

Mapeamento de cobertura e uso da terra no bioma Pampa utilizando diferentes sensores orbitais e classificador *Random Forest*

Land cover and land use mapping in the Pampa biome using different orbital sensors and Random Forest classifier

Patricia Michele Pereira Trindade
Coordenação Espacial Sul (COESU)/Instituto Nacional de Pesquisas Espaciais (INPE)
patriciatrindade.sr@gmail.com

Daniela Wancura Barbieri Peixoto
Coordenação Espacial Sul (COESU)/Instituto Nacional de Pesquisas Espaciais (INPE)
daniwbarbieri@gmail.com

Greice Vieira Silveira
Coordenação Espacial Sul (COESU)/Instituto Nacional de Pesquisas Espaciais (INPE)
greice.v.silveira@gmail.com

Tatiana Mora Kuplich
Coordenação Espacial Sul (COESU)/Instituto Nacional de Pesquisas Espaciais (INPE)
tatiana.kuplich@inpe.br

Igor da Silva Narvaes
Coordenação Espacial Sul (COESU)/Instituto Nacional de Pesquisas Espaciais (INPE)
igor.narvaes@inpe.br

Resumo

A proposta da presente pesquisa foi testar o algoritmo *Random Forest* para classificação do uso e cobertura da terra em uma área com grande variação de declividade no Bioma Pampa a partir de dados ópticos/Sentinel 2, dados termais/Landsat 8 e Modelo Digital de Elevação ALOS PALSAR. O município de Caçapava do Sul foi definido como área piloto para desenvolvimento da pesquisa, sendo considerado a “capital gaúcha da geodiversidade”. A metodologia proposta seguiu cinco etapas principais: 1 - segmentação de imagens, 2 - treinamento, 3 - cálculo de estatísticas zonais para cada segmento, 4 - classificação e 5 - validação. Foram geradas doze classificações de cobertura da terra com diferentes combinações de dados. A partir do cruzamento entre amostras de referência e mapas classificados foi possível gerar as métricas de acurácia, dentre elas a acurácia global (AG). O melhor desempenho geral verificou-se na classificação realizada a partir da combinação de bandas ópticas e MDE, com 84,59% de AG com diferença estatisticamente significativa entre as demais classificações. Neste sentido, destaca-se a importância do modelo digital de elevação combinado aos dados ópticos para o mapeamento de uso e cobertura da terra de regiões de maiores variações de relevo.

Palavras-chave: Sentinel 2, Infravermelho Termal, ALOS, aprendizado de máquina.

Abstract

The objective of the present research was to test the Random Forest algorithm for the classification of land use and land cover in an area with great declivity variation in the Pampa biome using optical

data/Sentinel 2, thermal data/Landsat 8, and Digital Elevation Model ALOS PALSAR. The town of Caçapava do Sul was chosen as the pilot area for the development of the research, being considered the “state’s geodiversity capital”. The proposed methodology followed five main steps: 1 – segmentation of images, 2 – training, 3 – calculation of zonal statistics for each segment, 4 – classification, and 5 – validation. Twelve classifications of land cover were generated with different combinations of data. Crossing reference samples and classified maps made it possible to generate the accuracy metrics, among them the Global Accuracy (GA). The best general performance was ascertained in the classification achieved with the combination of optical bands and DEM, with 84,59% of GA, with the statistically significant difference among the other classifications. In this sense, we highlight the importance of the digital terrain elevation model combined with the optic data for the mapping of land use and land cover of regions with greater relief variation.

Keywords: Sentinel 2, Thermal Infrared, ALOS, Machine Learning.

Introdução

A disponibilidade de informações espaciais sobre as mudanças e atual cenário do uso e cobertura da terra dos biomas brasileiros é essencial para formulação de políticas públicas para planejamento e gestão dos recursos naturais. Por muitos anos, as iniciativas para preservação eram focadas no bioma com maior representatividade de área e vegetação natural arbórea, a Amazônia. Enquanto isso, os demais biomas também sofriam alteração na sua cobertura natural, no entanto, as iniciativas de mapeamentos são mais recentes (SOUZA JR *et al.*, 2020). Monitoramentos globais também vêm sendo produzidos, tais como o GFW (GLOBAL FOREST WATCH, 2023), para o monitoramento de florestas, e o *Forest Early Warning System in the Tropics* (JICA-JAXA, 2023)), voltado para a região dos trópicos, por exemplo.

No Brasil, algumas iniciativas de mapeamentos de cobertura da terra e supressão da vegetação ocorridas podem ser citadas: Inventário Nacional de 1994, 2002 e 2010 (ALMEIDA FILHO *et al.*, 2005), mapas do Instituto Brasileiro de Geografia e Estatística (IBGE) de 2000, 2010, 2012, 2014, 2016 e 2018, mapas gerados pelo projeto MAPBIOMAS, com série histórica de 1985 a 2022 (SOUZA JUNIOR *et al.*, 2020), monitoramentos realizados pelo Instituto Nacional de Pesquisas Espaciais (INPE), como o PRODES, desde 1988, o DETER, a partir de 2004 (ALMEIDA *et al.*, 2022) e, recentemente, o Monitoramento Ambiental dos Biomas Brasileiros por satélites: Mata Atlântica, Caatinga, Pampa e Pantanal (INPE, 2019).

Recentemente, o mapeamento dos desmatamentos e supressão da vegetação nativa ocorridos em todos os biomas brasileiros é realizado pelo INPE e pode ser consultado na plataforma *TerraBrasilis* (ASSIS *et al.*, 2019), sendo que o incremento das áreas desmatadas é disponibilizado anualmente. Os mapas são gerados a partir da combinação de

pré-processamento de um conjunto de imagens Landsat 8, seguido de vetorização manual baseada na interpretação visual (ALMEIDA *et al.*, 2022).

A evolução de metodologias que auxiliem na automatização do processo de mapeamento, tanto da supressão da vegetação nativa quanto das mudanças de uso e cobertura da terra, tornam-se necessárias e um dos classificadores baseados em *machine learning* que vem demonstrando grande aderência entre os pesquisadores, principalmente na literatura internacional, é o *Random Forest* (BELGIU; DRAGUT, 2016). A partir dele, é possível obter classificações confiáveis com previsões derivadas de um conjunto de árvores de decisão (BREIMAN, 2001), apresentando excelentes resultados e rápido processamento (RODRIGUEZ-GALIANO *et al.*, 2012; DU *et al.*, 2015). Diversos estudos em biomas distintos investigaram a utilização do *Random Forest* para classificação a partir de diferentes conjuntos de dados, como: sensor OLI Landsat 8 (SOTHE *et al.*, 2017; PAN *et al.*, 2020); sensor MSI Sentinel 2 (DINIZ *et al.*, 2020; WASNIEWSKI *et al.*, 2020; MA *et al.*, 2021; PEREIRA *et al.*, 2023); dados do Modelo Digital de Elevação - MDE (GISLASON *et al.*, 2006; CORCORAN *et al.*, 2013); e infravermelho termal (SUN; SCHULZ, 2015; EISAVI *et al.*, 2015).

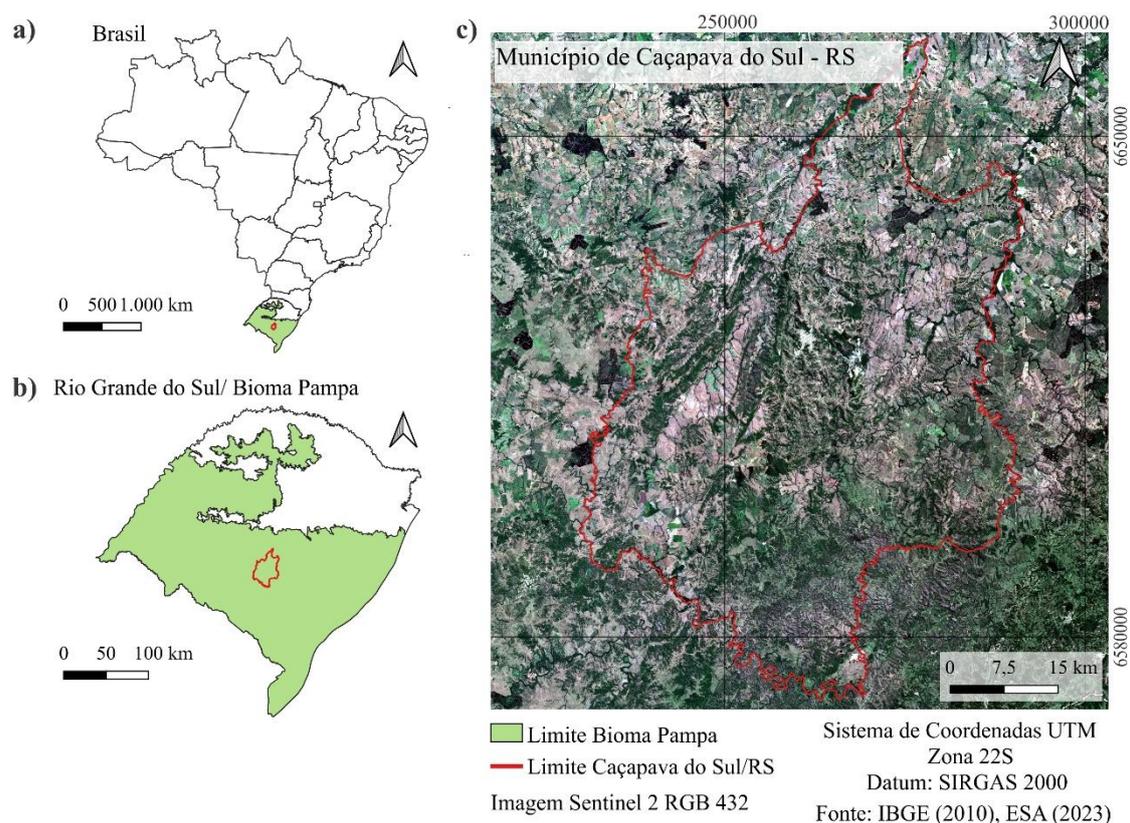
Neste sentido, parte-se do pressuposto de que os dados termais e de altimetria associados às imagens de reflectância de superfície podem trazer valores mais elevados de acurácia para mapas de cobertura da terra gerados a partir do *Random Forest*. A presente pesquisa teve por objetivo utilizar técnicas de classificação baseadas em *machine learning*, em específico o classificador *Random Forest*, a partir da combinação de dados ópticos, termais e altimetria para classificação da cobertura da terra no município de Caçapava do Sul, localizado no bioma Pampa. Com isto, pretende-se contribuir para aumentar o processo de automatização dos mapeamentos que já vem sendo realizados em relação ao monitoramento ambiental dos biomas brasileiros, em especial, o bioma Pampa.

Caracterização da área de estudo

O município de Caçapava do Sul localiza-se na metade sul do Estado do Rio Grande do Sul (Figura 1), integralmente no bioma Pampa; apresenta aproximadamente 3.048 km² de área total e população estimada de 33.476 habitantes (IBGE, 2022). O município encontra-se na microrregião Serra do Sudeste, em uma transição das regiões geomorfológicas da Depressão Central e Planalto Sul-riograndense, sendo encontradas quatro unidades geomorfológicas

(CPRM, 2006; STRECK *et al.*, 2008): Depressão do Rio Jacuí, Planície Alúvio-Coluvionar, Planalto Rebaixado Marginal e Planaltos Residuais Canguçu-Caçapava do Sul.

Figura 1. Localização do município de Caçapava do Sul. a) Caçapava do Sul no Brasil; b) Caçapava do Sul no Rio Grande do Sul; c) Composição colorida RGB 432 para o município de Caçava do Sul.



Fonte: Elaboração própria.

A região apresenta uma grande diversidade geológica, o que, associado a relevos ondulados a forte ondulados e cerros, que se destacam como altos topográficos, deu origem a solos como Neossolos Regolíticos e Neossolos Litólicos (STRECK *et al.*, 2008). De modo geral, os topos são convexos, com algumas ocorrências de topos que configuram cristas. As vertentes são íngremes e muitas se apresentam com afloramentos rochosos (CPRM, 2010), sendo uma característica que garante à região um ambiente propício para o desenvolvimento de espécies endêmicas da flora e da fauna do bioma Pampa.

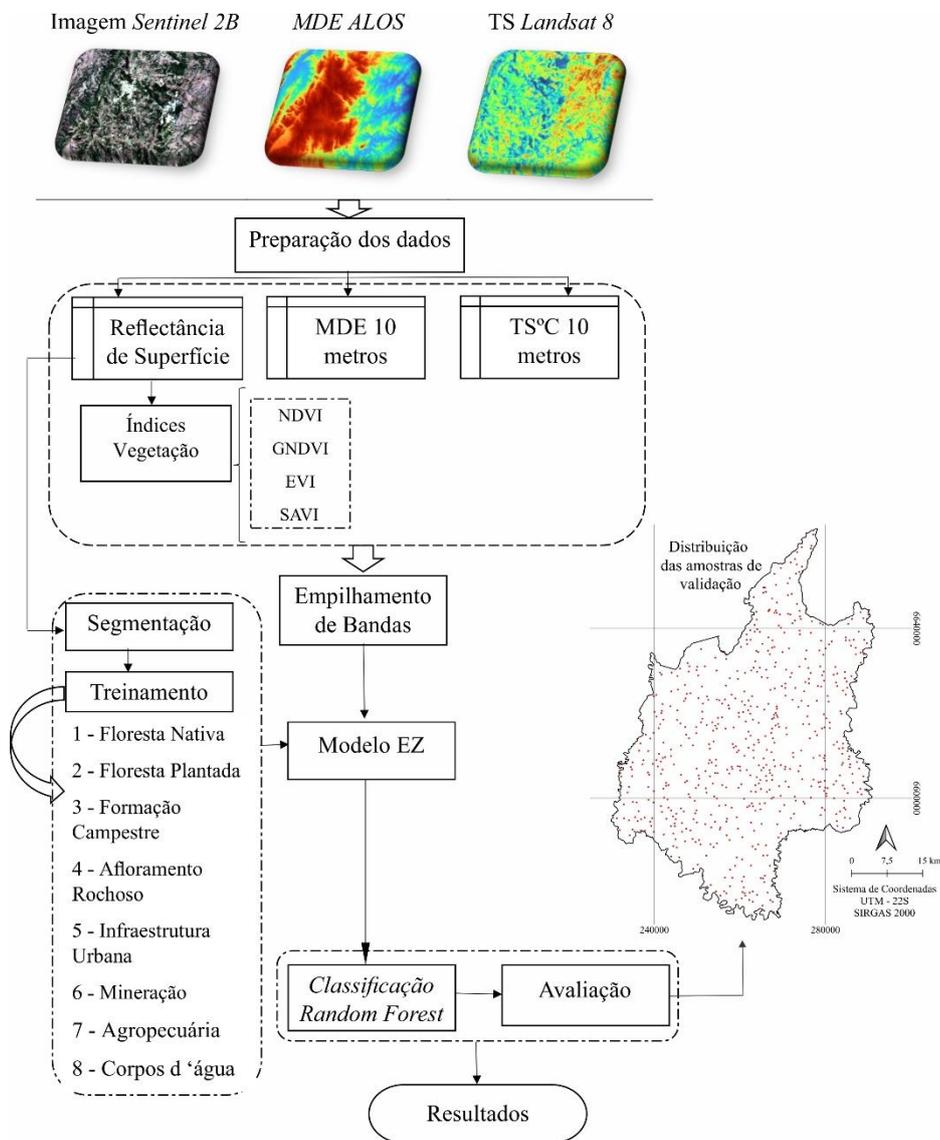
Dadas estas características, o reconhecimento da geodiversidade e geopatrimônio do município garante o potencial turístico da região, com destaque para práticas como trilhas, escaladas, observação de aves e geoturismo. O município recebeu o título oficial de “Capital

Gaúcha da Geodiversidade” através da Lei Ordinária Estadual número 14.708 de 15 de julho de 2015, a qual contou com a iniciativa de inúmeros profissionais das geociências e Assembleia Legislativa do Rio Grande do Sul (BORBA, 2017).

Metodologia

A Figura 2 apresenta a sequência de processos empregados para a elaboração de classificação de uso e cobertura da terra da área de estudo a partir da combinação de diferentes dados de entrada nos modelos gerados com classificador *Random Forest*.

Figura 2. Fluxograma dos procedimentos metodológicos



Fonte: Elaboração própria.

A disponibilidade de imagens de diferentes estações do ano foi consultada (período 2021-2022) para verificar as diferenças que a sazonalidade poderia causar nos mapas gerados. Deste modo, os critérios para seleção das imagens ópticas e termais foram: disponibilidade de imagens sem nuvens e datas de aquisição próximas entre os satélites. Assim, foi possível selecionar duas imagens, referentes a maio/outono e fevereiro/verão. Os processamentos das imagens foram realizados no *software* QGIS 3.22.10, Aplicativos Sen2Cor v2.11 (SEN2COR, 2022) e OTB 6.6.1. Os principais dados utilizados foram:

- Imagens *Sentinel 2B* (resolução espacial 10 m) – bandas 2 (azul), 3 (verde), 4 (vermelho) e 8 (infravermelho próximo) do sensor óptico *MSI (Multispectral Imager)* para as datas de 02/05/2021 e 16/02/2022;
- Imagens *Landsat 8* – banda 10 (infravermelho termal) em temperatura de superfície (TS) do sensor *TIRS (Thermal Infrared Sensor)* para as datas 07/05/2021 e 19/02/2022 com resolução espacial original de 100 m reamostradas para 30 m;
- Modelo Digital de Elevação *ALOS PALSAR* com resolução espacial de 12,5 m;
- Dados complementares (para etapa de validação): Polígonos de desmatamento *TerraBrasilis*; classificação *MapBiomas*; imagens de alta resolução espacial do *Google Earth* e *PlanetScope*.

Preparação dos dados

As imagens Sentinel foram adquiridas no nível de processamento 1C, ou seja, em reflectância no topo da atmosfera. As correções para reflectância de superfície foram realizadas na ferramenta *Sen2cor* disponibilizada pela European Space Agency (ESA) para correção atmosférica das imagens. Os produtos finais foram imagens de reflectância de superfície, nível 2A.

A partir das imagens Sentinel corrigidas para reflectância de superfície, os índices de vegetação *Normalized Difference Vegetation Index* – NDVI (ROUSE *et al.*, 1974), *Green NDVI* – GNDVI (SOUZA; PONZONI, 1998), *Enhanced Vegetation Index* - EVI (JUSTICE *et al.*, 1998) e *Soil – Adjusted Vegetation Index* – SAVI (HUETE, 1988) foram calculados.

As imagens TS e ALOS foram reamostradas para que coincidissem com a resolução de 10 m das imagens Sentinel 2. Para isso, utilizou-se o método de reamostragem *bilinear* recomendado para dados contínuos como elevação e temperatura (BABOO; DEVI,

2010). Este processo não implica em uma melhoria da resolução do dado, mas somente numa nova divisão, visando equivalência espacial da dimensão do pixel com dados de melhor resolução. Após este processo, foi realizado o empilhamento de todos os dados para facilitar os processos seguintes. Desta forma, foram considerados dois empilhamentos de bandas diferentes: o primeiro correspondeu ao empilhamento somente de bandas do satélite Sentinel 2; o segundo correspondeu ao empilhamento de bandas de todos os dados considerados (bandas Sentinel 2 e índices de vegetação supracitados, TS Landsat e MDE ALOS).

Para a segmentação de imagens utilizou-se o empilhamento das bandas ópticas 2, 3, 4 e 8 e o método baseado em regiões *mean-shift* (COMANICIU; MEER, 2002) o qual utiliza três entradas: um raio, no qual todos os pixels dentro do raio devem ser levados em consideração; um valor de diferença, no qual, de todos os pixels dentro do raio, apenas os que apresentarem valores iguais ou menores a esta diferença serão levados em consideração e; tamanho mínimo da região. Após testar diferentes valores, os parâmetros utilizados foram: raio igual a 5, diferença de 200 e região mínima de 20, estabelecidos de acordo com a melhor delimitação possível de cada alvo pelo usuário.

Treinamento, Modelos de Estatísticas Zonais e Classificação

O treinamento foi realizado em duas etapas. A primeira etapa consistiu na interpretação visual dos alvos presentes na imagem a partir de características como forma, textura e cor. Para tanto, foram consideradas as composições coloridas RGB 432 e RGB 843, e, assim, foram definidas as classes de cobertura da terra para a coleta das amostras de treinamento. A segunda etapa consistiu na coleta de amostras para cada classe a partir dos polígonos gerados na segmentação da imagem Sentinel 2. Foram coletadas amostras para a classe Floresta Nativa (FN – n = 53), Floresta Plantada (FP – n = 46), Formação Campestre (FC – n = 63), Afloramento Rochoso (AR n = 25), Infraestrutura Urbana (IU - n = 41) Mineração (MI - n = 41), Agricultura (AG – n = 65) e Corpos d'água (CA – n = 57).

A partir da ferramenta “Estatísticas Zonais” do *software* QGIS, foi possível calcular estatísticas descritivas (Contagem, Soma, Média, Mediana, Desvio Padrão, Mínimo, Máximo, Intervalo, Minoria, Maioria, Variabilidade e Variância) para cada polígono gerado na segmentação com base nas imagens de entrada. As imagens de entrada foram: as bandas ópticas para as datas de maio/2021 e fevereiro/2022, índices de vegetação calculados a partir das bandas de maio/2021 e fevereiro/2022, banda termal de maio/2021 e fevereiro/2022 e MDE ALOS. Já que a estrutura original da ferramenta possibilita somente o cálculo por banda individual, foram

gerados modelos de “Estatísticas Zonais – EZ” criados para automatizar o processo e calcular as variáveis para todos os polígonos gerados na segmentação com os dados de uma única vez. No total, foram criados doze modelos de EZ a partir de diferentes combinações entre os dados de entrada geradas para as 12 classificações propostas (Quadro 1).

Quadro 1: Combinação de dados de entrada nos modelos

| Data imageamento | Classificação | Combinação de Dados | Nº dados de entrada |
|------------------|---------------|--|---------------------|
| Maio/2021 | CM1 | Bandas 2, 3, 4 e 8 em reflectância de superfície (RS) do Sentinel 2 | 4 |
| | CM2 | Bandas 2, 3, 4 e 8 em reflectância de superfície (RS) do Sentinel 2 e Temperatura de superfície (TS) do Landsat 8 | 5 |
| | CM3 | 2, 3, 4 e 8 em reflectância de superfície (RS) do Sentinel 2 e modelo digital de elevação ALOS | 5 |
| | CM4 | Bandas 2, 3, 4 e 8 em reflectância de superfície (RS) do Sentinel 2, Temperatura de superfície (TS) do Landsat 8 e modelo digital de elevação ALOS | 6 |
| | CM5 | Bandas 2, 3, 4 e 8 em reflectância de superfície (RS) do Sentinel 2 e Índices de vegetação (I) NDVI, GNDVI, EVI e SAVI | 8 |
| | CM6 | Bandas 2, 3, 4 e 8 em reflectância de superfície (RS) do Sentinel 2, Índices de vegetação (I), Temperatura de superfície (TS) do Landsat 8 e modelo digital de elevação ALOS | 10 |
| Fevereiro/2022 | CF7 | Bandas 2, 3, 4 e 8 em reflectância de superfície (RS) do Sentinel 2 | 4 |
| | CF8 | Bandas 2, 3, 4 e 8 em reflectância de superfície (RS) do Sentinel 2 e Temperatura de superfície (TS) do Landsat 8 | 5 |
| | CF9 | 2, 3, 4 e 8 em reflectância de superfície (RS) do Sentinel 2 e modelo digital de elevação ALOS | 5 |
| | CF10 | Bandas 2, 3, 4 e 8 em reflectância de superfície (RS) do Sentinel 2, Temperatura de superfície (TS) do Landsat 8 e modelo digital de elevação ALOS | 6 |
| | CF11 | Bandas 2, 3, 4 e 8 em reflectância de superfície (RS) do Sentinel 2 e Índices de vegetação (I) NDVI, GNDVI, EVI e SAVI | 8 |
| | CF12 | Bandas 2, 3, 4 e 8 em reflectância de superfície (RS) do Sentinel 2, Índices de vegetação (I), Temperatura de superfície (TS) do Landsat 8 e modelo digital de elevação ALOS | 10 |

Fonte: Elaboração própria.

As classificações foram geradas no aplicativo Orfeo Toolbox (OTB), caracterizado como um projeto de código aberto para sensoriamento remoto de última geração com uma ampla variedade de aplicações disponíveis, considerando as fases treino e classificação: desde ortorretificação ou *pansharpening*, até classificação, processamento SAR, entre outros, estando instalado no QGIS como um *plugin* (<https://www.orfeo-toolbox.org/>). O treino (“*Train Vector Classifier*”) consistiu na criação dos modelos onde foram inseridas as variáveis de entrada (estatísticas zonais dos dados CM1 até CF12), amostras de treinamento e

definição dos parâmetros do classificador *Random Forest*. Ao executar este classificador, é necessário definir dois parâmetros principais, o número de árvores (Ntree) e o número de atributos selecionados aleatoriamente (Mtry). Por definição, o Mtry é calculado a partir da raiz quadrada do número de variáveis de entrada (BELGIU; DRAGUT, 2016). Em relação ao número de árvores, o Ntree utilizado foi 100, valor já apresentado em Guan *et al.* (2013). A partir dos modelos, foram geradas as 12 classificações (“*Vector classifier*”) correspondentes a cada modelo criado.

Avaliação das classificações

As doze classificações geradas foram submetidas a uma etapa de validação para verificar o desempenho de cada uma considerando as diferentes combinações de dados. A etapa de validação foi realizada a partir da definição de amostras de referência seguindo protocolo de “boas práticas” encontradas em Pontius e Millones (2011), Olofsson *et al.* (2014), Stehman e Foody (2019) e Congalton e Green (2019). Neste sentido, foram consideradas as etapas de desenho amostral, desenho resposta e análises. O desenho amostral escolhido foi a amostragem aleatória simples. Assim, para o cálculo do tamanho amostral foi utilizada a função multinomial, intervalo de confiança de 95% ($z = 1,96$) (CONGALTON; GREEN, 2019). Desta forma, foram analisados 636 pontos amostrais (FN – $n = 117$), (FP – $n = 26$), (FC – $n = 220$), (AR – $n = 41$), (IU – $n = 25$), (MI – $n = 15$), (AG – $n = 119$), (CA – $n = 13$). No desenho resposta, cada ponto foi interpretado a partir de imagem de alta resolução *Google Earth* e *PlanetScope*, com auxílio das imagens *Sentinel* e dados da plataforma *TerraBrasilis*. Todo processo de validação foi gerado no aplicativo *AcaTaMa* instalado no *QGIS*, onde foi possível interpretar cada ponto selecionado com auxílio de diferentes imagens.

A etapa de análises da qualidade da classificação foi realizada a partir das matrizes de erros de cada classificação: acurácia global (AG) e erro global (EG), acurácia do produtor (AP), acurácia do consumidor (AC), erros de omissão (EO) e comissão (EC). A partir das matrizes de erros foi possível calcular também os erros de alocação (EA) ou (*Allocation Disagreement – AD*) e os erros de quantidade (EQ) ou (*Quantity Disagreement – QD*) (OLOFSSON *et al.*, 2014; PONTIUS; MILLONES, 2011). O cálculo desses erros permite identificar falhas na classificação, separando a natureza das discordâncias em quantidade e alocação. Se o valor de EQ for relativamente alto, infere-se diferenças substanciais nas quantidades de amostras categorizadas na classificação ou na referência. Se o valor de EA for elevado, existe diferença na alocação espacial das categorias. Os erros que superam a margem de 10% são considerados elevados.

Como recomendado na literatura, as amostras de treino e validação foram coletadas de forma independente, pois as necessidades do treinamento são diferentes da etapa de validação (OLOFSSON *et al.*, 2014; STEHMAN; FOODY, 2019; CONGALON; GREEN, 2019). Por exemplo, na etapa de treinamento, a amostragem proposital de seleção local é aceitável para a classificação, entretanto, uma amostragem probabilística é um requisito de “boa prática” para a etapa da validação dos mapas gerados (STEHMAN; FOODY, 2019).

Por fim, para testar a significância das diferenças entre os desempenhos das classificações geradas, neste caso, as diferenças de AG, foi utilizado o teste estatístico Z para um intervalo de confiança de 95%, conforme definições apresentadas em Ma e Redmond (1995) e Congalton e Green (2019).

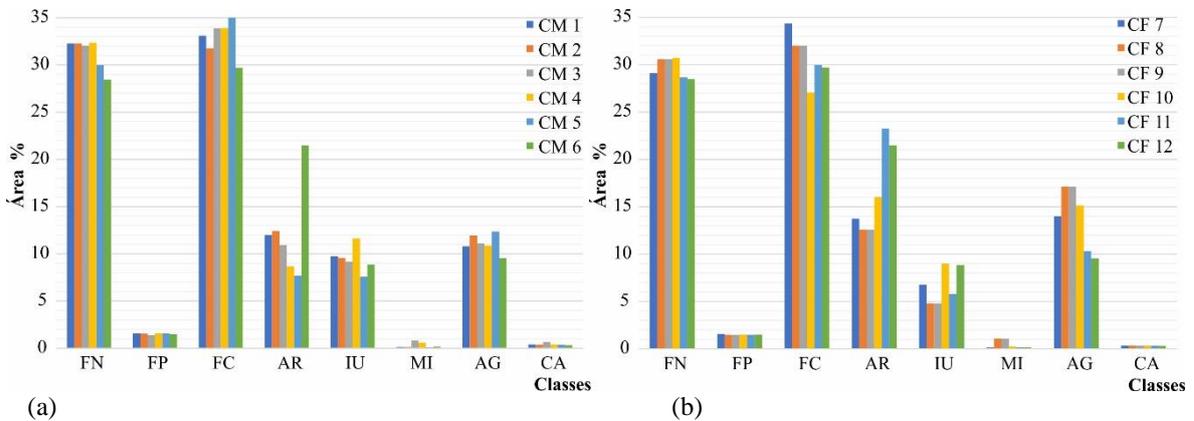
Resultados e Discussões

Representatividade de cada classe (área %)

Ao analisar as áreas das classes (Figura 3), verificou-se que as classes Formação Campestre (média de 30% da área total do município na classificação de fevereiro e 33,9% na classificação de maio) e Floresta Nativa (média de 29% em fevereiro e 32,16% para maio) apresentaram maior predominância na área de estudo. A média para classe Floresta Plantada foi de 1,5% em fevereiro e 1,56% em maio, de Afloramento Rochoso 14% em fevereiro e 9,86% em maio, Infraestrutura Urbana 6,28% em fevereiro e 9,36% em maio, Mineração 0,21% em fevereiro e 0,15% em maio, Agricultura 14,56% em fevereiro e 11,35% em maio e Corpos d’água a classe, com menor área (0,34% em fevereiro e 0,40% em maio).

As paisagens naturais pouco alteradas e diversidade geológico-geomorfológica de Caçapava do Sul garantem notável beleza cênica ao município, sendo um dos centros de estudos na área de geociências da região sul do Brasil (BORBA, 2015). Nesta região, ao considerar a soma das áreas das classes floresta nativa, formação campestre e afloramento rochoso (coberturas naturais – área média das classificações geradas) pode-se dizer que, aproximadamente, 75% da área do município ainda se encontram preservadas. Tal resultado é similar ao visualizado nos resultados dos mapeamentos do *Mapbiomas* – Coleção 7.1 (<https://mapbiomas.org/>), com aproximadamente 77% de área ainda natural e aos dados de monitoramento de supressão vegetal produzidos pelo INPE disponibilizados na plataforma *TerraBrasilis* (<http://terrabrasilis.dpi.inpe.br/app/map/deforestation?hl=pt-br>), com 66%.

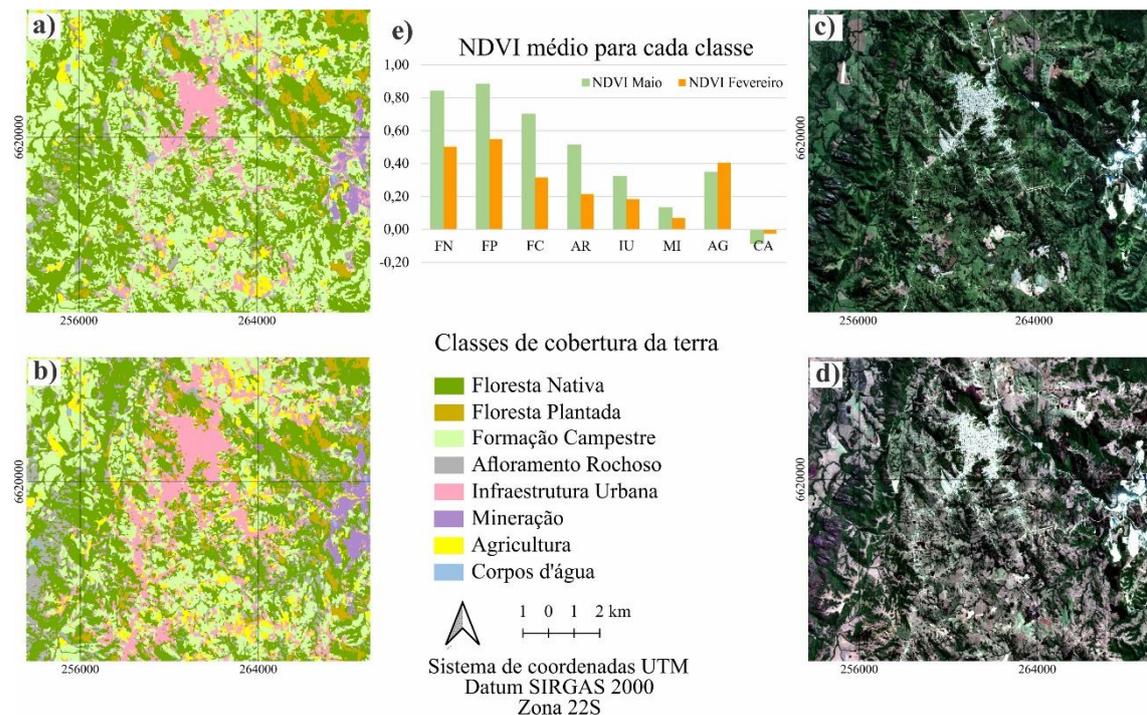
Figura 3. Porcentagem de área para cada classe nas classificações geradas (%) para maio de 2021(a) e fevereiro de 2022 (b)



(a) Fonte: Elaboração própria.

Ao comparar a classificação CF9 (maior AG para a imagem de fevereiro/2022 – Figura 4a) e CM6 (maior AG para a imagem de maio – Figura 4b) é possível verificar as principais diferenças visuais nas classes de cobertura da terra para um recorte da área de estudo. Em CF9 foi identificada maior ocorrência de áreas de afloramento rochoso, em contrapartida, CM6 verificou-se menor área de afloramento rochoso.

Figura 4. Exemplos de diferentes classificações geradas – *Random forest* e NDVI: a) Classificação CM6 (bandas Sentinel, índices de vegetação, TS e ALOS); b) Classificação CF9 (bandas Sentinel 2, 3, 4 e ALOS; c) RGB 432 de maio de 2021; d) RGB432 de fevereiro de 2022; e) NDVI médio para cada classe de cobertura da terra



Fonte: Elaboração própria.

As diferenças de área encontradas podem ser justificadas pela sazonalidade da área de estudo. As estações de outono e inverno apresentam os maiores índices pluviométricos e, à medida que o verão se aproxima, são verificados os menores valores de precipitação (SARTORI, 2016). Assim, o vigor da cobertura vegetal pode variar de acordo com a distribuição das chuvas ao longo do ano e, conseqüentemente, variações no comportamento espectral das diferentes superfícies naturais. Este fato pode ser claramente identificado nas áreas de campo e afloramento rochoso, pois na imagem de maio/2021 (Figura 4c) as áreas de afloramento rochoso não estavam tão expostas quanto na imagem de fevereiro/2022 (Figura 4d). Na mesma linha de análise, foi possível verificar uma superestimação da classe Infraestrutura Urbana (no recorte apresentado na Figura 4d) para a classificação de fevereiro/2022, sendo que este fato também pode ser justificado pela variação espectral da vegetação entre as estações do ano, fazendo com que ocorresse maior confusão entre as classes Infraestrutura Urbana e Formação Campestre.

A maior diferença dos valores médios do índice de vegetação NDVI (Figura 4e) para a maioria das classes corroboram as afirmações supracitadas, sendo que, foram verificados valores mais elevados de NDVI em maio de 2021, padrão evidenciado quando da verificação dos dados de NDVI produzidos pelo satélite MODIS para os sensores Terra e Aqua para a região de Caçapava do Sul-RS do Sistema de Análise Temporal da Vegetação (SATVEG, 2023) da Embrapa. Jin *et al.* (2021) também comprovaram que a variação do NDVI ao longo do ano pode ser considerado um indicador que reflete o ambiente regional. Ao estudarem a distribuição espaço-temporal do NDVI na China, os autores verificaram que a umidade do solo (dependente dos volumes pluviométricos) foi o principal fator que influenciou os valores de NDVI.

Erros de Quantidade e Alocação

Para as classificações de maio/2021, os erros de quantidade (EQ) variaram de 6,68% a 8,33% e alocação (EA) de 13,68% a 17,45% (Tabela 1). A classificação realizada com bandas Sentinel 2 e Alos (CM3) apresentou o menor valor para EQ (6,68%). Valor mais elevado foi verificado na classificação realizada com combinação de bandas Sentinel 2 e índices de vegetação (8,33%) – CM1. O menor erro de alocação (13,68%) foi identificado na classificação CM6 (Bandas Sentinel, índices, TS e ALOS), já o maior erro foi verificado na classificação CM4 (Bandas Sentinel e índices de vegetação).

Tabela 1. Erros de quantidade (EQ) e alocação (EA)

| Classificação | Erros % | | Classificação | Erros % | |
|---------------|---------|-------|---------------|---------|-------|
| | EQ | EA | | EQ | EA |
| CM1 | 7,86 | 16,19 | CF7 | 6,60 | 13,52 |
| CM2 | 7,55 | 17,45 | CF8 | 6,76 | 14,31 |
| CM3 | 6,68 | 16,12 | CF9 | 5,66 | 9,75 |
| CM4 | 8,33 | 16,82 | CF10 | 10,93 | 12,81 |
| CM5 | 8,33 | 13,99 | CF11 | 11,48 | 11,16 |
| CM6 | 7,55 | 13,68 | CF12 | 12,58 | 11,64 |

Fonte: Elaboração própria.

Para as classificações de fevereiro/2022, os erros de quantidade (EQ) variaram de 5,66% a 12,58% e alocação (EA) de 9,75% a 13,52% (Figura 5). A classificação que utilizou as bandas Sentinel 2 e Alos (CF9) apresentou os menores valores para EQ (5,66%) e EA (9,75%). Em contrapartida, os maiores erros foram verificados na classificação realizada com a combinação de dados Sentinel 2, Índices de vegetação, TS e ALOS (CF12), com EQ de 12,58% e EA de 11,64%.

Considerando as definições apresentadas por Pontius e Millones (2011), as classificações geradas apresentaram maiores problemas na alocação espacial das classes, sendo que a única classificação que apresentou valores inferiores a margem limite de erros EQ e EA (menor que 10%) foi a CF9 (combinação de bandas Sentinel e Alos – fevereiro/verão). Ao utilizarem o *Randon Forest* com a combinação de imagens ópticas e dados de elevação, Gislason *et al.* (2006) também comprovaram que a elevação foi uma variável topográfica com importância para classificar as diferentes coberturas da terra em uma região de relevo movimentado em Colorado. Os autores ainda destacaram que foi difícil distinguir tipos de coberturas vegetais utilizando apenas dados ópticos. Em contrapartida, Pan *et al.* (2020) alcançaram bom desempenho para classificação de uso e cobertura da terra com diferentes combinações somente de dados ópticos a partir do algoritmo *Randon Forest* na plataforma *Google Earth Engine* e verificaram que as imagens da estação do verão também apresentaram maiores valores de acurácia.

Acurácia do consumidor (AC) e Produtor (AP)

Para as classificações de maio/2021 (Tabela 2), os maiores valores de AC foram verificados para as classes FN (média 94,2%), MI (91,2%) e CA (100%), já os menores valores de AC foram identificados para a classe IU em todas as classificações (média 34,5%). A classe FP apresentou média AC de 73,5%, a classe FC com média de 83,9%, a classe AR

apresentou valor médio de 51,2%, a classe MI com média de 89,3% e AG apresentou média de 71,5%. Os erros de comissão variaram de 3% (FN) a 72,9% (IU).

Tabela 2. Acurácia do consumidor (AC) e acurácia do produtor (AP) – maio/2021

| Classes | CM1 | | CM2 | | CM3 | | CM4 | | CM5 | | CM6 | |
|---------|-------|-------|-------|-------|------|-------|-------|-------|-------|-------|-------|--------|
| | AC% | AP% | AC% | AP% | AC% | AP% | AC% | AP% | AC% | AP% | AC% | AP% |
| FN | 91,6 | 93,8 | 95,6 | 92,7 | 93,4 | 95,5 | 92,1 | 92,1 | 97,0 | 92,1 | 95,3 | 92,09 |
| FP | 67,9 | 73,1 | 84,6 | 73,1 | 80,8 | 80,8 | 62,5 | 76,9 | 75,0 | 80,8 | 70,0 | 80,77 |
| FC | 81,4 | 80,0 | 82,2 | 77,7 | 81,1 | 80,0 | 80,4 | 78,2 | 77,6 | 86,8 | 77,0 | 86,82 |
| AR | 41,4 | 63,4 | 66,0 | 58,5 | 49,1 | 65,9 | 55,8 | 58,5 | 47,9 | 56,1 | 46,8 | 53,66 |
| IU | 28,8 | 56,0 | 35,7 | 60,0 | 31,9 | 60,0 | 27,1 | 76,0 | 34,9 | 60,0 | 48,6 | 68,00 |
| MI | 92,9 | 93,3 | 85,7 | 86,7 | 82,4 | 93,3 | 93,3 | 93,3 | 92,9 | 86,7 | 100,0 | 100,00 |
| AG | 70,7 | 46,2 | 65,5 | 48,7 | 71,8 | 47,1 | 70,8 | 42,9 | 72,4 | 46,2 | 77,6 | 49,58 |
| CA | 100,0 | 100,0 | 100,0 | 100,0 | 86,7 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 92,9 | 100,00 |

Fonte: Elaboração própria.

Nas classificações de maio/2021, foram verificados maiores valores de AP para as classes FN (média 93,03%) e CA (média 100%), e menores na classe AG (média 46,78%). A média de AP para a classe FP foi de 77,56%, para a classe FC a média foi de 81,59%, para a classe AR de 59,35%, para a classe IU 63,33% e para MI 92,2%. Os erros de omissão (EO) variaram de 4,5% (FN) a 57,14% (AG).

Nas classificações das imagens de fevereiro/2022, os maiores valores de AC foram verificados nas classes FN (média 96,1%) e CA (100%) em todas as classificações (Tabela 3), e os menores valores para as classes IU (média 40,5% nas classificações CF7, CF11, CF10 e CF12) e AR (média 48% nas classificações CF9 e CF11). Para a classe FP, identificou-se valor médio de 82,4%, a classe FC com média de 83,9%, a classe MI verificou-se valor médio de 89,3% e a classe AG apresentou média de 71,9%. Os valores dos erros de comissão variaram de 2,9% (FN) a 64,9% (IU).

Tabela 3. Acurácia do consumidor (AC) e acurácia do produtor (AP) – fevereiro/2022

| Classes | CF7 | | CF8 | | CF9 | | CF10 | | CF11 | | CF12 | |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| | AC% | AP% |
| FN | 97,1 | 94,4 | 95,6 | 97,2 | 96,1 | 96,6 | 95,5 | 96,0 | 95,4 | 93,8 | 97,0 | 92,7 |
| FP | 81,5 | 84,6 | 84,6 | 84,6 | 84,6 | 84,6 | 76,9 | 76,9 | 87,0 | 76,9 | 80,0 | 76,9 |
| FC | 82,7 | 78,2 | 82,2 | 79,5 | 86,1 | 78,6 | 85,7 | 68,2 | 84,5 | 76,8 | 82,1 | 75,0 |
| AR | 54,5 | 87,8 | 66,0 | 80,5 | 59,7 | 97,6 | 46,8 | 87,8 | 36,3 | 90,2 | 37,1 | 80,5 |
| IU | 53,7 | 88,0 | 35,7 | 80,0 | 75,0 | 96,0 | 37,5 | 84,0 | 48,6 | 68,0 | 35,1 | 80,0 |
| MI | 92,3 | 80,0 | 85,7 | 80,0 | 88,2 | 100,0 | 92,9 | 86,7 | 76,5 | 86,7 | 100,0 | 93,3 |
| AG | 66,7 | 53,8 | 65,5 | 46,2 | 78,4 | 67,2 | 63,9 | 52,1 | 79,2 | 47,9 | 77,9 | 44,5 |
| CA | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 | 100,0 |

Fonte: Elaboração própria.

As classes FN e CA apresentaram os maiores valores de AP, em contrapartida, os menores valores foram identificados na classe AG (média 52%). A média de AP para a classe FP foi de 80,8%, para a classe FC 76,1%, para a classe AR 87,4%, para IU 82,7% e para MI 87,8%. Os erros de omissão (EO) variaram de 2,8% (FN) a 55,5% (AG). De modo geral, as classes que apresentaram maior confusão foram AR/FC, IU/FC, AG/FC e AG/IU, sendo que, para a classificação de fevereiro/2022, cinco das oito classes apresentaram concordância superior a 80%, e, em maio/2021, apenas três das oito classes apresentaram acerto maior que 80%. A confusão gerada entre as classes supracitadas pode ser justificada pela variação da resposta espectral das coberturas entre as diferentes estações do ano, neste sentido, uma análise de séries temporais poderia trazer maior qualidade aos mapas gerados. Este processo pode ser otimizado a partir do pacote SITS (*Satellite Image Time Series Analysis*), por exemplo. O SITS abrange um conjunto de ferramentas para classificação e análise de séries temporais que podem ser visualizadas e analisadas a partir da criação de cubos de imagens (<http://www.brazildatacube.org/sits-satellite-image-time-series/>).

Como foi observado nos resultados referentes às acurácias de consumidor e produtor, os menores erros foram verificados nas classes Floresta Nativa e Corpos d'água, sendo que os valores mais elevados foram identificados nas classes Infraestrutura Urbana e Agricultura, resultados semelhantes aos encontrados em pesquisas anteriores (MA *et al.*, 2021; PEREIRA *et al.*, 2023; PATSALIDIS *et al.*, 2019, CHO *et al.*, 2021) que também utilizaram o *Random Forest* para classificação de imagens.

Ao realizarem classificação dos tipos de vegetação a partir de imagens Sentinel 2 e utilizando o *Random Forest*, Ma *et al.* (2021) também verificaram resultados satisfatórios para as classes com cobertura arbórea, com valores elevados de AC e AP, já as áreas de agricultura apresentaram valores baixos de AC, de até 54%.

Ao compararem classificações realizadas a partir de imagens Landsat 8 e Sentinel 2 utilizando o algoritmo *Random forest*, Pereira *et al.* (2023) verificaram acurácias globais de 76% a 84%. Entretanto, os autores verificaram que não ocorreu diferença estatisticamente significativa (nível de significância de 95%) entre elas e verificaram bom desempenho para as classes Formações Florestais, Corpos Hídricos e Formações Campestres, com taxas de acerto superiores a 80% e melhores resultados na utilização de imagem Sentinel.

Ao comparar o desempenho dos classificadores *Maxver* e *Random Forest*, Patsalidis *et al.* (2019) verificaram maior acurácia nos mapas gerados com o segundo. Os

autores também indicaram a classe de corpos hídricos com maior desempenho e áreas urbanas com menores valores de acurácia. Em adição, Sheykhmousa *et al.* (2020) realizaram uma ampla revisão de literatura de trabalhos científicos, com 251 publicações em revistas internacionais. Eles demonstraram que, de maneira geral, o algoritmo de classificação *Random forest* possui acurácia global similar em comparação ao *Support Vector Machine* na análise de sete diferentes tipos de aplicações, desde desastres naturais e áreas urbanas à classificação de uso e cobertura da terra, por exemplo, além do *Random Forest* ser o mais utilizado como algoritmo de *machine learning*.

A utilização de imagens ópticas e modelo digital de elevação para classificação a partir do *Random Forest* também apresentou resultados satisfatórios na classificação de uso e cobertura da terra realizada por Cho *et al.* (2021) para regiões do bioma Cerrado. Os autores verificaram acurácia global de 84%, entretanto, também foi verificada uma superestimação para as classes agricultura e área urbana.

Acurácia global – AG

A AG da classificação de maio/2021 variou entre 74,84% a 78,77%, sendo que a menor AG foi identificada na classificação realizada com a combinação das bandas ópticas, termal e ALOS – CM4. O melhor desempenho geral (78,77%) foi verificado na classificação realizada a partir da combinação de todas as bandas (Sentinel 2, 3, 4, índices de vegetação, termal e ALOS) – CM6, com EG de 21,23% (Tabela 4).

Tabela 4. Acurácia Global e Erro Global para cada classificação

| | Diferença entre as classificações | | | | | | | | | | | AG | EG |
|------|-----------------------------------|------|------|------|------|------|------|------|------|------|------|-------|-------|
| | CM1 | CM2 | CM3 | CM4 | CM5 | CM6 | CF7 | CF8 | CF9 | CF10 | CF11 | | |
| CM1 | - | - | - | - | - | - | - | - | - | - | - | 75,94 | 24,06 |
| CM2 | 0,94 | - | - | - | - | - | - | - | - | - | - | 75,00 | 25,00 |
| CM3 | 1,26 | 2,20 | - | - | - | - | - | - | - | - | - | 77,20 | 22,80 |
| CM4 | 1,10 | 0,16 | 2,36 | - | - | - | - | - | - | - | - | 74,84 | 25,16 |
| CM5 | 1,73 | 2,67 | 0,47 | 2,83 | - | - | - | - | - | - | - | 77,67 | 22,33 |
| CM6 | 2,83 | 3,77 | 1,57 | 3,93 | 1,10 | - | - | - | - | - | - | 78,77 | 21,23 |
| CF7 | 3,93 | 4,87 | 2,67 | 5,03 | 2,20 | 1,10 | - | - | - | - | - | 79,87 | 20,13 |
| CF8 | 2,99 | 3,93 | 1,73 | 4,09 | 1,26 | 0,16 | 0,94 | - | - | - | - | 78,93 | 21,07 |
| CF9 | 8,65 | 9,59 | 7,39 | 9,75 | 6,92 | 5,82 | 4,72 | 5,66 | - | - | - | 84,59 | 15,41 |
| CF10 | 0,31 | 1,25 | 0,95 | 1,41 | 1,42 | 2,52 | 3,62 | 2,68 | 8,34 | - | - | 76,25 | 23,75 |
| CF11 | 1,41 | 2,35 | 0,15 | 2,51 | 0,32 | 1,42 | 2,52 | 1,58 | 7,24 | 1,10 | - | 77,35 | 22,65 |
| CF12 | 0,16 | 0,79 | 1,42 | 0,94 | 1,89 | 2,99 | 4,08 | 3,14 | 8,80 | 0,46 | 1,56 | 75,79 | 24,21 |

Fonte: Elaboração própria.

Para as classificações geradas com a imagem de fevereiro/2022, a acurácia global variou entre 75,79% a 84,59% (Figura 6), sendo que a menor AG foi identificada na classificação realizada com a combinação de todas as bandas (Sentinel 2, 3, 4 e 8, índices de vegetação, TS e ALOS) – CF12. O melhor desempenho geral (84,59%) foi verificado na classificação realizada a partir da combinação de bandas ópticas e MDE (Sentinel 2, 3, 4 e ALOS) – CF9, com baixo EG de 15,41%.

Ao comparar todas as classificações, verificaram-se diferenças de 0,15% a 9,75% (Figura 6c), entretanto, só ocorreu diferença estatisticamente significativa (Z 95%) quando comparada a classificação CF9 (bandas ópticas e altimetria – imagem fevereiro/2022) com as demais classificações. Isso significa que a combinação de dados ópticos e de altimetria com imagens de fevereiro (verão) apresentou resultados de acurácia significativamente diferentes das demais combinações de dados e, por consequência, valor de AG mais elevado, semelhantes aos valores encontrados por Corcoran *et al.* (2013). Estes autores verificaram que associar dados do terreno (curvatura e elevação) e imagens ópticas (sensor TM/Landsat5) adquiridas na primavera garantiu melhor desempenho na classificação dos diferentes tipos de cobertura da terra. Resultados similares também foram encontrados por Wasniewski *et al.* (2020) no noroeste do Gabão, utilizando imagens Sentinel 2, índices de vegetação e modelo digital de elevação, tendo verificado maior precisão para as classificações geradas com a combinação das bandas espectrais Sentinel e MDE.

Em contrapartida, a partir de segmentação de imagens em diferentes cenários de classificação e combinação de bandas Sentinel 2B, índices de vegetação e modelo linear de mistura espectral, Diniz *et al.* (2020) encontraram resultados divergentes ao presente estudo. Os autores concluíram que não ocorreram diferenças significativas quando comparadas as acurácias das classificações quando utilizada todas as combinações possíveis de dados de entrada.

A combinação de imagens ópticas e termais também foi uma das propostas de análise na presente pesquisa, tendo sido possível verificar acurácia global inferior a 79% tanto na classificação de fevereiro/2022, quanto na de maio/2021. Diferentes pesquisas também analisaram a combinação de bandas ópticas e termais para classificação de imagens (SUN; SCHULZ, 2015; EISAVI *et al.*, 2015; TRINDADE *et al.*, 2021). Ao utilizar bandas do visível e infravermelho próximo associadas às imagens termais, Sun e Schulz (2015) verificaram um aumento de 5% a 6% no desempenho geral na classificação. Trindade *et al.* (2021) analisaram o desempenho de classificação supervisionada considerando bandas

ópticas e termais em áreas selecionadas no bioma Pampa e identificaram um aumento moderado (até 3%) de acurácia geral com a inclusão de banda termal nas classificações geradas.

Considerações finais

De modo geral, as classificações geradas mostraram-se com bom desempenho, com acurácia global de 74% a 84,5%. Os menores erros foram verificados nas classes Floresta Nativa e Corpos d'água, sendo que os valores mais elevados foram identificados nas classes Infraestrutura urbana e Agricultura. As áreas com coberturas naturais predominaram em aproximadamente 75% da área do município, resultado similar ao visualizado nos mapeamentos do *Mapbiomas* (Coleção 7.1) e dados de desmatamento/supressão vegetal do *TerraBrasilis*.

Conforme as análises geradas, a combinação de imagens Sentinel e MDE/Alos para imagem de fevereiro (verão/2022) apresentaram melhores resultados para classificação da área de estudo, com diferença estatisticamente significativa em relação às demais classificações. Neste sentido, destaca-se a importância da combinação de dados ópticos e de elevação para o mapeamento de uso e cobertura da terra de regiões com diferentes configurações de relevo. Embora os resultados desta pesquisa sejam promissores, ainda existe espaço para melhoria adicional, com modificações nas variáveis de entrada e de parâmetros do algoritmo *Random Forest*, como, por exemplo, a elevação e a declividade obtidos do MDE.

Outras imagens de diferentes estações do ano também podem ser utilizadas como entrada nos modelos para, assim, também ser possível comparar a dinâmica de supressão da vegetação nativa ao longo dos ciclos de plantio das culturas na área de estudo. Esta abordagem poderia contribuir para diminuir a confusão entre as classes Afloramento Rochoso e Formação Campestre, Áreas Agrícolas e Formação Campestre, Infraestrutura Urbana e Áreas Agrícolas. Neste sentido, a utilização de técnicas mais robustas como análises de séries temporais via SITS (*Satellite Image Time Series Analysis*) auxiliaria na otimização e melhoria nas acurácias da classificação gerada.

A continuidade da presente pesquisa seguirá na linha de análise a partir da utilização do pacote SITS, com a geração de resultados para todo o bioma pampa, pela sua alta capacidade de análise de séries temporais a partir de diferentes plataformas. Além disso, o SITS é um pacote de código aberto, passível de visualização e análise dos resultados da classificação geradas para algoritmos de classificação baseados tanto em *machine learnig*

(*Randon Forest, Support Vector Machine*) como em *deep learning* (*Artificial Neural Network*) e comparadas aos resultados gerados pelo algoritmo *Random Forest*.

Agradecimentos

Agradecimento ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq processos: 422354/2023-6 (Monitoramento e avisos de mudanças de cobertura da terra nos biomas brasileiros – capacitação e semiautomatização do Programa BIOMASBR); 301423/2024-5; n. 384905/2023-4; n. 384906/2023-0; n. 409336/2021-1, apoiados pelo Instituto Nacional de Pesquisas Espaciais – INPE.

Referências bibliográficas

ALMEIDA, C. A. de; MAURANO, L. E. P.; VALERIANO, D. M.; CÂMARA, G.; VINHAS, L.; MOTTA, M.; GOMES, A. R.; MONTEIRO.; SOUZA, A. A. A.; MESSIAS, C. G.; RENNÓ, C. D.; ADAMI, M.; ESCADA, M. I. S.; SOLER, L. S.; AMARAL, S. *Metodologia utilizada nos sistemas PRODES e DETER*. 2º Edição: INPE, São José dos Campos, 2022.

ALMEIDA-FILHO, R.; ROSENQVIST, A.; SHIMABUKURO, Y.E.; dos SANTOS, J.R. Evaluation and Perspectives of Using Multitemporal L-Band SAR Data to Monitor Deforestation in the Brazilian Amazônia. *IEEE Geosci. Remote Sensing*, v. 2, p. 409–412, 2005.

ASSIS, L. F. F. G, FERREIRA, K. R., VINHAS, L., MAURANO, L., ALMEIDA, C., CARVALHO, A., CAMARGO, C. TerraBrasilis: A Spatial Data Analytics Infrastructure for Large-Scale Thematic Mapping. *ISPRS International Journal of Geo-Information*, v. 8, n. 11, 513, 2019.

BABOO, S. S; DEVI, M. R. An analysis of different resampling methods in Coimbatore, District. *Global Journal of Computer Science and Technology*, v. 10, n. 15, p. 61 – 66, 2010.

BELGIU, M.; DRAGUT, L. Random Forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 114, p. 24 – 31, 2016.

BDC – Brazil Data Cube. Satellite Image Time Series Analysis - SITS Disponível em: <<http://www.brazildatacube.org/sits-satellite-image-time-series/>>. Acesso: out. 2023.

BORBA, A. W. Proposta de uma (Geo)identidade visual para Caçapava do Sul, “Capital Gaúcha da Geodiversidade”. *Geographia Meridionalis*, v. 1, n. 2, 405–411, 2015.

BORBA, A. W. Um geopark na região de Caçapava do Sul (RS, Brasil): uma discussão sobre viabilidade e abrangência territorial. *Geographia Meridionalis*, v. 03, n. 01, p. 104–133, 2017

BREIMAN, L., 2001. Random forest. *Machine Learning*, v. 45, p. 5 – 32, 2001.

CHO, D. F.; SCHWAIDA, S. F.; CICERELLI, T. A.; RAMOS, A. P. M. R.; SANO, E. E. Desempenho do Algoritmo de Classificação de Imagens Random Forest para Mapeamento do Uso e Cobertura do Solo no Cerrado Brasileiro. *Anuário do Instituto de Geociências*, v. 44, 37979, p. 1 - 11, 2021.

COMANICIU, D.; MEER, P. Mean shift: a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 5, p. 603–619, 2002.

CONGALTON R. G.; GREEN, K. *Assessing the Accuracy of Remotely Sensed Data*. 3ª ed. Taylor & Francis Group, 2019.

CORCORAN, J., KNIGHT, J., GALLANT, A. Influence of multi-source and multitemporal remotely sensed and ancillary data on the accuracy of random forest classification of wetlands in northern Minnesota. *Remote Sensing*, v. 5, p. 3212–3238, 2013.

CPRM. COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. Mapa Geológico do Rio Grande do Sul. Escala 1: 750.000. Legenda expandida. CPRM/SGB. Porto Alegre, 2006.

CPRM. COMPANHIA DE PESQUISA DE RECURSOS MINERAIS. *Geodiversidade do Estado do Rio Grande do Sul – Programa Geologia do Brasil – Levantamento da Geodiversidade*. VIERO, A. C. & SILVA, D. R. A. (orgs). Porto Alegre, 2010. 250p

DINIZ, J. M. F. de S.; MACIEL, D. A.; GAMA, F. F.; ADAMI, M. Avaliação do Sentinel-2, NDVI e MLME para Mapeamento do Uso e Cobertura da Terra. *Anuário do Instituto de Geociências – UFRJ*, v. 43, n.2, p. 381 – 391, 2020.

DU, P., SAMAT, A., WASKE, B., LIU, S., LI, Z. Random Forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. *ISPRS J. Photogramm. Remote Sensing*, v. 105, p. 38–53, 2015.

EISAVI, V.; HOMAYOUNI, S.; YAZDI, A.M.; ALIMOHAMMADI, A. Land cover mapping based on random forest classification of multitemporal spectral and thermal images. *Environmental Monitoring and Assessment*, v. 187, n. 5, p. 1–14, 2015.

GISLASON, P.O., BENEDIKTSSON, J.A., SVEINSSON, J.R. Random forests for land cover classification. *Pattern Recogn. Lett.*, v. 27, p. 294–300, 2006.

GLOBAL FOREST WATCH. Monitoramento de Florestas Projetado para a Ação. Disponível em: <<https://www.globalforestwatch.org/>>. Acesso: out. 2023.

GUAN, H., Li, J., CHAPMAN, M., DENG, F., JI, Z., YANG, X. Integration of orthoimagery and lidar data for object-based urban thematic mapping using random forests. *Int. J. Remote Sensing*, v. 34, 5166–5186, 2013.

HUETE, A. R. A soil adjusted vegetation index (SAVI). *Remote Sens. Environ.*, v. 25, p. 295-309, 1988.

IBGE. Município de Caçapava do Sul. Disponível em: <<https://cidades.ibge.gov.br/brasil/rs/cacapava-do-sul/panorama>>. Acesso: 04 mai. 2023.

INPE. INSTITUTO NACIONAL DE PESQUISA ESPACIAIS – Coordenação Geral de Observação da Terra. Programa de Monitoramento da Amazônia e demais Biomas e Fundação de Ciência, Aplicações e Tecnologia Espaciais – FUNCATE. ALMEIDA *et al.*, 2019. *Monitoramento Ambiental dos biomas brasileiros por satélite: mata atlântica, caatinga, pampa e pantanal* - Relatório de Referência metodológica dos subprojetos 1 a 4 (versão 1).

JICA-JAXA Forest Early Warning System in the Tropics. Watching on Deforestation. Disponível em: <<https://www.eorc.jaxa.jp/jjfast/>>. Acesso: out. 2023.

JIN, H; CHEN, X; WANG, Y; ZHONG, R; ZHAO, T; LIU, Z; TU, X. Spatio-temporal distribution of NDVI and its influencing factors in China. *Journal of Hydrology*, v. 603, 127129, 2021.

JUSTICE, C. O.; VERMOTE, E.; TOWNSHEND, J. R. G.; DEFRIES, R.; ROY, D. P.; HALL, D. K.; SALOMONSON, V.V.; PRIVETTE, J. L.; RIGGS, G.; STRANHLER, A. The moderate resolution imaging spectroradiometer (MODIS): land remote sensing for global change research. *IEEE Transactions on Geoscience and Remote Sensing*, v. 36, n. 4, p. 1228 – 1249, 1998.

MA, Z., REDMOND, R.L. Tau coefficients for accuracy assessment of classification of remote sensing data. *Photogrammetric Engineering and Remote Sensing*, v. 61, n. 4, p. 453 – 439, 1995.

MA, M.; LIU, J.; LIU, M.; ZENG, J.; LI, Y. Tree Species Classification Based on Sentinel-2 Imagery and Random Forest Classifier in the Eastern Regions of the Qilian Mountains. *Forests*, v. 12, 1736, p. 1 – 12, 2021.

OLOFSSON, P.; FOODY G. M.; HEROLD, M.; STEHMAN S. V.; WOODCOCK, C. E.; WULDER, M. A. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, v. 148, p. 42 – 57, 2014.

PAN, T. N.; KUCH, V.; LEHNERT, L. W. Land Cover Classification using Google Earth Engine and Random Forest Classifier—The Role of Image Composition. *Remote Sensing*, v. 12, 2411, p. 1 – 22, 2020.

PATSALIDIS, S.; AGAPIOU, A.; HADJIMITSIS, D. G. Random Forest classification analysis of Sentinel-2 and Landsat-8 images over semi-arid environment in the Eastern Mediterranean. In: International conference on geographic information, 2019. *Proceedings...* Amsterdam: Springer, 2020.

PEREIRA, P. R. M.; OLIVEIRA, M. M. N.; BOLFE, E. L.; MACARRINGUE, L. S. Comparação da classificação do uso e cobertura da terra em imagens Landsat-8 e Sentinel-2 no Cerrado Maranhense. *GEOUERJ*, n. 42, e66306, 2023.

PONTIUS Jr., R. G.; MILLONES, M. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, v. 32, n. 15, p. 4407 – 4429, 2011.

RODRIGUEZ-GALIANO, V.F., GHIMIRE, B., ROGAN, J., CHICA-OLMO, M., RIGOL-SANCHEZ, J.P. An assessment of the effectiveness of a random forest classifier for landcover classification. *ISPRS J. Photogramm. Remote Sensing*, v. 67, p. 93–104, 2012.

ROUSE, J. W., ROUSE, J. W., HAAS, R. W., SCHELL, J. A., DEERING, D. W., HARLAN, J. C., ROUSE, J. W., HAAS, R. H., SHELL, J. A., DEERING, D. W., HARLAN, J. C., ROUSE, J. W., ROUSE, J., HASS, RH. *Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation*. Greenbelt: NASA, 371p. 1974.

SARTORI, M.G.B. *O Vento Norte*. Santa Maria: Gráfica e Editora Palloti, 2016.

SATVeg. *Sistema de Análise Temporal da Vegetação*. Disponível em: <<https://www.satveg.cnptia.embrapa.br/satveg/login.html>>. Acesso: out. 2023.

SEN2COR. *Science Toolbox Exploitation Platform (STEP)*. European Space Agency. Disponível em: <<http://step.esa.int/main/third-party-plugins-2/sen2cor/>>. Acesso: set. 2022)

SHEYKHMUSA, M., MAHDIANPARI, M., GHANBARI, H., MOHAMMADIMANESH, F., GHAMISI, P., HOMAYOUNI, S. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 13, 6308-6325, 2020.

SOTHE, C.; LIESENBERG V.; ALMEIDA, C. M.; SACHIMALSKI, M. B. Abordagens para classificação do estágio sucessional da vegetação do Parque Nacional de São Joaquim empregando imagens Landsat-8 e Rapideye. *Bulletin of Geodetic Sciences*, v. 23, n. 3, p. 389 – 404, 2017.

SOUZA, C. L.; PONZONI, F. J. Relação entre índice de área foliar, estimado através de sensoriamento remoto, e parâmetros dendrométricos em floresta implantada de Pinus spp. In: Simpósio Brasileiro de Sensoriamento Remoto, 9; 1998, Santos. *Anais...* Santos: INPE/SELPER, p. 1549-1560, 1998.

SOUZA JR., 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, p. 2 – 27, 2020.

STEHMAN S. V.; FOODY G. M. Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment*, v. 231 p. 1 – 23, 2019.

STRECK, E.V.; KÄMPF, N.; DALMOLIN, R.S.D.; KLAMT, E.; NASCIMENTO, P.C.; SCHNEIDER, P.; GIASSON, E. & PINTO, L.F.S. *Solos do Rio Grande do Sul*. 2.ed. Porto Alegre, EMATER/RS-ASCAR, 222p, 2008.

SUN, L., SCHULZ, K. The improvement of land cover classification by thermal remote sensing. *Remote Sensing*, v. 7, 8368, 2015.

TRINDADE, P. M. P.; PEIXOTO, D. W. B.; KUPLICH, T. M.; ALMEIDA, C. A. Desempenho da classificação supervisionada com a inclusão da banda termal em áreas selecionadas do bioma Pampa, Brasil. *In: Congresso Brasileiro de Agroinformática*, 13; 2021. *Anais...* p. 62-70, 2021.

WASNIEWSKI, A.; HOSCITO, A.; ZAGAJEWSKI, B.; MOUKÉTOU-TARAZEWICZ, D. Assessment of Sentinel-2 Satellite Images and Random Forest Classifier for Rainforest Mapping in Gabon. *Forests*, v. 11, p. 1 – 17, 2020.