito de rio seco e grandes edificações em zonas rurais. Caracteriza-se por baixa rugosidade.
Mineração	Áreas de destruição mineral com solo exposto e baixa cobertura vegetal. Apresenta textura lisa e respostas espectrais variadas.
Aquicultura	Áreas com lagos artificiais, atividades aquícolas e/ou de salicultura. Apresenta formas geométricas bem definidas e presença de água.
Apicum	Áreas de transição entre manguezais e terra firme, sujeitas a inundações periódicas pelas marés. Vegetação herbácea esparsa ou ausente, solo exposto.
Rio, Lago e Oceano	Corpos d'água naturais como rios, lagos, represas, reservatórios e oceanos. Apresenta baixa rugosidade e com pouca variação entre o período seco e chuvoso.

Fonte: Elaboração ou Organização: os autores, 2024.

2.3 PRÉ-PROCESSAMENTO E ANÁLISE DOS DADOS

O pré-processamento e a análise dos dados foram realizados em duas etapas principais: extração das amostras via Google Earth Engine (GEE) e posterior análise e visualização em Python. O primeiro passo envolveu a extração das amostras de referência do MapBiomas, adaptando o código oficial disponível no repositório MapBiomas/accuracy¹. A modificação principal foi a inclusão de um limite espacial específico à área de estudo, garantindo que apenas as amostras dentro dessa região fossem consideradas. Dentro do ambiente GEE, realizaram-se etapas de filtragem das amostras com base na área delimitada, exclusão de classes indesejadas ou inconsistentes, ajuste de probabilidades e pesos de amostragem para assegurar representatividade estatística por estrato, vinculação das amostras às classificações anuais do MapBiomas e exportação dos resultados em arquivos CSV diretamente pelo GEE.

Com os dados exportados do GEE, foi desenvolvido um notebook Python que organiza o fluxo de avaliação em etapas claras: aquisição e organização dos dados, limpeza e filtragem,

¹ O repositório oficial do Mapbiomas: <https://github.com/mapbiomas-brazil/accuracy>



reclassificação e agregação das classes de uso e cobertura da terra, cálculo de métricas de acurácia (como acurácia global, matriz de confusão e erros de alocação) e geração de visualizações em gráficos e mapas para interpretação dos resultados. O código completo está disponível em repositório público (LAMBDA GEO, 2025), garantindo transparência e reprodutibilidade das análises. Os notebooks seguem o conceito de Programação Literária (literate programming), integrando código, documentação e comentários em formato narrativo (Knuth, 1984), apoiando os princípios da ciência aberta, promovendo acesso livre a dados, métodos e resultados, e favorecendo a reprodutibilidade e descoberta científica (Murray-Rust, 2008).

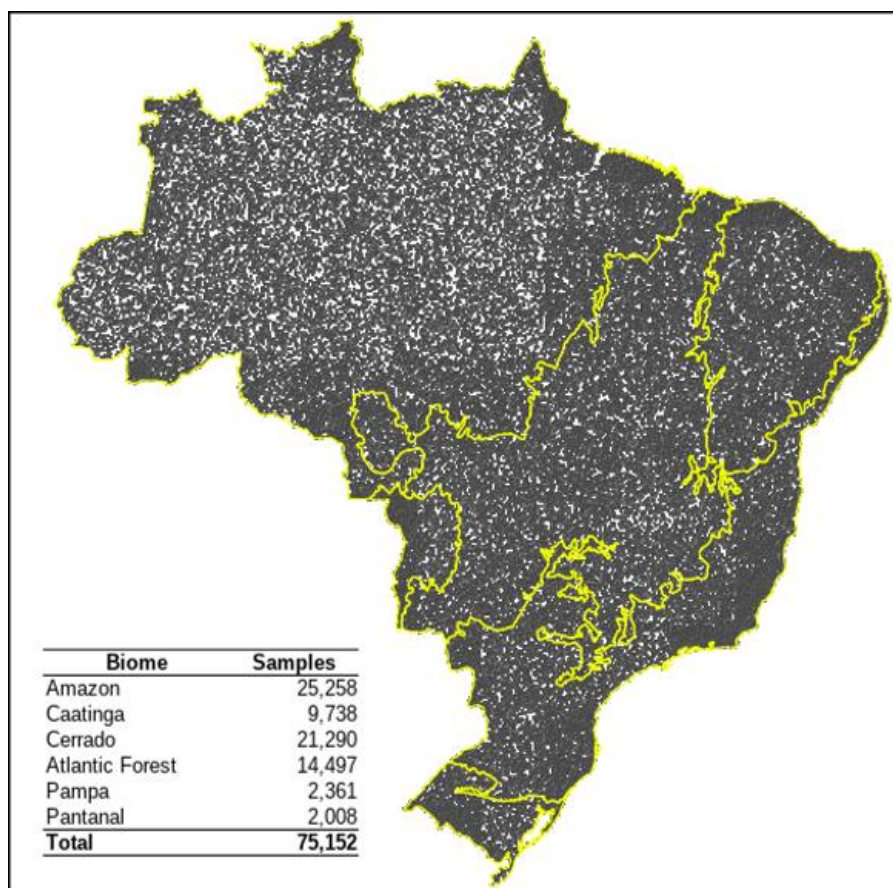
2.4 AVALIAÇÃO DE ACURÁCIA

O MapBiomias utiliza métricas de acurácia desenvolvidas para avaliar a qualidade de seus mapeamentos, fundamentando-se em alguns trabalhos (Pontius e Millones, 2011; Stehman *et al.*, 2014; Stehman e Fody, 2019). O Documento de Base Teórica do Algoritmo (ATBD) do projeto descreve todo o processo de desenvolvimento e produção dos mapas, incluindo a metodologia de avaliação de acurácia. Um estudo relevante neste contexto é o trabalho de Parente *et al.* (2021), no qual os autores analisaram a transformação da paisagem brasileira, destacando que a ocupação da terra, o desmatamento em larga escala e a rápida expansão agrícola desempenham papéis cruciais nesse processo. Essas observações são essenciais para garantir a confiabilidade dos dados espaciais, que são utilizadas em diversas aplicações, desde o monitoramento ambiental até o planejamento urbano.

Para avaliação de acurácia, o projeto MapBiomias gerou um conjunto de 85.000 amostras espacialmente independentes para todo o país. Destas, 10.000 amostras foram utilizadas como amostras de treinamento para o bioma Amazônia. Cada amostra foi avaliada por três intérpretes distintos, de modo que a classe final foi definida por uma votação majoritária (ou seja, a classe com dois ou três votos). Para amostras com discordância completa (ou seja, atribuição de três classes diferentes), a classe final foi estabelecida a *posteriori* por um quarto supervisor-intérprete (PARENTE *et al.*, 2021).



Figura 2 - Amostras aleatórias independentes.

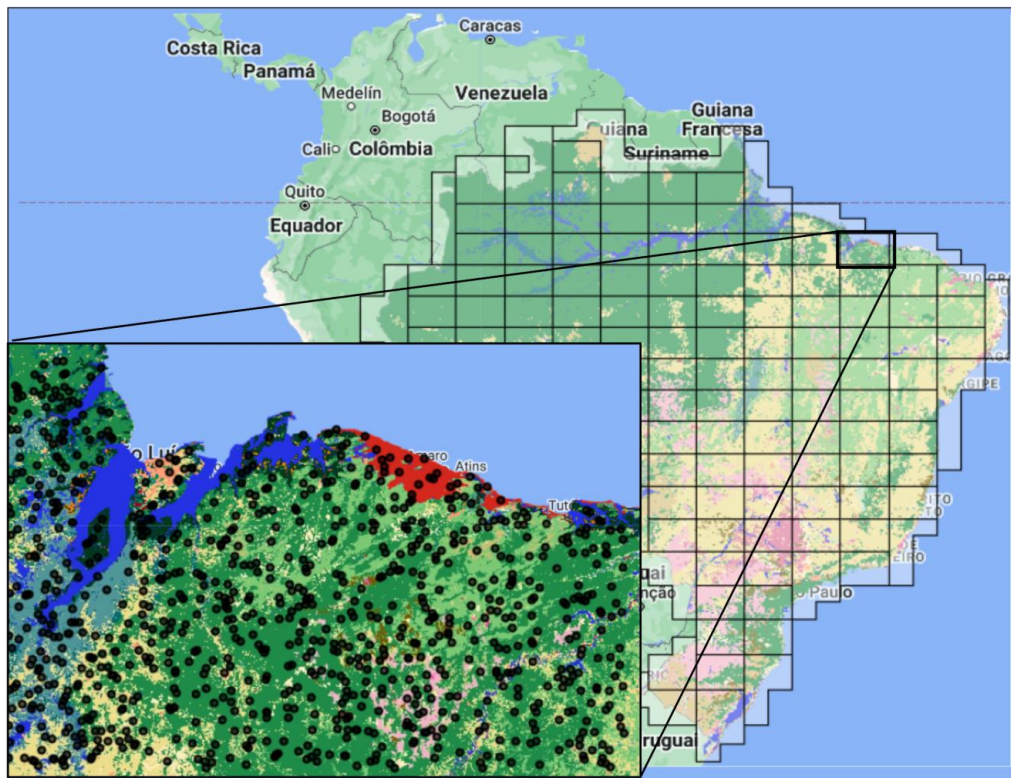


Fonte: Documento de Base Teórica do Algoritmo (ATBD), Coleção 8 (2023). Elaboração ou Organização: os autores, 2024.

De acordo com Parente *et al.* (2021), o número de amostras dentro de cada grupo de cartas do IBGE foi controlado de modo a garantir uma margem de erro máxima de 5% e um nível de confiança de 95%. Para manter esse nível de confiança, optou-se por utilizar integralmente um destes agrupamentos de cartas que cobrissem o maior percentual da área de estudo. Logo, para este trabalho será analisada a acurácia com base nas 712 amostras distribuídas no agrupamento de cartas SA-23-Z, como mostra a Figura 3. O agrupamento SA-23-Z engloba 72% da área de estudo, garantindo assim uma boa representatividade.



Figura 3 - Amostras espacialmente distribuídas no Grupo de Cartas SA-23-Z, visualizados no Google Earth Engine.



Fonte: MapBiomias (2022). Elaboração ou Organização: os autores, 2024.

Para a geração dos dados utilizados na avaliação de acurácia, foi adaptado o código disponibilizado pelo projeto MapBiomias, executado no Google Earth Engine. A adaptação teve como objetivo restringir as amostras à área do agrupamento SA-23-Z, modificando apenas as linhas responsáveis pelo filtro espacial.

O código resultou em uma planilha contendo os dados de referência e as classificações para qualquer ano entre 1985 e 2022. Para este trabalho, foram analisados apenas os dados referentes ao ano de 2022. A partir desta planilha foram calculadas algumas métricas de acurácia. Entre as principais, destaque-se a Acurácia Global, que estima a proporção de acertos em relação ao total de classificações, e a Acurácia do Consumidor, que mede uma fração de pixels mapeados para cada classe que está corretamente classificada, associando-se ao erro de comissão. Por outro lado, a Acurácia do Produtor refere-se à fração amostral de pixels de cada classe que foram corretamente atribuídas pelos classificados, estando relacionados ao erro de omissão. Tanto a acurácia do produtor quanto a do consumidor, ajudam a compreender como cada classe foi corretamente classificada.

O MapBiomias dá ênfase maior na Discordância de Quantidade e Discordância de Alocação, pois permitem uma análise mais aprofundada dos erros de classificação (Pontius e Millones, 2011). A Discordância de Quantidade refere-se à diferença entre as proporções de cada classe no mapa



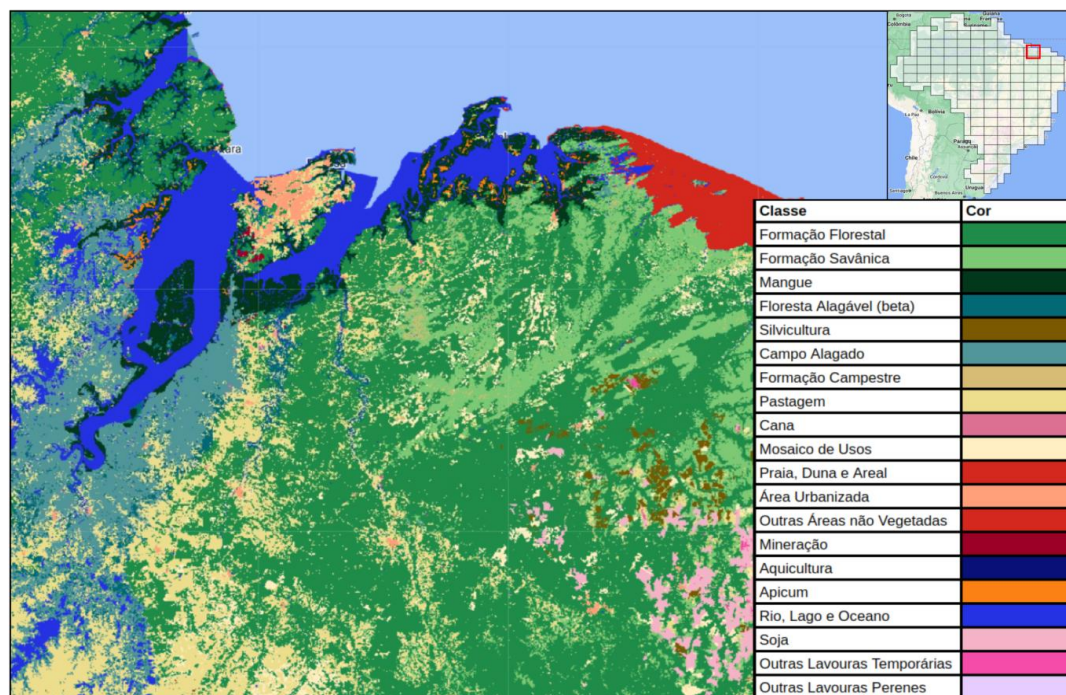
classificado e no mapa de referência, enquanto a Discordância de Alocação mede a diferença entre a proporção corretamente alocada e a proporção total discordante. Essa abordagem fornece uma compreensão mais rica das falhas na classificação digital, distinguindo entre erros de quantidade e alocação, e contribuindo para a melhoria contínua dos processos de mapeamento (Pontius e Millones, 2011).

Além dessas análises, o índice Kappa foi incorporado à análise. O Kappa mede a concordância entre o mapa classificado e o mapa de referência, oferecendo uma perspectiva adicional sobre a precisão da classificação. No entanto, é importante notar que Pontius e Millones (2011) criticam o uso do Kappa, apontando suas limitações e indicando que as métricas de discordância são mais informativas. Portanto, ao incluir o Kappa, buscamos uma abordagem mais abrangente, complementando as análises de discordância com a avaliação de concordância, permitindo uma avaliação mais robusta da acurácia dos mapeamentos.

3. RESULTADOS

Conforme descrito na metodologia, este estudo concentrou-se na avaliação da acurácia para o ano de 2022, considerando exclusivamente um agrupamento de cartas do IBGE entre as 127 definidas pelo MapBiomias. A partir de agora, esse agrupamento será referido como 'área de estudo', uma vez que todos os resultados apresentados se restringem a essa área. A Figura 4 apresenta as 20 classes de uso e cobertura da terra identificadas em todos os pixels da área de estudo.

Figura 4 - Classes de uso e cobertura encontradas na área de estudo.



Fonte: Elaboração ou Organização: os autores, 2024.



As 20 classes de uso e cobertura da terra foram resumidas em 18 classes de uso da terra no nível 2, com a classe Agricultura englobando: Cana, Soja, Outras Lavouras Temporárias e Outras Lavouras Perenes.

Dentre as 75 mil amostras de validação disponibilizadas pelo MapBiomias, 712 estão localizadas dentro da área de estudo. No nível 2, elas cobrem 17 diferentes classes de uso e cobertura da terra. A Tabela 1 sintetiza a informação sobre as classes de nível 2 identificadas na imagem e nas amostras.

Tabela 1 - Síntese das classes de nível 2.

ID	Classe	Quantidade de Amostras		Quantidade de Pixels
		Classificada	Referência	
3	Formação Florestal	290	300	27.903.480
4	Formação Savânica	104	117	9.358.030
5	Mangue	32	17	2.608.078
6	Floresta Alagável (beta)	29	0	2.076.737
9	Silvicultura	3	4	413468
11	Campo Alagado	59	55	4.610.585
12	Formação Campestre	12	26	794.098
15	Pastagem	92	106	7.939.154
18	Agricultura	13	13	1.071.574
21	Mosaico de Usos	18	10	2.124.698
23	Praia, Duna e Areal	14	14	1.106.692
24	Área Urbanizada	10	10	489.633
25	Outras Áreas não Vegetadas	5	5	335.708
30	Mineração	0	0	28328
31	Aquicultura	0	0	239
32	Apicum	5	3	213.594
33	Rio, Lago e Oceano	26	20	4.818.666
50	Restinga Herbácea/Arbustiva	0	12	0
Total		712	712	65.892.762

Fonte: Elaboração ou Organização: os autores, 2024.

Observa-se pela Tabela 1 que nenhuma amostra está localizada na área de Mineração e Aquicultura. Outro ponto a ser observado é com relação a classe Restinga Herbácea/Arbustiva, que existe nas amostras, contudo nessa região nenhum pixel foi classificado com este uso.

Com base nesses dados, o número de amostras foi reduzido para 655 durante o pré-processamento. Isso ocorreu porque foram removidas as classes com menos de 10 amostras de



referência; por exemplo, Silvicultura, que tinha apenas 4 amostras de referência. Seguindo o mesmo critério, as classes Apicum, Floresta Alagável e Outras Áreas Não Vegetadas também foram excluídas. Além disso, a classe Restinga Herbácea/Arbustiva foi removida por não conter nenhum pixel classificado, o que pode indicar que ela ainda não foi incluída no algoritmo de classificação. A Tabela 2 mostra a distribuição final das 655 amostras nas 11 classes consideradas para a análise de acurácia, após o pré-processamento.

Tabela 2 - Distribuição das 655 amostras em relação às 11 classes.

ID	Classe	Quantidade de Amostras	
		Classificada	Referência
3	Formação Florestal	289	282
4	Formação Savânica	93	117
5	Mangue	32	15
11	Campo Alagado e Área Pantanosa	59	47
12	Formação Campestre (Campo)	10	25
15	Pastagem	92	102
18	Agricultura	13	13
21	Mosaico de Usos	18	10
23	Praia e Duna	14	14
24	Área Urbanizada	9	10
33	Rio, Lago e Oceano	26	20
Total		655	655

Fonte: Elaboração ou Organização: os autores, 2024.

A partir dessas classes, foi elaborada a matriz de confusão para avaliar a acurácia. A Tabela 3 apresenta a matriz para as 11 classes consideradas neste estudo, com linhas e colunas rotuladas pelos códigos das classes. Em uma matriz de confusão, é importante observar a diagonal principal, pois ela indica o número de amostras que o algoritmo classificou corretamente com base na referência. Por exemplo, das 282 amostras da classe 3 (Formação Florestal), 225 foram classificadas corretamente. Em contraste, das 25 amostras da classe 12 (Formação Campestre), apenas 1 foi classificada corretamente.

**Tabela 3** - Matriz de confusão.

IDs		Referência											Total
		3	4	5	11	12	15	18	21	23	24	33	
	3	225	34	1	0	3	23	1	0	0	0	2	289
	4	12	69	0	0	4	8	0	0	0	0	0	93
	5	18	0	14	0	0	0	0	0	0	0	0	32
	11	8	0	0	38	7	6	0	0	0	0	0	59
	12	0	6	0	0	1	2	0	0	0	0	1	10
	15	16	4	0	3	6	61	0	0	0	2	0	92
	18	0	0	0	0	0	1	12	0	0	0	0	13
	21	1	4	0	0	3	0	0	10	0	0	0	18
	23	0	0	0	0	0	0	0	0	14	0	0	14
	24	0	0	0	0	0	1	0	0	0	8	0	9
	33	2	0	0	6	1	0	0	0	0	0	17	26
Total		282	117	15	47	25	102	13	10	14	10	20	655

Fonte: Elaboração ou Organização: os autores, 2024.

Neste trabalho, optou-se por apresentar na matriz de confusão o número de amostras ao invés da proporção. Essa abordagem nos ajuda a visualizar melhor as possíveis influências do número de amostras reduzidas de algumas classes nos indicadores que serão apresentados a seguir.

Para cada classe, a Tabela 4 apresenta as acurácias do consumidor e produtor e os respectivos erros.



Tabela 4 - Dados da acurácia.

ID	Classe	Erro		Acurácia	
		Comissão	Omissão	Consumidor	Produtor
3	Formação Florestal	22%	20%	78%	80%
4	Formação Savânica	26%	41%	74%	59%
5	Mangue	56%	7%	44%	93%
11	Campo Alagado e Área Pantanosa	36%	19%	64%	81%
12	Formação Campestre (Campo)	90%	96%	10%	4%
15	Pastagem	34%	40%	66%	60%
18	Agricultura	8%	8%	92%	92%
21	Mosaico de Usos	44%	0%	56%	100%
23	Praia e Duna	0%	0%	100%	100%
24	Área Urbanizada	11%	20%	89%	80%
33	Rio, Lago e Oceano	35%	15%	65%	85%

Fonte: Elaboração ou Organização: os autores, 2024.

A classe Formação Florestal apresentou uma alta acurácia do consumidor e produtor, respectivamente 78% e 80%. Essa alta acurácia teve um impacto positivo na acurácia global, dado que mais de 40% das amostras são desta classe.

Dentre as classes de interesse na área de estudo que ainda não estavam disponíveis no momento da escrita deste trabalho, destaca-se a classe Mangue. Esta classe apresentou um baixo erro de omissão, de apenas 7%, uma vez que o classificador acertou 14 das 15 amostras de Mangue. No entanto, apresentou um alto erro de comissão de 56%, pois 18 amostras que pertenciam à formação florestal foram incorretamente classificadas como Mangue. Esta classe é de grande importância para a área de estudo e necessita de uma análise mais aprofundada em trabalhos futuros, incluindo a incorporação de mais amostras e a revisão por especialistas com experiência na área de estudo.

Outra classe de interesse, cuja acurácia não estava disponível, é a Praia/Duna. Para esta classe, o algoritmo conseguiu classificar corretamente todas as amostras, alcançando 100% de acurácia tanto do ponto de vista do produtor quanto do consumidor. No entanto, é importante



observar que o número de amostras dessa classe é relativamente baixo, com apenas 14 entre as 655 amostras. Adicionalmente, destaca-se que estes pontos estão localizados nos lençóis maranhenses, que por suas características, apresentou baixo risco de confusão entre classes.

A classe Mosaico de Usos apresentou alto erro de comissão. Das 18 amostras classificadas com este uso, 8 não pertenciam a essa classe. Por outro lado, todas as 10 amostras classificadas com essa classe foram corretamente classificadas. Deste modo, a acurácia do produtor foi de 100%.

Por fim, a classe Formação Campestre apresentou um alto erro de comissão e omissão. Dado que o algoritmo acertou apenas uma amostra dentre as 25 e classificou erroneamente outras 9 amostras como pertencentes desta classe. A acurácia desta classe é pior dentre as classes de nível 2 que estão disponíveis no MapBiomas (2024). Em nível nacional, a acurácia do produtor é de aproximadamente 9%, segundo o MapBiomas (2024). É um valor acima dos 4% na área de estudo deste trabalho. Deste modo, esta é uma classe que deverá ser analisada em mais detalhes em trabalhos futuros.

A Tabela 5 apresenta um comparativo dos resultados para a área de estudo e os dados publicados no portal do MapBiomas para todo o Brasil. Importante destacar que essa comparação é apenas para ter uma referência, dado que eram esperadas diferenças. Primeiramente, por ser um estudo mais localizado e que em nível nacional, aparentemente, não foi incluído diversas classes de uso e cobertura da terra em nível 2. Adicionalmente, os resultados publicados no painel de estatísticas de acurácia do MapBiomas (2023) são relacionados à área estimada e seguem a metodologia proposta por Stehman (2014). O presente trabalho não incluiu a estimativa de área, todos os índices foram gerados a partir do quantitativo de amostras.

Tabela 5 – Comparativo dos resultados.

	Acurácia Global	Erro Global	Erros de Quantidade	Erros de Alocação	Índice kappa
Área de Estudo	72%	28%	8%	20%	0,62
Brasil	85.8%	14.2%	4.8%	9.4%	-

Fonte: Elaboração ou Organização: os autores, 2024.

A acurácia global na área de estudo foi de 72%, abaixo dos 85.8% em escala nacional. O erro global de 28% foi separado em erros de quantidade e erros de alocação. O erro de alocação na área de estudo foi muito superior ao esperado, enquanto nacional está um pouco acima de 9% na área de estudo foi de 20%.

Quanto ao índice Kappa, o projeto do MapBiomas optou por focar nos indicadores de discordância de quantidade e alocação. Deste modo, não é apresentado o índice Kappa devido



algumas críticas apresentadas em Pontius e Millones (2011). Porém calculou-se esse índice para a área de estudo, que apesar das suas limitações é ainda considerado melhor que a acurácia global simples, que só considera os pontos corretamente classificados (CONGALTON & GREEN, 1999). Para essa área de estudo, obteve-se um índice de 0,62, que, de acordo com Landis e Koch (1977) (Tabela 6), pode ser considerado como ótima.

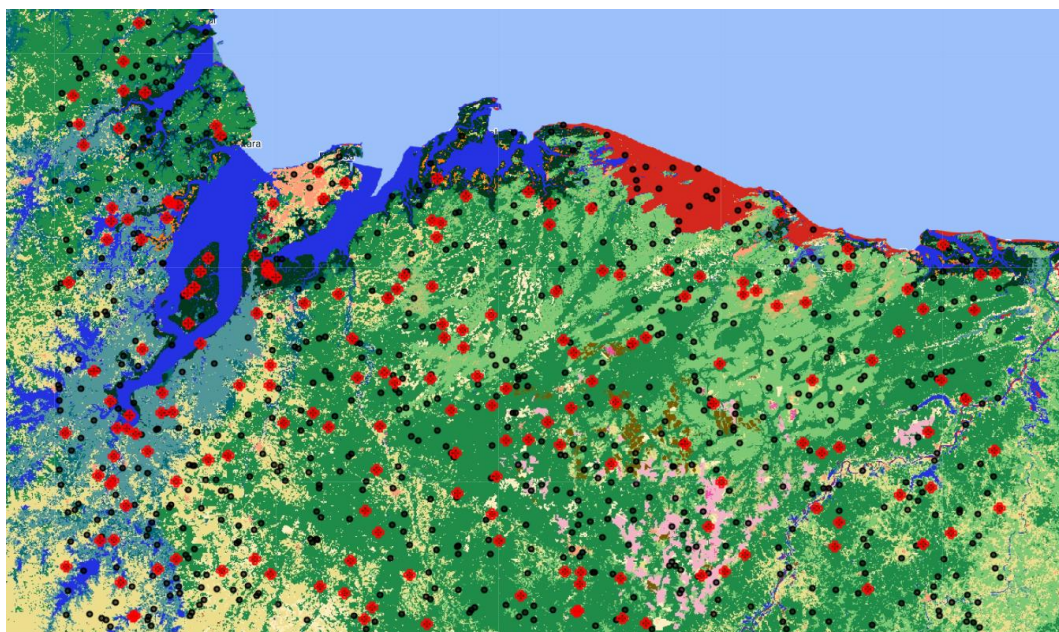
Tabela 6 - Interpretação do Índice Kappa

Índice Kappa	Interpretação
0.00 - 0.20	Baixa concordância (ou nenhum acordo)
0.21 - 0.40	Concordância moderada
0.41 - 0.60	Concordância substancial
0.61 - 0.80	Concordância ótima
0.81 - 1.00	Concordância quase perfeita

Fonte: Adaptado de Landis & Koch, 1977. Elaboração ou Organização: os autores, 2024.

Com relação ao erro, os 28% são resultado dos 183 pontos amostrais onde ocorreram discordância entre o classificado e a referência. A Figura 5 apresenta visualmente estes pontos, indicados por uma cruz vermelha. A análise visual tende a indicar que a discordância ocorreu de modo aleatório no espaço, ou seja, não estão concentradas em nenhuma sub-região espacial.

Figura 5 - Visualização dos pontos de discordância.



Fonte: Elaboração ou Organização: os autores, 2024.

5. CONCLUSÃO

Este trabalho avaliou a acurácia das classes de nível 2 em um agrupamento de cartas do IBGE, cobrindo mais de 70% da área de estudo. A análise não só forneceu informações sobre a acurácia para uma área específica, mas também incluiu classes não presentes no painel de estatísticas de acurácia do MapBiomias, como Mangue e Praia e Duna.



A classe Mangue apresentou um alto erro de comissão: das 32 amostras classificadas como Mangue, 18 eram, na verdade, Formação Florestal. Esse resultado destaca o desafio significativo de classificar corretamente essa cobertura, como também evidenciado por diversos estudos na literatura, indicando a necessidade de pesquisas futuras. Além disso, seria importante incluir a acurácia da classe Apicum, que não pôde ser avaliada neste trabalho.

Quanto à classe Praia e Duna, o algoritmo obteve 100% de acurácia tanto do ponto de vista do produtor quanto do consumidor. Contudo, é relevante observar que as 14 amostras estavam localizadas na região dos Lençóis Maranhenses, o que pode ter contribuído para essa alta acurácia.

Em relação ao erro global, a área de estudo apresentou um erro de 28%, com um erro de alocação de 20%, superior aos 9,4% publicados no painel de estatísticas de acurácia do MapBiomas (2023). No entanto, apesar do erro superior ao publicado, a classificação mostrou uma acurácia geral satisfatória, com um índice Kappa de 0,62.

Em resumo, a classificação realizada pelo MapBiomas demonstra uma boa acurácia mesmo em níveis locais, embora seja inferior à acurácia observada em escala nacional. Essa discrepância deverá ser investigada em estudos futuros, que devem gerar novas amostras e validá-las com especialistas na área de estudo. Esse conjunto de amostras deve ser elaborado para representar melhor a acurácia das classes em áreas costeiras, como Mangue, Apicum e Praia e Duna.

REFERÊNCIAS

AZEVEDO, T.; MATIAS, L. F. Dinâmica da alteração do uso e ocupação agrícola na Amazônia Maranhense: uma análise a partir de dados do MapBiomas. *In: ENCONTRO EM COMEMORAÇÃO AOS 20 ANOS DO PROGRAMA DE PÓS-GRADUAÇÃO EM GEOGRAFIA (IG-UNICAMP)*, 1., 2022, Campinas. **Anais do Evento em Comemoração aos 20 anos do Programa de Pós-Graduação em Geografia**. Campinas: IG-UNICAMP, v. 1, n. 1, p. 89-103, 2022.

BARBOSA DE SOUZA, K.; ROSA DOS SANTOS, A.; PEZZOPANE, J. E. M.; DIAS, H. M.; FERRARI, J. L.; PELUZIO, T. M. O.; TOLEDO, J. V.; CARVALHO, R. C. F.; MOREIRA, T. R.; ARAÚJO, E. F.; *et al.* Modelagem da dinâmica de uso e cobertura da terra e sua projeção futura para o bioma Amazônia. **Florestas**, v. 14, n. 7, p. 1281, 2023. DOI: <https://doi.org/10.3390/f14071281>.

CONGALTON, R. G.; GREEN, K. *Assessing the accuracy of remotely sensed data: principles and practices*. Boca Raton: Lewis Publishers, 1999.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Cidades e estados: Maranhão**. Brasília, DF: IBGE, 2024. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/ma.html>. Acesso em: 23 jul. 2024.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Estimativas da população**. Brasília, DF: IBGE, 2024. Disponível em: <https://www.ibge.gov.br/estatisticas/sociais/populacao>. Acesso em: 23 jul. 2024.



INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Manual técnico de uso da terra**. 3. ed. Rio de Janeiro: IBGE, 2013. Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv81615.pdf>. Acesso em: 2 ago. 2024.

KNUTH, D. E. Literate Programming. **The Computer Journal**, v. 27, n. 2, p. 97–111, 1984. DOI: 10.1093/comjnl/27.2.97.

LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p. 159–174, 1977. DOI: <https://doi.org/10.2307/2529310>.

MAPBIOMAS. **Códigos de legenda**. São Paulo: MapBiomias, 2024. Disponível em: <https://brasil.mapbiomas.org/codigos-de-legenda/>. Acesso em: 24 jul. 2024.

MAPBIOMAS. **Coleções MapBiomias**. São Paulo: MapBiomias, 2024. Disponível em: <https://brasil.mapbiomas.org/colecoes-mapbiomas/>. Acesso em: 24 jul. 2024.

MAPBIOMAS. **Projeto MapBiomias – Coleção 8 da Série Anual de Mapas de Cobertura e Uso da Terra do Brasil, período 1985-2022**. São Paulo: MapBiomias, 2023. Disponível em: <http://brasil.mapbiomas.org/>. Acesso em: 20 jul. 2024

MAPBIOMAS. **Visão geral da metodologia**. São Paulo: MapBiomias, 2024. Disponível em: <https://brasil.mapbiomas.org/visao-geral-da-metodologia/>. Acesso em: 24 jul. 2024

MURRAY-RUST, P. Open data in science. *Nature Precedings*, p. 1–1, 2008.

PARENTE, L.; MATTOS, A. P.; BAUMANN, L.; *et al.* Shaping the Brazilian landscape: a process driven by land occupation, large-scale deforestation, and rapid agricultural expansion. **Research Square**, Preprint, 2021. DOI: <https://doi.org/10.21203/rs.3.rs-819697/v1>.

PONTIUS, R. G.; MILLONES, M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. **International Journal of Remote Sensing**, v. 32, n. 15, p. 4407–4429, 2011. DOI: <https://doi.org/10.1080/01431161.2011.552923>.

ROCHA, Thomas Victor de Sousa Malheiros; COSTA, Sérgio Souza; CORREIA, Luís Fernando Cirqueira da Silva; BEZERRA, Denilson da Silva. **AValiação da Acurácia na Classificação de Uso e Cobertura da Terra: Um Estudo de Caso na Região Intermediária de São Luís com o MapBiomias**: código-fonte. GitHub, 2024. Disponível em: https://github.com/LambdaGeo/acuracia_slz. Acesso em: 22 jul. 2024.

SOUZA, C. M., Jr.; ROSA, M. R.; PARENTE, L. L.; ALENCAR, A.; RUDORFF, B. F. T.; HASENACK, H.; MATSUMOTO, M.; FERREIRA, L. G.; SOUZA-FILHO, P. W. M.; *et al.* (2020). Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. **Remote Sensing**, 12(17), 2735. <https://doi.org/10.3390/rs12172735>.

STEHMAN, S. V. Estimating area and map accuracy for stratified random sampling when the strata are different from the map classes. **International Journal of Remote Sensing**, v. 35, n. 13, p. 4923–4939, 2014. DOI: <https://doi.org/10.1080/01431161.2014.930207>.

TAMIMINIA, H.; SALEHI, B.; MAHDIANPARI, M.; QUACKENBUSH, L.; ADELI, S.; BRISCO, B. Google Earth Engine for geo-big data applications: a meta-analysis and systematic



review. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 165, p. 15-27, 2020. DOI: <https://doi.org/10.1016/j.isprsjprs.2020.04.001>.

TURNER, M.; ROMME, W. Landscape dynamics in crown fire ecosystems. **Landscape Ecology**, v. 9, p. 59–77, 1994. DOI: <https://doi.org/10.1007/BF00135079>.