

## **COMPARAÇÃO ENTRE DADOS COM E SEM CORREÇÃO ATMOSFÉRICA NA CLASSIFICAÇÃO DA COBERTURA DA TERRA DE UMA ÁREA DA CAATINGA UTILIZANDO O GOOGLE EARTH ENGINE**

*Comparison between Data with and without Atmospheric Correction on Land  
Cover Classification of an Area in the Brazilian Biome Caatinga using Google  
Earth Engine*

**Khalil Ali Ganem<sup>1</sup>, Gustavo Macedo de Mello Baptista<sup>1 e 2</sup>, Washington de  
Jesus Sant'Anna da Franca Rocha<sup>2</sup>, Rodrigo Nogueira Vasconcellos<sup>2</sup>,  
Marcos Reis Rosa<sup>3</sup> & Deorgia Tayane Mendes de Souza<sup>1 e 2</sup>**

**<sup>1</sup>Universidade de Brasília – UnB  
Instituto de Geociências**

Campus Universitário Darcy Ribeiro, 70910-900, Caixa Postal 4465, Brasília, DF – Brasil  
khalilganem@hotmail.com; gmbaptista@unb.br; deorgiasouza.geo@gmail.com

**<sup>2</sup>Universidade Estadual de Feira de Santana – UEFS**

**Programa de Pós-Graduação em Modelagem em Ciências da Terra e do Ambiente**  
Km 3, BR 116, Módulo 5 – Campus Universitário, 44031-460, Feira de Santana, BA - Brasil  
francarocha@gmail.com; rnvuefsppgm@gmail.com

**<sup>3</sup>Universidade de São Paulo – USP**

**Departamento de Geografia**  
Avenida Professor Lineu Prestes, 338, Caixa Postal 72042, 05508-000, São Paulo, SP – Brasil  
marcosrosa@usp.br

*Recebido em 7 de Abril, 2017/ Aceito em 24 de Setembro, 2017  
Received on April 7, 2017/ Accepted on September 24, 2017*

### **RESUMO**

O Mapeamento Anual da Cobertura e Uso do Solo no Brasil – MapBiomias – é um projeto voltado para a classificação da cobertura da terra em imagens dos satélites da série Landsat. Em sua coleção 2, que envolve classificações geradas de 2000 a 2016, o MapBiomias utilizou, de 2013 em diante, imagens do Landsat-8 em reflectância no topo da atmosfera (*TOA*). As classificações foram geradas baseadas em uma árvore de decisão definida empiricamente pela equipe do projeto. No entanto, é importante verificar como as classificações se comportam após correção atmosférica pelos algoritmos *Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH)* e *Quick Atmosphere Correction (QUAC)*. Assim, a partir da seleção da carta topográfica SD-23-X-D (Bom Jesus da Lapa – BA), as classificações foram geradas pelo *Code Editor* do *Google Earth Engine*, plataforma de processamento de imagens de satélite em nuvem, que permite a interpretação e análise de uma ampla gama de informações. Em seguida, foi feita uma avaliação da acurácia à fim de identificar o dado que melhor representou as classes mapeadas. O dado em reflectância *TOA* e os dados corrigidos pelo *FLAASH* e *QUAC*, apresentaram, respectivamente, valores do *Tau* de 55,45%, 68,90% e 64,89%, *Kappa* de 48,28%, 64,60% e 61,55%, e, exatidão global de 62,88%, 74,09% e 70,74%. Ao final, concluiu-se que a correção atmosférica impactou positivamente a classificação da cobertura da terra da carta SD-23-X-D. A diferença de aproximadamente 4%

a menos do *QUAC* em relação ao *FLAASH*, torna-o mais apropriado, dada sua menor complexidade e maior rapidez no tempo de processamento.

**Palavras-chave:** SD-23-X-D, Landsat-8, *FLAASH*, *QUAC*, MapBiomias.

## ABSTRACT

The Brazilian Annual Land Use and Land Cover Mapping – MapBiomias – is a project focused on the land cover classification in images of the Landsat series satellites. In its collection 2, involving classifications generated from 2000 to 2016, MapBiomias used, from 2013 onwards, Landsat-8 top-of-atmosphere (TOA) reflectance images. The classifications were generated based on a decision tree empirically defined by the project team. However, it is important to verify how classifications behave after atmospheric correction by *Fast Line-of-sight Atmospheric Analysis of Hypercubes* (FLAASH) and Quick Atmosphere Correction (QUAC) algorithms. Thus, after the selection of the SD-23-X-D (Bom Jesus da Lapa – BA) topographic chart, the classifications were generated by the Code Editor of Google Earth Engine, a cloud-based satellite image processing platform, which enables the interpretation and analysis of a wide range of information. Then, an accuracy assessment was performed to identify the data that best represented the mapped classes. The TOA reflectance data and the data corrected by FLAASH and QUAC showed, respectively, Tau values of 55.45%, 68.90% and 64.89%, Kappa of 48.28%, 64.60% and 61.55%, and, overall accuracy of 62.88%, 74.09% and 70.74%. In the end, it was concluded that atmospheric correction positively impacted the land cover classification of the SD-23-X-D chart. The difference of approximately 4% less of QUAC compared to FLAASH makes it more appropriate, given its lower complexity and faster processing time.

**Keywords:** SD-23-X-D, Landsat-8, FLAASH, QUAC, MapBiomias.

## 1. INTRODUÇÃO

A Caatinga é o bioma menos estudado dentre as regiões naturais brasileiras, sendo também a única cujos limites restringem-se inteiramente ao território nacional. A maior parte do esforço científico em relação ao bioma está voltado para regiões específicas do seu contexto territorial, de onde se destacam as principais capitais nordestinas, por exemplo. Além disso, a Caatinga é a região natural brasileira menos protegida, com menos de 2% de seu território coberto por unidades de conservação, sem considerar o extenso processo de degradação ambiental provocado pelo uso insustentável dos seus recursos naturais (LEAL & SILVA, 2003).

Isso faz com que os esforços para a compreensão da dinâmica de uso e ocupação da terra devam ser intensificados no bioma, a fim de identificar o padrão de distribuição espacial de cada classe de cobertura da terra e as possíveis causas associadas à sua maior ou menor predominância no bioma ao longo do tempo. Daí vem a importância da classificação automatizada, uma vez que favorece a conservação ambiental a partir da categorização dos pixels da imagem analisada em diferentes classes de uso e ocupação da terra, permitindo identificar seu padrão de distribuição espacial.

Para melhor avaliar a classificação da cobertura da terra, torna-se necessário compreender a diferença entre uso e cobertura, tendo em vista o fato de cada um remeter a uma diferente escala de abrangência na área a ser mapeada. De acordo com o Instituto Brasileiro de Geografia e Estatística, o uso da terra é uma extensão do território apropriada e usada, e sua intensidade é determinada pelo contexto econômico, social e político da região avaliada. A cobertura da terra, por sua vez, envolve os elementos da natureza tais como vegetação, seja ela nativa ou não, água, rocha e demais construções artificiais que recobrem a superfície (IBGE, 2006). A classificação da cobertura da terra é considerada como pré-requisito para aplicações em sensoriamento remoto, tais como mapeamento temático, monitoramento ambiental, dentre outras (LV *et al.*, 2013).

A interferência dos efeitos atmosféricos pode impactar a imagem no sentido de tornar alguns de seus dados “mascarados”. Consequentemente, devido a esse fato, e buscando um aumento na precisão dos dados de sensoriamento remoto, torna-se necessário tomar medidas para corrigir esses efeitos (RICHARDS & JIA, 2006). Dentre essas medidas, destaca-se a conversão dos dados de radiância para reflectância. Porém, devido à mistura do sinal da atmosfera com o sinal da superfície, o valor

obtido é denominado reflectância aparente, planetária ou no topo da atmosfera. Por outro lado, quando técnicas apropriadas são utilizadas com o objetivo de atenuar a contaminação promovida pela atmosfera no sinal captado, é obtida a reflectância de superfície (ANTUNES, 2003).

Atualmente, existem vários modelos que são utilizados na correção atmosférica de imagens multiespectrais. A correção atmosférica pode alterar os valores de reflectância em algumas bandas de diferentes formas, pois os efeitos de absorção e espalhamento atuam em menor ou maior escala, e este fato depende do comprimento de onda ou das bandas em que o sensor registra a imagem. Para a reflectância na superfície, a precisão do método de transferência radiativa e o modelo atmosférico têm influência significativa (ROJAS, 2005).

Um dos projetos nacionais no âmbito da classificação da cobertura da terra é o Mapeamento Anual da Cobertura e Uso do Solo no Brasil – MapBiomas, o qual envolve os cinco biomas brasileiros e tem passado por constantes adaptações a fim de tentar obter o melhor produto (PROJETO MAPBIOMAS, 2017a).

As diversas classes abordadas no âmbito do projeto estão sendo identificadas e mapeadas com base em imagens das séries Landsat 5, 7 e 8 (sensores TM, ETM+ e OLI, respectivamente), sendo que, atualmente, para o Landsat 5 e 7, são utilizadas coleções de imagens em reflectância de superfície ortorretificadas (L1T), enquanto que, para o Landsat-8, os dados utilizados são de reflectância no topo da atmosfera. Isso ocorre porque os produtos disponibilizados pela plataforma do Google Earth Engine, os quais estão vinculados à biblioteca do United States Geological Survey (USGS), não possuem dados em reflectância de superfície ortorretificados para o Landsat-8. No entanto, os dados em reflectância de superfície podem ser obtidos a partir do dado bruto e dos algoritmos presentes em alguns softwares de processamento de imagens, como o ENVI (EXELIS VISUAL INFORMATION SOLUTIONS, 2017), por exemplo, dentre os quais se destacam o Fast Line-of-sight Atmospheric Analysis of Hypercubes (FLAASH) (ADLER-GOLDEN *et al.*, 1998) e o Quick Atmosphere Correction (QUAC)

(BERNSTEIN *et al.*, 2004).

Com isso, a partir da seleção prévia de uma área da Caatinga, representada pela carta topográfica SD-23-X-D (Bom Jesus da Lapa – BA), o presente trabalho objetivou comparar dados do Landsat-8/OLI, com a atmosfera corrigida por meio dos algoritmos FLAASH e QUAC, e o produto em reflectância no topo da atmosfera (LC8\_L1T\_TOA), disponibilizado pelo USGS e utilizado para gerar as classificações da coleção 2 do MapBiomas no âmbito da Caatinga. A partir de análises visuais e da aplicação de testes estatísticos foi identificado o dado que melhor se ajusta ao mapeamento da cobertura da terra no bioma.

### **1.1 MapBiomas**

Em março de 2015, o Sistema de Estimativas e Emissões de Gases de Efeito Estufa (SEEG), o qual é uma iniciativa do Observatório do Clima (OC), realizou um seminário na cidade de São Paulo e convidou especialistas em sensoriamento remoto e mapeamento de vegetação para se reunirem e verificarem a possibilidade de se produzir mapas anuais de cobertura e uso do solo para todo o Brasil de forma significativamente mais barata, rápida e atualizada, comparativamente aos métodos e práticas atuais, e que possibilitem recuperar o histórico das últimas décadas. Tal iniciativa deu origem ao Projeto de Mapeamento Anual da Cobertura e Uso do Solo no Brasil (MapBiomas).

Algumas das condições para o desenvolvimento do projeto foi a existência de uma capacidade de processamento sem precedentes e um alto grau de automatização do processo, além da participação de uma comunidade de especialistas em cada bioma e temas transversais. Assim, a partir deste diagnóstico foi estabelecido um contato com o Google que gerou um termo de cooperação técnica para desenvolver a iniciativa tendo como base a plataforma Google Earth Engine (GEE). Especificamente, o projeto teve início em julho de 2015, com um treinamento da equipe do MapBiomas sobre o *Google Earth Engine* na cidade de *Mountain View*, na Califórnia. Informações mais detalhadas sobre o projeto podem ser obtidas no portal do projeto na internet ([mapbiomas.org](http://mapbiomas.org)).

Os primeiros trabalhos desenvolvidos

pelo MapBiomias objetivaram, basicamente, diferenciar áreas vegetadas de não vegetadas a fim de identificar o grau de cobertura vegetal presente nos biomas brasileiros. Essa etapa também é referida como Coleção 1. A Coleção 2, definida em 2016 e já incorporada por este trabalho, busca aumentar o número de classes, para especificar melhor as áreas vegetadas e não vegetadas. A metodologia utilizada pelo MapBiomias serviu como subsídio para o desenvolvimento deste trabalho.

A unidade de mapeamento do MapBiomias é a Carta Internacional do Mundo ao Milionésimo na escala 1:250.000, cobrindo uma área de 1°30' de longitude por 1° de latitude. Como o próprio nome sugere, trata-se de cartas de todas as regiões do globo terrestre e a sua utilização, de acordo com Nozawa (1987), restringe-se a trabalhos que envolvam grandes áreas.

Um total de 558 cartas (ou folhas) cobre o território nacional. Os mapas de referência utilizados para avaliar os resultados baseiam-se nos mapas do Projeto de Conservação e Utilização Sustentável da Diversidade Biológica Brasileira (PROBIO/MMA, 2006) divulgados pela Secretaria de Biodiversidade e Florestas do Ministério do Meio Ambiente (MMA), em parceria com uma série de órgãos, dentre os quais destacam-se o IBGE e a Universidade Estadual de Feira de Santana (UEFS).

## 1.2 Google Earth Engine (GEE)

O *Google Earth Engine* – GEE é uma plataforma utilizada em ampla escala com o objetivo principal de analisar dados ambientais. Uma das grandes vantagens dessa plataforma é o fato dela reunir mais de 40 anos de imagens globais de satélites, de décadas anteriores e atuais (GORELICK *et al.*, 2017). Essa plataforma também fornece ferramentas e uma vasta capacidade computacional para a realização de análises e a exploração de uma elevada amplitude de dados. Dentre suas principais aplicações destacam-se a detecção de desmatamento, a classificação da cobertura da terra e suas alterações correspondentes, assim como as estimativas de biomassa florestal e de carbono.

A ferramenta integra um conjunto de dados geoespaciais com cobertura global, capacidade

de armazenamento e processamento altamente elevados correspondentes ao ambiente de computação da nuvem Google, e um ambiente de desenvolvimento integrado suportando a implementação de algoritmos de análise e processamento de dados geoespaciais nas linguagens *Java Script* e *Python* (GORELICK *et al.*, 2017).

Dentre os dados geoespaciais disponíveis no catálogo do GEE é possível destacar vários produtos com cobertura regional e/ou global de diversos sensores (ASTER, MODIS) e satélites (LANDSAT, NOAA), assim como produtos *MeaSURES* (novos registros de imagens da superfície terrestre que partem do princípio da junção de diferentes dados com o objetivo de formar uma série temporal única, consistente e coerente).

## 1.3 Landsat-8

O Landsat-8 foi lançado em 11 de fevereiro de 2013 e tem como objetivo principal coletar, arquivar, processar e distribuir dados científicos de forma consistente com o registro de dados existente dos demais sensores da série Landsat (IRONS *et al.*, 2012). O satélite é referenciado ao sistema WRS-2 e opera em uma órbita polar sol síncrona próxima de 705 km de distância da superfície terrestre e cruzando a linha do Equador aproximadamente às 10:13, horário médio local durante o nó descendente de cada órbita.

O Landsat-8 possui resolução radiométrica de 16 bits, resolução temporal de 16 dias e 185 km de faixa imageada e destaca-se pela presença de dois instrumentos imageadores: um óptico, *Operational Land Imager* (OLI), e outro abrangendo a faixa do infravermelho termal, *Thermal Infrared Sensor* (TIRS), o qual apresenta duas bandas espectrais de ondas longas (bandas 10 e 11) que equivalem ao comprimento de onda antes coberto por uma única banda do sensor TM, a banda 6 (USGS, 2012). A resolução espacial do OLI é de 30 metros e a do TIRS, que é originalmente de 100 metros, foi reamostrada pelo USGS para 30 metros para fins de comparação com dados do OLI.

## 2. MATERIAL E MÉTODOS

Para realização do presente trabalho foi necessário desenvolver as etapas metodológicas descritas na Figura 1.



Fig. 1 – Fluxograma metodológico do MapBiomias adaptado para o presente trabalho. Legenda: *FMASK* – algoritmo *Function of Mask*; *NDFI* – Índice Normalizado por Diferença de Fração.

## 2.1 Caracterização Climático-Temporal

O intervalo temporal definido pelo MapBiomias para a geração dos mosaicos temporais no bioma Caatinga foi baseado numa caracterização climático-temporal, a qual se destacou pela coleta de dados pluviométricos. Esses dados serviram como subsídio para a definição da janela temporal mais adequada a ser aplicada para o contexto desse bioma, uma vez que a sazonalidade é um dos fatores de maior impacto na classificação da cobertura da terra nessa região.

A partir disso, definiu-se que a janela temporal mais apropriada é de seis meses, devendo abranger, prioritariamente, os meses de janeiro a junho, fevereiro a julho e, em último caso, março a agosto. Isso ocorre porque o primeiro semestre é marcado pela maior ocorrência de chuvas no bioma.

As etapas de geração do mosaico temporal, modelo de mistura espectral e classificação da cobertura da terra foram todas desenvolvidas pelo GEE a partir de scripts pré-definidos, os quais detêm as informações e parâmetros necessários para gerar o dado final.

Após a geração do mosaico temporal e a realização da correção atmosférica a partir do *FLAASH* e *QUAC*, por meio do software *ENVI* 5.1, as imagens foram reinseridas no GEE para o processo de classificação. Ao final, a etapa de pós-classificação foi desenvolvida por meio do software *ArcGIS* 10.1, tanto para a produção dos mapas temáticos, como para a extração dos dados para posterior avaliação da acurácia.

## 2.2 Seleção da Área de Estudo (Carta SD-23-X-D)

Como o sensor a ser utilizado foi o Landsat-8/OLI, logo as imagens disponíveis são posteriores a fevereiro de 2013. A definição da carta mais apropriada e do melhor ano para análise deu-se a partir de uma avaliação da qualidade dos mosaicos. Essa avaliação foi realizada com base na verificação do percentual de nuvens e presença de linhas e/ou pixels ruidosos, que poderiam ter dificultado a visualização dos alvos presentes no mosaico. Assim, foi feita uma análise da qualidade dos mosaicos buscando enquadrá-los em três diferentes categorias, com base no percentual de nuvens: boa (<10%), regular (entre 10 e 30%) e ruim (> 30%).

Nos anos considerados (de 2013 a 2016), a carta SD-23-X-D (Bom Jesus da Lapa – BA) foi uma das poucas que apresentou qualidade boa, dada a janela temporal de janeiro a junho. Entretanto, o ano de 2015 apresentou presença quase nula de ruídos ou nuvens, o que justificou sua escolha para a geração das classificações e posterior avaliação da acurácia.

A referida carta está limitada pelas coordenadas 42°00' a 43°30' de longitude oeste e 13° a 14° de latitude sul, e está localizada na região sudoeste da Bahia, totalizando uma área de cerca de 18 mil km<sup>2</sup>. Vale ressaltar que aproximadamente 8,75% da área total da carta pertence ao bioma Cerrado. Seguindo a metodologia do MapBiomias, cartas que

abranjam mais de um bioma são processadas com parâmetros de ambos. No entanto, como mais de 90% da área total da carta pertence à Caatinga, foram considerados apenas os parâmetros desse bioma na árvore de decisão que gerou a classificação da carta SD-23-X-D.

Ao observar o mapa de referência da vegetação produzido pelo MMA, em parceria com diversos órgãos (Figura 2), a carta caracteriza-se por apresentar as seguintes formações vegetais, cuja nomenclatura está de acordo com o Manual

Técnico da Vegetação do IBGE (2012): Áreas de Tensão Ecológica; Floresta Estacional Decidual; Floresta Estacional Semidecidual; Savana; Savana-Estépica; Formação Pioneira com Influência Fluvial e/ou Lacustre; Vegetação Secundária. As demais classes não vegetais identificadas correspondem a corpos d'água, áreas de exploração agropecuária e áreas com influência urbana. Esse mapa foi utilizado como um dos elementos de validação da classificação gerada pela metodologia do MapBiomias.

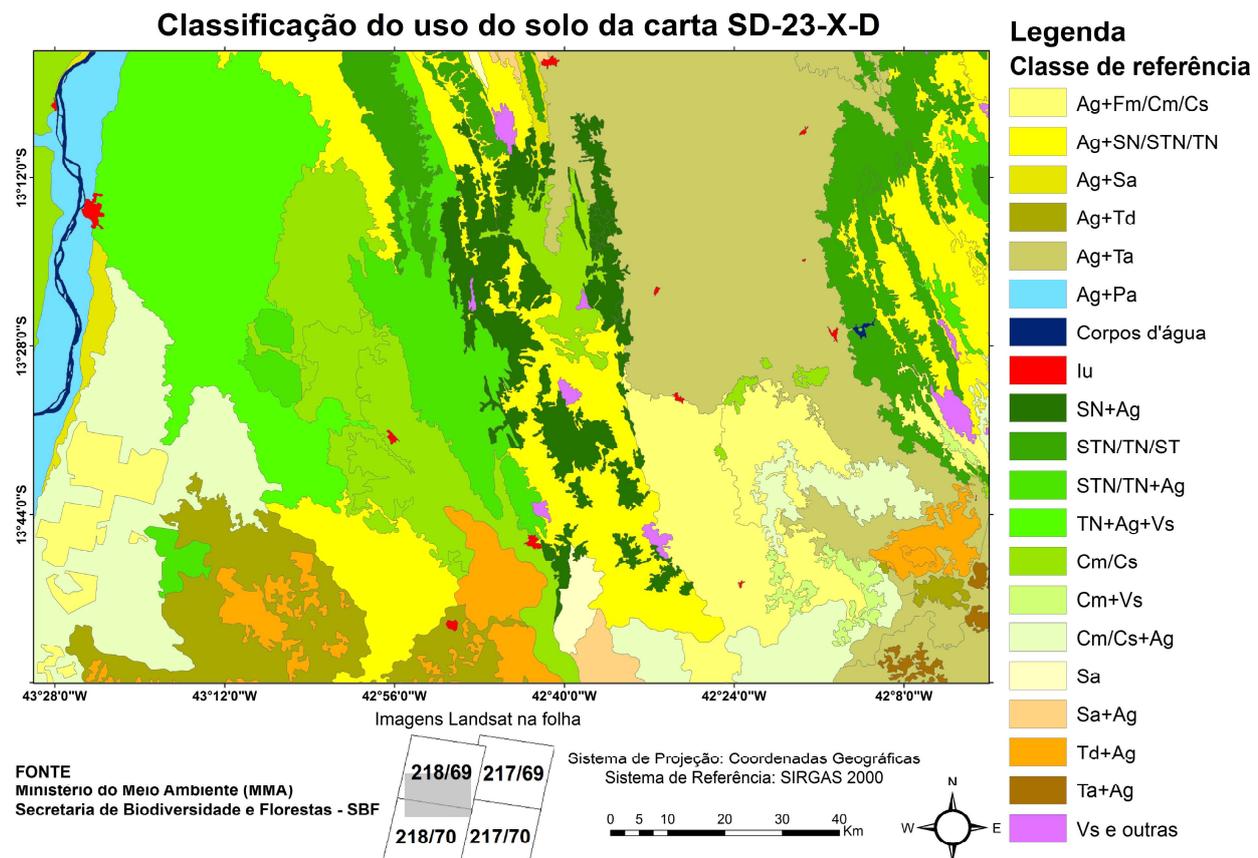


Fig. 2 – Mapa de referência da cobertura da terra da carta SD-23-X-D. Fonte: Adaptado de Ministério do Meio Ambiente (2006). Legenda: **Ag** – Agropecuária; **Cm** – Floresta Estacional Decidual Montana; **Cs** – Floresta Estacional Decidual Submontana; **Fm** – Floresta Estacional Semidecidual Montana; **Fs** – Floresta Estacional Semidecidual Submontana; **Iu** – Influência Urbana; **Pa** – Formação Pioneira com influência fluvial e/ou lacustre; **Sa** – Savana Arborizada; **SN** – Savana/Floresta Estacional; **ST** – Savana/Savana-Estépica; **STN** – Savana/Savana-Estépica/Floresta Estacional; **Ta** – Savana-Estépica Arborizada; **Td** – Savana-Estépica Florestada; **TN** – Savana-Estépica/Floresta Estacional; **Vs** – Vegetação Secundária.

### 2.3 Geração do Mosaico Temporal

O período considerado para a geração do mosaico temporal foi de 01/01/2015 a 30/06/2015 e o percentual de nuvens considerado foi de 1%. Isso resultou num mosaico de alta qualidade em termos visuais, composto por um total de 10 cenas dos seguintes pontos/órbitas: 217/70

(1 cena), 218/69 (2 cenas); 218/70 (2 cenas); 219/69 (2 cenas); e 219/70 (3 cenas). Para obter o mosaico temporal, inicialmente, removeu-se a borda da cena da imagem Landsat para suprimir eventuais ruídos. São considerados ruídos os pixels das cenas Landsat que não possuem dados em todas as bandas.

O próximo passo caracterizou-se pela aplicação de um algoritmo de detecção e remoção de nuvens, o *Function of Mask (FMASK)*, proposto por Zhu e Woodcock (2012). O algoritmo *FMASK* é um método para detecção de nuvens e as sombras provocadas pelas mesmas em imagens Landsat (ZHU e WOODCOCK, 2012).

O algoritmo requer como entrada dados de reflectância no topo da atmosfera e temperatura de brilho. A partir de critérios baseados nas propriedades físicas das nuvens, os pixels considerados como nuvens potenciais são separados dos pixels de céu claro.

Em seguida, por meio da combinação das probabilidades dos parâmetros temperatura, variabilidade espectral e brilho dos alvos, é produzida uma “máscara” para nuvens sobre superfícies terrestres e aquáticas, separadamente.

Posteriormente, os pixels considerados como nuvens potenciais, juntamente com a “máscara” criada, são utilizados para gerar uma provável camada de nuvens. Zhu & Woodcock (2012) ainda destacam que o efeito do sombreamento gerado pelas nuvens, na região do infravermelho, é utilizado para gerar uma camada de sombra potencial. Além disso, informações como o ângulo de visada do satélite e a iluminação são utilizadas como forma de prever possíveis sombras nos locais onde as nuvens são detectadas. Os valores da “máscara” criada pelo algoritmo são: 0 – céu claro; 1 – água; 2 – sombra; 3 – neve; 4 – nuvem; e 255 – não identificado.

Após a aplicação do *FMASK*, os pixels marcados como nuvem ou sombra de nuvem são removidos da imagem original, mantendo apenas pixels com valores válidos. Por fim, uma redução das cenas consideradas no intervalo temporal pré-definido pelo valor da mediana dos pixels válidos gera o mosaico da imagem final da carta. O filtro da mediana é importante por remover os valores extremos em cada pixel, que podem ser ruídos do sensor ou contaminação por névoa. A mediana é superior à média por excluir os extremos. No entanto, o valor existente do pixel de alguma imagem, dentre as consideradas, é mantido, sem modificação.

Enquanto a mediana preserva as transições rápidas, a média, por considerar os valores da janela no cálculo do valor final de cada pixel, atenua estas transições. A janela temporal de, no máximo, seis meses buscou ser mantida para reduzir mudanças significativas na cobertura florestal e aumentar a área da superfície observada a partir do processo de filtragem de nuvens descrito anteriormente.

#### 2.4 Etapa de Processamento no ENVI

Primeiramente, os dados brutos foram convertidos para radiância no topo da atmosfera (*TOA*). De acordo com Chander *et al.* (2009), a calibração radiométrica dos sensores da série Landsat envolve o reescalonamento dos números digitais brutos transmitidos pelo satélite para números digitais calibrados, os quais possuem a mesma escala radiométrica para todas as cenas processadas na superfície terrestre em um dado período.

A conversão dos dados de radiância para reflectância de superfície foi feita por meio dos métodos de correção atmosférica *FLAASH* e *QUAC*, ambos presentes no software *ENVI 5.1*. Para ambos os métodos de correção atmosférica, considerou-se o dado de radiância como entrada. No caso do *FLAASH*, utilizou-se, basicamente, fator de escala (*single scale factor*) 10, altura de escala do aerossol (*aerosol scale height*) de 1,5 km e taxa de CO<sub>2</sub> de 400 ppm. O modelo atmosférico selecionado foi o tropical. A visibilidade pode ser estimada pela própria imagem, quando nela são detectados pixels escuros. Entretanto, antes de processar o *FLAASH*, é importante definir este parâmetro. Para isso, utilizou-se o valor médio entre as duas equações mais consolidadas na literatura para a estimativa da visibilidade: a de Deschamps (1981) e a de Koschmieder (1924). O resultado de visibilidade obtido foi de 21 km. Os valores utilizados para os demais parâmetros foram os pré-definidos pelo *FLAASH* (default).

O fluxograma da Figura 3 mostra as etapas realizadas por este trabalho após serem incorporadas à metodologia do MapBiomass, identificando também as ferramentas utilizadas para os processamentos.



As imagens fração dos componentes puros mencionados também foram necessárias para calcular o *Savannah Ecosystem Fraction Index (SEFI)*, desenvolvido pela equipe do MapBiomias, do bioma Pantanal (PROJETO MAPBIOMAS, 2017b). O *SEFI* foi desenvolvido a partir de um ajuste do índice *NDFI* e também é utilizado na Caatinga, conforme apresentado pela Equação 5. Já a equação 6 apresenta o cálculo da variável *GVNPVs*.

$$SEFI = \frac{GVNPVs - (SOIL + CLOUD)}{GVNPVs + SOIL + CLOUD} \quad (5)$$

$$GVNPVs = \frac{(GV + NPV)}{(100 - SHADE)} \quad (6)$$

Esse ajuste foi feito com o objetivo de minimizar o efeito da sazonalidade da caatinga no *NDFI*, uma vez que o período seco afeta a vegetação no sentido de promover uma maior predominância da vegetação seca em detrimento da vegetação verde, diminuindo os resultados que indicam maior predominância de solo exposto em períodos de seca, ou seja, reduzindo a variação do *NDFI*.

## 2.6 Classificação da Cobertura da Terra

As imagens fração obtidas a partir do Modelo de Mistura Espectral foram utilizadas na classificação da cobertura da terra para as seguintes classes: Agropecuária, Água, Áreas Não Vegetadas, Área Úmida, Floresta Aberta, Floresta Densa, e Outras Formações Nativas Não Florestais.

É importante fazer algumas ressalvas: a classe Agropecuária envolve tanto cultivos agrícolas quanto áreas de produção pecuária, como pastagens. A classe Outras Formações Nativas Não Florestais envolve fitofisionomias que não possuem um aspecto florestal, ou seja, com um menor porte e um maior espaçamento entre as espécies arbóreo-arbustivas. Com isso, tendo como base uma abordagem bastante genérica e abrangente, essa classe será denominada, deste ponto em diante, na legenda dos mapas e nos textos que seguem, de Savana, para ser melhor associada com esse padrão de vegetação inerente ao bioma Caatinga: a Savana-Estépica.

Tabela 1: Caracterização das classes de cobertura da terra mapeadas

CLASSES	DESCRIÇÃO
Agropecuária	Terra utilizada para a produção de alimentos. Incluem-se, também, terras em descanso, lavouras temporárias, permanentes e pastagens plantadas.
Água	Incluem-se todas as classes de água, tais como: rios, riachos, lagos naturais e reservatórios artificiais, entre outros.
Áreas Não Vegetadas	Cobertura da terra de natureza variada, tais como áreas urbanas, afloramentos rochosos e solo exposto.
Floresta Aberta	Baseando-se no Manual Técnico de Vegetação do IBGE (2012): Savana-Estépica Arborizada.
Floresta Densa	Baseado no Manual Técnico de Vegetação do IBGE (2012): Floresta Estacional Decidual e Semidecidual, e Savana-Estépica Florestada e Savana-Estépica Arborizada.
Savana	Baseado no Manual Técnico de Vegetação do IBGE (2012): Savana-Estépica Parque e Savana-Estépica gramíneo-lenhosa.

Assim, em termos gerais, a classe de Floresta Densa buscou englobar áreas com frações vegetais vigorosas, com elevada densidade e atividade fotossintética. A classe Floresta Aberta, por sua vez, englobou áreas da Savana-Estépica Arborizada, com maior espaçamento entre a vegetação. A classe Savana englobou um padrão característico das vegetações do tipo Savana-Estépica Parque e Savana-Estépica gramíneo-lenhosa, com um padrão ainda maior de espaçamento entre a vegetação, sendo essas predominantemente de pequeno porte.

Outro ponto que merece ressalva é o fato das referidas classes apontadas no parágrafo anterior serem adotadas pelo MapBiomias e aplicadas, com a mesma nomenclatura que foram apresentadas, para outros biomas brasileiros. Com isso, torna-se necessário e fundamental que haja uma descrição do que é abrangido por cada uma das classes adotadas, especialmente, as que envolvem formações vegetais.

De antemão, é importante que fique claro que o principal intuito é diferenciar essas classes

quanto ao nível de abundância da vegetação, de onde infere-se densidade, proximidade e tipos das formações florestais e campestres. A Tabela 1 apresenta uma descrição do que foi considerado para cada classe.

Após a caracterização das classes a serem mapeadas, fez-se a classificação seguindo alguns passos. Primeiro, foi aplicada uma “máscara” de nuvens para mapear áreas cobertas por

nuvens que não puderam ser filtradas durante o processo de geração do mosaico temporal da carta. Esse procedimento foi feito por meio do algoritmo *FMASK*. Em seguida, foi aplicada uma árvore de decisão (Figura 4), cuja estrutura, parâmetros e valores dos nós foram definidos empiricamente para mapear cada classe (PROJETO MAPBIOMAS, 2017a).



Fig. 4 – Árvore de decisão para classificação da cobertura da terra na Caatinga com os parâmetros utilizados e seus respectivos valores parametrizados para separar as classes que constam nos quadros coloridos. Legenda: *GV* – Vegetação Verde; *GVs* – Fração da vegetação verde normalizada pela sombra; *NPV* – Vegetação não fotossinteticamente ativa; *SEFI* – Índice de Fração do Ecossistema Savana; *SHADE* – Sombra; *SLOPE* – Declividade; *SOIL* – Solo.

O ponto de partida para a interpretação da árvore é o algoritmo *FMASK*. Conforme já foi explicitado anteriormente, o valor do algoritmo igual a 1 representa água. A escolha desse parâmetro como raiz parte do princípio de que a água é um alvo facilmente identificado e que se confunde menos com os principais alvos presentes na cobertura da terra. Por isso, a criação dessa “máscara” de água como parâmetro inicial da árvore de decisão.

Entretanto, verificou-se empiricamente, que também é possível considerar como raiz da árvore de decisão, o parâmetro *SHADE* (sombra), com o valor a partir de 85. Em seguida, os ramos que dão origem aos nós *SLOPE* (declividade) e índice *SEFI* produzem um forte impacto na distinção dos demais alvos, especialmente os vegetais.

Com base nisso, para a árvore de decisão apresentada, foi feito um reescalonamento dos

valores do *SEFI*, os quais passaram a variar de 0 a 200, sendo que, quanto maior a presença de vegetação verde e sombra, mais próximos de 200 ficaram os valores. O reescalonamento foi feito por meio da soma do pixel após a aplicação do referido índice a 1, seguida da sua multiplicação por 100. Esse nó é responsável pela distinção entre um pixel que pode ser classificado como Floresta Densa ou Floresta Aberta, de um pixel que pode ser classificado como outra formação nativa não florestal, como o caso da Savana.

O parâmetro *GVs*, conforme apresentado anteriormente, refere-se à vegetação verde normalizada pela sombra, proposta por Souza Jr. *et al.* (2005). Os parâmetros *GV*, *NPV*, *SOIL* e *SHADE* referem-se à vegetação verde, vegetação não fotossinteticamente ativa, solo e sombra, respectivamente, e seus valores são derivados do modelo de mistura espectral, o qual calcula uma fração para cada alvo. Assim, trata-se de valores que podem variar de 0 a 100%.

A sombra (*SHADE*) do Modelo de Mistura Espectral é calculada da seguinte forma:  $100 - (GV + NPV + SOIL + CLOUD)$ . Todos esses parâmetros são informados pelo *MME*, inclusive o de nuvem. Além disso, é extremamente importante que fique claro que a sombra calculada pelo Modelo de Mistura Espectral não possui relação com a sombra removida pelo *FMASK*, que são oriundas de nuvens.

Conforme pode ser observado na árvore de decisão, o parâmetro sombra (*SHADE*) aparece em dois nós. É importante que o valor do nó desse parâmetro que caracteriza a classe Savana seja sempre maior que o valor do nó anterior, que caracteriza a classe Agropecuária. Isso ocorre porque esse último nó classifica apenas pixels com *SEFI* baixo e, conseqüentemente, acaba sendo mais eficiente na diferenciação de áreas com pouca sombra, que, geralmente, são pastagem e áreas com aspectos vegetais gramíneo-herbáceo. Já Áreas Úmidas (alagadas) costumam ser interpretadas como sombra, o que justifica valores maiores para esse parâmetro no nó que diferencia Áreas Úmidas de Savana. No entanto, para a carta em questão, essa classe não foi identificada.

A informação com relação à declividade (*SLOPE*) foi proveniente dos dados *Shuttle Radar Topography Mission (SRTM)*, disponibilizados pelo USGS, e foi utilizado para fazer correções

mais simples, como vegetações com maior padrão de espaçamento entre si. Áreas com alto *GV* costumam ter pouca sombra e por isso são, normalmente, classificadas como áreas de agricultura. Entretanto, cabe fazer uma ressalva. Vertentes de alta declividade podem ter muita iluminação solar e, por isso, tendem a ser classificadas como agricultura, mesmo que se trate de floresta densa. Isso faz com que a interferência da iluminação solar aliada à declividade seja determinante para diferenciar ambas as classes.

## 2.7 Pós-Classificação

A avaliação da acurácia temática para as classificações foi feita a partir das seguintes etapas: definição do número de amostras necessárias para garantir a melhor qualidade possível da avaliação; distribuição aleatória dessas amostras nas imagens Landsat em composição colorida; checagem por interpretação visual; verificação da exatidão por meio do uso da matriz de confusão e cálculo de índices de concordância de forma geral e por classe.

A classe que apresentou maior predominância na carta SD-23-X-D foi a Floresta Aberta em todos os resultados analisados. Para o cálculo do número de amostras, foi utilizado o método de distribuição multinomial (CONGALTON & GREEN, 2008), que leva em consideração o número de classes e a proporção da classe com maior área na classificação. Além disso, a unidade de amostragem utilizada para este trabalho foi o pixel, uma vez que a similaridade entre as informações espectrais de cada um serviu como embasamento para gerar a classificação de cada carta. Essa forma de validação, com base na estrutura de imagem de satélite, é bastante utilizada por permitir um levantamento ágil e de boa precisão (EPIPHANIO *et al.*, 2002; MALDONADO, 2005; VALENTE, 2001), especialmente se comparado aos levantamentos de campo.

Definido o tamanho da amostra para a carta, realizou-se uma amostragem aleatória simples para a imagem Landsat, por meio da ferramenta *Hawths Analysis Tools* disponível no *ArcGIS 10.2*, o qual utilizou, como espaço amostral, um arquivo de pontos do centroide de cada pixel das imagens de cada carta.

O passo seguinte caracterizou-se pela atribuição das classes correspondentes a cada amostra por meio de interpretação visual na tela do computador, utilizando a imagem ao fundo, além do auxílio do aplicativo *Google Earth Engine*. Esses pontos foram checados e classificados com rigor para então serem utilizados como dados de referência (verdade de campo) para construir a matriz de confusão, a qual permitiu computar a exatidão global e os coeficientes de concordância total e por classe.

## RESULTADOS E DISCUSSÃO

A informação derivada dos dados de sensores remotos, de acordo com Jensen (2009), é geralmente apresentada como uma imagem realçada, um mapa-imagem, mapas temáticos, arquivos espaciais em banco de dados, estatísticas ou gráficos. No caso deste trabalho, os resultados

foram apresentados em mapas e análises estatísticas.

### 3.1 Classificação da Cobertura da terra a partir do TOA

Analisando, visualmente, o mapa de classificação da cobertura da terra, a partir do dado de reflectância no topo da atmosfera, para a carta SD-23-X-D do ano de 2015, verifica-se um maior predomínio das classes de vegetação, especialmente, as classes Floresta Aberta e Floresta Densa (Figura 5).

A área representada por essa carta apresenta baixa interferência antrópica comparada com outras cartas que abrangem o bioma. Além disso, a Água, que é uma classe pouco presente na carta, também está bem representada. A Tabela 2 apresenta a matriz de confusão para a classificação oriunda do dado sem correção atmosférica para a referida carta.

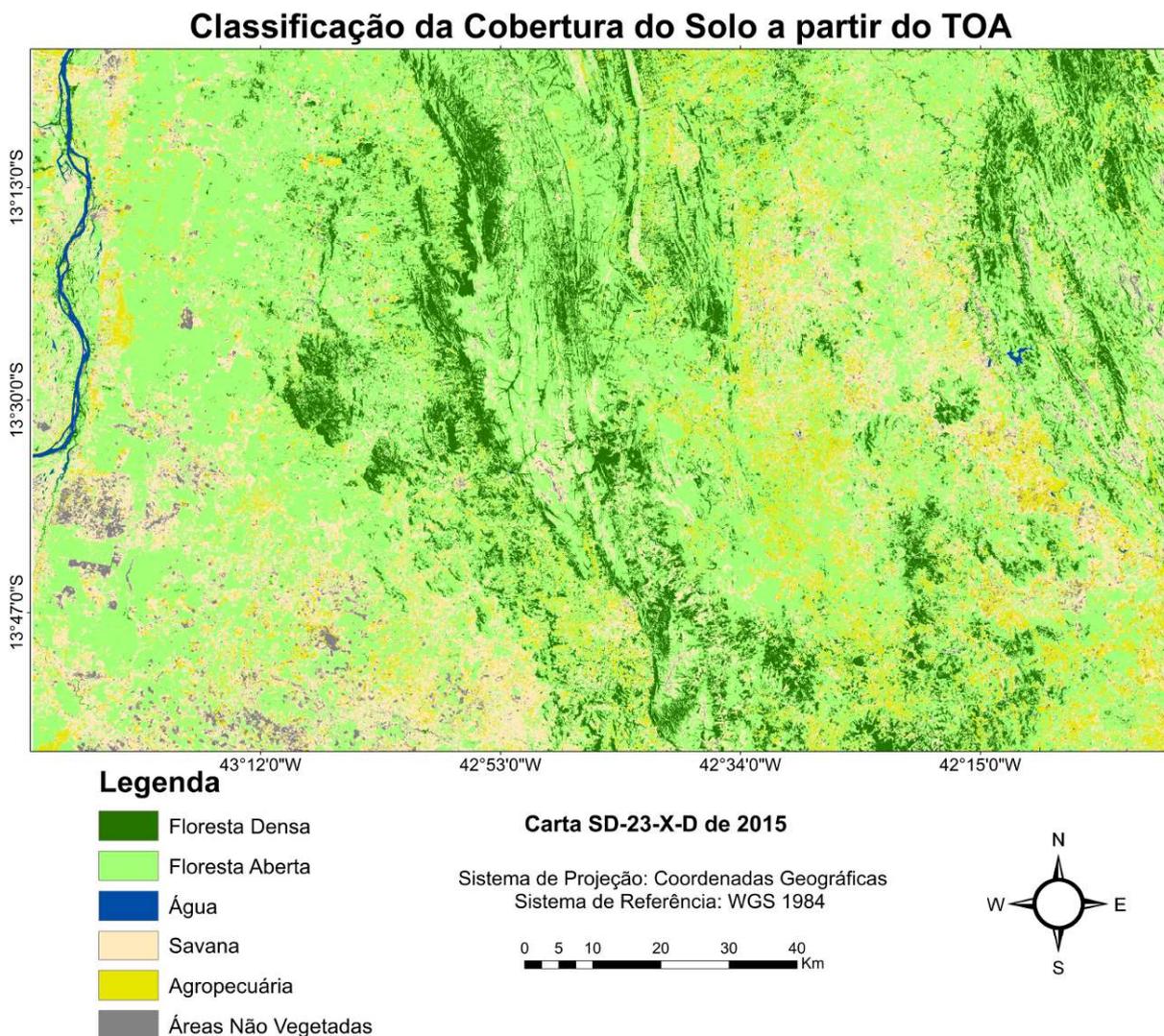


Figura 5 – Mapa da classificação da cobertura da terra do mosaico temporal da carta SD-23-X-D de 2015 a partir do dado de reflectância no topo da atmosfera (aparente) processado pelo *Google Earth Engine*.

Tabela 2: Matriz de Confusão da classificação oriunda a partir do produto LC8\_L1T\_TOA para a carta SD-23-X-D do ano de 2015

Referência (verdade de campo)  Classificação	Agropecuária	Água	Áreas Não Vegetadas	Floresta Aberta	Floresta Densa	Savana	Total
Agropecuária	45	-	8	2	-	5	<b>60</b>
Água	-	3	1	-	-	-	<b>4</b>
Áreas Não Vegetadas	5	-	7	-	-	2	<b>14</b>
Floresta Aberta	44	-	2	232	53	42	<b>373</b>
Floresta Densa	4	-	-	11	68	2	<b>85</b>
Savana	55	-	12	7	-	77	<b>151</b>
Total	<b>153</b>	<b>3</b>	<b>30</b>	<b>252</b>	<b>121</b>	<b>128</b>	<b>687</b>

Observando a matriz de confusão é possível constatar que, de um total de 687 pontos distribuídos aleatoriamente sobre a carta e que representam os dados de referência (verdade de campo), 432 se mostraram concordantes com a classificação. Isso representa uma exatidão global de 62,88%, *Kappa* de 48,28%, e *Tau* de 55,45%. Considerando a escala de qualidade de Landis e Koch (1977), para cada classe individualmente aplicada aos valores do *Kappa* condicional, a classificação oriunda do dado de reflectância aparente pode ser considerada muito boa para a Água, Agropecuária e Floresta Densa, e boa para as classes Áreas Não Vegetadas, Floresta Aberta e Savana. De modo geral, é possível notar que não houve variações muito extremas entre as classes, ou seja, da mesma forma que não houve qualidade excelente, não houve também ruim ou péssima.

A Água apresentou exatidões de 100% e 75% para o produtor e usuário, respectivamente. Já o pior resultado observado foi para a classe referente às Áreas Não Vegetadas, que apresentaram erros de omissão e comissão de, respectivamente, 76,66% e 50%. No entanto, comparando a quantidade de amostras dessa classe com as demais, observa-se uma quantidade bem reduzida, o que mostra menor predominância da mesma na carta. A classe Água, apesar de ter apenas três pontos amostrais,

apresentou elevada concordância em função da facilidade de diferenciação desse alvo em relação aos demais, uma vez que a água apresenta propriedades espectrais únicas.

Dentre as classes de vegetação, a que apresentou melhores resultados foi a Floresta Aberta, com exatidão do produtor de 92,06% e do usuário de 62,19%, seguida da classe Floresta Densa, com 56,19% e 80%, e Savana, com 60,15% e 50,99% para as exatidões do produtor e usuário, respectivamente. Do ponto de vista do produtor, a Savana foi muito confundida com a Floresta Aberta. Já sob a perspectiva do usuário, houve bastante confusão com a classe Agropecuária. Por fim, a classe Agropecuária apresentou erro de omissão elevado, de 70,58%, o que quer dizer que menos de 30% dos pixels foram de fato atribuídos corretamente à classe. Já em termos de erro de comissão, apresentou 25%, ou seja, cerca de um quarto dos pixels foram incluídos em uma classe à qual não pertencem de fato.

### 3.2 Classificação da Cobertura da terra com Correção a partir do QUAC

Ao avaliar visualmente o mapa de classificação da cobertura da terra a partir do dado em reflectância de superfície gerado pelo QUAC (Figura 6), ainda é possível observar o mesmo padrão de domínio de áreas vegetadas

na carta. Entretanto, em comparação com o dado sem correção atmosférica, nota-se que há um aumento da Savana e consequente redução das áreas de Floresta Aberta. Em relação às demais classes, mais uma vez, a Água foi bem

representada e houve um aumento considerável da Agropecuária. As Áreas Não Vegetadas aparecem mais, porém com um leve aumento em relação ao dado *TOA*. A Tabela 3 apresenta a matriz de confusão para essa classificação.

### Classificação da Cobertura do Solo com correção a partir do QUAC

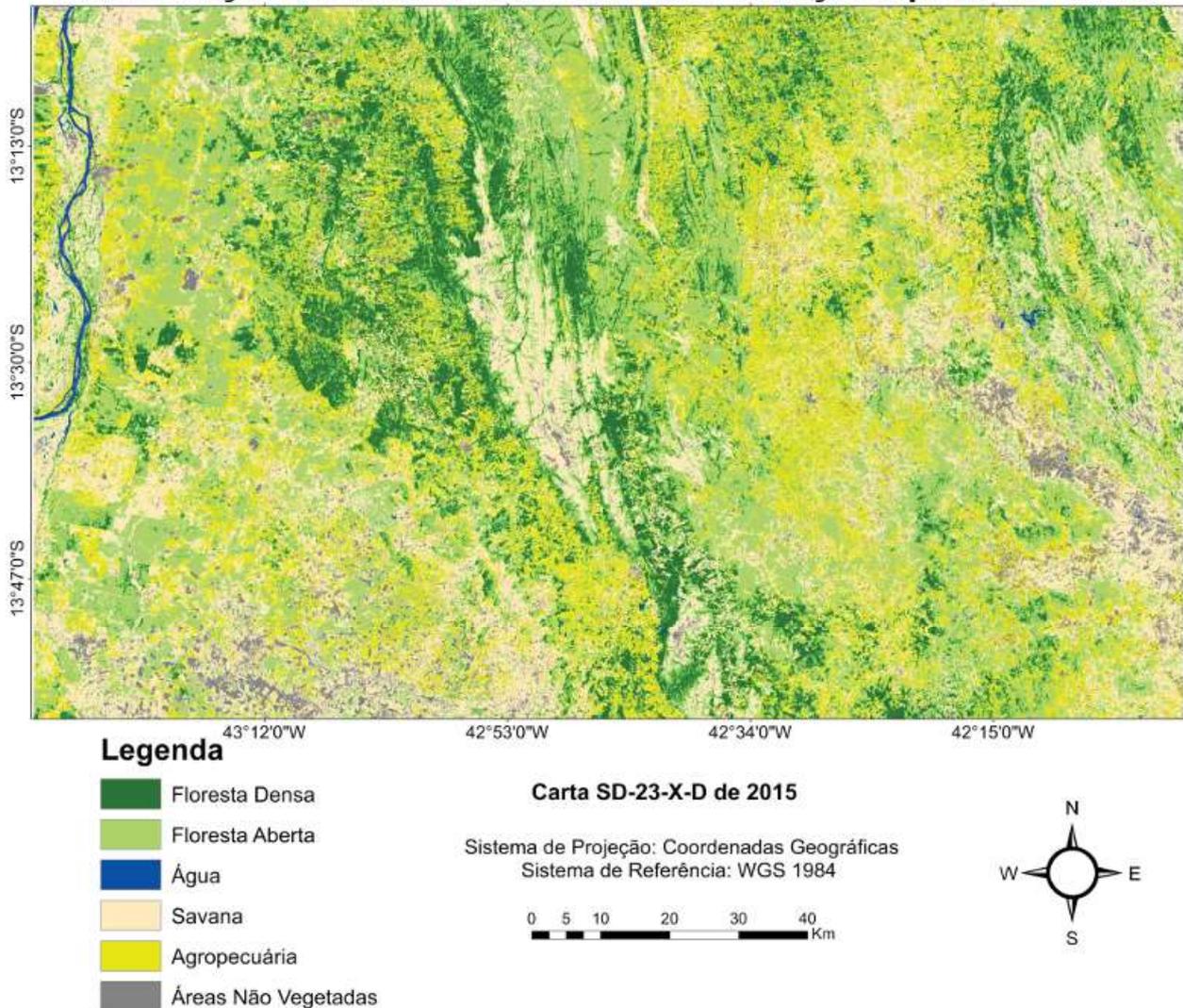


Fig. 6 – Mapa da classificação da cobertura da terra do mosaico temporal da carta SD-23-X-D de 2015 processado pelo *Google Earth Engine* a partir do dado corrigido pelo *QUAC*.

A matriz de confusão mostra que 486 pixels de um total de 687 se mostraram concordantes com os resultados oriundos da classificação, o que representa uma exatidão global de 74,09%, *Kappa* de 64,60%, considerado muito bom, e *Tau* de 68,90%. A classificação de Landis & Koch (1977), proposta para o índice *Kappa*, ao considerar para os valores do *Kappa* condicional, o dado com correção atmosférica, a partir do *QUAC* foi excelente para as classes Água e

Áreas Não Vegetadas, muito boa para a Floresta Densa, Savana e Agropecuária, e boa para a Floresta Aberta. Não houve classes consideradas razoável, ruim ou péssima sob o ponto de vista do *Kappa* condicional, o que demonstra o melhor desempenho dessa classificação em comparação ao dado sem correção atmosférica.

A classe Água apresentou nível de concordância de 100% com a classificação, tanto sob a perspectiva do produtor como do

usuário, o que resultou em erros de comissão e omissão nulos para a classe. No entanto, não é possível apontar qual a classe apresentou o pior resultado, uma vez que houve grandes variações em termos dos valores de exatidão do produtor e do usuário, e seus respectivos erros de omissão e comissão. De modo geral, de todas as classificações avaliadas e comparadas, essa foi a que apresentou menores erros, não tendo nenhuma classe apresentado erro superior a 45%, nem de omissão, nem de comissão.

Dentre as classes de vegetação, os pixels de referência que foram classificados corretamente, ou seja, que apresentaram maior exatidão do produtor, foram os correspondentes às classes Savana, Floresta Densa e Floresta Aberta, respectivamente, com 82,81%, 71,07% e

67,85%. Já as classes que apresentaram maior exatidão do usuário foram a Floresta Aberta, com 81,04%, Floresta Densa, com 71,66% e Savana, com 59,55%. Comparando com a classificação sem correção atmosférica, percebe-se uma diminuição da representatividade da Floresta Aberta para a carta nesta classificação.

A classe Áreas Não Vegetadas apresentou exatidão do produtor de 70%, com erro de omissão de 30% associado. Já a exatidão do usuário foi mais baixa, sendo de 58,33%. Isso se deve ao fato de, dos pixels restantes, boa parte ter sido considerado Savana, classe responsável em maior escala pelo erro de comissão de 41,66%, o maior observado dentre as classes. A classe Agropecuária apresentou exatidões do produtor e do usuário de 64,70% e 71,22%, respectivamente.

Tabela 3: Matriz de confusão da classificação oriunda a partir dos dados de reflectância de superfície gerados pelo QUAC.

Referência (verdade de campo) Classificação	Agropecuária	Água	Áreas Não Vegetadas	Floresta Aberta	Floresta Densa	Savana	Total
Agropecuária	99	-	2	26	-	12	139
Água	-	3	-	-	-	-	3
Áreas Não Vegetadas	6	-	21	1	1	7	36
Floresta Aberta	3	-	1	171	34	2	211
Floresta Densa	8	-	-	25	86	1	120
Savana	37	-	6	29	-	106	178
Total	153	3	30	252	121	128	687

### 3.3 Classificação da Cobertura da terra com Correção a partir do FLAASH

A classificação da cobertura da terra a partir do dado em reflectância de superfície gerado pelo FLAASH (Figura 7) diminuiu a presença de Floresta Densa e aumentou a presença de Agropecuária em comparação com o dado corrigido pelo QUAC, muito embora o predomínio de áreas vegetadas se mantenha na carta, havendo apenas certa confusão espectral

entre as referidas classes ao se comparar os três métodos. As Áreas Não Vegetadas diminuíram em comparação com o dado corrigido pelo QUAC, mostrando padrão visual mais similar à classificação a partir de dados TOA para essa classe. A Água, assim como os outros resultados, foi bem classificada. A Tabela 4 apresenta a matriz de confusão para a classificação da carta SD-23-X-D do ano de 2015 a partir de dados corrigidos pelo FLAASH.

### Classificação da Cobertura do Solo com correção a partir do FLAASH

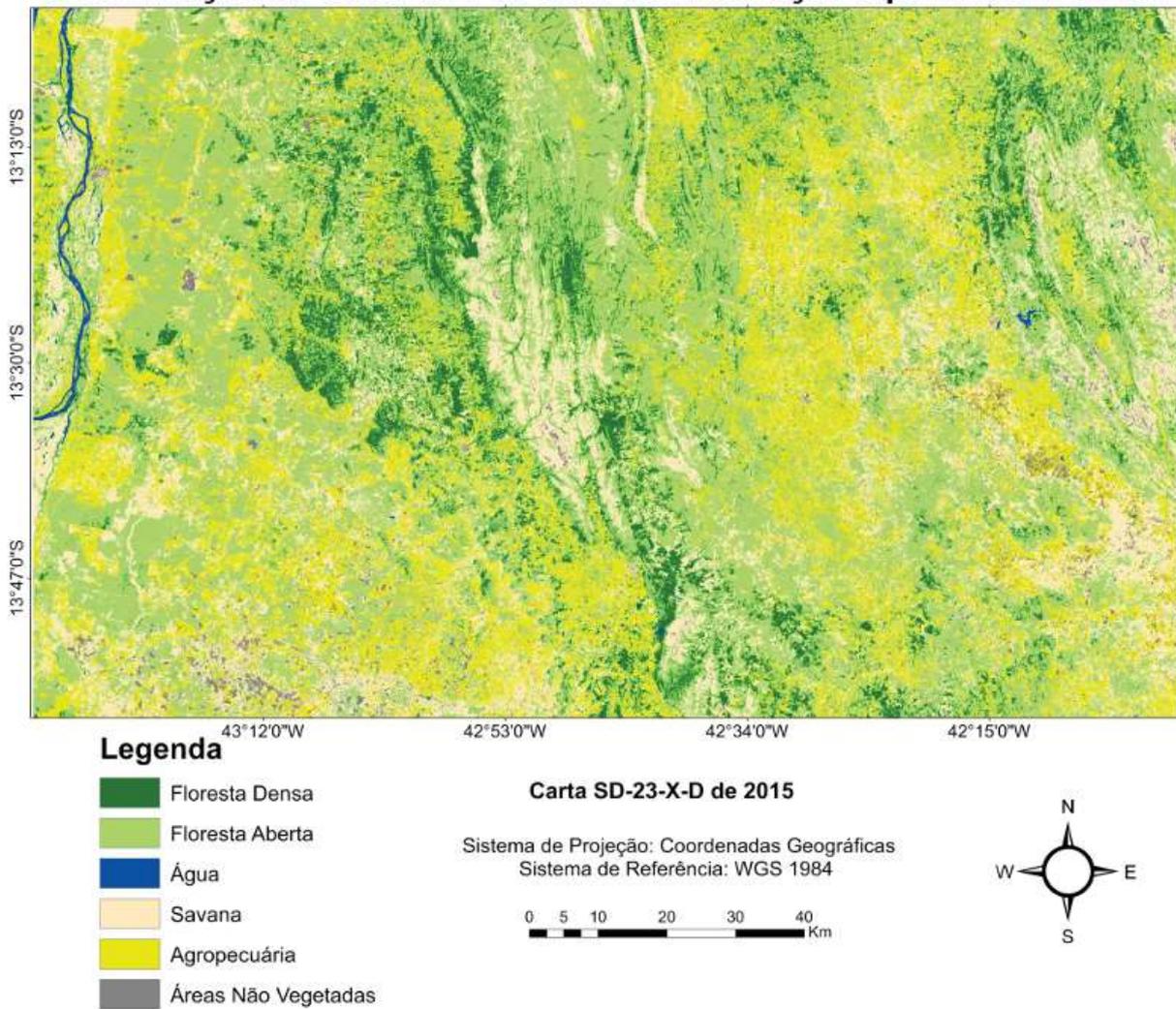


Fig. 7 – Mapa da classificação da cobertura da terra do mosaico temporal da carta SD-23-X-D de 2015 processado pelo *Google Earth Engine* a partir do dado corrigido pelo *FLAASH*.

Tabela 4: Matriz de confusão da classificação oriunda a partir dos dados de reflectância de superfície gerados pelo *FLAASH*

Referência (verdade de campo)							
Classificação	Agropecuária	Água	Áreas Não Vegetadas	Floresta Aberta	Floresta Densa	Savana	Total
Agropecuária	124	-	5	10	-	19	158
Água	-	3	-	-	-	-	3
Áreas Não Vegetadas	1	-	15	-	-	-	16
Floresta Aberta	13	-	1	212	58	15	299
Floresta Densa	7	-	-	7	62	1	77
Savana	8	-	9	23	1	93	134
Total	153	3	30	252	121	128	687

A matriz de confusão permite constatar que um total de 509 pixels concordaram entre si em relação à totalidade dos pontos considerados como referência. Isso representa uma exatidão global de 74,09%, *Kappa* de 64,60%, considerado muito bom, e *Tau* de 68,90%. Considerando a escala de qualidade de Landis & Koch (1977), para cada classe individualmente aplicada aos valores do *Kappa* condicional, o dado corrigido pelo *FLAASH* pode ser considerado excelente para as classes Água e Áreas Não Vegetadas, muito boa para a Agropecuária, Floresta Densa e Savana, e boa para a Floresta Aberta. O conjunto de dados informados nesse parágrafo já são suficientes para apontar o dado corrigido pelo *FLAASH* como sendo o que apresentou melhor desempenho para a carta.

A Água apresentou alto nível de concordância, tanto sob a perspectiva do produtor como do usuário, sendo 100% de exatidão para ambos. Assim como no dado corrigido pelo *QUAC* para a mesma carta, não houve uma classe que tenha se destacado por apresentar os piores resultados. Em termos de exatidão do produtor, com exceção das classes Floresta Densa e Áreas Não Vegetadas, todas as classes apresentaram valores de exatidão superiores a 70%. O mesmo ocorreu para a exatidão do usuário, não sendo observados, porém, erros de comissão maiores que 30% para nenhuma das seis classes.

Dentre as classes de vegetação, a que apresentou melhores resultados foi a Floresta Aberta, com exatidão do produtor de 84,12% e do usuário de 70,90%, seguida da classe Savana, com 72,65% e 69,40%, e Floresta Densa, com 51,23% e 80,51% para as exatidões do produtor e usuário, respectivamente. A classe Floresta Densa, inclusive, confundiu muito com a classe Floresta Aberta.

A classe correspondente às Áreas Não Vegetadas, por sua vez, apresentou resultados muito favoráveis sob a perspectiva da exatidão do usuário, com 93,75%. Já a exatidão do produtor apresentou 50% de concordância, o que mostra que vários pixels foram atribuídos a classes que não pertencem na realidade, como por exemplo, da classe Savana. Entretanto, assim como já foi dito, essa classe não apresentou muitos pontos amostrais, o que mostra sua baixa predominância na carta. Os erros de omissão e comissão para a classe Agropecuária foram,

respectivamente, de 18,95% e 21,51%, ou seja, de modo geral, é possível afirmar que a classe foi bem representada no sentido de que maior parte dos seus pixels foram classificados adequadamente.

### 3.4 Comparação entre os resultados obtidos a partir dos coeficientes de concordância globais e por classe

A Tabela 5 apresenta uma síntese dos resultados estatísticos a partir do uso dos índices de concordância globais para as classificações baseadas nos três diferentes dados de reflectância utilizados para proceder com a avaliação. A carta SD-23-X-D obteve melhor desempenho a partir da classificação a partir dos dados corrigidos pelo *FLAASH*. Os três índices utilizados, apesar de variarem os valores entre si, não apresentaram divergências quanto ao resultado apontado.

As diferenças encontradas entre a exatidão global e o índice *Kappa* são normais (MATHER, 2004), visto que o segundo resume a informação proveniente da matriz de erro, enquanto o primeiro é a média da acurácia individual das classes. Brites (1996) afirma que, na formulação do índice *Tau*, o valor da concordância casual é estabelecido *a priori*, evitando falhas que incorram da exatidão global e do índice *Kappa*. Isso faz com que o *Tau*, de acordo com o autor, talvez seja o índice que represente com maior fidedignidade a exatidão dos processos de classificação.

Tabela 5: Coeficientes de concordância global para classificações geradas a partir dos dados sem correção atmosférica (TOA) e corrigidos pelo *FLAASH* e *QUAC*

COEFICIENTE	MÉTODO		
	FLAASH	QUAC	TOA
Exatidão Global	74,09%	70,74%	62,88%
<i>Kappa</i>	64,60%	61,55%	48,28%
<i>Tau</i>	68,90%	64,89%	55,45%

Se for feita uma análise isolada a partir do coeficiente *Kappa*, é possível verificar que, se associarmos os valores obtidos com a avaliação do índice proposta por Landis e Koch (1977) para a carta SD-23-X-D, as duas classificações oriundas de dados com correção atmosférica (*FLAASH* e *QUAC*) podem ser consideradas muito boas, pois seus resultados para o índice se encontram no intervalo entre 60% e 80%.

Analisando as classificações pelos dados de reflectância utilizados, verifica-se que os dados corrigidos pelo QUAC apresentaram um bom desempenho com variação entre 3% e 4% comparados com os dados corrigidos pelo FLAASH para os três coeficientes de concordância. Isso representa uma grande vantagem para o método, considerando sua maior praticidade (de onde infere-se maior rapidez) e menor complexidade para corrigir os efeitos atmosféricos.

De um modo geral, a análise sistêmica das diferentes classificações por meio do uso dos coeficientes de concordância global contribuiu, sem qualquer dúvida, para que seja possível chegar a diferentes conclusões sobre o impacto da correção atmosférica na classificação da cobertura da terra na Caatinga. No entanto, para que se tenha um panorama ainda melhor acerca do desempenho dessas classificações, uma análise mais detalhada por meio dos coeficientes de concordância por classe faz-se necessária, pois assim é possível identificar, por exemplo, as classes que apresentam maior confusão entre si e aquelas que podem impactar em diferentes intensidades as diferentes classificações. Assim, as Tabelas 6 e 7 apresentam os resultados das exatidões do produtor e do usuário para cada classe avaliada. Enquanto a exatidão do produtor indica a probabilidade de acerto da classificação (erro de omissão), a exatidão do usuário indica a probabilidade de um pixel classificado na imagem de fato corresponder com o dado de referência (verdade de campo), sendo, assim, associada ao erro de comissão.

Tabela 6: Exatidão do Produtor de cada classe das classificações geradas a partir dos dados sem correção atmosférica (TOA) e corrigidos pelo FLAASH e QUAC

CLASSE	MÉTODOS		
	FLAASH	QUAC	TOA
Agropecuária	0,81	0,64	0,29
Água	1,00	1,00	1,00
Áreas Não Vegetadas	0,50	0,70	0,23
Savana	0,72	0,82	0,60
Floresta Aberta	0,84	0,67	0,92
Floresta Densa	0,51	0,71	0,56

Para todos os dados de reflectância, corrigidos ou não, a Água foi a classe que apresentou melhor desempenho, com 100% de exatidão do produtor. As demais classes variaram bastante, o que pode ser um forte indicio de que fatores, tais como condições climáticas e disposição da paisagem, possam ter interferido de maneira considerável nos resultados observados. As Áreas Não Vegetadas apresentaram o pior resultado considerando todos os três dados comparados. Já as classes de vegetação apresentaram valores de exatidão do produtor maiores, o que está de acordo com a realidade da carta, cuja predominância dos alvos é de áreas vegetais.

Tabela 7: Exatidão do Usuário de cada classe das classificações geradas a partir dos dados sem correção atmosférica (TOA) e corrigidos pelo FLAASH e QUAC

CLASSE	MÉTODOS		
	FLAASH	QUAC	TOA
Agropecuária	0,78	0,71	0,75
Água	1,00	1,00	0,75
Áreas Não Vegetadas	0,93	0,58	0,50
Savana	0,69	0,59	0,51
Floresta Aberta	0,70	0,81	0,62
Floresta Densa	0,80	0,71	0,80

Os resultados de exatidão do usuário também foram bastante elevados para a Água, com uma única exceção no dado de reflectância no topo da atmosfera da carta SD-23-X-D, que ficou em 75%. A classe Agropecuária, por sua vez, foi igualmente classificada, não havendo grandes diferenças.

No contexto da carta SD-23-X-D, a classificação gerada, a partir de dados corrigidos pelo FLAASH, apresentou maiores probabilidades de concordância entre os pixels para todas as classes, com exceção da Floresta Aberta, que apresentou exatidão do usuário de 70% contra 81% do dado corrigido pelo QUAC. Entretanto, as diferenças entre ambos os dados que foram submetidos à correção atmosférica não são tão grandes em comparação aos dados de reflectância aparente, ou seja, sem correção atmosférica. A Tabela 8 apresenta um comparativo entre as classes sob a ótica do índice Kappa condicional.

Tabela 8: Kappa Condicional de cada classe das classificações geradas a partir dos dados sem correção atmosférica (TOA) e corrigidos pelo FLAASH e QUAC

CLASSE	MÉTODO		
	FLAASH	QUAC	TOA
Agropecuária	0,72	0,63	0,68
Água	1,00	1,00	0,75
Áreas Não Vegetadas	0,93	0,56	0,48
Savana	0,62	0,50	0,40
Floresta Aberta	0,54	0,70	0,40
Floresta Densa	0,76	0,66	0,76

Fazendo uma avaliação mais apurada do *Kappa* condicional para a carta SD-23-X-D, os melhores resultados observados para cada classe em função das classificações geradas foram: excelente para a Água, nos dois métodos com correção atmosférica, e para as Áreas Não Vegetadas, no método corrigido pelo *FLAASH*, muito bom para as classes Agropecuária e Floresta Densa, nos três dados de reflectância. Para a Savana o *Kappa* condicional foi considerado muito bom apenas no dado corrigido pelo *FLAASH*, enquanto que para a Floresta Aberta, apenas no dado corrigido pelo *QUAC*.

Observando os resultados dos três coeficientes de concordância para cada classe individualmente, é possível concluir o impacto dos dados com e sem correção atmosférica, em cada classe da carta avaliada. Para a carta SD-23-X-D, o dado corrigido pelo *FLAASH* apresentou melhores resultados para a classe Agropecuária, apresentando resultados superiores a 70% para todos os três coeficientes de concordância. A classe Água apresentou 100% de exatidão para ambos os dados com correção atmosférica para os três coeficientes. A classe que equivale às Áreas Não Vegetadas apresentou maior exatidão para os dados submetidos à correção atmosférica, superando 90% para o *FLAASH* sob a ótica da exatidão do usuário e do *Kappa* condicional, porém, com maior exatidão do produtor registrada pelo *QUAC*, com 70%. Comportamento similar foi observado para a classe Savana.

A Floresta Aberta foi a única com maior exatidão observada para um dos coeficientes para o dado sem correção atmosférica, com 92% de exatidão do produtor. Entretanto, para os demais

coeficientes, apresentou melhores resultados para o dado corrigido pelo *QUAC*. A classe Floresta Densa teve bastante equivalência entre os índices para todos os dados, tendo apresentado similaridades entre dados com e sem correção atmosférica. Porém, comparando a exatidão da classe sob a ótica dos três coeficientes, o dado com correção atmosférica apresentou os maiores resultados, tendo o *QUAC* sido melhor sob o ponto de vista da exatidão do produtor e o *FLAASH* para os outros dois coeficientes.

Galletti e Myint (2014), ao fazerem o mapeamento da cobertura da terra, em um contexto similar ao da Caatinga, uma região árida no Arizona, alegaram que a confusão espectral é um aspecto inevitável, especialmente em classes como áreas agrícolas exploradas, que são facilmente confundidas com áreas com aspecto desértico, solo exposto ou ainda áreas rurais pouco povoadas. Além disso, Sano *et al.* (2007) também alegaram o fato da confusão espectral entre certas classes vegetais serem grandes desafios no mapeamento temático para o bioma Cerrado, o qual também apresenta certa similaridade com a Caatinga. Os autores também destacaram a sazonalidade marcante da vegetação natural como sendo outro fator que exerce grande influência no processo de mapeamento do bioma.

Ainda sob uma perspectiva climática, outro aspecto que merece ser abordado é quanto à influência dos fenômenos El Niño e La Niña na classificação da cobertura da terra em ambientes semiáridos. A influência desses fenômenos no comportamento das classificações deve ser melhor avaliada. A área da classe Agropecuária aumentou nas classificações oriundas do dado de reflectância superfície em comparação ao dado TOA, o que pode ser um forte indício de esses eventos climáticos serem mais evidentes em dados com correção atmosférica.

Uma comparação maior e mais detalhada dos dados permite verificar que o dado de reflectância no topo da atmosfera, apesar de ter apresentado valores estatísticos mais baixos em relação aos dados em reflectância de superfície, não é um dado que mereça ser descartado. Além disso, considerando a janela temporal definida para o estudo, é possível constatar que o mosaico temporal carrega inevitavelmente consigo características do período de seca. Por

isso cabe enfatizar a necessidade de se fazer ajustes nos *endmembers* e/ou nos nós da árvore de decisão, a fim de se obter um dado com melhores resultados.

Carvalho & Baptista (2015), ao realizarem uma comparação entre dados *TOA* e *FLAASH* em cenas RapidEye no cálculo do Índice de Vegetação Melhorado (*Enhanced Vegetation Index – EVI*), observaram, por meio de análises estatísticas, que não há diferenças significativas entre proceder ou não com a correção atmosférica. Sob a condição da pesquisa realizada, os autores constataram a desnecessidade de se proceder com a correção no cálculo do *EVI*. O fato de o estudo desenvolvido pelos autores ter se dado em uma área do bioma Cerrado, o qual possui maior semelhança com a Caatinga dentre os demais biomas brasileiros, serve como reforço quanto à necessidade de se avaliar melhor o uso do dado *TOA*, especialmente em classificações da cobertura da terra.

Um outro fator que pode reduzir os valores estatísticos observados nos resultados das classificações dessa natureza é a resolução espacial das imagens Landsat. Considerando-se que um pixel de 30 metros compreende uma quantidade relativa de informação, muitas vezes um pixel que está localizado em uma zona de transição, por exemplo, entre uma faixa coberta por vegetação e uma área não vegetada, acabará sendo vinculado à classe prevalecente nas dimensões desse pixel. Considerando que os pontos amostrais não compreendem toda a dimensão do pixel, o processo de avaliação estatística também acaba, inevitavelmente, vindo com erros associados, uma vez que representa a realidade sob uma perspectiva mais abrangente e não muito específica. Devido à amplitude de espaçamentos que áreas consideradas vegetadas podem ter em termos de distribuição de espécies arbóreas, um determinado pixel da resolução do Landsat-8 (30 metros) tende a contribuir para o erro no resultado final da classificação.

## CONCLUSÃO

Os dados com correção atmosférica apresentaram maior acurácia para a classificação da cobertura da terra da carta SD-23-X-D, tendo o *FLAASH* apresentado melhor desempenho em comparação com o *QUAC*, com diferença variando entre 3% e 4% de acordo com os índices *Kappa*, *Tau* e Exatidão Global.

Dada a baixa diferença estatística observada entre o *FLAASH* e o *QUAC*, este último método revela-se uma opção mais viável, tendo em vista a maior rapidez no tempo de processamento e menor complexidade em comparação com o *FLAASH*. O fato da Caatinga ser coberta por 68 cartas reforça ainda mais a escolha do *QUAC* para proceder com a correção atmosférica no contexto do MapBiomias.

O dado de reflectância no topo da atmosfera, por sua vez, superestimou algumas classes em detrimento de outras, não representando bem as classes Savana, Agropecuária e Áreas Não Vegetadas. Esse dado apresentou *Tau* com diferença de 13% a menos em comparação com o dado corrigido pelo *FLAASH* e 9% a menos em relação ao dado corrigido pelo *QUAC*. No entanto, esse dado pode vir a apresentar uma consistência em comparação com a verdade terrestre caso a árvore de decisão seja parametrizada com valores que sejam melhor adequados ao dado.

Além disso, salienta-se que o trabalho de executar o pré-processamento fora do ambiente em nuvem para depois inserir os dados atmosféricamente corrigidos na plataforma talvez não se justifique, pois, a diferença obtida dos dados corrigidos com o dado *TOA* não são suficientes para que o dado seja descartado, principalmente, levando-se em conta que o Projeto MapBiomias abrange a extensão do território nacional e pretende se expandir para outros países.

Como sugestão para trabalhos futuros, no âmbito do MapBiomias, destaca-se o aprimoramento da árvore de decisão contemplando especificidades do bioma Caatinga, tais como realização de novos testes em relação à janela temporal definida para geração do mosaico, aprofundamento da análise da variação temporal interanual cruzando com dados de variação climática (El Niño/La Niña), além da realização de testes envolvendo outros índices de vegetação, como o Índice de Vegetação ajustado ao Solo (*SAVI*), que considera a interferência do solo na detecção da vegetação.

## AGRADECIMENTOS

Os autores agradecem à Coordenação Geral e Técnica do MapBiomias, Tasso Azevedo e Carlos Souza, a João Victor Siqueira (Terras

App.) e César Diniz (Solved). Finalmente, ao Laboratório de Sensoriamento Remoto e Análise Espacial, do Instituto de Geociências (IG) da Universidade de Brasília, por ter disponibilizado os computadores com a licença dos softwares ENVI 5.0 e ArcGIS 10.2, bem como ao Laboratório de Processamento de Imagens da Universidade Estadual de Feira de Santana pelo suporte na utilização do GEE.

## REFERÊNCIAS BIBLIOGRÁFICA

ADAMS, J. B.; SMITH, M. O.; JOHNSON, P. E. Imaging spectroscopy: Interpretation based on spectral mixture analysis. In C. M. Pieters, & P. A. J. Englert (Eds.), **Remote geochemical analysis: Elemental and mineralogical composition**. Cambridge, England: Cambridge University Press, 1993, p. 145-164.

ADLER-GOLDEN, S.; BERK, A.; BERNSTEIN, L. S.; RICHTSMIEIER, S.; ACHARYA, P. K.; MATTHEW, M. W.; CHETWYND, J. H. FLAASH, a MODTRAN4 atmospheric correction package for hyperspectral data retrievals and simulations. In: **Proceedings 7th Ann. JPL Airborne Earth Science Workshop**. Pasadena, CA: JPL, 1998. p. 9-14.

ANTUNES, A. F. B. **Classificação de ambiente ciliar baseada em orientação a objeto em imagens de alta resolução espacial**. Tese (Doutorado em Ciências Geodésicas) - Universidade Federal do Paraná, Curitiba. 2003, 148p.

BERNSTEIN, S. M.; ADLER-GOLDEN, R. L.; SUNDBERG, R. Y.; LEVINE, T. C.; PERKINS, A.; BERK, A.; RATKOWSKI, A.; HOKE, M. L. A new method for atmospheric correction and aerosol optical property retrieval for VIS-SWIR multi- and hyperspectral imaging sensors: QUAC (Quick atmospheric correction),” **AVIRIS Workshop 2004**, vol. 5, pp. 3549–3552, 2004.

BRITES, R.S.; SOARES, P.S.; RIBEIRO, C. A. A. S. Comparação de desempenho entre três índices de exatidão aplicados a classificação de imagens orbitais. In: **Anais Simpósio Brasileiro de Sensoriamento Remoto**, 8, Salvador: INPE, 1996. 14-19p.

BRASIL. Ministério do Meio Ambiente (MMA). **Mapas de cobertura vegetal dos biomas brasileiros [2006]**. Brasília, 2006. Disponível em:

<<http://mapas.mma.gov.br/mapas/aplic/probio/datadownload.htm>>. Acesso em: 25 de outubro de 2017.

CARVALHO, A. A.; BAPTISTA, G. M. M. Influência da correção atmosférica na conversão dos dados de radiância para reflectância para o cálculo do EVI em imagem Rapideye. **Revista Brasileira de Cartografia**, 67 (7), p. 1371-1390, 2015.

CHANDER, G.; MARKHAM, B. L.; HELDER, D. L. Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. **Remote Sensing of Environment**, 113 (5), p. 893-903, 2009.

CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and applications**. Lewis Publishers, Boca Raton, 2008. 183p.

DESCHAMPS PY, H. M. TANRE D. Influence de l’atmosphère en télédétection des ressources terrestres. Modélisation et possibilités s de correction. In: **Colloque International Signatures Spectrales d’objets en Télédétection**. Avignon. Proceedings. Versailles: Institut National de la Recherche Agronomique, 1981, 543–558.

EPIPHANIO, J. C. N.; LUIZ, A. J. B.; FORMAGGIO, A. R. Estimativa de áreas agrícolas municipais, utilizando sistema de amostragem simples sobre imagens de satélite. **Bragantia**, 61(2), p. 187-197, 2002.

EXELIS VISUAL INFORMATION SOLUTIONS. Boulder, Colorado: Exelis Visual Information Solutions, 2017.

GALLETTI, C. S.; MYINT, S. W. Land-use mapping in a mixed urban-agricultural arid landscape using object-based image analysis: A case study from Maricopa, Arizona. **Remote Sensing**, 6(7), p. 6089-6110, 2014.

GORELICK, N; HANCHER, M.; DIXON, M.; ILYUSHCHENKO, S; THAU, D.; MOORE, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, 202, p. 18–27, 2017. <http://dx.doi.org/10.1016/j.rse.2017.06.031>.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Mapa Técnico e Uso da Terra**. Rio de Janeiro, 2006. Disponível em:

- <<http://biblioteca.ibge.gov.br/visualizacao/livros/liv81615.pdf>>. Acesso em: 20 nov. 2016.
- INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Manual Técnico da Vegetação Brasileira**. 2ª ed. Rio de Janeiro: IBGE, 2012. 275p.
- IRONS, J. R.; DWYER, J. L.; BARSÍ, J. A. The next Landsat satellite: The Landsat data continuity mission. **Remote Sensing of Environment**, v. 122, p. 11-21, 2012.
- JENSEN, J. R., **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. Tradução José Carlos Neves Epiphânio (coordenador) São José dos Campos – SP: Parêntese, 2009. 598p.
- KOSCHMIEDER, H. Theorie der horizontalen Sichtweite, Beiert. Phys. frei. **Atoms.**, 12, p. 171–181, 1924.
- LANDIS, J. R; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**, 33(1), p. 159-174, 1977.
- LEAL, I. R.; SILVA, J. M. C. **Ecologia e conservação da Caatinga**. Recife: Editora Universitária UFPE, 2003.
- LV, Z.; YU, X.; ZHANG, Z.; WANG, G. Automatic Remote Sensing Image Classification Method Based on Spectral Angle and Spectral Distance. **Geoscience and Remote Sensing Symposium (IGARSS) IEEE International**, p. 3140-3143, 2013.
- MALDONADO, F.D. **Desenvolvimento e avaliação de uma metodologia de detecção de mudanças na cobertura vegetal do semiárido**. 2005. Tese (Doutorado em Sensoriamento Remoto) – Programa de Pós-Graduação em Sensoriamento Remoto, INPE, São José dos Campos – SP. 177p.
- MATHER, P.M. **Computer processing of remotely sensed images: an introduction**. 3ª Ed. Chichester: John Wiley & Sons, 2004. 324p.
- NOZAWA A. Carta do mundo ao milionésimo: subdivisões e localizações. **GEOGRAFIA (Londrina)**, 4, p. 49-53, 1987.
- PROJETO MAPBIOMAS, 2017a – Coleção 2 da Série Anual de Mapas de Cobertura e Uso do Brasil, Disponível em: <http://mapbiomas.org>, acesso em 6 de fevereiro de 2017.
- PROJETO MAPBIOMAS, 2017b – Coleção 2 da Série Anual de Mapas de Cobertura e Uso do Solo do Brasil, Disponível em: <https://s3.amazonaws.com/mapbiomas-ecostage/Base+de+dados/Metodologia?ATDB-R++Bioma+Pantanal+-+2017-03-17.pdf> acesso em 22 de março de 2017.
- RICHARDS, J. A.; JIA, X. **Remote sensing digital image analysis-hardback**. Springer, 2006, 439p.
- ROJAS, E. H. M. Síntese Genética de Redes Neurais Artificiais ART2 na classificação de imagens ASTER para mapeamento de uso e cobertura da terra na região Norte de Mato Grosso. 2005. **Dissertação** (Mestrado) – Instituto Nacional de Pesquisas Espaciais – INPE. São José Dos Campos, 2005, 176p.
- SANO, E.E.; FERREIRA, L.G.; ASNER, G.P.; STEINKE, E.T. Spatial and temporal probabilities of obtaining cloud-free Landsat images over the Brazilian tropical savanna. **International Journal of Remote Sensing**, 28, p. 2739-2752, 2007.
- SMITH, M. O., JOHNSON, P. E., & ADAMS, J. J. Quantitative determination of mineral types and abundances from reflectance spectra using principal components analysis. **Journal of Geophysical Research**, 90, p. 797–804, 1985.
- SOUZAJR., C. M., ROBERTS, D.; COCHRANE, M. A. Combining Spectral and Spatial Information to Map Canopy Damages from Selective Logging and Forest Fires. **Remote Sensing of Environment**. 98. 329-343, 2005.
- UNITED STATES GEOLOGICAL SURVEY (USGS). **Landsat data continuity mission**. U.S. Department of the Interior, U.S. Geological Survey. 2012–3066. July 2012.
- VALENTE, R. de O. A. **Análise da Estrutura da Paisagem na Bacia do Rio Corumbataí**. Dissertação (mestrado) – Escola Superior de Agricultura Luiz de Queiroz da Universidade de São Paulo, 2001. 144p.
- ZHU, Z. WOODCOCK, C.E. Object-based cloud and cloud shadow detection in Landsat imagery. **Remote Sensing of Environment**, 118, p.83-94, 2012.