

MAPEAMENTO MULTITEMPORAL UTILIZANDO FLORESTAS ALEATÓRIAS NO PERÍODO SECO NOS SERTÕES CEARENSES

Adinan Marzulo Maia Martins¹

Carla Bernadete Madureira Cruz²

RESUMO

Este trabalho tem por objetivo a análise das mudanças ambientais a partir de processamentos em nuvem na plataforma Google Earth Engine (GEE). Foram gerados 4 mapeamentos dos Sertões Cearenses por meio do classificador Random Forest (RF) no período seco para as faixas temporais 2004-2005, 2013-2014, 2019-2020 e 2023-2024 a partir de dados ópticos dos sensores Landsat 5, 8 e 9 e do Modelo Digital de Elevação (MDE) NASADEM. As mudanças mais expressivas na região foram a conversão de Vegetação para Solo com Alto Grau de Exposição (5.918,2 km²), seguida pela transição inversa (4.670,2 km²), indicando ciclos de degradação e recuperação. Esses resultados evidenciam transformações preocupantes, como a perda de áreas vegetadas para coberturas arenosas e a redução da cobertura hídrica, que contribuem para a desertificação e degradação regional.

Palavras-Chave: GEE; RF; RFE Semiárido; Seca.

MULTI-ITEM MAPPING USING RANDOM FORESTS IN THE DRY PERIOD IN THE SERTÕES OF CEARÁ

ABSTRACT

This article aims to analyze environmental changes through cloud-based processing on Google Earth Engine's (GEE). Four mappings of the Sertões Cearenses region were generated with the Random Forest (RF) classifier during the dry season, for 2004-2005, 2013-2014, 2019-2020 and 2023-2024 timeframes, using Landsat 5, 8 and 9's optical data and the Digital Elevation Model (DEM) NASADEM. The conversion from Vegetation to Highly Exposed Soil (5,918.2 km²) were the region's most significant changes, followed by the inverse transition (4,670.2 km²), pointing to cycles of degradation and recovery. Pressing matters are highlighted by these results, such as the loss of vegetated areas for sandy surfaces and the declining water surfaces, which contribute to regional desertification and degradation.

Keywords: GEE; RF; RFE, Semiarid; Drought.

¹Mestre em Geografia pela Universidade Federal do Rio de Janeiro - UFRJ, Email: adinanmaia@gmail.com

²Doutora em Geografia e Prof(a). Titular do Departamento de Geografia pela Universidade Federal do Rio de Janeiro, UFRJ, Email: carlamad@gmail.com

MAPEO MULTI-ÍTEMS UTILIZANDO BOSQUES ALEATORIOS EN EL PERÍODO SECO EN LOS SERTÕES DE CEARÁ

RESUMEN

Este artículo tiene como objetivo analizar los cambios ambientales a través del procesamiento basado en la nube en el motor de Google Earth (GEE). Se generaron cuatro mapas de la región de Sertões Cearenses con el clasificador Random Forest (RF) durante la estación seca, para los periodos 2004-2005, 2013-2014, 2019-2020 y 2023-2024, utilizando datos ópticos de Landsat 5, 8 y 9 y el Modelo Digital de Elevación (DEM) NASADEM. La conversión de la vegetación a suelo muy expuesto (5.918,2 km²) fue el cambio más significativo en la región, seguido por la transición inversa (4.670,2 km²), lo que indica ciclos de degradación y recuperación. Estos resultados ponen de relieve cuestiones apremiantes, como la pérdida de zonas de vegetación en favor de superficies arenosas y la disminución de las superficies de agua, que contribuyen a la desertificación y la degradación regionales.

Palabras-Clave: GEE; RF; RFE, Semiárido; Sequía.

INTRODUÇÃO

O mapeamento do bioma Caatinga, cuja variabilidade paisagística é afetada pelas condições naturais no clima, representa atualmente um dos grandes desafios para os geocientistas. As geotecnologias impulsionadas por métodos de classificação de aprendizado de máquina no semiárido têm uma tarefa primorosa para conseguir mapear as classes naturais que tanto requerem atenção devido às mudanças climáticas no bioma Caatinga (NOBRE e OYAMA, 2003).

Estudos através de mapeamentos multitemporais de uso e cobertura da terra poderiam ser a chave para entender quais áreas no semiárido brasileiro, no espaço e no tempo, efetivamente requerem uma maior atenção sob os aspectos da degradação ambiental e desertificação. Explorar as concepções ferramentais da Inteligência Artificial, como o Random Forest (RF), é um aliado importante para uma melhor caracterização do semiárido em termos de aspectos naturais e da influência humana.

Os objetivos deste trabalho são diversos. Primeiro, buscou-se estabelecer métricas de seleção de modelos para o mapeamento do período seco para as faixas temporais 2004-2005, 2013-2014, 2019-2020 e 2023-2024 a partir da lógica Recursive Feature Elimination (REF) para classificação de imagens adaptadas ao Google Earth Engine (GEE). A seleção de variáveis

importantes conforme Zhang e Yang (2020) pode trazer benefícios quanto a simplicidade e a eficiência computacional.

O empilhamento excessivo de informações por grandes quantidades de variáveis preditoras ao modelo pode levar à redundância de informações e ao aumento do tempo de computação (CHEN et al., 2024). Além disso, a identificação de padrões quanto às variáveis preditoras influentes nos 4 modelos escolhidos para a classificação, em um contexto de grande variabilidade paisagística, pode ser essencial para entender a dinâmica da cobertura terrestre nos Sertões Cearenses.

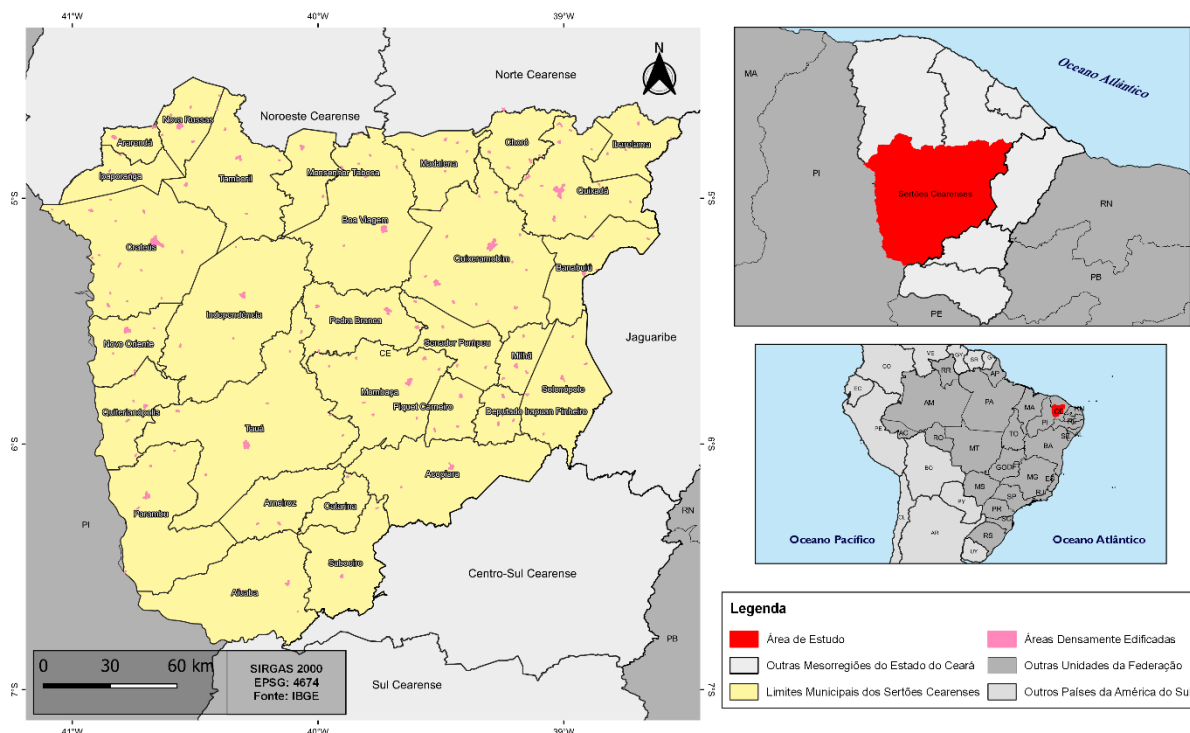
Todo este processamento em nuvem possibilitou a caracterização e a dinâmica da cobertura da terra no período seco, gerando também, mapas auxiliares como o mapa de persistência de classes e a detecção de mudanças. Contudo, isto permitiu a identificação de possíveis focos de degradação ambiental e a expansão de áreas desertificadas na área de estudo.

MATERIAIS E MÉTODOS

Área de Estudo

A área de estudo corresponde à mesorregião dos Sertões Cearenses que possui a área de 46.266,71 km² (IBGE, 2022). Nesta área de planejamento são compostas por 30 municípios (Figura 1) inseridos no semiárido político sendo caracterizados por longos períodos de estiagem e chuvas irregulares no tempo e no espaço que geram impactos socioeconômicos (TAVARES et al., 2019; BOTELHO, 2011; FERNANDES et al., 2009).

Figura 1: Localização da mesorregião dos Sertões Cearenses.

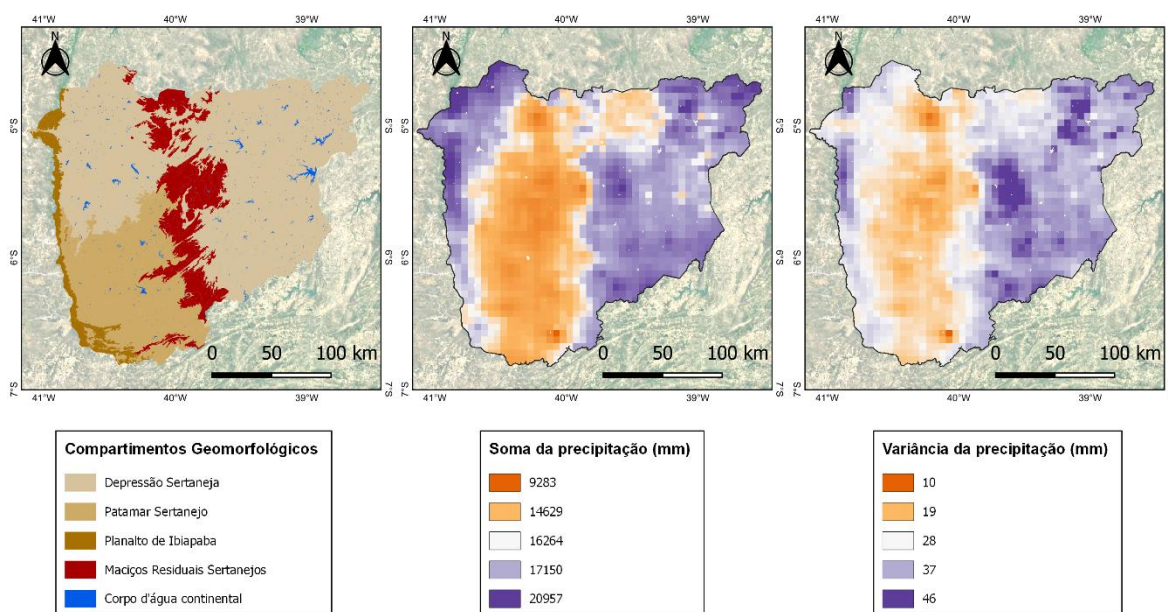


Fonte: Elaborado pelos Autores com dados do IBGE (2022).

O fenômeno de secas, característica essencial do semiárido brasileiro, está relacionado com os fenômenos El-Niño e Dipolo do Atlântico (MELO, 1999). O aquecimento dos mares do Pacífico pelo enfraquecimento da célula de Walker e do Atlântico Norte tornam a região do nordeste mais seca.

As condições geomorfológicas regionais podem ser também relacionadas com a intensificação dos eventos de secas. Nos Sertões Cearenses são observados a predominância da Depressão Sertaneja, caracterizada por áreas baixas formadas pela acumulação de sedimentos e da erosão, favorecendo o aplainamento nessas áreas. Os Maciços Residuais situam-se na porção central dos sertões cearenses, compartimento de maiores elevações, separa as unidades Depressão Sertaneja, Patamar Sertanejo e Planalto da Ibiapaba. Além disso, este compartimento por resistentes à erosão, podem atuar como bloqueios topográficos de umidade. Esta condição regional associada a essas áreas mais elevadas acabam por influenciar no quantitativo e na variabilidade da precipitação (Figura 2), formando extensões de coberturas mais secas representadas em tons alaranjados como mostram os dados do CHIRPS de 2000 a 2024.

Figura 2: Compartimentos Geomorfológicos e dados de precipitação do CHIRPS.



Fonte: Elaborado pelos Autores com dados do IBGE (2022) e CHIRPS (2000 – 2024).

Preparação dos dados

Para analisar os últimos 20 anos (2004 - 2024) de dados de cobertura da terra foi necessário gerar diversas classificações RF para a área de estudo. Os dados orbitais processados são oriundos da série temporal Landsat de resolução espacial de 30 metros em dados de refletância de superfície. Esse processamento foi efetuado na plataforma em nuvem GEE que permite o processamento de um grande contingente de dados em servidores da Google sem necessitar do processamento de máquina local.

A escolha para esta faixa temporal deu-se pela interferência do sistema meteorológico Zona de Convergência Intertropical (ZCIT) na faixa espectral óptica que podem resultar em perdas de dados ao Landsat, o que pode prejudicar as análises que comparam diferentes mapas para a faixa temporal principalmente nas regiões Norte e Nordeste do Brasil. Tal realidade e associada aos ruídos quanto ao processo de imageamento do sistema sensor do Landsat 5 na área de estudo impossibilitaram o uso para datas mais antigas permitindo apenas a análise da cobertura terrestre no período seco de 2004 a 2024.

A fim de gerar uma série temporal de dados contínuos Landsat foi necessário renomear as bandas espectrais para que não ocorresse alguma incompatibilidade na geração dos cálculos de índices espectrais, uma vez que os sensores mais recentes como é o caso do OLI e OLI 2 dos satélites Landsat 8 e 9 possuem a banda COASTAL BLUE (AZUL COSTEIRO) como identificador “SR_B1” enquanto o sensor TM possui “SR_B1” representada pela banda do BLUE (Azul). Foram utilizadas e devidamente renomeadas as seguintes bandas: B2 (BLUE - Azul), B3 (GREEN - Verde), B4 (RED - Vermelho), B5 (NIR – Infravermelho próximo), B6 (SWIR 1 - Infravermelho de ondas curtas 1) e B7 (SWIR 2 – Infravermelho de ondas curtas 2).

Foram também utilizados algoritmos de máscara de nuvens a fim de não influenciar os valores de reflectância. Outro detalhe importante é a utilização de filtros de cenas por valores de cobertura de nuvens menores que 30%, e do filtro espacial, que selecionará cenas para o recorte espacial de interesse. A partir do critério adotado referente à precipitação, foram utilizados 8 filtros temporais para a seleção das cenas que comporão os 4 mapeamentos RF (Quadro 1), as quais foram delimitadas para a área de estudo. É importante mencionar que as áreas edificadas presentes na própria base do IBGE (2022) foram excluídas do recorte espacial dos Sertões Cearenses. Além disso, como apoio ao critério de seleção de dados no período seco foi utilizado de precipitação da série temporal do ERA LAND 5 considerando apenas os meses com precipitação mensal inferior a 7 mm. Desta forma, as imagens utilizadas na classificação são frutos de composições de média do período seco.

Quadro 1: Filtros temporais e quantidade de cenas utilizadas.

MAPEAMENTOS RF	Primeiro Filtro Temporal	Segundo Filtro Temporal	Sensores	Quantidade de Cenas
2004 - 2005	1 de agosto a 30 de novembro de 2004	1 de julho a 30 de novembro de 2005	LANDSAT 5 - TM	64
2013 - 2014	1 de agosto a 30 de setembro de 2013	1 de agosto a 30 de setembro de 2014	LANDSAT 8 - OLI	35
2019 - 2020	1 de setembro a 30 de novembro de 2019	1 de agosto a 30 de dezembro de 2020	LANDSAT 8 - OLI	49
2023 - 2024	1 de julho a 30 de	1 de julho a 30 de	LANDSAT 8 - OLI	137

	outubro de 2023	outubro de 2024	LANDSAT 9 - OLI 2	
--	-----------------	-----------------	-------------------	--

Fonte: Elaborado por autor na plataforma GEE.

Para essas faixas temporais foram efetuados cálculos de índices espectrais de água (NDWI), vegetação (SAVI), solo exposto (MBI) e de secas (NDDI) totalizando 10 bandas e índices espectrais para cada imagem da área de estudo a ser classificada. Outro conjunto de dados importantes são advindos do Modelo Digital de Elevação (MDE) NASADEM fruto do reprocessamento do SRTM dispondo de dados de elevação para a região estudada. Foram efetuados cálculos de derivadas de relevo a partir do algoritmo *Terrain Analysis* presente no GEE como a Declividade, Sombreamento e Aspecto, que foram utilizadas no processo de classificação RF.

As classes definidas tiveram por base outros trabalhos na região estudada, como os mapas do IBGE (2022), FUNCEME (2018), Martins et al. (2022) e De Lima et al. (2019). Assim, foram definidas as classes Vegetação de Altitude, Vegetação Densa, Vegetação Seca, Vegetação de Rios, Solo com alto grau de exposição, Areia Exposta e Água sendo ao todo 7 classes. Foram coletadas amostras para treinamento e teste de cada classe a partir de suas características espectrais e visuais, como cor, textura, sombreamento e elevação. Abaixo, é possível observar a chave de interpretação para cada classe e quantidade de amostras de cada mapeamento utilizado (Quadro 2). A quantidade de amostras foi definida seguindo as recomendações de Congalton (1991) que indica a quantidade mínima de 75-100 amostras para coberturas acima de 500 km².

Quadro 2: Classes e quantidade de amostras na classificação RF.

Classes	Crítérios	Amostras (2004-2005)	Amostras (2013-2014)	Amostras (2019-2020)	Amostras (2023-2024)
Vegetação de Altitude	Pixels gerados através de uma máscara SAVI > 0,25 e com a elevação acima de 400 metros.	153	151	157	152
Vegetação Densa	Pixels gerados através de uma máscara SAVI > 0,25 e com a elevação inferior a 400 metros.	152	152	155	154

Vegetação Seca	Pixels em tons de verdes terrosos, normalmente densos, que se destacam na composição colorida falsa cor (SWIR 1, NIR e RED) nos valores min: 0,0924 e max: 0,4.	174	180	176	154
Vegetação de Rios	Pixels em tons de verdes brilhantes ou escuros nas margens de rios que se destacam na composição colorida falsa cor (SWIR 1, NIR e RED) nos valores min: 0,0924, max: 0,4 e gamma:0,16	83	84	80	86
Solo com Alto Grau de Exposição	Pixels em tons de marrom sinalizando solo descoberto que se destacam na composição colorida falsa cor (SWIR 1, NIR e RED) nos valores min: 0,0924 e max: 0,4.	195	197	197	196
Areia	Pixels em tons de amarelos brilhantes que se destacam em áreas de solos expostos que se destacam na composição colorida falsa cor (SWIR 1, NIR e RED) nos valores min: 0,0924, max: 0,4 e gamma:0,16	82	85	92	86
Água	Pixels tons azul marinho ou escuros com formatos circulares (lagos) ou lineares (rios) que se destacam na composição colorida falsa cor (SWIR 1, NIR e RED) nos valores min: 0,0924 e max: 0,4.	129	133	131	86

Fonte: Elaborado pelo autor na plataforma GEE.

Além disso calculou-se bandas e índices estatísticos por redutores (média, mediana, desvio padrão e variância) considerando todas as cenas presentes em cada faixa temporal mapeada, incluindo imagens de períodos úmidos. Tal abordagem gerou 74 bandas preditoras para a classificação supervisionada desenvolvida neste trabalho.

Classificação Random Forest no período seco

O classificador RF segundo Breiman (2001) o define como “uma combinação de preditores de árvores tais que cada árvore depende dos valores de um vetor aleatório amostrado de forma independente e com a mesma distribuição para todas as árvores da floresta, e cada árvore lança um voto unitário para a classe mais popular para a classificação de uma entrada”. A possibilidade da utilização de grandes extensões de florestas aleatórias para tomada de decisão busca preencher a lacuna científica deixada pela classificação CART (Ziegler e König, 2014). Presente na plataforma GEE, o RF foi utilizado para a classificação de imagens para a área estudo levando em consideração o guia prático de Boulesteix et al., (2012) a partir da separação dos dados em 70% para treino e 30% para teste, sendo esta última para a avaliação do modelo.

A redução de dimensionalidade dos dados no processo classificatório foi efetuada a partir do método RFE adaptado na plataforma GEE. Segundo Chen et al., (2024) o objetivo do RFE é eliminar iterativamente características com pequenas contribuições para a classificação, reduzindo assim a redundância e aprimorando o desempenho do classificador.

O índice de Gini foi incorporado para avaliar a quantidade relativa que cada variável contribuiu para a separabilidade entre as classes na classificação supervisionada (SILVA et al., 2021). De início há o treinamento com todas as variáveis preditoras e com base nas variáveis de importância são excluídas aquelas que menos contribuem para a distinção entre as classes.

Validação e escolha dos modelos RF

A partir da construção da matriz de confusão e o uso dos dados de teste foram aplicadas diversas métricas para a avaliação da performance do modelo RF. A acurácia global (1) é representada pela razão entre o total das predições corretas e o total de observações.

$$Acurácia\ Global = \frac{VP + VN}{VP + VN + FP + FN} (1)$$

É importante mencionar que as variáveis presentes na fórmula e na matriz de confusão são definidas como: Verdadeiro Positivo (VP) é representado pela classificação correta onde a observação real era de fato positiva. O Verdadeiro Negativo (VN) é o oposto da situação

anterior sendo a predição negativa para uma determinada classe onde a observação real era de fato negativa. O Falso Negativo (FN) é o caso em que o modelo prevê a classe equivocadamente, mas a observação real é positiva. O Falso Positivo (FP) é o caso em que o modelo prediz positivamente a classe e a observação real é negativa. Com base no cálculo da acurácia (1) é possível calcular o índice Kappa (2) que mensura o quanto a acurácia observada excede em relação a acurácia esperada.

$$Kappa = \frac{Acurácia\ Observada - Acurácia\ Esperada}{1 - Acurácia\ Esperada} (2)$$

Outra métrica importante utilizada é o *recall* (sensibilidade) que avalia a proporção de VP em relação a soma entre VP e FN (3). Basicamente, indica em porcentagem de resultados positivos corretamente identificados pelo modelo dentre todos os resultados que deveriam ter sido identificados como positivos (DE SOUZA et al., 2024).

$$Recall = \frac{VP}{VP + FN} (3)$$

A métrica de validação *precision* (precisão) avalia a proporção entre VP e a soma entre VP e FP (4). Este avaliador indica a porcentagem de resultados positivos corretamente identificados pelo modelo dentre todos os classificados como positivo (DE SOUZA et al., 2024).

$$Precision = \frac{VP}{VP + FP} (4)$$

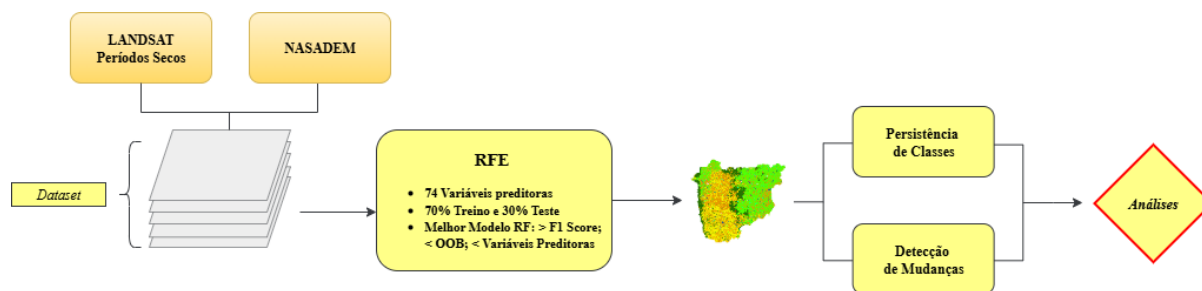
O F1-Score (5) é definido como a média harmônica entre *recall* e a *precision* indicando um único valor para a qualidade geral do modelo (DE SOUZA et al., 2024). Por fim, o erro Out-of-Bag (OOB), produto interno do modelo RF, também foi uma métrica utilizada para entender o comportamento do modelo. É importante mencionar que para fins de comparação com as outras métricas em uma mesma escala foi adotado 1-OOB, sendo assim, quanto maior desta operação o valor menor o erro OOB.

$$F1 = 2 \times \frac{precision \times recall}{precision + recall} (5)$$

A redução de dimensionalidade dos dados no processo classificatório foi efetuada a partir do método Recursive Feature Elimination (RFE) adaptado na plataforma GEE. O índice de Gini foi incorporado para avaliar a quantidade relativa que cada variável contribuiu para a

separabilidade entre as classes na classificação supervisionada (SILVA et al., 2021). De início há o treinamento com todas as variáveis preditoras e com base nas variáveis de importância são excluídas aquelas que menos contribuem para a distinção entre as classes. A partir disso são avaliadas as melhores métricas como Acurácia, Kappa, F1-Score, menor erro OOB a partir dos dados de validação. Por fim, foram definidos os critérios de escolha de melhor modelo: Modelo com maior F1, caso empate, será escolhido o modelo com o menor erro OOB, novamente caso empate, será determinado pela menor quantidade de variáveis preditoras restante no modelo.

A partir dos modelos escolhidos e a classificação dos 4 mapas a posteriori foram efetuadas álgebras de mapas para a geração do mapa de persistência de classes e a detecção de mudanças. A primeira abordagem metodológica consiste no empilhamento dos 4 mapas classificados a fim de entender quais classes persistiram ao longo do tempo. Tal cálculo foi efetuado pelo *Raster Calculador* presente no ArcMap 10.8 a partir desta operação através da condicional $Con((2004 - 2005 == 2013 - 2014) \& (2013 - 2014 == 2019 - 2020) \& (2019 - 2020 == 2023 - 2024), 2004 - 2005, -9999)$. Já o segundo mapa é baseado na operação matemática entre o mapa mais antigo com o mais recente pelo cálculo $((2023 - 2024) \times 10) + (2004 - 2005)$, sendo método que permite compreender as mudanças em uma cobertura terrestre (SINGH, 1999). Como abordagem metodológica, devido às diferenças de estágios da vegetação pela intensidade da seca, entre as duas coberturas, as classes representadas pela vegetação foram unidas através da função *reclassify*. Entretanto, trabalhar com uma classe única de vegetação em detecção de mudanças pode limitar análises mais detalhadas, como estudos de regeneração ou degradação entre classes diferentes. Esse desafio é particularmente relevante para os pesquisadores que atuam com sensoriamento remoto na Caatinga brasileira, pois a transição entre as classes de vegetação pode gerar ambiguidades, dificultando a distinção entre degradação e variações naturais em resposta à seca, ou entre regeneração e respostas naturais ao período chuvoso. Podemos observar na Figura 3 abaixo o fluxograma destacando os principais pontos abordados na metodologia para a elaboração dos resultados.

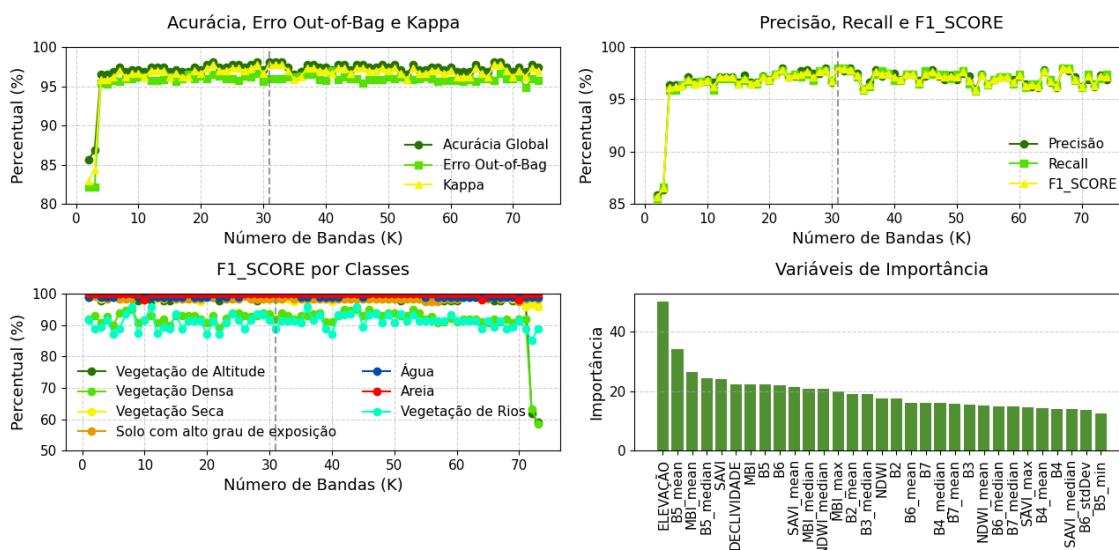
Figura 3: Fluxograma Metodológico.

Fonte: Elaborado pelos Autores.

RESULTADOS E DISCUSSÕES

Validação dos modelos RF

Para a validação referente aos dados de teste do modelo RF, utilizando dados da faixa óptica do Landsat-5 e o MDE NASADEM, do primeiro mapa representativo ao período seco de 2004-2005 a partir dos critérios estabelecidos pelo RFE foi escolhido o modelo com 31 variáveis mais importantes no processo classificação da cobertura terrestre (Figura 4). Foram obtidos resultados satisfatórios para Acurácia Global, Kappa, Erro OOB, *Recall*, *Precision* e F1 sendo obtido pelos valores respectivamente 98,07%; 98,01%; 96,01%; 97,99%; 98,04%; 97,72%.

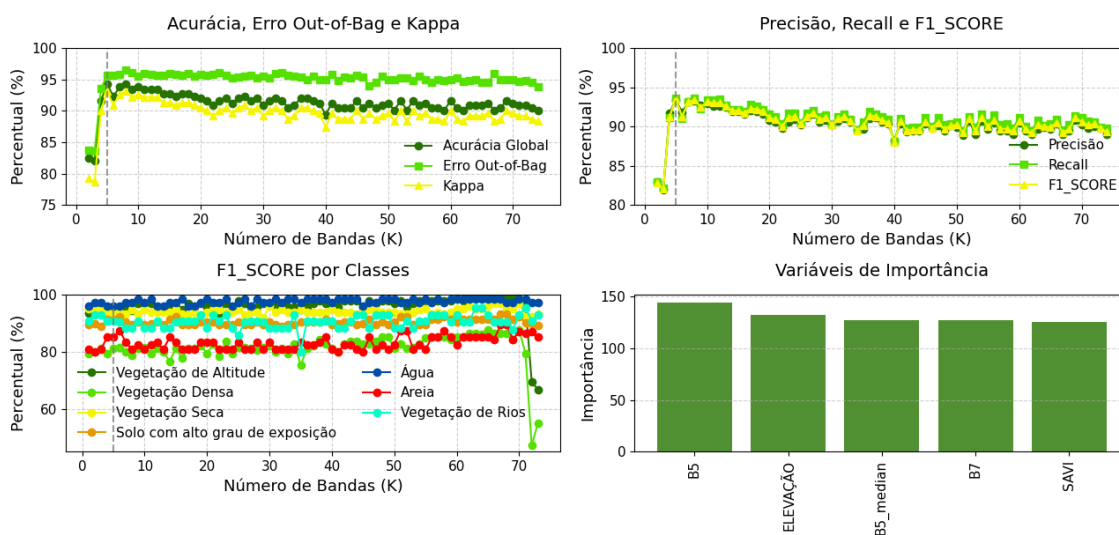
Figura 4: Modelo RF (2004 – 2005) e validação pelos dados de teste.

Fonte: Elaborado pelos autores.

Nota-se que dentre todas as variáveis preditoras a banda elevação relacionada ao relevo assume o protagonismo para a distinção de classes. Outras bandas como o infravermelho próximo (média e mediana) importantes para a detecção de vegetação e o índice de solo exposto MBI também contribuem, mas acabam apresentando um papel secundário em relação à elevação. Além disso, as outras variáveis preditoras apresentam um comportamento de importância mais bem distribuído e parecido em comparação àquelas citadas anteriormente.

Diferentemente do modelo anterior, por questões de disponibilidade dos dados ópticos e melhorias no sistema sensor, foi optado o uso dos dados ópticos do Landsat-8 para o mapeamento de 2013-2014 (Figura 5). É importante mencionar que foi escolhido o melhor modelo com 5 variáveis preditoras para o mapeamento apresentando como valores das métricas de validação Acurácia Global, Kappa, Erro OOB, *Recall*, *Precision* e F1 sendo respectivamente, 94,16%; 93,06%; 95,61%; 93,64%; 93,41%; 93,51%. São valores similares à classificação anterior, porém, com valores inferiores, além disso, apresentam menor robustez em termos de uso de variáveis preditoras.

Figura 5: Modelo RF (2013 – 2014) e validação pelos dados de teste.



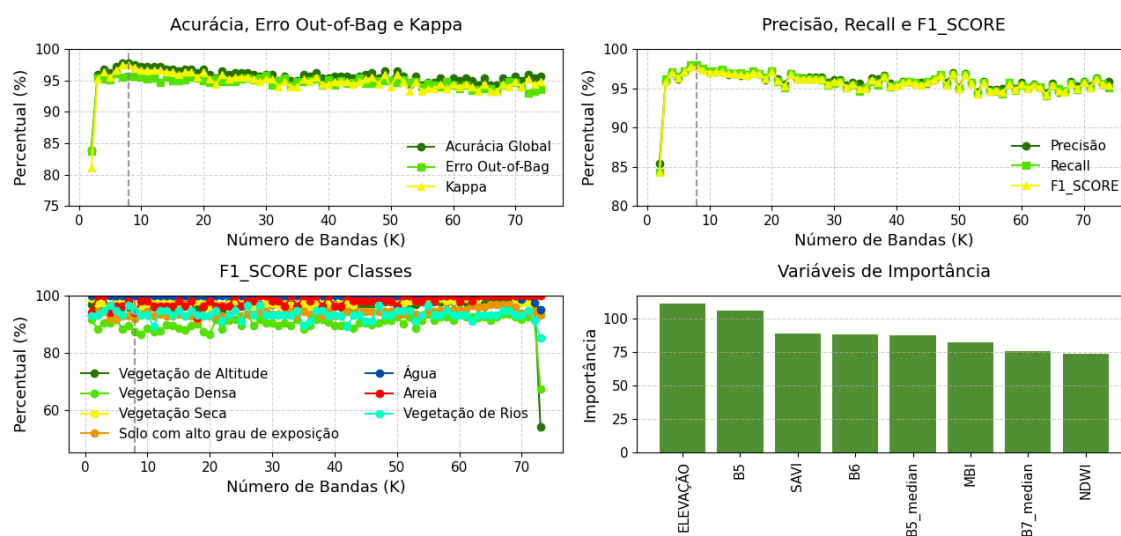
Fonte: Elaborado pelos autores.

As 5 variáveis preditoras selecionadas para o processo de classificação de imagem apresentam valores em magnitude superiores ao processo de modelagem anterior. A banda B5

infravermelho próximo e a sua estatística mediana, o relevo pela elevação e o índice de vegetação SAVI acabam por serem os responsáveis pela distinção das classes mapeadas. A banda B7 infravermelha de ondas curtas 1 que antes não apresentava um protagonismo na modelagem anterior agora apresenta importância significativa ao modelo.

O modelo para o mapa de 2019-2020, que possui os mesmos sensores do modelo anterior, selecionou apenas 8 variáveis preditoras para o processo de classificação (Figura 6). As métricas de validação Acurácia Global, Kappa, Erro OOB, *Recall*, *Precision* e F1 para este modelo tiveram como resultado 97,79%; 97,39%; 95,67%; 97,96%; 97,72%; 97,73%. Comportamentos similares observados nos processamentos anteriores, porém, com valores superiores ao modelo selecionado para o ano de 2013-2014.

Figura 6: Modelo RF (2019 – 2020) e validação pelos dados de teste.



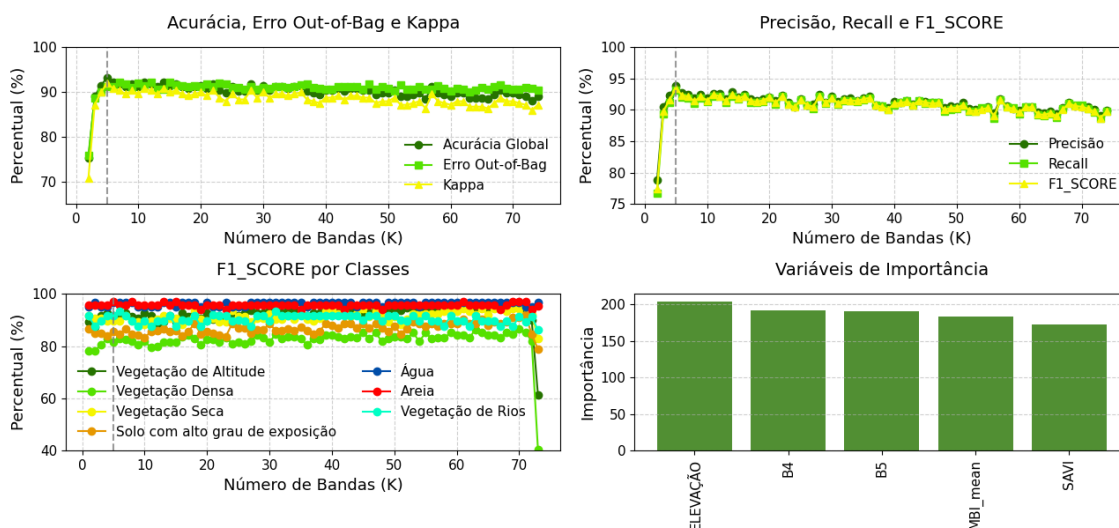
Fonte: Elaborado pelos autores.

As 8 variáveis preditoras destacam-se novamente a elevação e o infravermelho próximo. Além disso, surge no modelo a influência do índice para a detecção de corpos hídricos NDWI e a Banda 7 correspondentes à mediana estatística do infravermelho de ondas curtas 2.

A diferença do último modelo classificado com os demais é a junção dos dados ópticos do Landsat 8 e 9. Neste caso o modelo escolhido selecionou apenas 5 variáveis preditoras na

classificação do mapa (Figura 7) apresentando em valores as métricas de validação Acurácia Global, Kappa, Erro OOB, *Recall*, *Precision* e F1 como resultado respectivamente 93,15%; 91,92%; 91,08%; 93,08%; 93,84%; 93,42%. De todos os modelos, este foi o que teve performance inferior aos demais, mesmo assim, alcançou resultados satisfatórios para a classificação de imagem na área de estudo. Assim como descrito em modelos anteriores, é notório a influência da variável preditiva elevação à classificação. Outras variáveis preditivas que têm forte relação pelas mudanças paisagísticas para a vegetação como o índice de vegetação SAVI, a banda vermelha B4, o infravermelho próximo e o surgimento de áreas com solos mais expostos como o índice MBI tiveram impacto para a classificação desta imagem.

Figura 7: Modelo RF (2023 – 2024) e validação pelos dados de teste.



Fonte: Elaborado pelos autores.

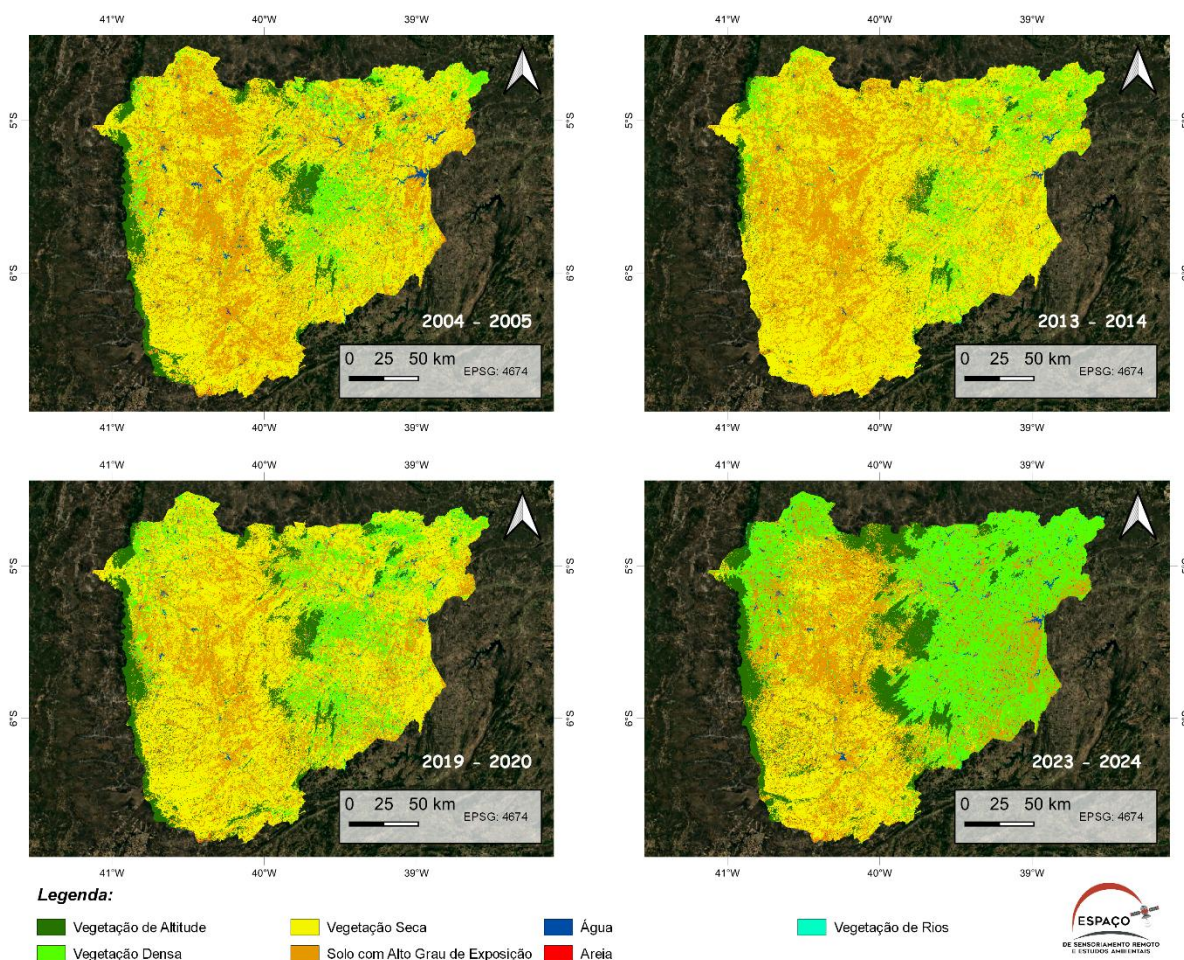
É importante salientar que em todos os processos iteração pelas variáveis predictoras na classificação de imagem RF (Figuras 3, 4, 5, 6), a F1 por classe mostrou dificuldades a distinção das classes de vegetação. Isso fica evidente quando observamos as mesmas tendências de quedas para todos os gráficos as quedas expressivas ao chegarmos em K=73 onde os modelos acabam perdendo a capacidade de detectar padrões importantes ainda que outras métricas se apresentem com valores altos. Este comportamento pode estar relacionado a um desequilíbrio entre as classes, além disso, devido à alta complexidade em que o modelo se assume pode evidenciar um possível processo de *overfitting*. Outro detalhe a destacar é a instabilidade dos

modelos para $K < 5$ para as métricas de validação gerais, porém, um modelo estável sob apenas essas métricas podem esconder eventuais problemas no modelo a partir da alta complexidade.

Mapas de Cobertura da Terra nos Sertões Cearenses

Os 3 primeiros mapas gerados pelos modelos RF (Figura 8) corroboram com os eventos de secas e os fatores de relevo que condicionam a intensificação destes eventos. Segundo autores Marengo et al., (2016) e Gomes e Zanella (2023), no semiárido brasileiro os fenômenos de secas no século XXI ocorreram nos anos 2001, 2003, 2005, 2010, 2012/2013, 2014/2015 e 2016/2017 sendo as primeiras décadas deste século com mais ocorrências quando comparamos com os registros históricos. No caso do mapa mais recente o avanço da vegetação densa sobre as áreas de vegetação seca reflete a classificação de seca fraca ou inexistente para a região estudada conforme o monitoramento para o ano de 2023 e 2024 (CEMADEN, 2024) sendo uma condição natural da vegetação adaptada aos grandes períodos sem chuva na região. Segundo Alves (2007) a vegetação do Bioma Caatinga, presente na região, possui como características mudanças morfológicas e fisiológicas permitindo em épocas mais secas o acúmulo de água em seus tecidos e raízes, a redução do tamanho das folhas e folíolos, esclerofilia e emurchecimento das folhas dando aspecto seco e esbranquiçada sendo estratégias para suportar condições de déficit hídrico.

Figura 8: Mapas de cobertura da terra dos Sertões Cearenses no período seco.



Fonte: Elaborado pelos autores.

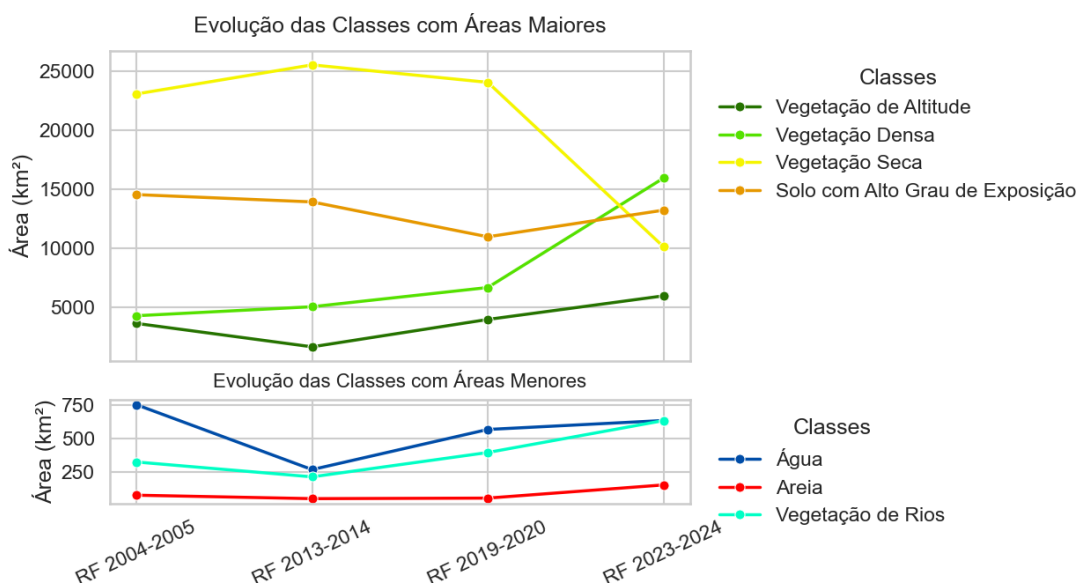
Comportamentos indicam variações entre as classes no período seco das classificações RF nos períodos estudados. É possível notar a presença de quatro classes predominantes na área de estudo Vegetação Seca, Solo com Alto Grau de Exposição, Vegetação Densa e Vegetação de Altitude enquanto as classes representadas pela Água, Vegetação de Rios e Areia são as menos proeminentes.

Há uma relação espacial evidente entre as classes temáticas geradas pelos modelos RF com os compartimentos geomorfológicos e com a precipitação na região. Tanto as vegetações mais secas, como também, os solos mais descobertos são localizados em áreas que menos chove cumulativamente. Outrossim, a vegetação de altitude está mais concentrada nas áreas de maior

elevação que representam os maciços residuais e o Planalto de Ibiapaba no extremo oeste da delimitação espacial dos Sertões Cearenses.

Na seca prolongada de 2012 a 2017, o mapa representativo para o período seco 2013-2014 foi o pico para as classes Vegetação Seca (25537,7 km²), ao menos tempo que houve as mínimas para as classes Água (267,3 km²), Vegetação de Rios (212,8 km²) e Vegetação de Altitude (1636,1km²). Este padrão muda com o mapa classificado 2023-2024 onde a Vegetação Densa acaba sendo a classe mais predominante para a área de estudo enquanto a classe Vegetação Seca cai drasticamente. Além disso, as classes Água e Vegetação de Rios crescem no último mapa em comparação com o mapa representativo 2013-2014 o que corrobora com o monitoramento de secas do CEMADEM pelo enfraquecimento da seca para este período. Mesmo com o aumento da classe água com o abrandamento da seca na região, ainda assim, é observável que não houve uma recuperação que alcançasse o pico para esta classe (755,1km²) presente no primeiro mapa 2004-2005 da série histórica. Apesar de haver tendência de crescimento para a Vegetação de Rios como forma de proteção de rios a sua interpretação é necessário cautela pois podem envolver espécies exóticas, como a algaroba, presente no semiárido brasileiro. Nesse sentido, análises locais são essenciais para a identificar se a vegetação corresponde a mata ciliar típica da região ou espécies invasoras. Na Figura 9 abaixo é possível observar as séries temporais das classes mapeadas.

Figura 9: Séries temporais das classes em km² nos Sertões Cearenses.



Fonte: Elaborado pelos autores.

A classe Areia em comparação do primeiro mapa com o mais recente duplicou a sua área (75 km² para 152 km²) sendo um indicativo da ampliação de núcleos de desertificação e degradação na região estudada sendo importante a discussão de planos de ação de tomadores de decisão para a recuperação ambiental nessas áreas.

Mapa de Persistência de Classes nos Sertões Cearenses

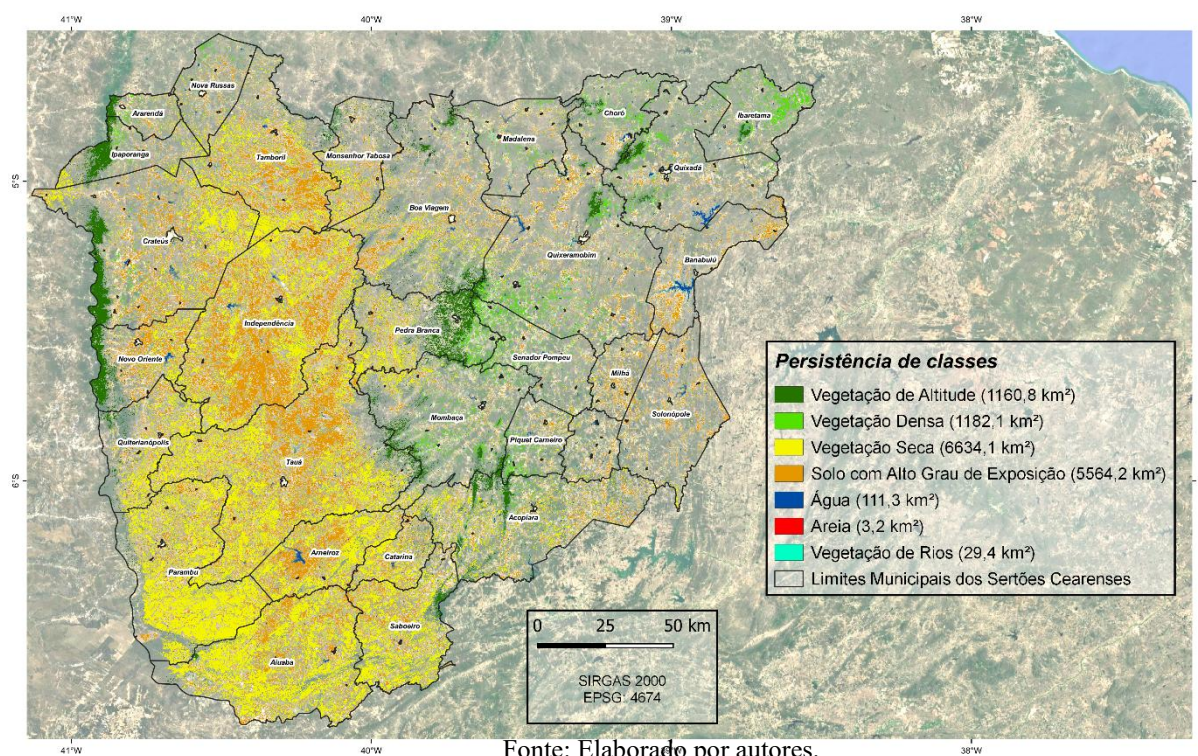
Nos Sertões Cearenses o mapa de persistências de classes mostra as classes mais resilientes em termos de variabilidade da cobertura da terra em termos de intra-mapas 2004-2005, 2013-2014, 2019-2020 e 2023-2024 (Figura 10). As partir do empilhamento dos dados é possível notar a predominância das classes Vegetação Seca (6634,1 km²) e Solo com Alto de Exposição (5564,2 km²) concentrando-se na parte oeste da área de estudo em compartimentos geomorfológicos das depressões e patamares sertanejos. Essa persistência da vegetação mais seca e solos mais descobertos coincidem com as áreas que precipitam menos como apresentada na área de estudo sendo coberturas de terras secas. Tal caracterização é essência da vegetação xerófila que sobrevivem por longos períodos com balanço hídrico negativo.

Além disso, é perceptível no mapa a manutenção da classe Vegetação de Altitude (1160,8 km²) em áreas mais elevadas como o Planalto de Ibiapaba no extremo oeste e nos

maciços e serras na parte central. A classe Vegetação Densa (1182,1 km²) normalmente tem sido situada, predominantemente, na porção leste principalmente sendo vizinhos aos limites de vegetação de altitude e em áreas que acabam apresentando certos graus de variabilidade em suas coberturas que não foram representadas no mapa ('espaços vazios preenchidos pelas imagens de satélite disponibilizados pela Google').

Grande parte da classe Água persistente concentram-se principalmente em áreas represadas como açudes sendo visível aquelas presentes nos municípios de Arneiroz (CE), Quixadá (CE) e Banabuiú (CE). As classes Vegetação de Rios (29,4 km²) e Areia (3,2 km²) acabam por serem coberturas já consolidadas diante dos mapas trabalhados.

Figura 10: Persistência de Classes nos Sertões Cearenses.



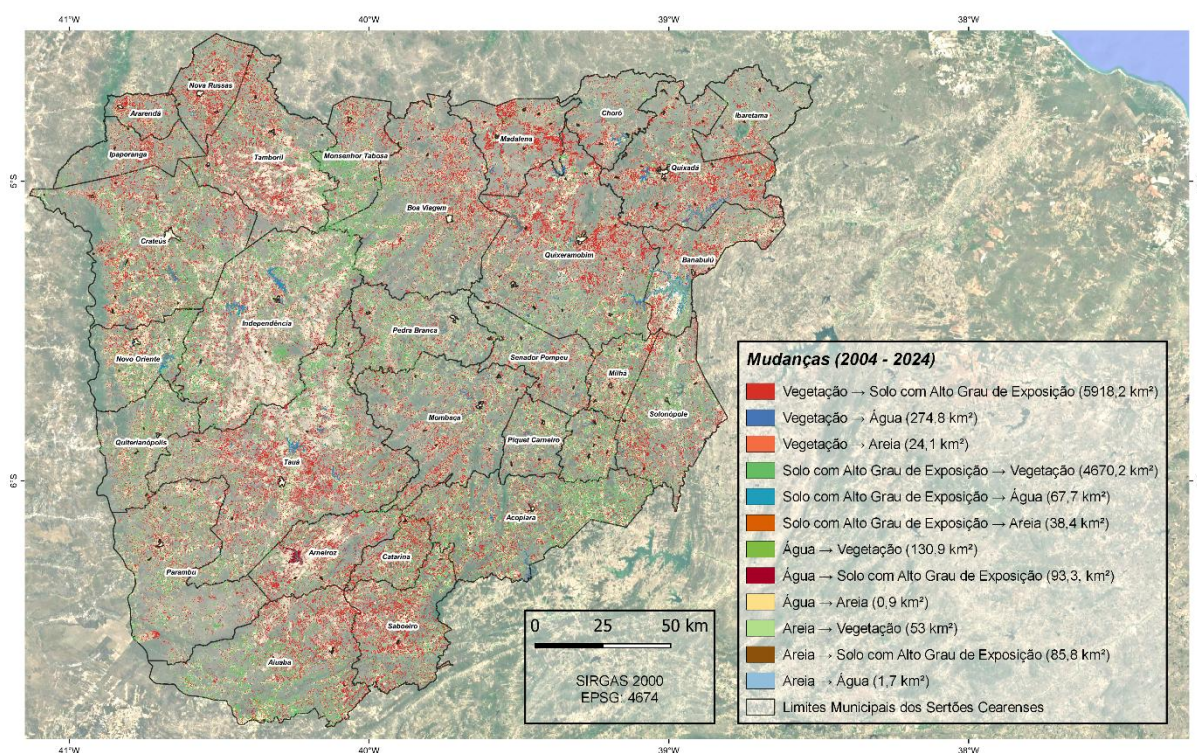
Deteção de Mudanças (2004-2005 e 2023-2024)

É perceptível as mudanças ocorridas quando comparamos os mapas mais antigo com o mais recente na região (Figura 11). Muitas mudanças podem ter relação com a sazonalidade, como as mudanças entre vegetação e solos com alto grau de exposição e as coberturas aquáticas

para outras e vice-versa. Por outro lado, mas o que tem se percebido são perdas da superfície aquática, perdas florestais, surgimento da classe Areia que podem ter relação com uma possível degradação ambiental e a ampliação de núcleos de desertificação na região.

A expansão da classe Vegetação para Solo com Alto Grau de Exposição (5918,2 km²) é a mais predominante nos Sertões Cearenses seguido da classe Solo com Alto Grau de Exposição para Vegetação (4670,2 km²) o que pode evidenciar recuperação ao longo dos anos na área de estudo. Outras mudanças despertam uma preocupação ambiental que podem desencadear problemas sociais na área de estudo como a transformação de áreas vegetadas em porções mais arenosas, e, perdas da cobertura aquática, o que pode comprometer a segurança e o gerenciamento hídrico na região.

Figura 11: Detecção de Mudanças nos Sertões Cearenses (2004 – 2024).



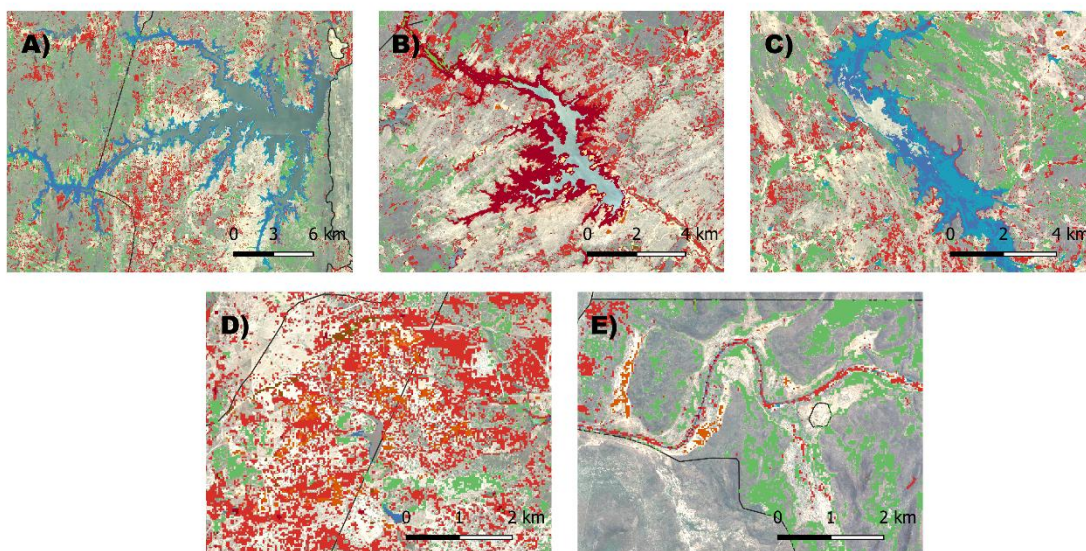
Fonte: Elaborado pelos autores.

Como exemplo, na Figura 12, é possível ver mudanças e algumas problemáticas já relatadas na área de estudo. O açude Castanhão presente no município de Banabuiú (CE) (Figura 12 – A) mostraram um relativo aumento da cobertura da água indicando uma

recuperação das ocorrências de secas à priori. Por outro lado, como observado nos mapas classificados há uma ocorrência de áreas mais descobertas próximas às margens do açude, outros autores como De Albuquerque et al., (2019), que estudaram a região notaram este comportamento na faixa temporal de 2004 a 2014 acendendo o alerta para o manejo adequado dos solos para fins de recuperação ambiental. No caso do açude em Independência (CE) (Figura 12 – C) ocorrem dois fenômenos interessantes como a transformação da vegetação e solos descobertos em água indicando recuperação das secas, porém, há perda da vegetação nos limites do corpo hídrico. Assim como em tantos rios na região como o Jaguaribe principal em termos de extensão na região, em Crateús (CE) o rio Poti (Figura 12 – E) tem ocorrências de perda da vegetação protetora podendo gerar focos de assoreamento no canal. Além disso, são observados núcleos de desertificação na região com o aumento de porções de areia.

Ainda mais alarmante, dentro desse contexto, como já estudado por Brasil et al., (2017), o açude Arneiroz II presente no município de Arneiroz (CE) mostram como a perda vegetal têm impacto significativo no aumento de processos erosivos na região, principalmente nas margens do corpo hídrico no contexto regional. As mudanças ocorridas na figura 12 – B relatam a expansão da classe representante aos solos mais descobertos com focos de classe areia podendo indicar degradação e surgimento de núcleos de desertificação. Isso não é diferente quando analisamos na divisa entre Milhã (CE) e Solonópole (CE) (Figura 12 – D) onde há a expansão de solos mais descobertos com areias ampliando áreas desertificadas ameaçando a produtividade agrícola, a segurança alimentar e hídrica regional.

Figura 12: Possíveis focos de degradação e desertificação em Sertões Cearenses



Fonte: Elaborado pelos autores.

Contudo, trabalhos de campo com ações de recuperação ambiental e monitoramento contínuo dessas áreas sensíveis são fundamentais para a manutenção da rica biodiversidade da Caatinga, a saúde dos rios, a redução dos focos de erosão e desertificação que tanto podem ser frutos de ações antropogênicas, como também, de efeitos já sentidos pelas mudanças climáticas na esfera regional.

CONSIDERAÇÕES FINAIS

Os modelos selecionados pelo modo iterativo do RFE reduziram significativamente o uso das variáveis preditoras no processo de classificação de imagem. Entretanto, quando combinamos o sensor TM do Landsat 5 com o MDE NASADEM observamos um valor relativamente alto de seleção ($K=31$ para o modelo de 2004-2005) padrão que não foi observado nos outros modelos utilizando os sensores do Landsat 8 e 9 ($K=5$, $K=8$ e $K=5$, para os modelos 2013-2014; 2019-2020; 2023-2024, respectivamente), requerendo mais investigação, neste caso, se o número de variáveis preditoras também tem relação com a qualidade dos dados e as resoluções do sistema-sensor.

O mapeamento multitemporal nos Sertões Cearenses trouxeram alguns padrões interessantes quanto as variáveis importantes do modelo RF. Em todos os modelos selecionados a elevação assumiu o protagonismo na distinção das classes. É possível observar também a influência do infravermelho próximo seguido de SAVI, pois tem relação com a variabilidade paisagística da vegetação na área de estudo. Outras variáveis preditoras como MBI, NDWI e infravermelho de ondas curtas para o processo de modelagem também foram observados sendo possivelmente influenciados pelas secas dada a retração das superfícies e a solos mais expostos diante da condição natural da vegetação da região.

Essa variabilidade paisagística fica muito bem demarcada quando comparamos as classes em séries temporais pelos quatro mapas. As mudanças ocorridas principalmente no mapa mais recente (2023-2024) mostram como é tão difícil mapear áreas com grande variabilidade onde muitas das vezes pode confundir a interpretação de especialistas da área. Houve, neste caso, a inversão de predominância entre as classes Vegetação Seca e Vegetação Densa que tem relação com o enfraquecimento das secas nesta região. Mesmo nesse contexto, com a tendência de crescimento da classe Água ainda não foi possível recuperar o patamar inicial apresentado no primeiro mapa (2004-2005). O crescimento das classes Areia e Solo com Alto Grau de Exposição observado também no período seco de 2023-2024 requer também atenção, uma vez que, para além da variabilidade regional, podem ter relação com a degradação ambiental e a ampliação de extensões de núcleos de desertificação.

O mapa de persistência de classes trouxe insights para estudos futuros na área de estudo devido à capacidade de fugir dessas variabilidades paisagísticas trazendo padrões importantes como a fixação da vegetação mais seca concentrados na parte oeste da área de estudo e a importância das formas de relevo como áreas mais elevadas para a fixação de tipos de coberturas existentes na região. O mapa de detecção de mudanças comparando o mapa mais antigo com o mais recente trouxe informações já corroboradas por outros autores como exemplo a degradação às margens de represamentos na região. Outras transições entre classes como as retrações das extensões da vegetação e perdas de superfícies aquáticas em detrimento de aumento de solos mais descobertos e as extensões mais arenosas podem indicar focos não apenas da questão da variabilidade regional, mas também, de possível ampliação de áreas degradadas e desertificadas na região.

AGRADECIMENTOS

Agradeço ao Laboratório ESPAÇO de Sensoriamento Remoto e Estudos Ambientais e à professora doutora Carla pelo suporte e a orientação ao longo deste trabalho. Também expressei meu agradecimento e o reconhecimento à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo financiamento da bolsa de doutorado ao aluno primeiro autor, o que permitiu o desenvolvimento de metodologias de baixo custo com potencial de contribuir para o monitoramento da cobertura da terra nos Sertões Cearenses.

REFERÊNCIAS BIBLIOGRÁFICAS

ALVES, J. J. Geoecologia da caatinga no semiárido do nordeste brasileiro. *CLIMEP – Climatologia e Estudos da Paisagem*, Rio Claro, SP, v. 2, n. 1, p. 58-71, jan./jun. 2007.

BOTELHO, C. L. *Revista do Instituto do Ceará*, Fortaleza, 1991.

BRASIL, J. B.; ROSA, G. Q.; RIBEIRO FILHO, J. C.; SILVA, C. V. T. Estimativa de perda de solo na bacia do açude Arneiroz II por ferramentas de geoprocessamento. *Revista Ambiência*, v. 13, n. 2, p. 503-517, 2017. Disponível em: <https://revistas.unicentro.br/index.php/ambiencia/article/view/4703/pdf>. Acesso em: 15 abr. 2023.

BREIMAN, L. Random Forests. *Machine Learning*, v. 45, p. 5–32, 2001. DOI: <https://doi.org/10.1023/A:1010933404324>.

CEMADEN. Mapa interativo de monitoramento de secas. Disponível em: <https://mapasecas.cemaden.gov.br/#iis3>. Acesso em: 7 jan. 2025.

CHEN, C.; LIANG, J.; SUN, W.; YANG, G.; MENG, X. An automatically recursive feature elimination method based on threshold decision in random forest classification. *Geo-Spatial Information Science*, p. 1–26, 2024. DOI: <https://doi.org/10.1080/10095020.2024.2387457>.

CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing of Environment*, v. 37, n. 1, p. 35–46, 1991.

COPERNICUS CLIMATE CHANGE SERVICE. ERA5-Land monthly averaged data from 1950 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Disponível em: https://developers.google.com/earth-engine/datasets/catalog/ECMWF_ERA5_LAND_MONTHLY_AGGR?hl=en.

DE ALBUQUERQUE, A. M. et al. Balanço hídrico como ferramenta de gerenciamento de recursos hídricos: aplicação na área de influência direta do açude Castanhão-CE. *Revista da Casa da Geografia de Sobral (RCGS)*, v. 21, n. 2, p. 454-466, 2019.

DE LIMA, V. R. P.; DA SILVA-SEABRA, V.; ALBUQUERQUE-XAVIER, R.; DA CONCEIÇÃO-DORNELLAS, P. Análise do padrão espacial do uso e cobertura do solo e da fragmentação e conectividade da vegetação no semiárido do Nordeste brasileiro: bacias dos rios Taperoá e Alto Paraíba-PB. *Revista de Estudios Andaluces*, n. 37, p. 25–49, 2019. DOI: <https://doi.org/10.12795/rea.2019.i37.02>. Disponível em: <https://revistascientificas.us.es/index.php/REA/article/view/5595>. Acesso em: 4 jan. 2025.

DE SOUZA, M. K. G.; FAITA PINHEIRO, M. M.; FURUYA, D. E. G.; OSCO, L. P.; JÚNIOR, J. M.; GONÇALVES, W. N.; RAMOS, A. P. M. Mapeamento de rios em imagens RGB com aprendizagem de máquina supervisionada. *Revista Tamoios*, São Gonçalo, v. 20, n. 1, 2024. DOI: <https://doi.org/10.12957/tamoios.2024.66545>. Disponível em: <https://www.e-publicacoes.uerj.br/tamoios/article/view/66545>. Acesso em: 1 mar. 2025.

FERNANDES, D. S.; HEINEMANN, A. B.; PAZ, R. L.; AMORIM, A. O.; CARDOSO, A. S. Índices para a quantificação da seca. Santo Antônio de Goiás, GO: EMBRAPA, 2009. (Documentos, 244). Disponível em: <https://www.infoteca.cnptia.embrapa.br/bitstream/doc/663874/1/doc244.pdf>. Acesso em: 6 set. 2024.

FUNCEME – FUNDAÇÃO CEARENSE DE METEOROLOGIA E RECURSOS HÍDRICOS. Mapa fitoecológico do Estado do Ceará. Fortaleza: FUNCEME, 2018. Escala 1:1.200.000. Disponível em: http://www.funceme.br/wp-content/uploads/2019/02/15-Mapa_CE_Fltoecologico_A2.pdf. Acesso em: 12 jun. 2025.

FUNK, C. et al. O clima ameaça a precipitação infravermelha com estações — um novo recorde ambiental para o monitoramento de extremos. *Scientific Data*, v. 2, 150066, 2015. DOI: <https://doi.org/10.1038/sdata.2015.66>.

GOMES, F. I. B. P.; ZANELLA, M. E. Histórico, causas e características da semiaridez do Nordeste do Brasil. *Geografares*, n. 37, 1 dez. 2023. Disponível em: <http://journals.openedition.org/geografares/10409>. Acesso em: 6 set. 2024.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Mapeamento de recursos naturais do Brasil. Escala 1:250.000, 2023.

MARENGO, J. A.; CUNHA, A. P.; ALVES, L. A seca de 2012-15 no semiárido do Nordeste do Brasil no contexto histórico. *Climanálise*, 2016.

MARTINS, M. M. A. et al. Mapeamento do uso e cobertura do solo do município de Saboeiro (CE) através de processamento na nuvem no GEE. *Revista Continentes*, n. 22, p. 66-85, out. 2023. DOI: <https://doi.org/10.51308/continentes.v1i22.410>. Disponível em: <https://www.revistacontinentes.revistacontinentes.com.br/index.php/continentes/article/view/410>. Acesso em: 3 jun. 2024.

MELO, J. C. O fenômeno El Niño e as secas no Nordeste do Brasil. *Raízes: Revista de Ciências Sociais e Econômicas*, v. 162, n. 20, p. 13-21, nov. 1999. DOI: <https://doi.org/10.37370/raizes.1999.v.162>.

NASA JPL. NASADEM Merged DEM Global 1 arc second V001 [Data set]. NASA EOSDIS Land Processes DAAC, 2020. DOI: https://doi.org/10.5067/MEaSURES/NASADEM/NASADEM_HGT.001.

OYAMA, M. D.; NOBRE, C. A. A new climate-vegetation equilibrium state for Tropical South America. *Geophysical Research Letters*, v. 30, n. 23, p. 2199, 2003.

SILVA, A. G. P.; CREMON, É.; BOGGIONE, G. A.; ALVES, F. C. Integração de imagens orbitais ópticas e SAR com processamento em nuvem no mapeamento da cobertura da terra no Cerrado. *Revista Geoaraguaia*, Barra do Garças, MT, v. 11, p. 85-106, ago. 2021.

SINGH, A. Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, v. 9, n. 6, p. 989-1003, 1988.

TAVARES, V. C.; ARRUDA, Í. R. P.; SILVA, D. G. Desertificação, mudanças climáticas e secas no semiárido brasileiro: uma revisão bibliográfica. *Geosul*, Florianópolis, v. 34, n. 70, p.