

MODELAGEM DA SITUAÇÃO URBANO-RURAL DO USO E COBERTURA DA TERRA POR REGRESSÃO LOGÍSTICA: O CASO DO DISTRITO FEDERAL NO CENSO BRASILEIRO 2022

Glaucia Guimarães Pereira

Universidade de Brasília, Centro de Desenvolvimento Sustentável, DF, Brasil
glauciagp23@gmail.com

Dácio José Cambraia Filho

Universidade de Brasília, Centro de Desenvolvimento Sustentável, DF, Brasil
daciocambraia@hotmail.com

Diego de Almeida Paim

Universidade de Brasília, Instituto de Geociências, DF, Brasil
diego.paim@aluno.unb.br

Roberto Mandetta Gandara

Sem filiação

rgandara@gmail.com

Gustavo Macedo de Mello Baptista

Universidade de Brasília, Instituto de Geociências e Centro de Desenvolvimento Sustentável, DF, Brasil
gmbaptista@unb.br

RESUMO

Até o Censo 2010, a situação urbano-rural era definida por limites administrativos e legislações municipais. Para o Censo 2022, o IBGE utilizou imagens orbitais de alta resolução na predefinição da situação, em campo pelos pesquisadores, integrando conhecimento empírico ao sensoriamento remoto. O presente estudo teve como objetivo modelar a situação urbano-rural a partir de categorias de uso e cobertura da terra provenientes de subprodutos de sensoriamento remoto. Tendo como área de análise o Distrito Federal, foi utilizada a regressão logística com a situação urbano-rural como variável dependente e percentuais de áreas de vegetação, agropecuária e não vegetada provenientes do MapBiomass como as variáveis independentes. O modelo desenvolvido, que teve acurácia de 97,3%, mostra que o percentual de área não vegetada tem o maior peso para a tipificação de um setor como urbano. Os resultados podem ser usados para prever a situação de áreas até a realização do próximo Censo, bem como para se planejar pesquisas amostrais em regiões que apresentam mudança no uso e cobertura da terra. O texto também discute a viabilidade de se usar aspectos do ambiente para previsão, em contraponto às práticas que tradicionalmente se valem de aspectos sociais para a definição da situação urbano-rural.

Palavras-chave: Sensoriamento remoto. Assentamentos humanos. Ordenamento territorial. MapBiomass. Geoprocessamento.

MODELING THE URBAN-RURAL CLASSIFICATION OF LAND USE AND LAND COVER USING LOGISTIC REGRESSION: THE CASE OF THE FEDERAL DISTRICT IN THE 2022 BRAZILIAN CENSUS

ABSTRACT

Until the 2010 census, Brazil's urban-rural classification was based on administrative boundaries and local legislation. For the 2022 census, the Brazilian Institute of Geography and Statistics (IBGE) used high-resolution satellite imagery for the preliminary classification. Researchers then refined this classification in the field by integrating empirical knowledge with remote sensing. This study aims to model the urban-rural classification using land use and land cover categories derived from remote sensing products. Focusing on the Federal District, we used logistic regression, treating the urban-rural classification as the dependent variable and the proportions of vegetated, agricultural, and non-vegetated areas from the MapBiomass dataset as the independent variables. The model achieved an accuracy of 97.3%, with the proportion of non-vegetated areas showing the strongest association with urban classification.

These findings can be used to estimate urban-rural status between censuses and support the planning of sample-based surveys in regions undergoing land use and land cover changes. This approach relies solely on data already available from MapBiomas. Additionally, the paper discusses the feasibility of using environmental characteristics for prediction instead of traditional approaches that primarily rely on social criteria to define urban and rural areas.

Keywords: Remote sensing. Human settlements. Territorial planning. MapBiomas. Geoprocessing.

INTRODUÇÃO

A distinção entre territórios urbanos e rurais é útil para orientar políticas públicas, distribuir recursos e planejar infraestruturas e serviços. Os estudos e debates sobre ordenamento do território não são novos; no entanto, a utilização de dados de sensoriamento remoto suscita aspectos ainda pouco explorados. Até o recenseamento de 2010, o Instituto Brasileiro de Geografia e Estatística (IBGE) utilizava os limites administrativos e a legislação de ordenamento do território para definir a classificação urbano-rural dos setores censitários. No Censo de 2022, foi incorporada à metodologia existente uma nova abordagem que utiliza imagens orbitais de alta resolução (IBGE, 2024).

A situação urbano-rural é definida como a categorização dicotômica (urbana ou rural) atribuída aos setores censitários pelo IBGE no processo de elaboração da malha territorial para a coleta de dados, em especial para o Recenseamento Geral da População e Habitação. No Recenseamento Geral de 2022, a situação urbano-rural é determinada pelo agrupamento de informações detalhadas: a situação urbana é composta por área urbana de alta densidade de edificações, área urbana de baixa densidade de edificações e núcleo urbano, ao passo que a situação rural é composta por povoado, núcleo rural, lugarejo, área rural (exceto aglomerados) e massas de água (IBGE, 2024).

Uma vez que as situações urbano-rurais são elaboradas com base em procedimentos analíticos de natureza morfológico-funcional que consideram aspectos políticos, econômicos, sociais e ambientais relacionados com os agregados estatísticos (IBGE, 2024), é importante que a metodologia da Base Territorial seja atualizada regularmente. De acordo com Souza et al. (2024), a revisão metodológica da Base Territorial efetuada para o Censo de 2022 visou harmonizar aspectos legais, conceituais e operacionais, mediante a aplicação de novos procedimentos metodológicos e tecnologias. Desta forma, mostrou-se necessária a atualização dos setores censitários face à sua situação urbano-rural, sobretudo nas localidades afetadas pela defasagem dos perímetros definidos anteriormente.

Para viabilizar esta atualização, foram utilizadas imagens de satélite de alta resolução, associadas ao conhecimento empírico dos analistas técnicos do IBGE, contribuindo para um processo de validação cruzada entre a observação de campo e os subprodutos de sensoriamento remoto. No entanto, tanto a nota técnica sobre a atualização da malha censitária como os artigos que descrevem os procedimentos metodológicos (IBGE, 2024; Souza et al., 2024) não apresentam pormenores técnicos sobre as características das imagens utilizadas, os processamentos aplicados e a sistematização dos elementos-chave utilizados na interpretação das imagens de satélite.

No contexto do Censo de 2010, Pera e Bueno (2016) já tinham identificado que as informações contidas nos cadernos metodológicos do IBGE não eram suficientes para compreender o método utilizado naquela ocasião e solicitaram mais informações ao setor técnico do instituto. As autoras concluíram que a elaboração das situações detalhadas tem um caráter híbrido, baseando-se tanto em legislação urbanística como na realidade existente, mas o IBGE não explicou como e em que municípios a realidade existente foi utilizada. No que se refere à utilização de imagens orbitais no Censo de 2022, a principal inovação parece residir na análise das imagens antes das vistorias no terreno. No entanto, não foi descrito como é feita a validação no terreno das características coletadas nas imagens.

É importante salientar que as discussões sobre o que é urbano e rural, ou cidade e campo, baseiam-se geralmente em características sociais, como a população, a densidade populacional e as relações de trabalho (Araújo; Soares, 2009; Santos; Silva, 2020). No entanto, ao incluir imagens orbitais e produtos resultantes de sensoriamento remoto, é possível incluir aspectos ambientais, bem como as características de utilização e cobertura do solo, na distinção entre urbano e rural. Neste contexto, o MapBiomas (Projeto Brasileiro de Mapeamento Anual de Uso e Cobertura do Solo) fornece informações sobre a dinâmica histórica destas transformações e os respectivos impactos nos biomas do país. Para

esse efeito, baseia-se em dados de sensoriamento remoto e séries temporais que têm sido amplamente utilizados para apoiar o desenvolvimento de projetos governamentais e iniciativas voltadas para a gestão territorial e ambiental (MapBiomass, 2015; Souza et al., 2020).

Não há consenso na literatura sobre a definição e a medição dos limites e da expansão dos assentamentos humanos (Dadashpour; Shahhossein, 2024), dificultando a realização de comparações e a elaboração de políticas públicas. Em países como Espanha, Portugal, Itália e Grécia, por exemplo, são consideradas rurais as localidades que, embora apresentem assentamentos humanos com mais de 10 mil habitantes, se encontram a uma distância considerável dos centros metropolitanos e das respectivas áreas de influência direta. Por outro lado, países como a França, a Argentina, a Bolívia e o México adotam critérios mais rígidos, considerando rurais as comunidades com menos de 2,5 mil habitantes (Abramovay, 2000).

No Brasil, a situação urbano-rural é o padrão adotado para a elaboração de estatísticas e políticas públicas. Estas definições são importantes para compreender os processos de urbanização, dado que, em 2010, 84,4% da população vivia em setores censitários classificados como urbanos, proporção que aumentou para 85,9% no censo de 2022 (IBGE, 2024). Por outro lado, Braga, Remoaldo e Fiúza (2014) aplicaram a metodologia definida pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE), que tem em conta a densidade populacional das regiões e dos seus centros urbanos, e concluíram, com base nos dados do recenseamento de 2010, que 87,5% da população do Brasil vivia em zonas urbanas naquele ano.

Skorupka (2017) propõe uma alteração à metodologia da OCDE, incluindo um terceiro tipo de território: as áreas naturais, locais com densidade populacional nula, presença de florestas, lagos, montanhas e pântanos, e um nível insignificante de intervenção humana. A autora propõe distinguir as áreas naturais das áreas rurais, reconhecendo que se trata de territórios com diferentes intenções e interações humanas. Tendo em conta que o principal objetivo deste campo de estudo é compreender os processos multidimensionais e os impactos dos assentamentos humanos (Dadashpour; Shahhossein, 2024), verifica-se necessário ir além da utilização da população ou da densidade populacional para distinguir o urbano do rural e considerar aspectos do ambiente natural e construído. A inovação consistiria na utilização de dados de sensoriamento remoto sobre o uso e a cobertura do solo, ou seja, em conjunto com a valorização do trabalho de validação de campo já realizado pelo IBGE, permitindo fazer previsões sobre a situação urbano-rural sem exigir novos esforços de campo.

Neste sentido, o presente trabalho tem como objetivo modelar a situação urbano-rural com base em categorias de utilização e cobertura do solo provenientes de subprodutos de observação da Terra. Para esse efeito, foi utilizada a regressão logística, que fornece a probabilidade de um setor censitário ser classificado como urbano ou rural, com base nas categorias de uso e cobertura da terra do MapBiomass (2024).

Do ponto de vista socioambiental, este estudo justifica-se, visto que a situação das áreas urbanas ou rurais influencia diretamente a formulação e a aplicação de políticas públicas, afetando a afetação de recursos, a infraestrutura e o acesso a serviços essenciais. A Agenda 2030, por meio do Objetivo de Desenvolvimento Sustentável 11 (ODS 11), destaca a importância de tornar as cidades e os assentamentos humanos inclusivos, seguros, resilientes e sustentáveis, com base no acompanhamento de indicadores e metas (Brasil, 2025). No entanto, a efetividade dessas metas depende da definição dos limites urbanos e rurais, uma vez que essa diferenciação influencia indicadores-chave, a alocação de recursos e as estratégias de planejamento territorial. A avaliação ex ante destas políticas permite analisar os seus potenciais impactos e mitigar distorções no desenvolvimento territorial, assegurando uma maior eficiência nos investimentos (Casa Civil; Ipea, 2018). A complexidade desta discussão ganha ainda mais relevância face aos desafios enfrentados pela sociedade contemporânea, como o acelerado processo de urbanização e os impactos das alterações climáticas, que aumentam a vulnerabilidade e os riscos, sobretudo em áreas com menor capacidade de adaptação (Dodman et al., 2023).

Do ponto de vista técnico-acadêmico, a novidade da utilização de imagens de satélite para o delineamento da recolha de dados pelo IBGE justifica um estudo para discutir abordagens indiretas e, por vezes, mais econômicas em termos de recolha e processamento de dados. Em termos de aplicação prática, o estudo pode contribuir para a identificação de uma área como urbana ou rural com base em dados de sensoriamento remoto, facilitando a elaboração e execução de inquéritos até o próximo recenseamento.

O trabalho está dividido em ‘Material e métodos’, sendo apresentados a área de estudo, os dados utilizados e as técnicas empregadas para a elaboração da regressão logística; em ‘Resultados e discussão’ são apresentadas as análises de dados, a modelagem e a discussão sobre os achados; o trabalho encerra-se com as considerações finais e perspectivas futuras.

MATERIAL E MÉTODOS

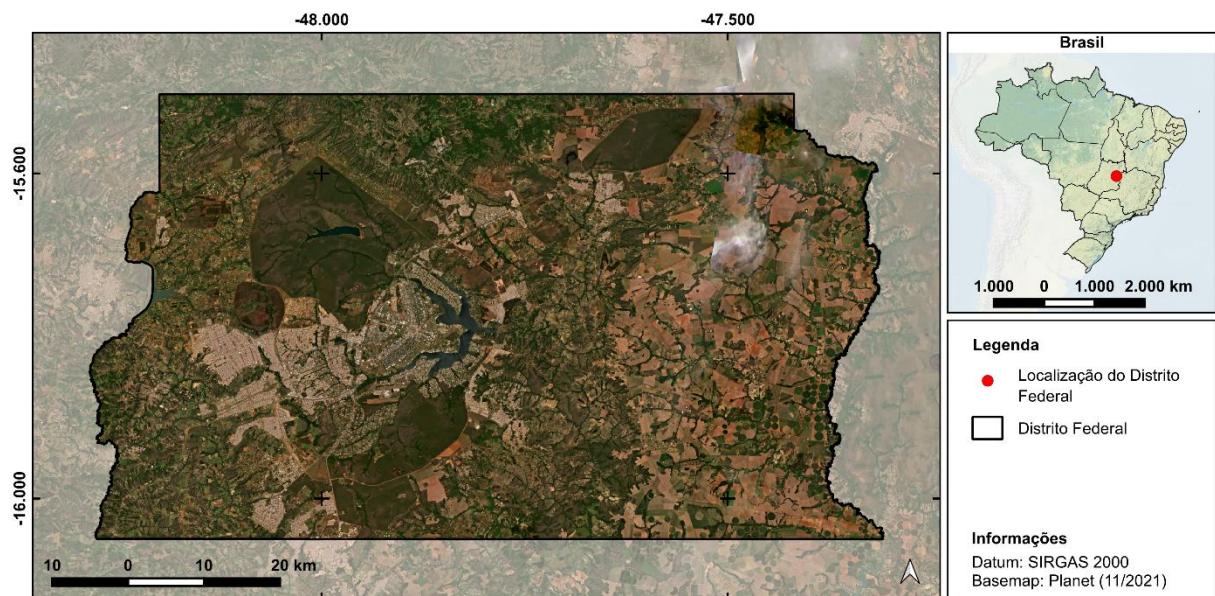
Este estudo adota uma abordagem quantitativa e um desenho não experimental, segundo a definição de Creswell e Creswell (2017). Em termos de métodos de investigação, caracteriza-se como uma investigação descritiva, de acordo com Gil (2002), uma vez que se concentra na descrição de um fenômeno e no estabelecimento da relação entre variáveis. No que se refere às práticas de Sistemas de Informação Geográfica (SIG), são utilizadas planimetrias e assinaturas, considerando-se o SIG um hiperespaço heurístico que permite obter informações empíricas sobre possíveis associações causais entre as variáveis ambientais (Xavier-da-Silva; Carvalho-Filho, 1993).

Área de estudo

Segundo o IBGE (2024) e Souza et al. (2024), a atualização da base cartográfica para o Recenseamento Geral da População e Habitação de 2022 foi conduzida de forma descentralizada, com a participação ativa das Superintendências do IBGE e das respectivas agências. Segundo os autores, este modelo permitiu uma representação mais precisa das referências geográficas que apoiam as operações de campo dos censos e inquéritos amostrais, garantindo uma correta identificação das estruturas territoriais através da integração de registros administrativos, imagens de sensoriamento remoto, interpretação da legislação e informações recolhidas no terreno.

Neste contexto, o Distrito Federal destaca-se como uma unidade federativa singular, na medida em que não possui subdivisões municipais e apresenta uma estrutura administrativa com competências legislativas reservadas aos estados e municípios (Brasil, 1988). Esta característica pode favorecer a adoção de abordagens mais homogêneas na definição da situação urbano-rural. Desta forma, para reduzir as fontes de variância no modelo proposto, optou-se por delimitar o Distrito Federal como área de estudo, conforme apresentado na Figura 1.

Figura 1 - Distrito Federal, localização da área de estudo



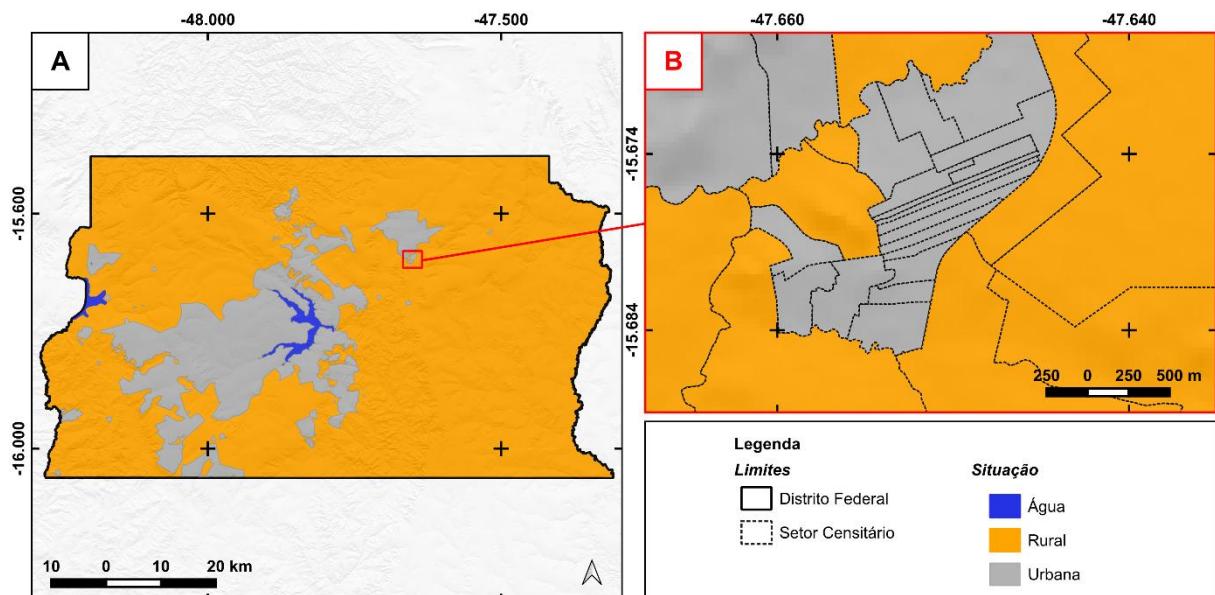
Fontes: IBGE, 2023; Planet, 2025. Elaboração: Os autores, 2025.

Com uma área de 5.761 km² e uma população estimada de 2.817.381 habitantes (IBGE, 2025), o Distrito Federal está localizado na região Centro-Oeste e está dividido em 35 regiões administrativas (SEGOV, 2025). Inserido no bioma do Cerrado, apresenta uma dinâmica espacial marcada pela intensa urbanização e crescimento populacional acelerado, impulsionados pelo desenvolvimento da capital federal. A sua área urbana expandiu-se cerca de 2983% entre 1960 e 2019, passando de 23 km² para 709 km², com um padrão de crescimento orientado do centro para o sudoeste e, em menor escala, para nordeste. A variação da taxa de crescimento ao longo dos últimos 60 anos aponta para uma aceleração da urbanização, com destaque para o crescimento a partir de 2013, o que indica que o Distrito Federal está longe de atingir a estagnação ou a diminuição do crescimento, conforme seria de esperar na maturidade dos centros urbanos (Chelotti; Sano, 2023).

Dados utilizados

A malha dos setores censitários do Distrito Federal utilizada no recenseamento de 2022 foi obtida a partir do repositório oficial do IBGE (IBGE, 2022). A Figura 2-A mostra a situação urbano-rural dos setores, ao passo que a Figura 2-B apresenta uma visualização detalhada dos limites dos setores censitários numa localidade específica. A malha é composta por 5.418 setores censitários, dos quais 5.021 são urbanos, 393 rurais e quatro sem situação urbano-rural definida. Estes últimos, por se encontrarem inseridos em massas de água e por não terem as respetivas situações censitárias, foram desconsiderados no modelo de análise. É importante salientar que esta malha foi atualizada em períodos anteriores à data de início do Recenseamento Geral de 2022, tendo sido adotado o ano de 2021 como referência temporal para o presente estudo (Souza et al., 2024).

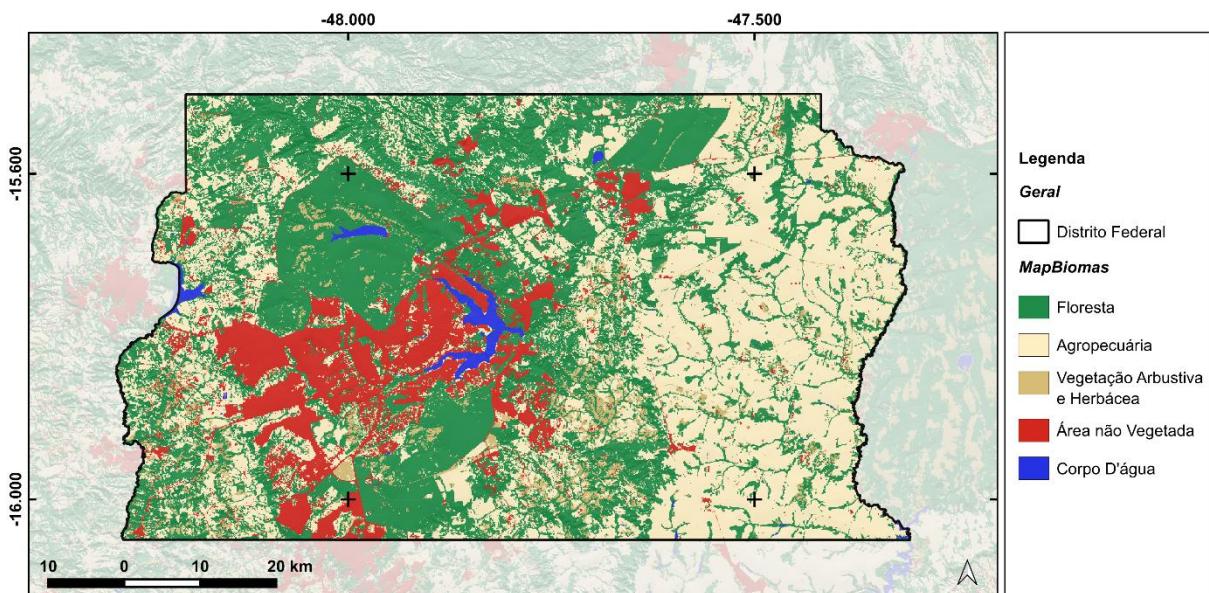
Figura 2 - Malha censitária do Distrito Federal com (A) Situação urbano-rural dos setores censitários; e (B) Representação detalhada dos limites dos setores censitários



Fontes: IBGE, 2022; 2023. Elaboração: Os autores, 2025.

Os dados de uso e cobertura da terra (Figura 3) foram adquiridos a partir da plataforma MapBiomas para o ano de 2021, especificamente a Coleção 9 com 30 metros de resolução espacial (MapBiomas, 2024; Souza et al., 2020). Foram consideradas somente as categorias correspondentes ao nível 1 de processamento da respectiva coleção, sendo: florestas, vegetação (formação natural não florestal), agropecuária, área não vegetada e corpos d'água.

Figura 3 - Uso e cobertura da terra do Distrito Federal em 2021



Fontes: MapBiomas, 2024; IBGE, 2023. Elaboração: Os autores, 2025.

Métodos

Para o processamento das bases de dados geoespaciais, foi utilizado o Sistema de Informação Geográfica (SIG) de código aberto QGIS 3.34.15 "Prizren" (QGIS, 2025). As bases de dados foram recortadas e ajustadas de acordo com os limites dos setores censitários, tendo sido posteriormente reprojetadas para o sistema de coordenadas Albers Equal Area Conic (Reference, 2025). Os dados de uso e cobertura da terra, originalmente em formato matricial, foram convertidos para o formato vetorial e unidos aos setores censitários para calcular a área das unidades vetoriais. Após a integração dos dados, estes foram exportados para um arquivo de tabela para tratamento em folha de cálculo. Neste ambiente, foram calculados as estatísticas descritivas e os testes de previsão do modelo.

Autores como Siddiqui et al. (2018) e Park et al. (2011) utilizam diferentes abordagens de modelação para analisar alterações no uso e na cobertura do solo com base em dados espaciais e algoritmos preditivos, incluindo a regressão logística. Este método estatístico, utilizado para modelar a relação entre uma variável dependente binária e um conjunto de variáveis independentes, foi utilizado neste estudo para estimar a probabilidade de um setor censitário ser classificado como rural ou urbano, com base na sua composição em termos de uso e cobertura do solo. Para esse efeito, a situação censitária foi adotada como variável dependente (0 para rural e 1 para urbano) e os percentuais das classes de utilização e cobertura do solo (vegetação, agropecuária e áreas não vegetadas) como variáveis independentes. Deve ser salientado que, devido à colinearidade perfeita entre estas variáveis, foi dedutivamente excluída a variável "floresta", por se considerar que, entre as variáveis, esta é a menos suscetível de ser confundida com área não vegetada. Esta decisão está em conformidade com Skorupka (2017), que se dedica a diferenciar as áreas naturais das áreas rurais, reconhecendo que ambas são distintas das áreas urbanas.

Para a regressão logística, o conjunto de dados foi dividido em amostras de treino e de validação, com 80% e 20%, respectivamente, tendo as análises sido conduzidas no programa informático RStudio 2024.12.0+467 (Posit, 2024), utilizando a linguagem de programação R (versão 4.4.2) (R Core Team, 2024). A importação dos dados foi realizada com recurso à biblioteca readxl (Wickham et al., 2023), ao passo que a análise exploratória foi efetuada com recurso à biblioteca GGally (Schlöerke et al., 2024). A modelação estatística baseou-se na biblioteca stats, utilizando o modelo de regressão logística através da função glm() (RStudio Team, 2019).

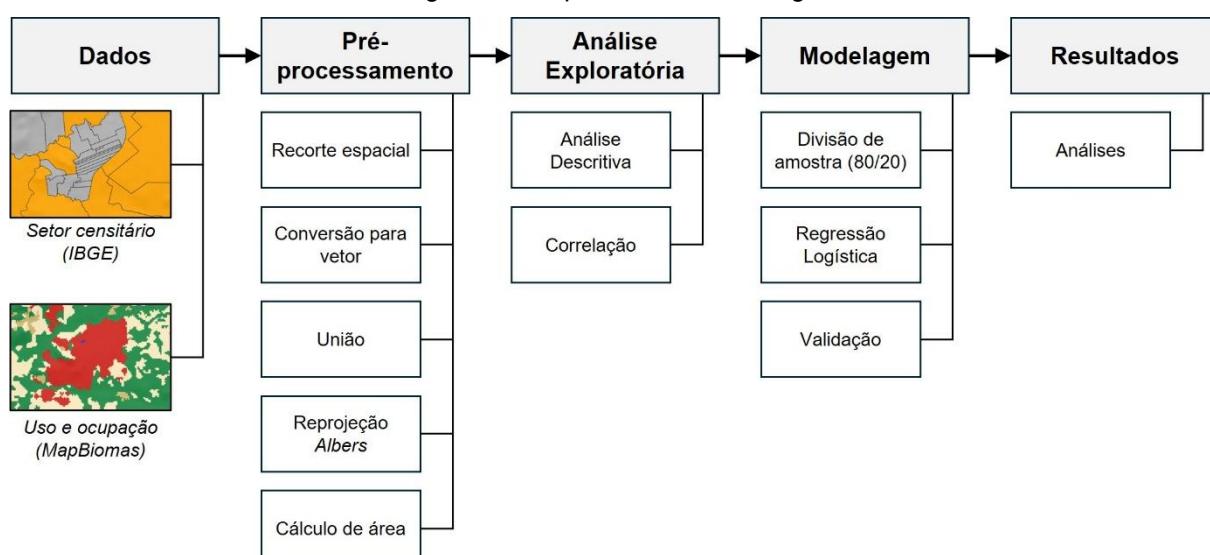
A avaliação da qualidade e do desempenho do modelo foi realizada com base em diferentes métricas estatísticas. Inicialmente, recorreu-se à biblioteca pscl (Jackman et al., 2024) para avaliar a qualidade do modelo, tendo sido extraídos os resultados do desvio nulo, do desvio residual e de uma adaptação

do coeficiente de determinação (R^2) de McFadden. Isto porque a regressão logística não possui uma única métrica de síntese da variância explicada pela variável dependente (Fernandes et al., 2020).

A validação do modelo, com um ponto de corte fixado nos 50%, incluiu a construção da matriz de confusão e o cálculo das métricas de acurácia geral, índice kappa, sensibilidade, especificidade, acurácia equilibrada e valores preditivos positivos e negativos, com recurso à biblioteca caret (Kuhn et al., 2024). Adicionalmente, o desempenho do modelo foi avaliado através da curva ROC (Receiver Operating Characteristic) e da área sob a curva (AUC) com a biblioteca pROC (Robin et al., 2023). Por fim, a visualização dos resultados foi elaborada com o apoio da biblioteca "ggplot2" (Wickham et al., 2024).

A Figura 4 apresenta um esquema da metodologia que ilustra as principais etapas do estudo, desde a aquisição e processamento dos dados geoespaciais, modelagem estatística e avaliação dos resultados.

Figura 4 - Esquema da metodologia



Elaboração: Os autores, 2025.

RESULTADOS E DISCUSSÃO

Esta seção apresenta a análise dos dados, a elaboração e a validação do modelo, conjuntamente com a discussão dos resultados.

Análise dos dados

De acordo com as diretrizes estabelecidas pelo IBGE (2024), a delimitação dos limites dos setores censitários tem em conta fatores preponderantes, como elementos da paisagem, pontos de referência para identificação e limites de estruturas territoriais. Por sua vez, o dimensionamento destes setores censitários visa cumprir critérios técnicos que equilibram a extensão territorial e a carga de trabalho dos recenseadores. Na prática, isto significa que a dimensão de cada setor está geralmente associada ao número de pessoas ou domicílios que os recenseadores precisam visitar durante o período de coleta de dados. A relação inversa entre a densidade populacional e a extensão territorial faz com que os setores urbanos com maior concentração de domicílios apresentem áreas menores, ao passo que os setores rurais, cuja população está mais dispersa, apresentem áreas mais extensas.

No Distrito Federal, os setores urbanos, que representam 92,7% do total de unidades censitárias, concentram-se em somente 16,6% da área total, o que resulta numa proporção aproximada de 13 setores urbanos para cada setor rural (Tabela 1). As métricas de área revelam contrastes entre os setores: os urbanos têm uma média de 0,188 hectares, variando entre 0,001 hectares e 17,482

hectares, ao passo que os rurais têm uma média de 12,126 hectares, com uma variação entre 0,002 hectares e 293,410 hectares. Esta disparidade reflete-se também nas distribuições das coberturas terrestres. Enquanto as áreas urbanas são dominadas por espaços não vegetados (83,2%), os setores rurais concentram quase a totalidade das florestas (91,1%), das áreas vegetadas (92,3%) e das atividades agropecuárias (95%). Tal padrão de ocupação do território está em conformidade com o que Sudhira, Ramachandra e Jagadish (2004) afirmam sobre a relação entre urbanização e alteração do ambiente por áreas construídas.

Tabela 1 - Situação e área de cobertura da terra em hectares

Situação	Setores (un)	Área (ha)	Floresta (ha)	Vegetação (ha)	Agropecuária (ha)	Não-vegetada (ha)
Rural	393 (7,3%)	475.155 (83,4%)	225.501 (91,1%)	25.405 (92,3%)	212.265 (95,0%)	11.983 (16,8%)
Urbana	5.021 (92,7%)	94.472 (16,6%)	21.990 (8,9%)	2.133 (7,7%)	11.068 (5,0%)	59.281 (83,2%)
TOTAL	5.414	569.627	247.492	27.538	223.333	71.264

Fontes: MapBiomass, 2024; IBGE, 2022. Elaboração: Os autores, 2025.

A matriz de correlação e dispersão apresentada na Figura 5 permite analisar as relações entre as variáveis e os respectivos padrões de distribuição. Neste sentido, foram avaliadas as distribuições entre as variáveis de uso e cobertura da terra, bem como as suas relações com a situação urbano-rural. A figura é uma matriz que contém gráficos e correlações das variáveis indicadas nas linhas e colunas. Na diagonal, encontram-se as distribuições de densidade de cada variável independente, divididas por rural e urbano. Na parte superior da diagonal, são mostradas as correlações entre as variáveis, no geral e também divididas por rural (0) e urbano (1). Na parte inferior, são apresentados gráficos de dispersão que demonstram a relação entre cada par de variáveis. Na Figura 5, os níveis de significância são indicados por **** para um valor de p inferior a 0,001, *** para um valor de p inferior a 0,01 e ** para um valor de p inferior a 0,1.

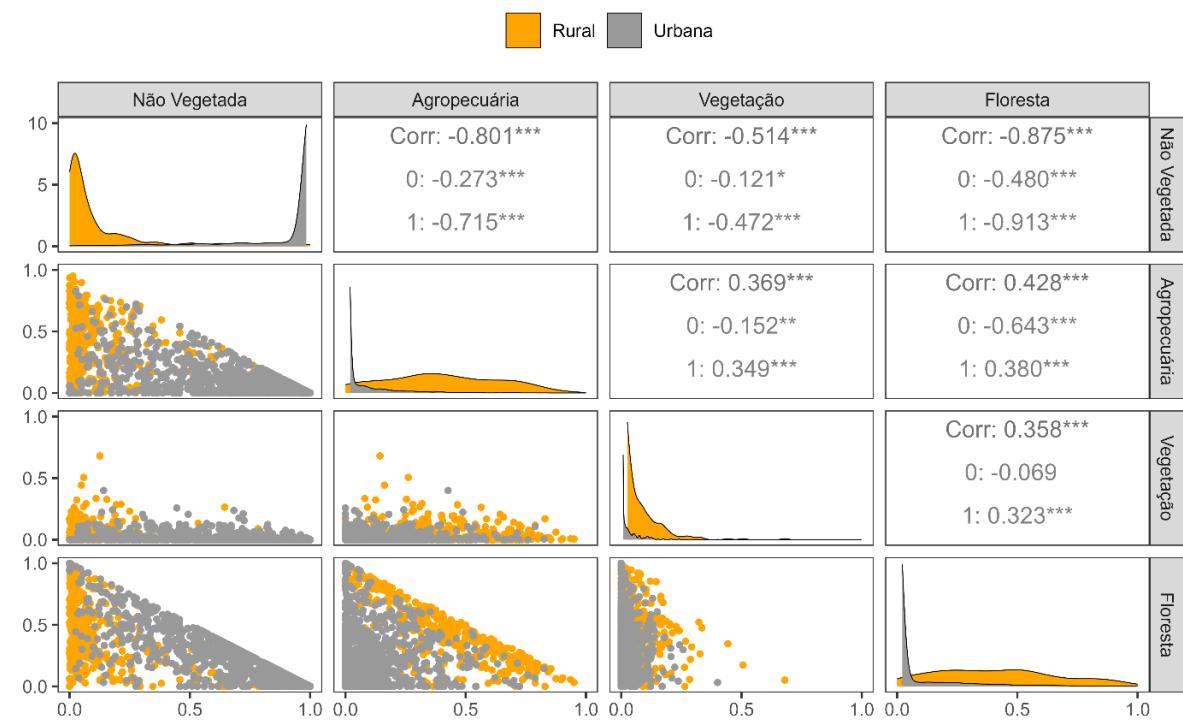
Ao analisar os gráficos da diagonal da Figura 5, é possível observar que as áreas não vegetadas apresentam uma distribuição bimodal, indicando a presença de dois grupos distintos de setores, um com uma proporção baixa e outro com uma proporção alta. A agropecuária exibe uma distribuição dispersa nos setores rurais, refletindo a diversidade de utilização neste contexto. Nos setores urbanos, a percentagem de ocorrência é reduzida. A vegetação apresenta uma concentração de valores inferior a 50%, o que indica que a maioria dos setores tem uma cobertura vegetal reduzida e reforça a tendência de fragmentação da vegetação na área de estudo. Por fim, as florestas apresentam uma distribuição semelhante à da agropecuária.

A análise das correlações entre as áreas não vegetadas e as demais categorias de cobertura terrestre mostra padrões consonantes com o tipo de cobertura. Em geral, as correlações entre as áreas não vegetadas e as vegetações são negativas e moderadas ($R = -0,514^{**}$), o que indica que os setores com maior cobertura vegetal têm proporcionalmente menos áreas não vegetadas. No entanto, ao analisar as florestas, as relações intensificam-se (R de $-0,875^{***}$), o que revela que as áreas florestais são sensíveis à expansão de superfícies não vegetadas. Ao segmentar as análises por categorias de cobertura (0 = rural, 1 = urbano), verifica-se que esta tendência é particularmente acentuada ($R = -0,913^{***}$), contrastando com a correlação mais moderada para as vegetações nessas mesmas áreas ($R = -0,472^{**}$). Nos setores rurais, as correlações entre vegetação e áreas não vegetadas são mais fracas (R de $-0,121^{*}$), ao passo que as correlações envolvendo florestas permanecem moderadas (R de $-0,480^{***}$), o que sugere que outros fatores podem influenciar a variação entre estas categorias.

No que se refere à agropecuária, as correlações com as áreas não vegetadas são negativas e fortes (R de $-0,801^{***}$), o que evidencia que o aumento de uma dessas categorias tende a reduzir a outra. Por outro lado, a relação entre a agropecuária e a vegetação mostra-se positiva, variando de fraca a moderada (R de $0,369^{**}$), sugerindo uma possível coexistência entre estas duas categorias. No contexto urbano, esta relação permanece positiva e moderada ($R = 0,349^{**}$), reforçando a possibilidade de existirem setores que ainda mantenham vestígios de vegetação junto a áreas de utilização agropecuária. Por outro lado, nos setores rurais, a correlação torna-se negativa e fraca ($R = -0,152^{**}$),

o que reflete que a expansão da agropecuária pode estar associada à conversão direta de vegetação em áreas produtivas. A Tabela 2 apresenta um resumo das comparações efetuadas e das respectivas interpretações dos resultados discutidos anteriormente.

Figura 5 - Matriz de dispersão e correlação entre setores censitários por situação e classes de uso e cobertura da terra



Fonte: Os autores, 2025.

Tabela 2 - Comparação e interpretação da correlação entre variáveis

Par de Variáveis	Correlação			Interpretação
	Total	Rural (0)	Urbano (1)	
Não Vegetada x Agropecuária	-0,801***	-0,273***	-0,715***	Relação inversa forte com a expansão de uma e redução da outra
Não Vegetada x Vegetação	-0,514**	-0,121*	-0,472***	Mais vegetação associada a menos área não vegetada com efeito mais forte em áreas urbanas.
Não Vegetada x Floresta	-0,875***	-0,480***	-0,913***	Florestas se mostram muito sensíveis à expansão não vegetada, especialmente em áreas urbanas
Agropecuária x Vegetação	0,369***	-0,152**	0,349***	Tendem a coexistir, sendo no rural uma relação negativa fraca com substituição de vegetação por agropecuária.
Agropecuária x Floresta	0,428***	-0,643***	0,380***	Setores urbanos ainda mantêm florestas, enquanto rurais tendem a substituí-las
Vegetação x Floresta	0,358***	-0,069	0,323***	Relação positiva moderada em urbano e quase nula no rural

Significância: *** 0,001; ** 0,01; * 0,1

Elaboração: Os autores, 2025.

No geral, as análises indicam relação negativa entre áreas não vegetadas e as demais coberturas. O modelo de regressão apresentado a seguir confirma este padrão.

Modelo de Regressão Logística

O modelo de regressão logística foi elaborado com a amostra de treino. Conforme apresentado na Tabela 3, o percentual de áreas não vegetadas ($P_{NãoVegetada}$) teve um impacto positivo e significativo no modelo, indicando que os setores censitários com maior cobertura desta categoria têm maior probabilidade de serem classificados como urbanos. Em contrapartida, os percentuais de áreas de agropecuária ($P_{Agropecuária}$) e de vegetação ($P_{Vegetação}$) tiveram efeitos negativos, sugerindo que os setores com maior cobertura destas categorias têm menor probabilidade de serem classificados como urbanos. A análise dos módulos dos coeficientes sugere que, por ordem de importância, as variáveis são: área não vegetada ($\beta = 7,1277$, $p < 0,001$), vegetação ($\beta = -5,4206$, $p = 0,005$) e agropecuária ($\beta = -3,6621$, $p < 0,001$).

Tabela 3 - Coeficientes do modelo

Informações	Estimativa	Erro padrão	Valor z	Valor p
Intercepto	-0,2673	0,2566	-1,042	0,29742
$P_{NãoVegetada}$	7,1277	0,4708	15,138	< 2e-16***
$P_{Agropecuária}$	-3,6621	0,5445	-6,726	1,75e-11***
$P_{Vegetação}$	-5,4206	1,9420	-2,791	0,00525**

Significância: *** 0,001; ** 0,01

Fonte: Os autores, 2025.

Os coeficientes das variáveis são significativos ao nível de 1% e revelam padrões na relação entre as categorias de utilização e cobertura do solo e a situação dos setores censitários. A qualidade do modelo foi avaliada com base no R^2 de McFadden. O desvio residual foi de 587,4 e o desvio nulo de 2.211,6, resultando num R^2 de McFadden de 0,734. Este valor demonstra que o modelo ajustado reduz a incerteza da variável resposta.

A equação 1 apresenta o modelo ajustado, em que p é a probabilidade de um setor ser urbano e $\ln(p)$ é o logaritmo neperiano de p . Para calcular a probabilidade de um setor ser urbano, é necessário multiplicar cada variável pelo respectivo coeficiente e somar esses valores. Em seguida, é aplicada a exponenciação com o número de Euler (e), gerando uma razão entre a probabilidade de um setor ser urbano e a probabilidade de não ser. Por fim, esta razão é dividida pela sua soma com uma unidade. Exemplos de resultados dos cálculos são apresentados na Tabela 4.

$$\ln \left(\frac{p}{1-p} \right) = -0,2673 + 7,1277 \times P_{NãoVegetada} - 3,6621 \times P_{Agropecuária} - 5,4206 \times P_{Vegetação} \quad (1)$$

Quanto à validação, o modelo confirmou uma elevada capacidade preditiva, com uma precisão geral de 97,3% na amostra de teste. Com um ponto de corte do modelo de 50%, o índice Kappa foi de 0,8216, indicando um elevado grau de concordância e reforçando a confiabilidade das previsões. Dos 5.414 setores censitários avaliados, o modelo aqui proposto prevê 129 distintamente das definições do IBGE, ou seja, 97,6% das tipificações são bem-sucedidas. Isto significa que o modelo é eficiente na designação da situação do setor censitário.

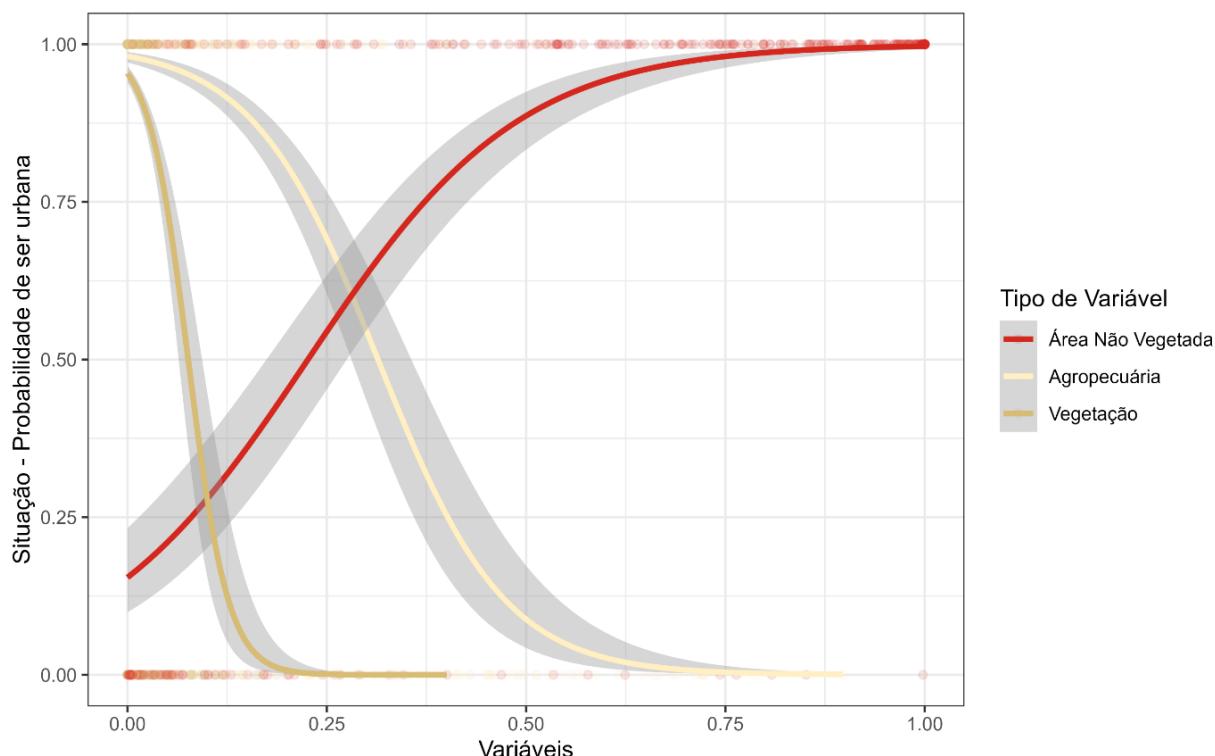
A sensibilidade de 85% demonstra a capacidade do modelo em classificar corretamente os setores urbanos, ao passo que a especificidade de 98,4% confirma a sua eficácia na identificação dos setores rurais. A acurácia equilibrada de 91,7% sugere que o modelo mantém um desempenho satisfatório mesmo em cenários com possíveis desequilíbrios entre as situações. Além disso, os valores preditivos positivos (82,2%) e negativos (98,7%) indicam uma taxa de erro reduzida nas previsões, minimizando o número de falsos positivos e falsos negativos. O valor da estatística de área sob a curva ROC (AUC) de 0,9852 reforça a capacidade discriminativa do modelo.

A Figura 6, elaborada com base na aplicação do modelo à amostra de validação, ilustra a distribuição das proporções das variáveis utilizadas no modelo (eixo horizontal) em função do tipo de setor censitário (eixo vertical), sendo que o valor zero (0) representa os setores rurais e o valor um (1)

representa os setores urbanos. Constatase que os setores urbanos tendem a apresentar valores mais elevados de área não vegetada, o que está em conformidade com o impacto positivo desta variável no modelo. Em contraste, os setores rurais apresentam proporções mais elevadas de agropecuária e de vegetação, o que está em conformidade com os coeficientes negativos destas variáveis.

Embora o modelo produza uma variável binária como resultado, as curvas apresentadas na Figura 6 indicam as diversas configurações possíveis entre os setores rurais e urbanos. A distinção entre estas categorias é abordada quantitativamente, mas também aborda as críticas à rigidez das classificações oficiais. Autores como Veiga (2004, 2013), Araújo e Soares (2009) e Santos e Silva (2020) argumentam que a divisão dicotômica simplifica realidades territoriais complexas, nas quais existem gradientes e hibridismos entre o urbano e o rural. Estes autores concordam com Lefebvre (1999), que questiona a distinção rígida entre o rural e o urbano, defendendo que os processos urbanos se estendem para além dos limites convencionais das cidades. Assim, a dicotomia rural-urbano é considerada uma simplificação da complexa realidade urbana contemporânea, a qual é caracterizada por uma urbanização generalizada do território (Buckley; Strauss, 2016). Nesse sentido, o modelo proposto oferece uma alternativa metodológica que capta a pluralidade de formas de ocupação do solo, traduzindo-as em probabilidades que refletem a intensidade das características urbanas ou rurais. A alta acurácia do modelo (97,3%) sugere que variáveis ambientais são *proxies* eficazes para essa categorização, validando a premissa de que a morfologia da paisagem, mensurada por sensoriamento remoto, dialoga com as dinâmicas socioespaciais criticadas pela literatura.

Figura 6 - Distribuição das proporções das variáveis preditoras em relação ao tipo de setor censitário



Fontes: Os autores, 2025.

A seguir, na Tabela 4, são apresentados quatro setores censitários selecionados intencionalmente (e não aleatoriamente) para ilustrar as diferentes predominâncias de uso e cobertura da terra em relação aos resultados do modelo proposto. Os setores em questão são os seguintes: A – rural, com predominância de áreas não vegetadas; B – rural, com predominância de áreas não vegetadas e agropecuária; C – rural, com predominância de agropecuária; D – urbano, com predominância de áreas não vegetadas e agropecuária. A tabela evidencia a situação censitária atual de cada um destes setores, as respectivas identificações, as percentagens de cobertura das categorias de utilização e

cobertura do solo e a probabilidade de classificação como área urbana, conforme o modelo proposto, discutido com a Figura 7.

Tabela 4 - Informações dos setores censitários demonstrados na Figura 5 e suas respectivas probabilidades de ser urbano conforme estimado pelo modelo

ID	Situação	Identificação	Não vegetada (%)	Agropecuária (%)	Vegetação (%)	Prob. Urb. (%)
A	Rural	530010805100441	98,1	1,9	0	99,9
B	Rural	530010805390021	49,5	34,2	3,5	89,0
C	Rural	530010805110393	00,5	74,4	2,1	5,7
D	Urbana	530010805400051	50,0	45,6	1,6	85,9

Fontes: Os autores, 2025.

A Figura 7 apresenta uma comparação visual entre a classificação temática de utilização e cobertura terrestre do MapBiomass e imagens de satélite PlanetScope de elevada resolução espacial (cerca de 3 metros por pixel), disponibilizadas pela iniciativa internacional norueguesa para o clima e as florestas (NCFI) em novembro de 2021 (Planet, 2025). As imagens identificadas como A1, B1, C1 e D1 oferecem uma visão detalhada da cobertura terrestre, permitindo comparar as condições observadas com as classificações temáticas apresentadas nas imagens A2, B2, C2 e D2. É importante salientar que as imagens de dezembro de 2021 não foram utilizadas, dado a ver nuvens sobrepostas aos setores censitários analisados, o que poderia comprometer as análises propostas.

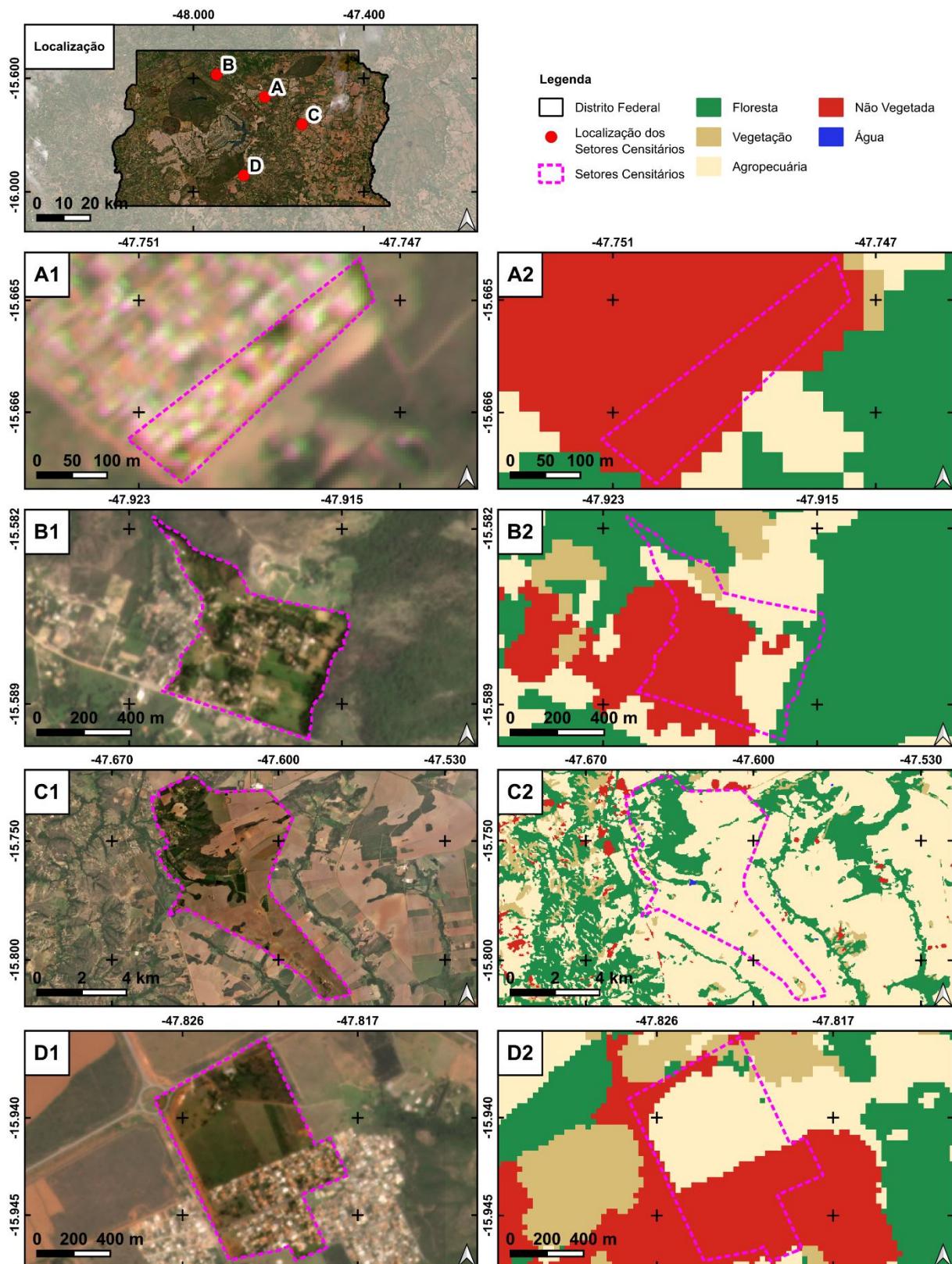
Os setores A e B são definidos como rurais no Censo de 2022, embora o modelo proposto indique uma elevada probabilidade de serem urbanos, respectivamente 99,9% e 89,0%. Esta discrepância pode ser explicada por características específicas destas áreas. O setor B aparentemente apresenta amplas áreas verdes, sem um ambiente totalmente construído, ao passo que o setor A, ainda que contíguo a uma área urbanizada com arruamento, poderá ter sido categorizado como rural devido à presença de campos de vegetação ou de atividades agropecuárias nas proximidades. Outra hipótese é a de que a não atualização destes setores no Recenseamento Geral de 2022 se tenha dado por motivos administrativos não divulgados. Independentemente das razões, a análise dos produtos resultantes da observação da Terra por satélite reforça a ideia de que estes setores são urbanos, corroborando as previsões do modelo.

É importante salientar que este modelo não tem em conta aspectos de vizinhança, ou seja, não incorpora a possibilidade de ser urbano em relação à situação dos setores próximos. Este aspecto é incomum nos estudos sobre espalhamento urbano, que geralmente avaliam a contiguidade e a forma da mancha urbana (Chettry, 2023). A vantagem de não utilizar a característica espacial é a possibilidade de prever áreas urbanas a partir de fragmentos não vegetados que surjam não adjacentes a áreas não vegetadas.

O setor C é rural e ilustra o funcionamento do modelo em áreas predominantemente agrícolas, com apenas 5,7% de probabilidade de ser urbano, o que está em conformidade com a análise visual. Por outro lado, o setor D, definido como urbano pelo IBGE, apresenta um caso interessante, visto possuir quantidades semelhantes de cobertura não vegetada (50,0%) e de cobertura de agropecuária (45,6%). Na imagem D1, é possível observar áreas bem delimitadas e o detalhe do MapBiomass na imagem D2 revela manchas não vegetadas próximas à rede viária. O modelo atribui uma probabilidade de 85,9% de ser urbano, o que está em conformidade com a classificação oficial do IBGE.

Os casos A e B, em que o modelo divergiu do IBGE, revelam precisamente zonas de transição onde os critérios ambientais foram diferenciados mediante sensoriamento remoto. Nesse sentido, e com base nas estatísticas apresentadas e nos casos exemplificados, o modelo evidencia-se pertinente para auxiliar na elaboração de pesquisas, visando sobretudo superar as fragilidades apontadas por Pera e Bueno (2016), nomeadamente os aspectos operacionais e financeiros da estrutura do IBGE para realizar o recenseamento a cada dez anos.

Figura 7 - Representação visual da localização e dos setores censitários analisados, destacando a situação em novembro de 2021 e as respectivas categorias de uso e cobertura da terra



Fontes: Planet, 2025; MapBiomas, 2024; IBGE, 2022. Elaboração: Os autores, 2025.

Por fim, este trabalho contribui para a distinção proposta por Santos (1993) entre a urbanização da sociedade e a urbanização do território, ao evidenciar as características urbanas a partir de suas manifestações materiais no espaço. Enquanto muitas metodologias tradicionais definem o urbano e o rural com base prioritariamente em critérios sociais, a utilização de subprodutos do sensoriamento remoto contribui com variáveis associando aspectos materiais e visíveis do espaço. Embora as relações sociais sejam centrais para a constituição do urbano, elas ocorrem em espaços fisicamente construídos, sendo captadas por imagens de satélites e vistorias técnicas. As características antrópicas são inferidas a partir do impacto ambiental no território. Nessa perspectiva, a área urbana aparece antes como um impacto ambiental provocado por ações humanas do que como um espaço de relações sociais. Em contrapartida, vale notar que comunidades tradicionais mantêm formas coletivas de vida sem gerar os mesmos impactos físicos sobre o ambiente.

CONSIDERAÇÕES FINAIS

O presente estudo caracterizou a situação urbano-rural dos setores censitários do Censo de 2022 do Distrito Federal, recorrendo à regressão logística e utilizando como variáveis independentes as categorias de utilização e cobertura do solo "não vegetado", "agropecuária" e "vegetação", provenientes do nível 1 da classificação de utilização e cobertura do solo do MapBiomas.

O estudo conclui que a situação urbano-rural pode ser prevista com base em dados de sensoriamento remoto. O estudo reforça a viabilidade desta abordagem para o acompanhamento territorial, salientando a forte ligação entre as áreas não vegetadas e a situação urbana. A escolha da regressão logística mostrou-se adequada para a elaboração do modelo, com uma taxa de previsão de 97,6%. Uma vantagem do método proposto é não depender de estatísticas de vizinhança, conseguindo identificar áreas não vegetadas como áreas urbanas, mesmo que estejam rodeadas por outras coberturas.

A metodologia implementada pelo IBGE no Censo de 2022, que teve em conta imagens de alta resolução para definir a situação urbano-rural e validou os tipos de cada setor censitário no terreno, abre novas potencialidades para a elaboração de estudos, ao ultrapassar os tradicionais limites administrativos e fornece uma malha validada com base nas características dos territórios e no conhecimento dos recenseadores. Pode-se dizer que o trabalho aqui apresentado traduziu essas características e conhecimentos num modelo que usa dados de sensoriamento remoto disponíveis. Tal pode ser utilizado para prever a situação de áreas até a realização do próximo recenseamento, bem como para planejar outras pesquisas amostrais em regiões que apresentem alterações no uso e na cobertura do solo.

A utilização dos dados de resolução espacial de 30 metros da Coleção 9 do MapBiomas permitiu ao modelo distinguir áreas urbanas e rurais, identificando padrões de utilização e cobertura do solo. Embora resoluções mais elevadas possam, no futuro, oferecer um maior nível de detalhe, os produtos de 30 metros revelaram-se adequados para detectar alterações na paisagem.

No que se refere a investigações futuras, há oportunidades para avaliar aspectos espaciais, temporais e de modelação. Podem ser avaliados níveis mais detalhados da variável "Situação" do IBGE, bem como o nível 2 da classificação do MapBiomas, que oferece uma maior especificidade para as categorias de utilização e cobertura do solo, para além de diferentes resoluções espaciais. Outro fator que pode ser melhor avaliado é a dimensão dos setores censitários, que tendem a ser menores quanto maior for a população e, consequentemente, a área não vegetada. Assim, os estudos podem valer-se de outros recortes espaciais para analisar o modelo. O trabalho tem também potencial para ser ampliado a outras regiões, podendo ser aplicado em diversos recortes, como municípios, regiões metropolitanas, estados e outros países onde haja cobertura do MapBiomas, por exemplo, parte da América Latina. Novas investigações podem igualmente explorar outros conjuntos de dados sobre o uso e a cobertura da terra, para além do MapBiomas, com especial enfoque nos assentamentos humanos. Quanto aos aspectos temporais, a disponibilidade histórica e consistente destes dados possibilita análises temporais para compreender a situação urbano-rural, a construção das cidades brasileiras e fazer previsões futuras. No modelo apresentado, foi adotado um ponto de corte de 50% na probabilidade logística e podem ser testados diferentes limiares para analisar cenários.

DISPONIBILIDADE DE DADOS E SCRIPTS

Os dados e scripts utilizados neste estudo estão disponíveis em repositório público, que pode ser acessado por meio do seguinte link: <https://github.com/daciocambraia/artigo-regressao-logistica-urbano-rural.git>.

AGRADECIMENTOS

Esta pesquisa foi apoiada pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por meio de bolsa de doutorado da autora Glaucia Guimarães Pereira (processo 88887.965945/2024-00).

CONTRIBUIÇÃO DOS AUTORES

Conceituação, Glaucia Guimarães Pereira, Dácio José Cambraia Filho; Curadoria de dados, Dácio José Cambraia Filho; Análise formal, Glaucia Guimarães Pereira, Dácio José Cambraia Filho; Investigação, Glaucia Guimarães Pereira, Dácio José Cambraia Filho; Metodologia, Glaucia Guimarães Pereira, Dácio José Cambraia Filho; Gustavo Macedo de Mello Baptista; Software, Dácio José Cambraia Filho; Supervisão, Gustavo Macedo de Mello Baptista; Visualização, Glaucia Guimarães Pereira, Dácio José Cambraia Filho; Redação – rascunho original, Glaucia Guimarães Pereira, Dácio José Cambraia Filho, Diego de Almeida Paim, Roberto Mandetta Gandara; Redação – revisão e edição, Glaucia Guimarães Pereira, Dácio José Cambraia Filho.

REFERÊNCIAS

- ABRAMOVAY, R. **Funções e medidas da ruralidade no desenvolvimento contemporâneo**. Rio de Janeiro: IPEA, 2000.
- ARAÚJO, F. A. V. de; SOARES, B. R. Relação cidade-capó: desafios e perspectivas. **Campo-Território: revista de geografia agrária**, v. 4, n. 7, p. 201-229, 2009. <https://doi.org/10.14393/RCT4711894>
- BRAGA, G. B.; REMOALDO, P. C.; FIÚZA, A. L. Aplicação da metodologia definida pela OCDE para o cálculo de população urbana e rural nos municípios do Brasil. In: COLÓQUIO IBÉRICO DE GEOGRAFIA, 14, 2014, Guimarães. **Anais [...]**, 2014. p. 2413-2418.
- BRASIL. **Constituição da República Federativa do Brasil de 1988**. 1988. Disponível em: https://www.planalto.gov.br/ccivil_03/constituicao/constituicao.html. Acesso em: 19 fev. 2025.
- BRASIL. **Objetivo de Desenvolvimento Sustentável 11: Cidades e comunidades sustentáveis**, 2025. Disponível em: <https://brasil.un.org/pt-br/sdgs/11>. Acesso em: 19 fev. 2025.
- BUCKLEY, M.; STRAUSS, K. With, against and beyond Lefebvre: Planetary urbanization and epistemic plurality. **Environment and Planning D: Society and Space**, v. 34, n. 4, p. 617–636, 2016.
- CASA CIVIL; IPEA - Instituto de Pesquisa Econômica Aplicada. **Avaliação de Políticas Públicas: Guia Prático de Análise Ex Ante**. 1. ed. Brasília: Ipea, 2018. Disponível em: <https://repositorio.ipea.gov.br/handle/11058/8285>. Acesso em: 17 fev. 2025.
- CHELOTTI, G. B.; SANO, E. Sessenta anos de evolução urbana do Distrito Federal analisadas por meio de métricas de paisagem. **Revista Gestão & Sustentabilidade Ambiental**, v. 12, p. 21, 2023. <https://doi.org/10.59306/rgsa.v12e12023e10279>
- CHETRY, V. A Critical review of urban sprawl studies. **Journal of Geovisualization and Spatial Analysis**, v. 7, n. 28, p. 13, 2023. <https://doi.org/10.1007/s41651-023-00158-w>
- CRESWELL, J. W.; CRESWELL, J. D. **Research Design: Qualitative, Quantitative, and Mixed Methods Approaches**. 6 ed. Los Angeles: SAGE Publications, 2017.
- DADASHPOOR, H.; SHAHHOSSEIN, G. Defining urban sprawl: A systematic review of 130 definitions. **Habitat International**, v. 146, p. 14, 2024. <https://doi.org/10.1016/j.habitatint.2024.103039>

DODMAN, D. et al. Cities, Settlements and Key Infrastructure. In: WORKING GROUP II CONTRIBUTION TO THE SIXTH ASSESSMENT REPORT OF THE INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (Coord.) **Climate Change 2022: Impacts, Adaptation, and Vulnerability**. Cambridge: Cambridge University Press, 2023. p. 907–1040.
<https://doi.org/10.1017/9781009325844.008>

FERNANDES, A. A. T. et al. Leia este artigo se você quiser aprender regressão logística. **Revista de Sociologia e Política**, v. 28, n. 74, p. e006, 2020.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Ed. Atlas, 2002.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Distrito Federal - Cidades e Estados**. Disponível em: <https://www.ibge.gov.br/cidades-e-estados/df.html>. Acesso em: 19 fev. 2025.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Malha de Setores Censitários**. Rio de Janeiro: IBGE, 2024.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Malha de Setores Censitários de 2022**. Disponível em: <https://www.i-bge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/26565-malhas-de-setores-censitarios-divisoes-intramunicipais.html>. Acesso em: 18 fev. 2025.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Malha Municipal 2023**. Disponível em: <https://www.ibge.gov.br/geociencias/organizacao-do-territorio/malhas-territoriais/15774-malhas.html?=&t=o-que-e>. Acesso em: 27 mar. 2025.

JACKMAN, S. et al. **Pscl: Political Science Computational Laboratory**, 2024. Disponível em: <https://cran.r-project.org/web/packages/pscl/index.html>. Acesso em: 20 fev. 2025.

KUHN, M. et al. **Caret**: Classification and Regression Training, 2024. Disponível em: <https://cran.r-project.org/web/packages/caret/index.html>. Acesso em: 20 fev. 2025.

LEFEBVRE, H. **A Revolução Urbana**. Belo Horizonte: Editora UFMG, 1999.

MAPBIOMAS. **MapBiomas**, 2015. Disponível em: <https://brasil.mapbiomas.org>. Acesso em: 2 abr. 2025.

MAPBIOMAS. **Collection 9 of the Annual Land Cover and Land Use Maps of Brazil (1985-2023)**, 2024. Disponível em: <https://data.mapbiomas.org/dataset.xhtml?persistentId=doi:10.58053/MapBiomas/XXUKA8>. Acesso em: 15 fev. 2025.

PARK, S. et al. Prediction and comparison of urban growth by land suitability index mapping using GIS and RS in South Korea. **Landscape and Urban Planning**, v. 99, n. 2, p. 104–114, 2011.
<https://doi.org/10.1016/j.landurbplan.2010.09.001>

PERA, C. K. L.; BUENO, L. M. de M. Revendo o uso de dados do IBGE para pesquisa e planejamento territorial: reflexões quanto à classificação da situação urbana e rural. **Cadernos Metrópole**, v. 18, n. 37, p. 722–742, 2016. <https://doi.org/10.1590/2236-9996.2016-3705>

PLANET. **Norway's International Climate and Forests Initiative Satellite Data Program (NICFI Program)**, 2025. Disponível em: <https://www.planet.com/nicfi/>. Acesso em: 21 fev. 2025.

POSIT, S. **RStudio**: Integrated Development Environment for R, 2024. Disponível em: <https://posit.co/download/rstudio-desktop/>. Acesso em: 22 fev. 2025.

QGIS. **QGIS Geographic Information System**, 2025. Disponível em: <https://qgis.org/>. Acesso em: 22 fev. 2025.

R CORE TEAM. **The R Project for Statistical Computing**, 2024. Disponível em: <https://www.r-project.org/about.html>. Acesso em: 19 fev. 2025.

REFERENCE, S. **ESRI:102033 - South_America_Albers_Equal_Area_Conic**, 2025. Disponível em: <https://spatialreference.org/ref/esri/102033/>. Acesso em: 21 fev. 2025.

ROBIN, X. et al. **pROC**: Display and Analyze ROC Curves, 2023. Disponível em: <https://cran.r-project.org/web/packages/pROC/index.html>. Acesso em: 20 fev. 2025.

- RSTUDIO TEAM. **stats (version 3.6.2)**, 2019. Disponível em: <https://www.rdocumentation.org/packages/stats/versions/3.6.2>. Acesso em: 20 fev. 2025.
- SANTOS, M. **A urbanização brasileira**. São Paulo: Editora HUCITEC, 1993.
- SANTOS, K. F. dos; SILVA, A. B. da. A urbanização do território e a relação entre o campo e a cidade. In: MAIA, D. S.; MARAFON, G. J. (Coord.) **Ensino superior e desenvolvimento regional: reconfigurando as relações entre as cidades e o campo**. Rio de Janeiro: EDUERJ, 2020. p. 148-171. <https://doi.org/10.7476/9786587949086>
- SCHLOERKE, B. et al. **GGally**: Extension to “ggplot2”, 2024. Disponível em: <https://cran.r-project.org/web/packages/GGally/index.html>. Acesso em: 20 fev. 2025.
- SEGOV - Secretaria de Estado de Governo do Distrito Federal . **Administrações Regionais**. Disponível em: <https://segov.df.gov.br/category/administracoes-regionais/>. Acesso em: 19 fev. 2025.
- SIDDQUI, A. et al. Urban growth dynamics of an Indian metropolitan using CA Markov and Logistic Regression. **The Egyptian Journal of Remote Sensing and Space Science**, v. 21, n. 3, p. 229–236, 2018. <https://doi.org/10.1016/j.ejrs.2017.11.006>
- SKORUPKA, M. Proposal for the modification of the OECD and EUROSTAT-based typologies for rural areas. **Geomatics, Landmanagement and Landscape**, n. 1, p. 121–146, 2017. <https://doi.org/10.15576/GLL/2017.1.121>
- SOUZA, C. M. et al. Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine. **Remote Sensing**, v. 12, n. 2735, p. 27, 2020. <https://doi.org/10.3390/rs12172735>
- SOUZA, A. L. de et al. Revisão metodológica da tipologia urbano-rural no Censo Demográfico 2022. **Ciência & Saúde Coletiva**, v. 49, n. 11, p. 12, 2024. <https://doi.org/10.1590/1413-812320242911.03062024>
- SUDHIRA, H. S.; RAMACHANDRA, T. V.; JAGADISH, K. S. Urban sprawl: metrics, dynamics and modelling using GIS. **International Journal of Applied Earth Observation and Geoinformation**, v. 5, n. 1, p. 29–39, 2004. <https://doi.org/10.1016/j.jag.2003.08.002>
- VEIGA, J. E. da. A atualidade da contradição urbano-rural. In: SUPERINTENDÊNCIA DE ESTUDOS ECONÔMICOS E SOCIAIS DA BAHIA. **Análise Territorial da Bahia Rural**. Salvador: 2004. https://sei.ba.gov.br/images/publicacoes/download/sep/sep_71.pdf. Acesso em: 5 abr. 2025.
- VEIGA, J. E. da. **Cidades Imaginárias**: o Brasil é menos urbano do que se calcula. 2. ed. Campinas: Armazém do Ipê, 2013. 304 p.
- WICKHAM, H. et al. **Ggplot2**: Create Elegant Data Visualisations Using the Grammar of Graphics, 2024. Disponível em: <https://cran.r-project.org/web/packages/ggplot2/index.html>. Acesso em: 20 fev. 2025.
- WICKHAM, H. et al. **Readxl**: Read Excel Files, 2023 Disponível em: <https://cran.r-project.org/web/packages/ggplot2/index.html>. Acesso em: 19 fev. 2025.
- XAVIER-DA-SILVA, J.; CARVALHO-FILHO, L. M. Sistemas de Informação Geográfica: uma proposta metodológica. In: CONFERÊNCIA LATINOAMERICANA SOBRE SISTEMAS DE INFORMAÇÃO GEOGRÁFICA, 4., 1993, São Paulo. **Anais [...]**. São Paulo, 2019. p. 609–629.

Recebido em: 06/04/2025

Aceito para publicação em: 20/08/2025