



# Revista Brasileira de Geografia Física

Homepage: <https://periodicos.ufpe.br/revistas/rbgfe>



## Identificação das Dunas do Atacama (Norte do Chile) a partir da avaliação de três algoritmos no Google Earth Engine

Leonel Enrique Sánchez Currihuinca<sup>1</sup>, Joselisa Maria Chaves<sup>2</sup>, Washington J. S. Franca Rocha<sup>2</sup>, Jocimara Souza Britto Lobão<sup>2</sup>, Plínio Martins Falcão<sup>3</sup>

<sup>1</sup> Geógrafo pela Universidade do Chile. Mestre do Programa de Pós-graduação em Modelagem em Ciências da Terra e do Ambiente (PPGM - UEFS), Avenida Transnordestina, s/n, Bairro Novo Horizonte, CEP: 44036-900, Feira de Santana-BA. (75) 3161-8807. e-mail: [leoneluefs@gmail.com](mailto:leoneluefs@gmail.com) (Autor Correspondente). <sup>2</sup> Professores, Doutores. Programa de Pós-graduação em Modelagem em Ciências da Terra e do Ambiente (PPGM - UEFS), Avenida Transnordestina, s/n, Bairro Novo Horizonte, CEP: 44036-900, Feira de Santana-BA. (75) 3161-8807, e-mail: [joselisa@uefs.br](mailto:joselisa@uefs.br); [wrocha@uefs.br](mailto:wrocha@uefs.br); [juci.lobao@gmail.com](mailto:juci.lobao@gmail.com). <sup>3</sup> Instituto Federal de Educação, Ciência e Tecnologia da Bahia – IFBA Departamento de Geografia / Campus Salvador, R. Emídio dos Santos, s/n - Barbalho, CEP: 40301-015, Salvador – BA. (71) 2102-9400. e-mail: [pliniomf@gmail.com](mailto:pliniomf@gmail.com).

Artigo recebido em 08/01/2021 e aceito em 19/10/2021

### RESUMO

As dunas correspondem a processos de sedimentação eólica, que podem estar tanto nas áreas costeiras marinhas, como no interior do continente com algumas diferenças na modelagem. No Sul do deserto do Atacama, no Norte do Chile, há um conjunto de seis campos de dunas intermontanhas chamadas Mar de Dunas do Atacama, as quais têm tipologias complexas de dunas do deserto, que podem ser ativas, semiativas ou estabilizadas. O seu monitoramento é conveniente para conhecer detalhes sobre a possível invasão de areias das dunas ao sul do rio Copiapó. Dessa forma, esta pesquisa tem como objetivo avaliar os métodos de classificação supervisionada Random Forest, CART e SmileCART através de duas metodologias de amostragens, aleatória e estratificada, numa imagem Landsat 5 na plataforma em nuvem Google Earth Engine, a fim de verificar qual método oferece o melhor resultado para o mapeamento do Mar de Dunas do Atacama. Para conseguir este objetivo, foram criados polígonos de classes para a realização da amostragem aleatória estratificada e chave de interpretação para amostragem aleatória simples. O processo de avaliação da acurácia foi feito através de imagem Sentinel 2 com a aplicação dos índices de Simultaneidade Geográfica, Erros de Comissão e Omissão, e Exatidão Global. Observou-se como resultados para os algoritmos testados, que os três algoritmos foram eficientes para o mapeamento das Dunas do Atacama, entretanto, a técnica de classificação supervisionada por CART, com a metodologia da amostragem aleatória simples, representou o melhor desempenho.

Palavras-chave: Sensoriamento Remoto, Processamento Digital de Imagens, Índices de Acurácia.

## Identification of the Atacama Dunes (Northern Chile) from the evaluation of three algorithms on Google Earth Engine

### ABSTRACT

The dunes correspond to wind sedimentation processes, which can be found both in marine coastal areas and in the interior of the continent with some differences in modeling. In the south of the Atacama desert, in northern Chile, there are a set of six inter-mountain dune fields called Mar de Dunas do Atacama, which have complex types of desert dunes, which can be active, semi-active or stabilized. Its monitoring is convenient to know details about the possible invasion of sand from the dunes south of the Copiapó River. Thus, this research aims to evaluate the supervised classification methods Random Forest, CART and SmileCART through two sampling methodologies, random and stratified, in a Landsat 5 image on the Google Earth Engine cloud platform, in order to verify which method offers the best result for mapping the Atacama Dunes Sea. In order to achieve this objective, class polygons were created to perform stratified random sampling and the interpretation key for simple random sampling. The accuracy assessment process was performed using a Sentinel 2 image with the application of the Geographic Simultaneity indices and the Commission and Omission Errors. It was observed as results for the tested algorithms, that the three algorithms were efficient for mapping the Atacama Dunes, however, the CART supervised classification technique, with the simple random sampling methodology, represents the best performance.

Keywords: Remote Sensing, Digital Image Processing, Accuracy Indexes.

## Introdução

As dunas são formações geomorfológicas constituídas de sedimentos não consolidados selecionados pela ação do vento. Para a formação desses ambientes é necessário uma série de pré-requisitos ambientais, como a existência de grande fonte de sedimentos (desembocaduras fluviais, largos lençóis de areia, praias, desertos, etc.), ventos com intensidade e direção constantes e um clima que alterne mais períodos áridos do que húmidos (Bird, 2008). As areias de uma duna, podem nascer, em quanto ao material litológico, no continente mesmo ou na costa marítima. No primeiro caso, são os rios os agentes de transporte mais ativo para levar os materiais que resultam da erosão da carga aluvial ao chegar ao mar. O mar pega as areias e as deposita no litoral próximo, de acordo à tendência de suas correntes costeiras. É assim que as dunas costeiras são fenômenos que se produzem, uma vez realizados outros processos, tais como erosão, sedimentação e transportes continentais, por conseguinte, as dunas definem um processo e não uma forma (Börgel, 1963). Dependendo da variação dos ventos, a paisagem pode se alterar ao longo do tempo. Por isso, as dunas são ecossistemas que estão em constante transformação, podendo alterar sua forma, se deslocarem, diminuir ou aumentarem.

No sul do deserto de Atacama, há um conjunto de seis campos de dunas intermontanas com tipologias complexas de dunas do deserto ativas, semiativas e estabilizadas. Elas são uma herança de condições morfoclimáticas passadas (Castro et al., 2012). A singularidade das formas destas dunas do deserto e seu grande volume se constituem num conjunto notável de paisagem cênica, localizado nas proximidades da cidade de Copiapó, no Norte do Chile.

O monitoramento da localização atual das Dunas do Atacama é conveniente para conhecer se existe invasão de areias das dunas ao sul do rio Copiapó, e assim definir se existe a presença de desertização, o que é típico nos limites do deserto (Mendonça, 1990). Nesse sentido, é interessante estudar e assim saber a localização do Mar de Dunas do Deserto de Atacama, que poderia estar avançando em direção ao sul do Chile, de modo que podem afetar à agricultura e áreas urbanas.

Segundo Lam et al., (2012), a maioria dos estudos de sensoriamento remoto em desertos se concentra apenas no monitoramento da vegetação. No entanto, a aplicação da invasão de dunas de areia em tais estudos melhoraria muito a precisão

dos critérios de predição para áreas propensas a desertização.

Os ambientes Dunares são amplamente pesquisados pela Geologia, Geomorfologia e Biologia, pois se tratam de ambientes de formação frágil e que podem ser facilmente comprometidos (Simões, 2014). Castro et al., (2012) constataram que as Dunas do Atacama são consideradas um geopatrimônio quem tem uma alta vulnerabilidade às atividades antrópicas.

O sensoriamento remoto (SR) é definido como a tecnologia que permite aquisição de informações sobre objetos sem manter contato físico com eles. De acordo com Novo (1989), sensoriamento remoto é a utilização conjunta de sensores, equipamentos para processamento de dados, entre outros, com o objetivo de estudar o ambiente terrestre através do registro e análise das interações entre a radiação eletromagnética e as diversas coberturas que compõem a superfície terrestre.

O uso de técnicas de sensoriamento remoto aplicadas, tem sido muito relevante para o estudo espacial e geomorfológico das dunas, pois permite uma melhor compreensão da estrutura e disposição dos vastos mares de areia (Castro et al., 2012). Desde os anos 70, autores (Fryberger e Dean, 1979; Breed e Grow, 1979) destacaram sua utilidade em classificar dunas em diferentes categorias, determinar sua distribuição e propriedades morfométricas, além de correlações entre comprimento, largura, diâmetro e comprimento de onda. Melhorias recentes na tecnologia de imagem de satélite e processamento de dados em nuvem aumentaram substancialmente nossa capacidade de mapear, quantificar e qualificar mudanças na cobertura de árvores em escalas globais, incorporando uma enorme quantidade de informações de imagens (Rosa *et al*, 2021).

Segundo Meneses e Almeida (2012), existem diversos métodos de classificação de imagens em uso, que podem ser divididos segundo diferentes critérios: classificação paramétrica e não-paramétrica, classificação espectral e espacial, e classificação supervisionada ou não-supervisionada. A maioria dos classificadores ou algoritmos executa a classificação por pixel, que se utiliza somente da informação espectral de cada pixel para encontrar regiões homogêneas, a partir de medidas de distâncias ou de probabilidades de um pixel pertencer a uma classe específica (Meneses e Almeida, 2012).

Segundo Campbell (1996), o processo de classificação supervisionada de imagens, a qual é usada nesta pesquisa, apresenta vantagens e desvantagens. A principal vantagem é que o analista tem maior controle sobre o processo, como por exemplo, pré-definir classes e identificar possíveis imprecisões graves pela análise das áreas de treinamento. A principal desvantagem é que o analista impõe uma determinada estrutura de classificação aos dados, através da definição prévia das classes ou categorias da informação. A classificação supervisionada se baseia na identificação de diferentes classes com comportamentos espectrais diferenciados. Para isso, algoritmos de classificação são adotados para extrair as feições de interesse em um espaço multidimensional (Bernardi et al., 2007).

O algoritmo de Random Forest (RF) ou Floresta Aleatória, introduzido por Breiman (2001), é um termo geral para métodos de conjunto de dados utilizando classificadores do tipo árvore. O RF constrói uma grande quantidade de árvores de decisão de subconjunto de dados a partir de um treinamento único definido. Tal treinamento é realizado usando um meta-algoritmo para melhorar a classificação e a regressão de modelos de acordo com a estabilidade e a precisão da classificação (bagging) (Drummond et al., 2017). Logo, em síntese pode-se afirmar que o RF é um algoritmo de aprendizagem supervisionada que cria várias árvores de decisão e as combina para obter uma predição com maior acurácia e mais estabilidade.

O classificador definido como Árvore de Classificação e Regressão (CART) é um classificador de árvore de decisão que é construído a partir de um conjunto de dados de treinamento usando o conceito de entropia de informações (Bishop, 2006). Em cada nó da árvore, um atributo dos dados que dividem seu conjunto de amostras em subconjuntos enriquecidos em uma classe ou outra é selecionado. Seu critério é o ganho de informação normalizado que resulta da escolha de um atributo para dividir os dados. O atributo com o maior ganho de informação normalizado é escolhido para tomar a decisão. O algoritmo então recorre às listas menores.

Para fazer a classificação supervisionada, se usou a plataforma *Google Earth Engine* (GEE), baseada em nuvem, a qual é uma ferramenta para análises geoespaciais em escala planetária, que coloca os recursos computacionais do Google em uma variedade de questões sociais de alto impacto, incluindo desmatamento, seca, desastre, doença, segurança alimentar, gerenciamento de água,

monitoramento climático e Proteção Ambiental (Gorelick et al., 2017). Outras capacidades de GEE é a integração com as linguagens de programação Python e JavaScript por meio da interface de programação usando o comando *Application Programming Interface* (API).

Ao obter os mapeamentos a partir dos classificadores (algoritmos), uma etapa importante é a validação ou avaliação de precisão dos mapas, sendo este um dos componentes integrantes da maioria dos projetos de mapeamento que incorporam dados detectados remotamente (Congalton e Green, 2001). Na década de 1980, o uso da matriz de erros se tornou uma ferramenta comum para representar a precisão de categorias de mapas individuais. É assim que, a avaliação ou validação da precisão é um componente essencial de qualquer projeto que emprega dados espaciais. Segundo Congalton e Green (2001), existem várias razões pelas quais essa avaliação é importante, por exemplo: capacidade de comparar quantitativamente diferentes metodologias de avaliação e a capacidade de usar as informações resultantes da sua análise de dados espaciais em algum processo de tomada de decisão.

O presente estudo teve como objetivo avaliar três algoritmos na plataforma GEE, a fim de verificar qual método oferece melhores resultados para o mapeamento do Mar de Dunas do Atacama, onde seu monitoramento é conveniente para conhecer detalhes sobre a possível invasão de areias das dunas ao sul do rio Copiapó, no norte do Chile.

## Material e Métodos

### Área de estudo

O Mar de Dunas do Atacama localiza-se entre 7.011.154 a 6.977.330 (Norte) e 324.761 a 391.327 (Leste) no sistema de coordenadas Universal Transversa de Mercator (UTM), datum WGS 84 hemisfério Sul, na região do Atacama, no Norte do Chile. As alturas da área de estudo variam de 91 a 1.842 m.s.n.m. As dunas estão localizadas a noroeste da cidade de Copiapó (Figura 1).

A cidade de Copiapó tem uma população de 158.261 habitantes, segundo o Censo do ano 2012 (INE, 2012). Também possui uma economia diversificada e potencial, mas o setor de mineração é a principal atividade econômica.

A área Central e Norte do Chile é um relevo gerado no contexto da subducção da placa oceânica de Nazca sob a placa continental sul-

americana e resulta em encurtamento cortical e formação de um arco vulcão-magmático durante o

Cenozóico (Coira et al., 1982). Numerosos estudos contribuem para o conhecimento sobre a caracte -

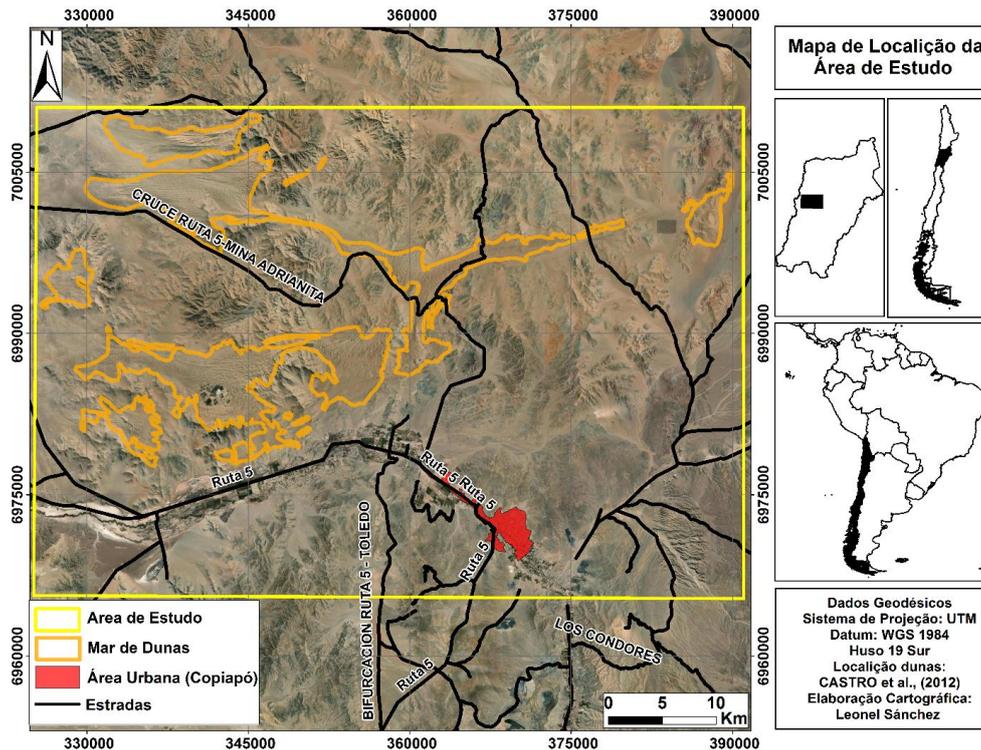


Figura 1. Mapa da área de estudo. Em polígonos de cor laranja se localiza Mar de Dunas do Atacama segundo CASTRO et al., (2012).

rização dos processos de construção do relevo andino e central do Chile (Oncken et al, 2006), bem como as condições climáticas durante o Neógeno-Quaternário (Dunai et al., 2005).

A geologia entre as 27° e 28° Latitude Sul da área de estudo, é caracterizada pela presença de afloramentos rochosos alongados na direção predominantemente NNE- SSW, formados por rochas com idade paleozoica superior (300 Ma) à recente. Os principais elementos estruturais são caracterizados pelos sistemas de falha de Atacama e La Ternera (Aguirre, 1999). Nos mapas geológicos de Caldera e Copiapó, elaborados por Godoy et al., (2003), as áreas com areias das dunas são representadas tanto em depressões como planícies e topografias altas, como no morro Medanoso. Esses autores indicam a origem do vento dos depósitos geralmente ativos e compostos por dunas lineares, depósitos mantiformes e campos de dunas parabólicas, em sucessões de cristas, bordas e famílias de formas de Holoceno e Pleistoceno.

A área de estudo representa uma climosequência na zona árida - hiperárida do Deserto do Atacama. O deserto do Atacama é uma região extrema entre as regiões áridas do planeta,

em função da atuação conjunta de diversos fatores determinantes de aridez, como proximidade a correntes oceânicas frias (Corrente de Humboldt), localização em uma sombra de chuvas da cordilheira dos Andes, e localização na subsidência atmosférica subtropical (Laity, 2008).

De acordo com a classificação climática de Köppen essa região enquadra-se em um clima BWk (Rube e Kottek, 2010). As precipitações são de origem frontal e seu regime é caracterizado por apresentar quantidades reduzidas de queda de água, concentradas no inverno (junho, julho e agosto), com média anual de 16,6 mm (Gómez, 2014). Entretanto, o clima da cidade de Copiapó é definido como deserto marginal baixo, pois as condições áridas são reduzidas pela influência do rio Copiapó, cujo vale marca a transição da extrema aridez do deserto de Atacama para a estepe quente (Juliá et al., 2008).

No trabalho de Castro et al., (2012) com base nos registros de direção do vento entre os anos 1954 até 2004 do Departamento de Meteorologia do Chile (2012), é possível destacar que o 82% dos ventos predominantes são provenientes do oeste e 7% do sudoeste, enquanto isso, a velocidade média anual do vento é de 3,7 m/s, com registros de

máximos próximos de 15 m/s. Os ventos predominantes no quadrante sudoeste têm uma distribuição muito complexa devido à influência da configuração montanhosa do relevo onde estão localizadas as dunas. Nestes locais grandes, as dunas são modeladas pelo vento, mas tendem a permanecer no lugar em vez de se mover para o interior porque a velocidade média anual é baixa para mover as areias segundo escala eólica de Beaufort (Tsoar e Arens; 2003).

As dunas do Atacama têm formas originais de larga escala que não têm uma fonte atual de alimentação de areia, dado que sua massa arenosa foi herdada de diferentes condições paleográficas das atuais (Paskoff et al., 2003). A geomorfologia da área de estudo segundo Börgel (1983) apresenta a Cordilheira da Costa, planície marina ou flúvio-marinha, planícies de sedimentação fluvial ou aluvial, pediplanos, glacis e Piedemonte. A cidade de Copiapó se localiza sobre a Planície de sedimentação fluvial – aluvial. O principal rio que existe na área de estudo é o rio Copiapó. É um rio que tem um caudal baixo, porém, com chuvas esporádicas pode levar a ter um caudal importante (Sánchez et al., 2015).

O Mar de Dunas é um conjunto de seis campos de dunas continentais do tipo erg, localizados nas montanhas e bacias da Cordilheira da Costa do Chile. As dunas cobrem parte dos

terraços marinhos mais altos, glacis, encostas e também estão dispostas cobrindo recheios aluviais e coluviais de vales e desfiladeiros locais (Castro et al., 2012).

As areias de origem fluvial, das dunas continentais, foram transportadas para o interior pelo vento, através de corredores transversais e vales interiores da cordilheira costeira, até uma distância aproximada de 60 Km da costa. Aqui, as dunas culminam no morro Medanoso, onde as dunas ativas alcançam 1.643 m.s.n.m. (Paskoff et al., 2003).

Em geral, os solos da região de Atacama são constituídos por um agregado de minerais não consolidados e partículas orgânicas produzidas pela ação combinada do vento, da água e dos processos de desintegração dos seres vivos. O conteúdo de matéria orgânica é o componente fundamental que possibilita seu uso agrícola (Ciren, 2013).

O deserto de Atacama desenvolve-se uma grande diversidade vegetativa, com a presença de diferentes formações vegetativas, como pradarias, matagais e florestas (Luebert e Plissock, 2017) e um elevado número de espécies de flora endêmica, que tem sido, em geral, muito pouco reconhecido (Plissock, et al., 2017).

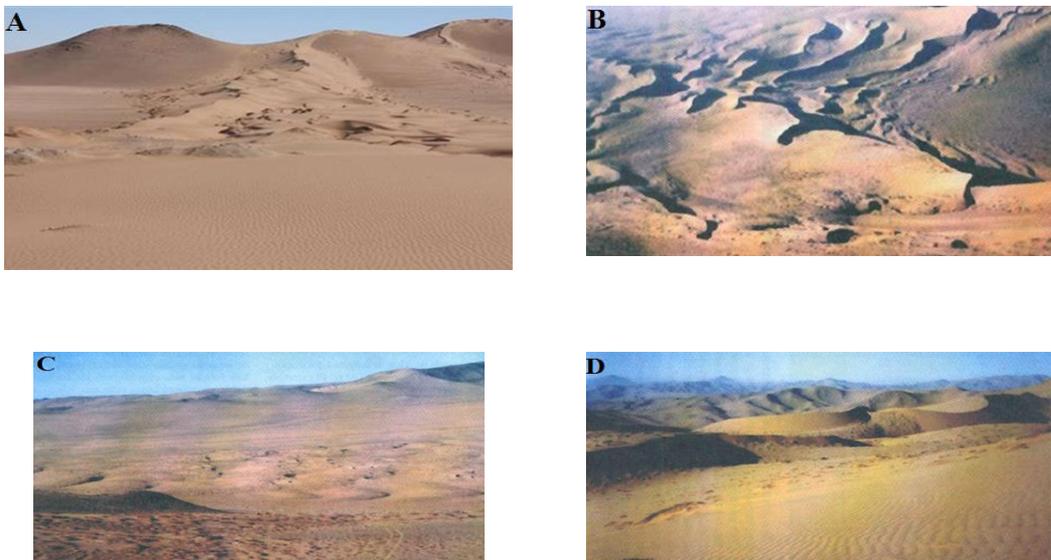


Figura 2. Dunas do Atacama. A: Duna longitudinal. B: Dunas transgressivas. C: Sistema Transverso-transgresivo. D: Dunas Parabólicas. Fonte: Griem (2009) e Araya-Vergara (2001).

### Métodos

O método geral consistiu no uso de processamento digital de imagens visando a realçar características espectrais associadas as litologias,

as estruturas geológicas e solos, e sua comparação com uma imagem Sentinel 2, além do uso estatístico da análise de erros de comissão e

omissão, como também de acurácia com os índices de exatidão global e simultaneidade geográfica.

Segundo Congalton (2009) não existe um procedimento único para realizar uma avaliação de precisão posicional ou temática. No entanto, todas as avaliações de precisão incluem três etapas fundamentais: desenho das amostras de avaliação de precisão, coleta dos dados para cada amostra e análise dos resultados.

Os procedimentos de amostragem para avaliação da precisão são projetados segundo as classes que foram determinados para esta pesquisa (Quadro 1) e as áreas de amostra no mapa (pontos de um pixel) são selecionadas de maneira amostras aleatórias simples e amostras aleatórias estratificadas. Para as análises, utiliza-se amostras da imagem Landsat 5 para dados de referência e amostras da imagem Sentinel 2 para dados de avaliação para cada classe.

**Tabela 1.** Dados da imagem utilizado na pesquisa.

Sistema Sensor	Resolução	Data	Fonte
Landsat 5 TOA	30 m	1991 – 02 - 02	NASA
SENTINEL 2	10 m	2020 – 12 - 31	ESA

De forma geral, as seguintes etapas de trabalho foram desenvolvidas nesse estudo (Figura 4)

- Criação de polígonos para delimitar as classes (categorias) para a amostragem aleatório estratificado através de *shapefiles* de instituições do governo do Chile. (Sernageomin, 2003; Ciren, 2013) e segundo Castro et al. (2012).
- Criação de chave de interpretação da imagem para delimitar as classes para a amostragem aleatório simples.
- Treinamento para a classificação supervisionada. Nesta etapa se trabalha dentro do GEE através de *Scripts* para a seleção da Imagem Landsat 5 do ano 2011 Ortorectificada com Cloud Cover 0 e as bandas B1, B2, B3, B4, B5, B6 e B7 do Landsat. Nesta etapa também se trabalha com a ferramenta *random points* para a criação de pontos aleatórios nas duas metodologias, com amostragem aleatório estratificado e amostragem aleatório simples, aplicação do *merge* nas classes, cálculo do NDVI, seleção das Bandas da Imagem para logo, a aplicação dos algoritmos. Em total, foram coletadas 4.925 amostras para amostragem aleatório estratificado e 4.017 amostras para amostragem aleatória simples. Nas duas metodologias, o 60% foram para treinamento e criação dos mapas supervisionados, e 40% para validação dos mapas.

Com a ferramenta *merge* no GEE as classes ficaram unidas e logo se aplicou a seleção

das bandas para assim ter as *samples regions*. Os scripts foram trabalhados na plataforma em nuvem de GEE com seu *code editor* na linguagem de *JavaScript* como também a entrada das amostras para amostragem aleatório simples e os polígonos para fazer os pontos para as amostras estratificadas até a criação da matriz de confusão.

- Resultados dos mapas da classificação supervisionada. Se obtém os três mapas com classificação supervisionada através dos algoritmos RF, CART e SmileCART.

- Processo de validação dos mapas é feita através dum mosaico da Imagem Sentinel 2 Cloud Cover 0 das datas entre 2018 -01-01 até 2020-12-31. Com a Imagem Sentinel 2 são criados pontos aleatórios simples e pontos aleatórios estratificados independentes das amostras de treinamento para cada classe. Se aplica o *merge* para as classes da validação e por último as *sample regions* para obter as três matrizes de erros para cada algoritmo baseado na amostragem aleatória estratificada e amostragem aleatório simples.

- Cálculo de medidas é índices de concordância. Nesta etapa da metodologia se calculam os índices de acurácia dos seis mapeamentos feitos com os algoritmos. Se aplica os índices de exatidão global, simultaneidade geográfica (SG) e os erros de omissão e comissão para determinar qual mapeamento tem melhor desempenho.

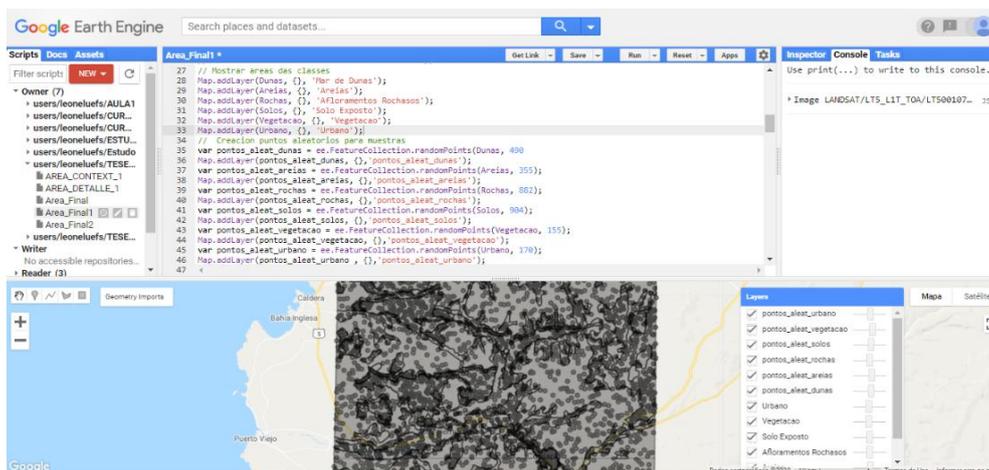


Figura 3. Imagem da consola de Code Editor da plataforma Google Earth Engine num processo de amostragem estratificada para a classificação supervisionada com os pontos das amostras aleatórias (áreas de treinamento).

**Quadro 1.** Características das classes de uso e cobertura da terra

Classe	Características
<b>Afloramento rochoso</b>	Rochas com idade paleozoica superior à recente.
<b>Solo Exposto</b>	Os solos são constituídos por minerais não consolidados e partículas orgânicas.
<b>Areais</b>	Sedimentos quaternários (Qm e Qe).
<b>Mar de Dunas</b>	As areias das dunas foram transportadas para o interior pelo vento, através de corredores transversais.
<b>Vegetação</b>	Áreas com cobertura vegetal regenerada ou em processo de regeneração.
<b>Área Urbana</b>	Estradas pavimentadas e ocupação urbana.

*Índices Espectrais*

Os índices de sensoriamento remoto são a derivação matemática das taxas de banda dos comprimentos de onda refletidos das imagens de satélite detectadas remotamente, com base em assinaturas espectrais das características da Terra. Esses índices também são conhecidos como índices normalizados, porque os valores resultantes estão sendo normalizados entre 0 e ± 1.

O Índice de Vegetação por Diferença Normalizada (NDVI) não é uma tecnologia nova, tendo mais de 40 anos desde a sua primeira aparição. O índice é muito utilizado não somente na agricultura, mas também em diversos estudos ambientais. Este índice é aplicado na identificação da presença de vegetação verde na superfície

(Ponzoni e Shimabukuro, 2007). O NDVI é obtido através da razão entre a diferença da refletância do infravermelho (NIR) e do vermelho (R) pela soma das mesmas variáveis (Equação 1).

$$NDVI = \frac{(NIR - R)}{(NIR + R)} \quad (1)$$

Os valores de NDVI variam entre -1 a 1, onde valores negativos representam corpos hídricos, nuvens e sombra. O solo exposto ou com vegetação rala, é representado por valores positivos próximos de zero. A vegetação densa, bem desenvolvida apresenta os maiores valores de NDVI, próximos a 1 (Poelking et al., 2007). Para interpretação do NDVI foi feito um fatiamento de acordo preconizado por Santos et al., (2008), conforme a Tabela 2

*Algoritmos para classificação*

Existem vários métodos de classificação que buscam, através de diversas abordagens, identificar com acurácia a informação de cada pixel da imagem, classificando-o em categorias (Queiroz et al., 2004). Para realizar a classificação são utilizados modelos matemáticos denominados algoritmos. De acordo com Vale et al., (2018), a classificação supervisionada requer que o analista tenha um conhecimento prévio das classes presentes na imagem. Em tal abordagem, o analista treina o algoritmo para poder distinguir as classes uma das outras, baseando-se em regras estatísticas pré-estabelecidas (Venturieri, 2007). Uma área da imagem que o analista identifica como representando uma das classes é chamada de área de treinamento.

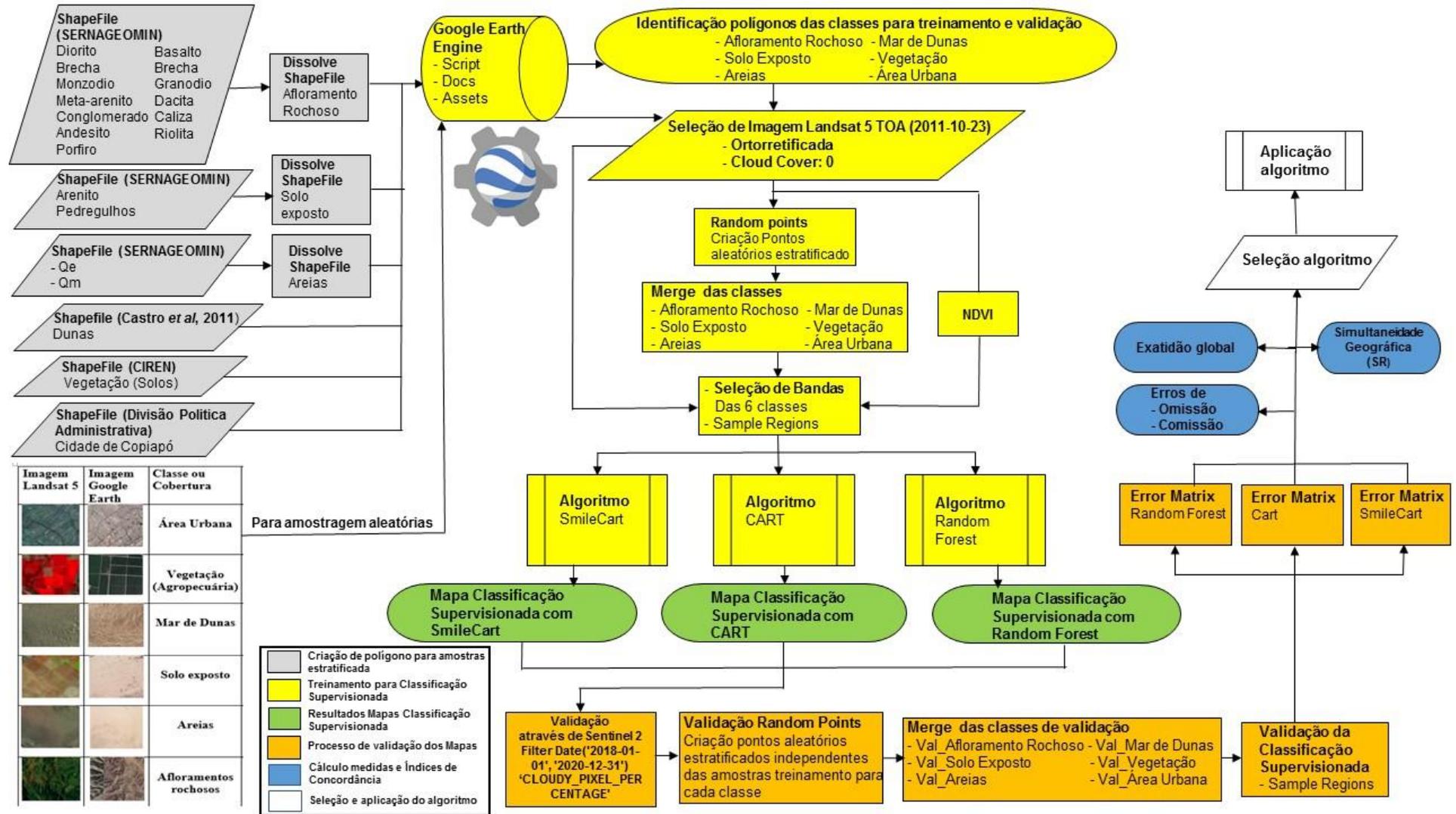


Figura 4. Fluxograma Metodológico do Desenvolvimento da Pesquisa. Ela tem cinco estados os quais são: criação dos polígonos para as classes (gris), treinamento para classificação (amarelo), resultados dos mapas classificados (verde), processo de validação dos mapas (laranja) e os cálculos de índices de concordância

**Tabela 2.** Fatiamento NDVI baseado em SANTOS et al., (2008).

Classes temáticas	Variação do NDVI	Características
Água e Sombra	-0,77 a -0,20	Corpos d' água, sombreamento por nuvens o relevo
Áreas não vegetadas	-0,20 a 0,18	Centros urbanos, povoados, vilas e os terrenos em preparo para atividade agropecuária.
Vegetação bastante esparsa	0,18 a 0,23	Áreas destinadas às atividades de criação de animais e os locais onde as plantações de culturas estão em fase inicial (recém plantadas).
Vegetação esparsa	0,23 a 0,44	Culturas em estágio de crescimento e vegetais de pequeno porte.
Vegetação densa	0,44 a 0,75	Vegetação com estrutura bem desenvolvida ou plantações em estágio próximo a colheita.
Vegetação muito densa	0,75 a 1	Vegetação densa, úmida e bem desenvolvida.

Um classificador de árvore de decisão é construído a partir de um conjunto de dados de treinamento usando o conceito de entropia de informações. Em cada nó da árvore, um atributo dos dados que dividem seu conjunto de amostras em subconjuntos enriquecidos em uma classe ou outra é selecionado. Os algoritmos de classificação supervisionada nesta pesquisa são: Árvore de Classificação e Regressão (CART), Classificador SmileCART (similar a CART) e Random Forest (RF)

A técnica de árvores de classificação e regressão (Classification and Regression Trees CART) introduzida por Breiman et al. (1984) serve de base para técnicas mais complexas. Este procedimento tem como objetivo discriminar, em diferentes regiões, observações com base em suas características.

A classificação por Random Forest (RF) é um método poderoso de aprendizagem de conjuntos que foi proposto por Breiman (2001). Este método dispensa suposições paramétricas e possui aplicabilidade em diversos problemas de predição. RF tem sido amplamente utilizado em muitos campos e exibiu bom desempenho (Calderoni et al., 2015; Chen et al., 2014; Hasan et al., 2014; Youssef et al., 2015). O algoritmo cria várias árvores de decisão e as combina para obter uma predição com maior acurácia e mais estável.

*Avaliação da Qualidade dos Dados*

Historicamente, medidas de concordância e índices de concordância têm sido propostos por vários autores para definir a qualidade dos mapeamentos realizados, tais como Cartwright

(1956), Haggard (1958), Cohen (1960), Everitt (1968), entre outros.

Independentemente da aplicação ou não de algum processo de agrupamento ou filtros para melhorar a classificação, é necessário avaliar os resultados da classificação para medir a eficácia do método utilizado e determinar o grau de erro do produto final. Isto pode ser feito com o uso da matriz de confusão ou por meio da estimativa de erros, a qual se destaca pela utilização de uma matriz composta por linhas e colunas e que indica a correspondência entre as classes na verdade de campo e no mapa (Quadro 2).

Quadro 2. Matriz de confusão (Congalton e Green, 2009).

		j = Colunas (Referência)			Total de Linhas
		1	2	k	$n_{i+}$
i = Linhas (Classificação)	1	$n_{11}$	$n_{12}$	$n_{1k}$	$n_{1+}$
	2	$n_{21}$	$n_{22}$	$n_{2k}$	$n_{2+}$
	k	$n_{k1}$	$n_{k2}$	$n_{kk}$	$n_{k+}$
Total de Colunas	$n_{+j}$	$n_{+1}$	$n_{+2}$	$n_{+1}$	$n$

*Exatidão Global (Overall accuracy)*

A Exatidão Global ou Acurácia Total é a razão entre o número de concordância e o total de píxeis da matriz de confusão do mapa, expressa em percentagem. A diagonal principal da matriz de confusão representa a concordância entre os dados de classificação e os dados de referência sendo representada por  $n_{ii}$ , onde  $i = j$  ( $i = 1, 2, \dots, k$ ) (Congalton e Green 2009). Portanto, a classificação correta é representada pelas amostras da diagonal principal da matriz de confusão e a classificação incorreta pelas amostras fora da diagonal principal (Couto, 2003; Stehman, 2006).

$$n_{i+} = \sum_{j=1}^k n_{ij} \quad (2)$$

$$n_{+j} = \sum_{i=1}^k n_{ij} \quad (3)$$

A Equação 2 corresponde à soma marginal de cada linha dos dados de classificação e a Equação 3 corresponde à soma marginal de cada coluna dos dados de referência.

A Exatidão Global ou overall accuracy ( $p_o$ ), é a precisão geral entre a classificação detectada remotamente e os dados de referência. Pode ser calculada da seguinte maneira (Equação 4):

$$p_o = \frac{\sum_{i=1}^k n_{ii}}{n} \quad (4)$$

*Erros de Omissão e de Comissão*

A Matriz de Confusão é composta por elementos na diagonal principal que expressam a quantidade de pixels que foram classificados corretamente ( $n_{ii}$ ), e elementos fora da diagonal que expressam tanto a quantidade de pixels que foram incluídos erroneamente em uma classe, erro de comissão (EC), assim como a quantidade de pixels que foram excluídos erroneamente de uma classe, erro de omissão (EO), (Ganem, 2017).

Em suma, enfatiza-se que um EC é simplesmente definido como a inclusão de uma área em uma categoria quando não pertence a essa categoria. Um EO está excluindo uma área da categoria à qual pertence. Todo erro é uma omissão da categoria correta e uma comissão para uma categoria errada (Congalton & Green, 2009) Equação 5:

$$EC = 1 - \frac{n_{ij}}{n_{i+}} \quad (5)$$

Para o cálculo de EO é preciso relacionar o total de concordância de cada atributo com o total de atributos encontrados nas linhas verticais da tabulação cruzada. O EO pode ser calculada por a Equação 6:

$$EO = 1 - \frac{n_{ij}}{n_{+j}} \quad (6)$$

*Simultaneidade Geográfica*

Definir os limites do uso de mapas através dos erros contidos neles é essencial. Existem numerosas fontes de erros os quais podem estar relacionados com data, posição e interpretação, entre outros. Independente da fonte dos mapas, eles precisam de validação. Os principais procedimentos para validação são o índice de Kappa (IK), PABAK e índice Tau (t). Muitas vezes estes índices apresentam inconsistências devidos a variações marginais (Silva et al., 2017).

Para remediar e corrigir as possíveis falhas nas interpretações do IK, PABAK e Tau, principalmente quando temos a necessidade de avaliar a correspondência entre mapas, mesmo aqueles que possuem um grande número de píxeis, Silva et al., (2017) propuseram um novo índice para validar resultados decorrentes de mapas temáticos oriundos de processamento digital de imagens o qual resolve paradoxos existentes em matrizes de confissão: a Simultaneidade Geográfica (SG).

O conceito da SG reflete tanto a quantidade de concordância geográfica entre píxeis existentes entre dois mapas quanto o posicionamento locacional dos píxeis, o que elimina as inconsistências resultantes das variações nas proporções marginais, tornando esse índice bastante confiável Silva (2018). A Tabela 3 apresenta a matriz de confusão genérica entre dois mapas temáticos.

**Tabela 3.** Matriz de confissão genérica entre dois mapas temáticos (SILVA et al., 2017).

		Interpreted Map (I)		
		1	Others	
Reference Map (R)	1	$\chi_{ii}$	$\chi_{i+} - \chi_{ii}$	$\chi_{i+}$
	Others	$\chi_{+i} - \chi_{ii}$	$N - [\chi_{ii} + (\chi_{i+} - \chi_{ii})] = (\chi_{+i} - \chi_{ii})$	$\{N - [\chi_{ii} + (\chi_{i+} - \chi_{ii}) + (\chi_{+i} - \chi_{ii})]\} = (\chi_{+i} - \chi_{ii})$
		$\chi_{+i}$	$\{N - [\chi_{ii} + (\chi_{i+} - \chi_{ii}) + (\chi_{+i} - \chi_{ii})]\}$	N

O SG é composto por duas partes, que são denominadas omissão, que representa, em percentagem, os erros cometidos entre um determinado atributo de um mapa interpretado e os erros cometidos desse referido atributo relacionado ao mapa de referência.

A soma entre omissão (Equação 8) e extensão (Equação 7) representa o SG, que varia entre 0 (em nenhuma correspondência) e 2 (total correspondência).

$$\text{Extensão} = 1 - \frac{(X_{i+} - X_{ii})}{X_{i+}} \quad (7)$$

$$\text{Omissão} = 1 - \frac{(X_{+i} - X_{ii})}{X_{+i}} \quad (8)$$

$$\text{SG} = \text{Omissão} + \text{Extensão}$$

Onde  $X_{ii}$  representa o total de concordância da categoria,  $(X_{i+} - X_{ii})$  representa o total de concordância da categoria com as outras,  $X_{i+}$  representa o total da linha,  $X_{+i}$  representa o total da coluna,  $(X_{+i} - X_{ii})$  representa o total da coluna menos o total da concordância da categoria e  $N$  é o total de píxeis.

A SG pode ser calculada em quaisquer circunstâncias, ou seja, mapas possuindo um número “n” de categorias, entretanto, somente têm validade quando os mapas envolvidos representarem exatamente a mesma área geográfica e possuírem a mesma resolução espacial, o que significa ter o mesmo número de linhas e colunas.

A SG varia 0 a 2, correspondendo a soma de omissão com extensão, quando tende para dois, de resultado consistente, e quando tende para zero, o resultado é imprestável. Após calculado o SG, esse resultado quando somado aos erros de omissão e comissão calculados segundo Gunter et al., (2006), sempre será igual a 2.

Para o cálculo do SG total, os dados são normalizados para cada categoria e, a seguir, é realizada a soma aritmética (Silva et al., 2017). A interpretação da correspondência entre dois mapas é apresentada na Tabela 4.

**Tabela 4.** Interpretação da correspondência entre dois mapas com índice de Simultaneidade Geográfica (SILVA, 2018).

SG	Grau de Acurácia
[1.7 a 2.0]	Excelente
[1.5 a 1.7]	Muito boa
[1.2 a 1.5]	Boa
[0.9 a 1.2]	Regular
[0.6 a 0.9]	Ruim
< 0.6	Imprestável

## Resultados

Os resultados do NDVI para auxiliar na identificação da classe vegetação se podem

observar na Figura 5. O fatiamento do NDVI está baseado segundo SANTOS et al., (2008) e feitos com Imagem Landsat 5 (2011 -10 -23). Em cores verdes claros ficam localizados as áreas destinadas às atividades de criação de animais e os locais onde as plantações de culturas estão em fase inicial ou recém-plantadas (0,18 até 0,23) e também em cor verde claro ficam as áreas com culturas em estágio de crescimento e vegetais de pequeno porte (0,23 até 0,44). A vegetação densa, bem desenvolvida apresenta os maiores valores de NDVI, próximos a 1.

### Mapas da classificação supervisionada

Com a aplicação dos algoritmos Random Forest, CART e SmileCART, as amostras de treinamento e a classificação supervisionada no ambiente do GEE se obtém seis mapas, três para amostragem aleatório estratificado e três para amostragem aleatório simples.

O mapeamento feito a través da amostragem aleatório estratificado ficou diferente ao mapeamento aleatório simples como se pode olhar na Figura 6. Principalmente a área que corresponde à classe de solo exposto tem mais área no mapeamento feito com amostragem estratificada devido a que amostras foram realizadas com o polígono derivado da geologia baseado no Sernageomin (2003), o qual tem uma superfície maior que a classe de solo exposto feita com amostragem aleatória simples, onde está última foi feito com o *shapefile* dos solos do Ciren (2007).

As outras cinco classes ficaram homogêneas no tamanho das áreas nos seis mapas feitos com os três algoritmos e as metodologias de amostragem aleatório simples e aleatório estratificado.

### Processo de validação dos mapas

Para realizar o processo de avaliação dos seis mapas da classificação supervisionada, se faz um processo de criação de novos pontos aleatórios para amostras de avaliação, as quais são independentes das amostras de treinamento (diferentes amostras para fazer avaliação e treinamento).

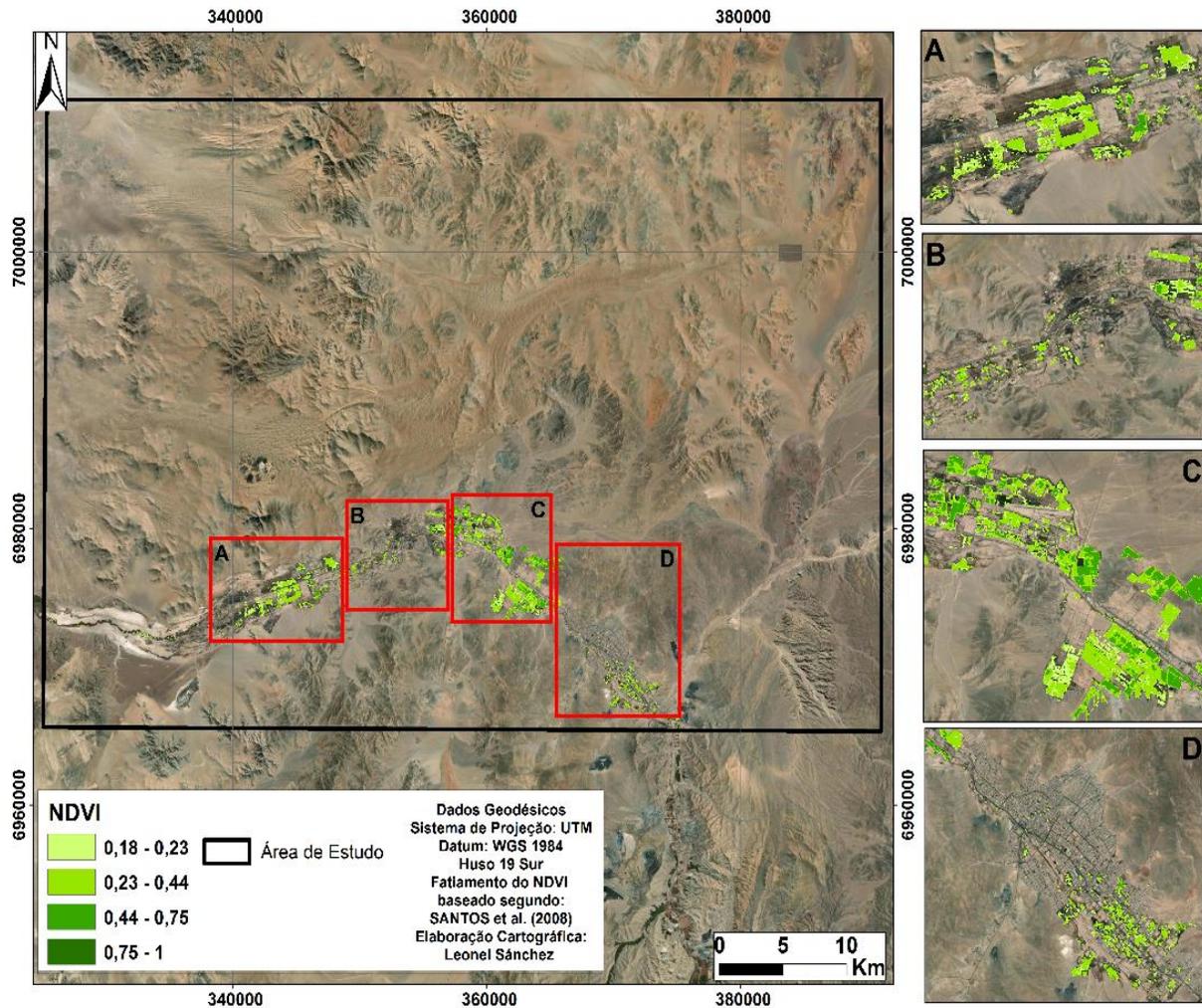


Figura 5. Índice de Vegetação por Diferença Normalizada (NDVI) para ano 2011.

Com um total de 1.967 e 1.537 amostras para avaliação da amostragem estratificado e amostragem aleatório simples respectivamente, se fizeram pontos aleatórios em cada polígono das classes na amostragem estratificada e em toda a área de estudo na amostragem aleatório simples.

Com as amostras de avaliação será possível obter pontos com as quais serão feitos uma sobreposição na imagem do Sentinel 2 e sobre os mapas classificados e com isso será possível saber se realmente tem correspondência verdadeira os pontos de validação com os pontos ou píxeis do mapa classificado. Para fazer esta avaliação é necessário realizar um geoprocesso chamado *merge*, aonde se unem os pontos segundo sua classe.

As análises da acurácia dos mapeamentos, com os três algoritmos foi realizada com a criação de uma variável avaliação no GEE (Quadro 3), que junta os mapas classificados com a coleção de pontos de validação e a dos mapas classificados. Com isso, se consegue obter a Matriz de Confusão para os seis mapeamentos baseados nos algoritmos aplicados e assim obter também os EC, EO, os índices de exatidão global e SG.

**Quadro 3.** Nomes das classes para processo de avaliação dos mapas com classificação supervisionada.

Nomes das classes	
Val_Afloramento rochoso	Val_Mar de Dunas
Val_Solo exposto	Val_Vegetação
Val_Areias	Val_Área urbana

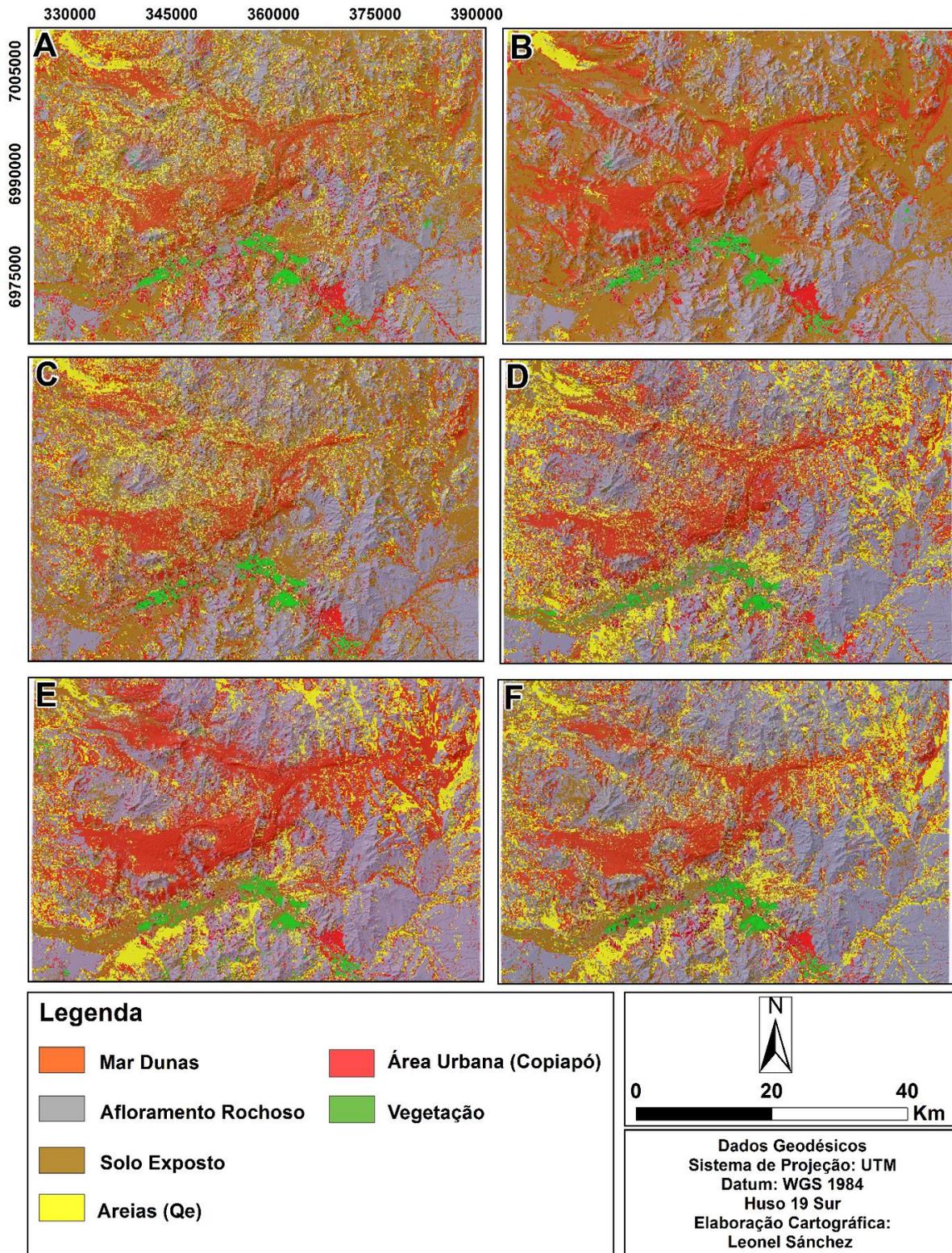


Figura 6. Resultados dos mapas com classificação supervisionada (2011). A: Random Forest com amostragem estratificada. B: Cart com amostragem estratificada. C: SmileCart com amostragem estratificada. D: Random Forest com amostragem aleatória simples. E: Cart com amostragem aleatória simples. F: SimleCart com amostragem aleatória simples

*EC e EO para amostragem aleatória estratificada*

Segundo a Figura 7, para a classe Dunas, os erros de omissão mais altos são para os mapas

com classificação Cart e SmileCART com 43 e 41% respectivamente. O erro de omissão é quando os pixels estão no terreno, mas no mapa da classificação não foram mapeados. O algoritmo

RF, Cart e SmileCART tem 44, 45 e 45% de EC respectivamente. O erro de comissão é quando os pixels que estão no mapa classificados, mas eles na realidade, não estão no terreno.

Na classe das Areias, os erros de omissão e comissão são 30 e 28% respectivamente na classificação feita no algoritmo RF, 20 e 71% em CART, e 51 e 53% em SmileCART. É assim que os EC foram pixels incorretamente classificados e onde os EO são pontos no terreno que pertencem à classe Areias, entretanto, na base de dados (pontos de validação) correspondem a categorias diferentes.

As classes de Dunas e Areias apresentaram confusão entre os classificadores principalmente porque estas duas classes têm semelhança espectral entre elas, isso porque o tipo de material geológico é o mesmo segundo o SERNAGEOMIN (Qe), por conseguinte, os EO e EC são altos.

O erro de omissão na classe de Afloramento Rochoso é alto no resultado do mapeamento com classificador SmileCART. Com SmileCart o EO é 31%, com RF tem 18%, e com CART é 22%. Os EO nesta classe são pixels que existem no terreno que não foram mapeadas. Os EC para a classe Afloramento Rochoso é pequena para o resultado do mapeamento com algoritmo RF (12%). Para os algoritmos SmileCART e CART é 21 e 18% respectivamente, portanto, são aquelas porcentagens de pixels foram classificados pelos algoritmos no mapa, mas eles não existem no terreno.

Na classe de Solo Exposto, o mapeamento com os algoritmos CART e SmileCART têm os EO mais altos resultados, com 36 e 32%. O EO para o mapeamento com RF é só de 18%. Os EO para a categoria Solo Exposto são pixels que não foram mapeados pelo classificador, entretanto, na realidade do terreno sim existem.

Os EC e EO para as classes de Afloramentos Rochoso e Solo Exposto, são relativamente altos, devido que possivelmente existe confusão com as classes de Dunas e Areias, já que, estas últimas, têm uma granulometria fina que pode ser trasladada pelo vento e cobrir a superfície das categorias de Afloramentos Rochoso e Solo Exposto e, por conseguinte, pode ter uma semelhança espectral entre elas.

Pode existir também confusão de classes na categoria do Solo Exposto, porque se adquiriram amostras dentro do polígono da classe Vegetação. Isso se dá devido ao fato de que em diferentes datas

do ano se tem cultivos em algumas áreas e em outras datas não, e nas amostras em que o terreno não fica como solo exposto já que há vegetação. Esta confusão faria aumentar os EC e EO na classe de Solo Exposto, devido à sobreposição entre as classes de Solo Exposto e Vegetação.

A classe vegetação nas três categorias têm um baixo nível de EC. Os mais baixos são os mapeamentos feitos com o algoritmo RF e CART com um 1%. Os EC para eles são pixels que foram incorretamente mapeados na categoria vegetação, no entanto, eles pertencem ao terreno de outra categoria. Os EO para a classe Vegetação são baixos. Nos três algoritmos o resultado é 3, 7 e 9% para RF, CART e SmileCART respectivamente. Com o qual, aqueles pixels foram omitidos pelos classificadores. Portanto, a classe vegetação apresentou excelente acurácia nos EC e EO pois tem baixa confusão de classes, assim, através do índice NDVI, se conseguiu uma boa detecção das áreas com vegetação.

Os resultados para os EC da Área Urbana têm bom resultado nos três classificadores: Cart, SmileCart e RF, com 15, 26 e 11% respectivamente. Essas porcentagens são pixels, que foram mapeados no mapa segundo aqueles algoritmos, porém, eles são pixels que não existem no terreno. Os EO na categoria Área Urbana, para os algoritmos RF, Cart e SmileCart são 23, 14 e 28% respectivamente. As porcentagens do EO são pixels, que estão no terreno, mas eles não foram mapeados pelos classificadores.

Em relação aos EO, os menores são para a classe Duna e acontece para o algoritmo RF, para a classe Areias é Cart, para Rochas é RF, para Solos é RF, para Vegetação é RF e para área Urbana é Cart. Para os EC, os mapas feitos pelos três classificadores, os menores EC em todas as classes ou categorias acontece no algoritmo RF.

As médias dos EO para as 6 classes para os algoritmos RF, Cart e SmileCart são 20, 24 e 32% respectivamente, e as médias dos EC para os algoritmos, RF, Cart e SmileCart para as 6 classes são 18, 29 e 31% respectivamente, portanto, entre os classificadores testados, a classificação por RF foi superior às demais técnicas de classificação segundo as médias dos EC e EO na metodologia aleatória estratificada.

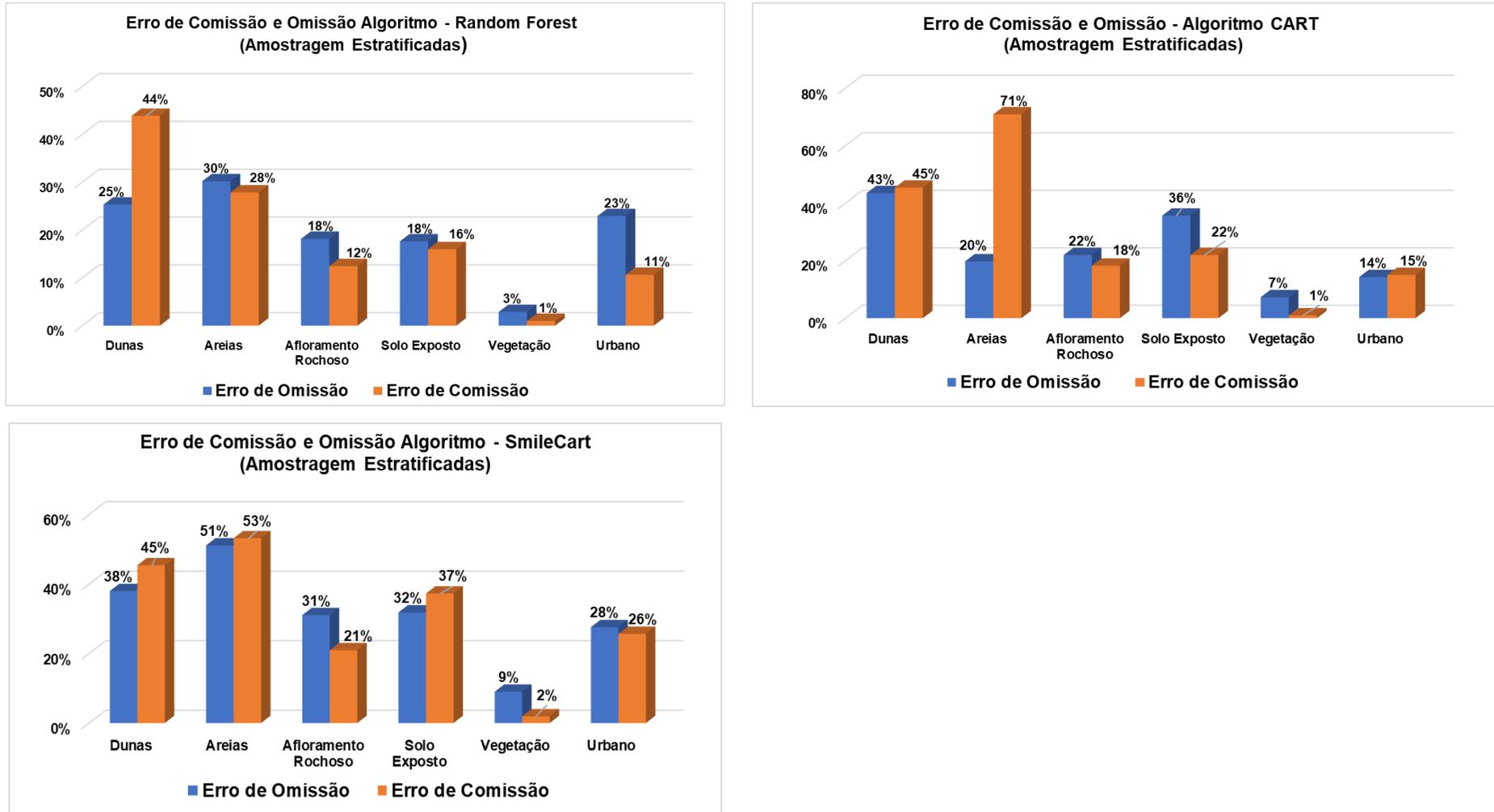


Figura 7. Resultados de Erros de Comissão e Omissão para os algoritmos RF, CART e SmileCART baseado na amostragem aleatório estratificado

*EC e EO para amostragem aleatório simples*

Segundo a Figura 8, para a classe Dunas na amostragem aleatório simples, o erro de omissão mais alto é para o mapa com classificação RF com 43%. O erro de omissão é quando os pixels estão no terreno, mas no mapa da classificação não foram mapeados. O algoritmo RF, CART e SmileCART tem 31, 12 e 24% de EC respectivamente. O erro de comissão é quando os pixels que estão no mapa classificado, mas eles na realidade, não estão no terreno.

Na classe das Areias, os erros de omissão e comissão são 55 e 59% respectivamente na classificação feita no algoritmo RF, 32 e 42% em CART, e 42 e 55% em SmileCART. As classes de Dunas e Areias também apresentaram confusão entre os classificadores principalmente porque estas duas classes têm semelhança espectral entre elas, isso porque o tipo de material geológico é o mesmo segundo o Sernageomin, (2003) (Qe), por conseguinte, os EO e EC são altos.

O erro de omissão na classe de Afloramento Rochoso é 31% no resultado do mapeamento com classificador RF. Com SmileCart o EO é 24%, com CART tem 11%. Os EO nesta classe são pixels que existem no terreno que não foram mapeadas. Os EC para a classe Afloramento Rochoso é 31% para o resultado do mapeamento com algoritmo RF. Para os algoritmos SmileCART e CART é 19 e 22% respectivamente.

Na classe de Solo Exposto, o mapeamento com os algoritmos RF e SmileCART têm os EO mais altos resultados, com 42 e 34%. O EO para o mapeamento com CART é de 28%. Os EO para a categoria Solo Exposto são pixels que não foram mapeados pelo classificador, entretanto, na realidade do terreno sim existem.

Os EC e EO para as classes de Areias e Solo Exposto, são relativamente altos, devido que possivelmente existe confusão com as classes.

A classe vegetação nas três categorias têm um baixo nível de EC. As mais baixas são os mapeamentos feitos com o algoritmo SmileCART e CART com um 0%. Os EO para a classe Vegetação são baixos. Nos três algoritmos o resultado é 12, 3 e 4% para RF, CART e SmileCART respectivamente.

A última classe analisada é a Área Urbana. Os resultados para os EC nos três classificadores, Cart, SmileCart e RF, são 6, 20 e 37% respectivamente. Os EO na categoria Área Urbana, para os algoritmos RF, Cart e SmileCart são 32, 12 e 13% respectivamente. As percentagens do EO são pixels, que estão no terreno, mas eles não foram mapeados pelos classificadores.

Em relação aos EO, os menores EO para a classe Duna acontece para o algoritmo SmileCART, para a classe Areias, Rochas, Solos e Vegetação é CART, e para área Urbana é RF e CART. Para os EC, os mapas feitos pelos três classificadores segundo a amostragem aleatório simples, os menores EC em todas as classes ou categorias acontece no algoritmo CART.

As médias dos EO para as 6 classes para os algoritmos RF, CART e SmileCART são 34, 18 e 23% respectivamente, e as médias dos EC para os algoritmos, RF, CART e SmileCART para as 6 classes são 33, 15 e 24% respectivamente, portanto, entre os classificadores testados, a classificação por CART foi superior às demais técnicas de classificação segundo as médias dos EC e EO baseado segundo a metodologia da amostragem aleatório simples.

*Simultaneidade Geográfica para amostragem aleatória estratificada*

Segundo a Figura 9, a classe Dunas tem bom desempenho para o algoritmo RF com um nível de 1,31, mas para CART e SmileCART é regular o desempenho com um nível de 1,11 e 1,17 respectivamente, segundo interpretação da Tabela 4. A classe Areias tem uma boa interpretação de correspondência para RF, e regular para CART e SmileCART com 1,42, 1,09 e 0,96 respectivamente.

Para a classe de Afloramento Rochoso, os resultados para RF 1,69 (muito boa), CART 1,60 (muito boa) e para SmileCART 1,48 (boa). A classe Solo Exposto tem 1,66 com RF, o que é muito bom, 1,42 com CART que é bom desempenho e com SmileCART tem o nível de SG é de 1,31 que é bom. A classe Vegetação tem o niveles de 1,96, 1,92 e 1,89, os são niveles excelentes para o índice de SG. A classe Urbano alcanço os niveles 1,66 para RF (muito boa), 1,71 para CART (Excelente) 1,47 para SmileCART (boa).

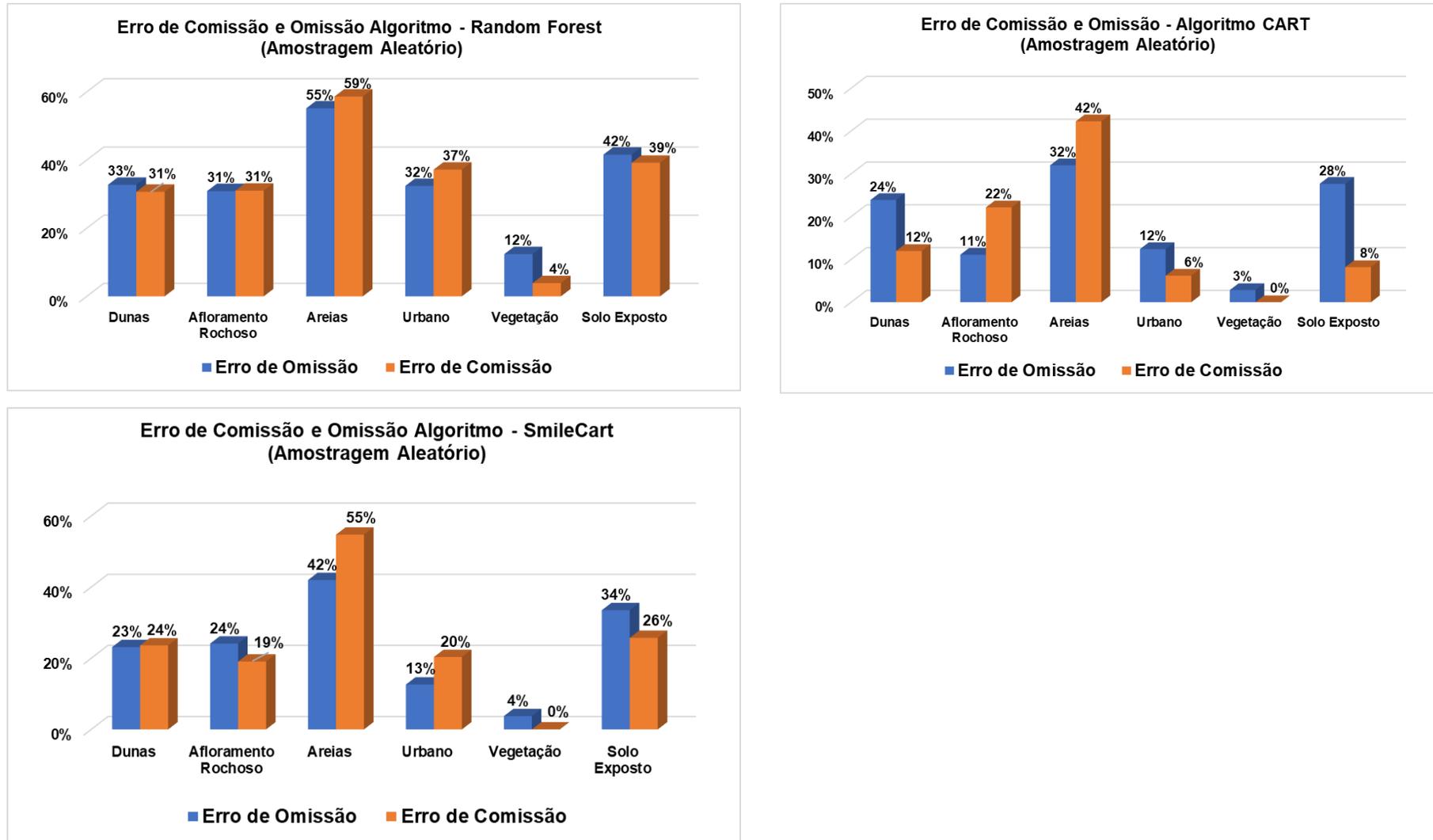


Figura 8. Resultados de Erros de Comissão e Omissão para os algoritmos RF, CART e SmileCART baseado na amostragem aleatório simples

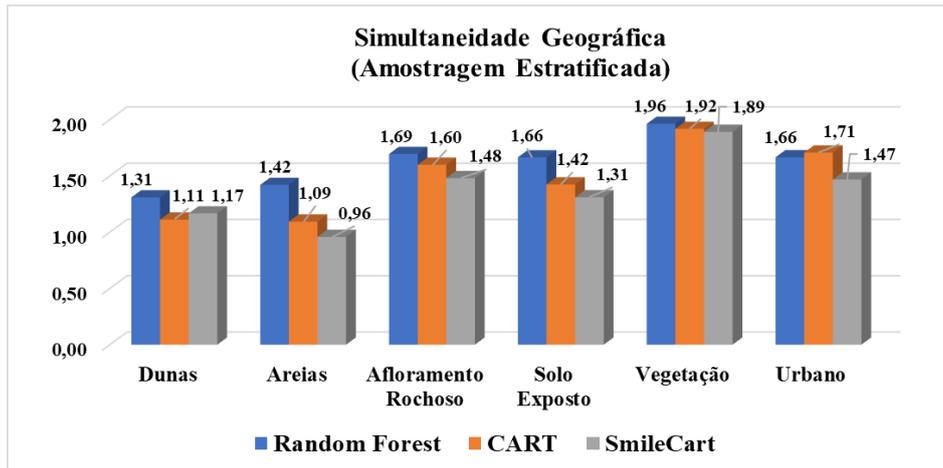


Figura 9. Resultados do índice de simultaneidade geográfica para amostragem aleatória estratificada.

*Simultaneidade Geográfica para amostragem aleatória simples*

Segundo a Figura 10, a classe Dunas tem muito bom desempenho para o algoritmo CART com um nível de 1,64, RF e SmileCART têm uma boa e muito bom desempenho com um nível de 1,37 e 1,53 respectivamente, segundo interpretação da Tabela 4. A classe Areias tem uma boa interpretação de correspondência para CART, ruim para RF e regular para SmileCART com 1,26, 0,86 e 1,03 para cada algoritmo respectivamente.

Para a classe de Afloramento Rochoso, os resultados para RF 1,38 (boa), CART 1,67 (muito boa) e para SmileCART 1,57 (muito boa). A classe Solo Exposto tem 1,19 com RF, o que é regular, 1,64 com CART que é muito boa desempenho e com SmileCART tem o nível de SG é de 1,41 que é boa. A classe Vegetação tem os níveis de 1,84, 1,97 e 1,96, os são níveis excelentes para o índice de SG. A classe Urbano alcanço os níveis 1,30 para RF (boa), 1,81 para CART (Excelente) 1,67 para SmileCART (muito boa).

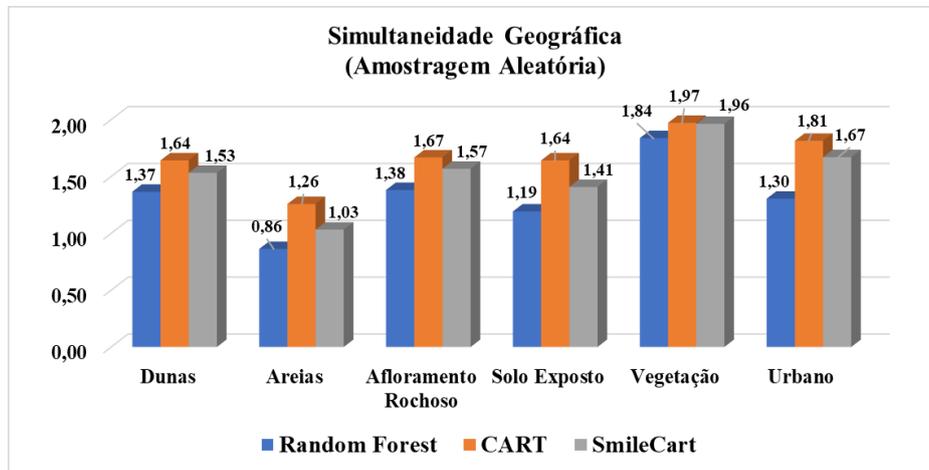


Figura 10. Resultados do índice de simultaneidade geográfica para amostragem aleatória simples.

*Resultados da classificação dos algoritmos com amostragem aleatória estratificada*

Baseado na Figura 11, o resultado do algoritmo RF para o índice Exatidão Global é muito bom, onde tem 0,8 (80%) de accuracy global que é a razão entre o número de concordância e o total de pixels da matriz de confusão. Para o algoritmo CART, o resultado do Exatidão Global é 0,71 (71%) e para SmileCART é 0,66 (66%).

O CART tem também uma boa acurácia, mas para SmileCART é relativamente baixa à concordância.

O índice de SG tem um nível de 1,62 para o algoritmo RF, o que é muito boa segundo a Tabela 4. O algoritmo CART e SmileCART conseguiram um nível de 1,48 e 1,38 respectivamente o que é bom. Em geral os três algoritmos têm uma boa resposta para a aplicação

dos índices de acurácia, porém, a classificação por RF foi superior às demais técnicas de classificação na metodologia da amostragem aleatória estratificada.

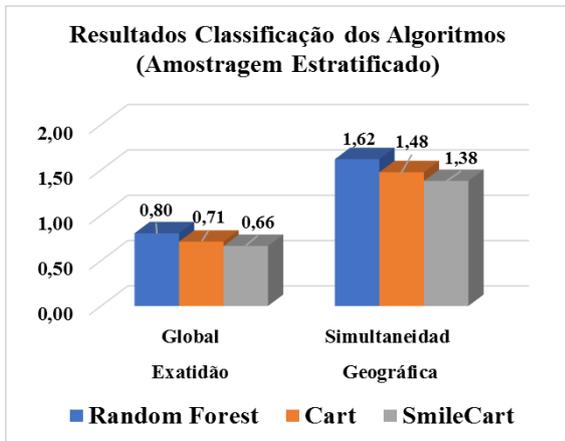


Figura 11. Comparação resultados dos índices de exatidão global e simultaneidade geográfica com a metodologia de amostragem aleatório estratificado.

*Resultados da classificação dos algoritmos com amostragem aleatório simples*

Baseado na Figura 12, o resultado do algoritmo RF para o índice Exatidão Global é bom, onde tem 0,65 (65%). Para o algoritmo CART, o resultado do Exatidão Global é 0,81 (81%) e para SmileCART é 0,75 (75%). O CART tem também uma muito boa acurácia, mais que o algoritmo RF.

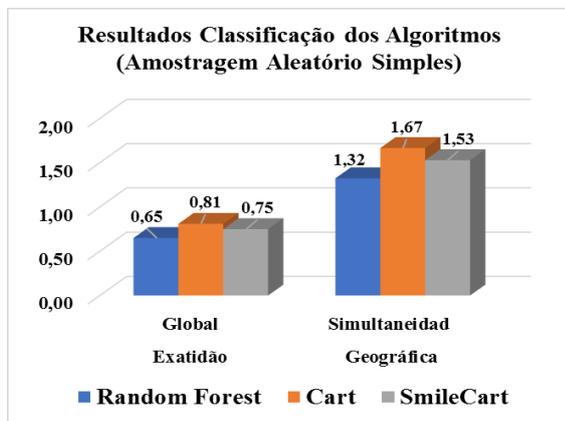


Figura 12. Comparação resultados dos índices de exatidão global e simultaneidade geográfica com a metodologia de amostragem aleatório simples.

O índice de SG tem um nível de 1,67 para o algoritmo CART, o que é muito boa segundo a Tabela 4. O algoritmo RF e SmileCART conseguiram um nível de 1,32 e 1,53 respectivamente o que é bom segundo a Tabela 4. Em geral os três algoritmos têm uma boa resposta para a aplicação dos índices de acurácia, porém, a classificação por CART foi superior às demais

técnicas de classificação na metodologia da amostragem aleatória simples.

**Discussão**

Com os resultados dos índices de SG, exatidão global e os EO e EC O fica como melhor algoritmo para a classificação supervisionada o mapeamento efetuado com o algoritmo CART com a metodologia aleatória simples. Em segundo lugar é RF com a metodologia de amostragem aleatória estratificada.

Com os resultados obtidos na Figura 6E, com a classificação aleatória simples e com o algoritmo CART, pode-se observar que as dunas ficam situadas ao norte do rio Copiapó e com poucos pixels das Dunas ao Sul do rio. Provavelmente, as areias das dunas ficariam fixas devido à topografia da área, não obstante, precisa-se um estudo temporal para verifica se há existido deslocamento das dunas para ao Sul do rio Copiapó no passado.

Na Figura 13 se pode observar o total das áreas (km<sup>2</sup>) que fica cada classe no mapeamento efetuado com o algoritmo CART com a metodologia aleatória simples.

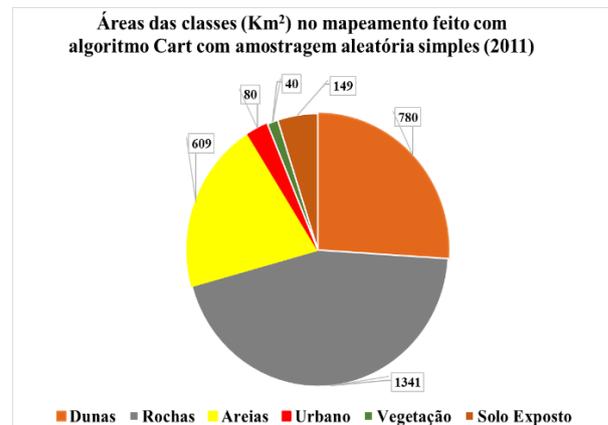


Figura 13. Área de cada classe no mapeamento feito com algoritmo CART com a metodologia aleatória simples

**Conclusão**

Os resultados obtidos neste trabalho de avaliação e comparação entre técnicas de classificação supervisionada e metodologias de amostragens aleatória estratificada e aleatória simples através de imagem Landsat TOA, junto com a plataforma em nuvem do Google Earth Engine, possibilitou o mapeamento do Mar de Dunas do Atacama, no norte do Chile

Para a metodologia de amostragem aleatória estratificada e a metodologia de amostragem aleatória simples, os diferentes

métodos de classificação supervisionada com os algoritmos CART, SmileCART e Random Forest, são eficientes para o mapeamento das dunas do Atacama e para as cinco classes pois, apresentaram, segundo o índice de Simultaneidade Geográfica níveis de 1,48, 1,38 e 1,62 respectivamente para amostragem aleatório estratificado e 1,67, 1,53 e 1,32 respectivamente para amostragem aleatório simples.

As classes Solo Exposto, Vegetação e Área Urbana obtiveram bom desempenho nas duas metodologias de amostragens (estratificada e aleatória) e em todas as técnicas de classificação aplicadas. As classes temáticas Vegetação (feita com o NDVI) e Área Urbana apresentaram excelente acurácia em todas as modalidades de classificação dos EC, EO, SG e simultaneidade geográfica. A classe Afloramento Rochoso também apresentou bom desempenho na sua exatidão.

A classe Mar de Dunas apresentou variação na exatidão e confusão entre os classificadores com a metodologia de amostragem aleatória estratificada. Isso aconteceu principalmente com as classes de Areia e Solo Exposto, devido à semelhança espectral podendo ocorrer a sobreposição em alguma delas e a habilidade de cada classificador em estabelecer o limite de decisão no espaço de feições. Na metodologia de amostragem aleatório simples a variação na exatidão e confusão foi menor.

Para o mapeamento das Dunas de Atacama, comparando as duas metodologias de amostragens, a técnica de classificação supervisionada por CART representa o melhor desempenho, já que entre os métodos testados, a classificação supervisionada feita pelo algoritmo CART, com a metodologia da amostragem aleatória simples, foi superior às demais técnicas de classificação.

### Agradecimentos

À Universidade Estadual de Feira de Santana (UEFS), em especial ao Programa de Pós-Graduação em Modelagem em Ciências da Terra e do Ambiente (PPGM), pela formação e estrutura oferecida para o desenvolvimento desta pesquisa. À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo financiamento do projeto de Mestrado. Por último, à Organização dos Estados Americanos (OEA) e ao Grupo Coimbra de Universidades Brasileiras (GCUB).

### Referências

- Aguirre, I. 1999. Hidrogeología do Valle do Río Copiapó entre Tranque Lautaro e Piedra Colgada. Memória para qualificar ao grau de Geólogo e Teses para optar al grau de Mestre em Geologia. Universidad de Chile.
- Araya-Vergara, J. F. 2001. Los ergs del desierto marginal de Atacama, Chile. *Invest. Geogr. Chile*, 2001, 35:27-66.
- Bernardi, H. V. F., Dzedzej, M., Carvalho, L. M. T., Acerbi Júnior, F. W. 2007. Classificação digital do uso do solo comparando os métodos “pixel a pixel” e orientada ao objeto em imagem QuickBird. In: Simpósio Brasileiro de Sensoriamento Remoto, 13., 2007, Florianópolis. Anais. São José dos Campos: INPE. p. 5595-5602.
- Bird, E.C.F. 2008. Coastal geomorphology: an introduction. Second edition. Wiley. 436p.
- Bishop, C. M. 2006. Pattern Recognition and Machine Learning. Berlin: Springer.
- Börgel, R. 1963. Dunas Litorales de Chile. Teoría y Aplicación. Instituto de Geografía de la Universidad de Chile. 10p.
- Börgel, R. 1983. Geomorfología. Instituto Geográfico Militar, Santiago de Chile. 25p.
- Breed, C., Grow, T. 1979. Morphology and distribution of dunes in sand seas observed by remote sensing. In: MCKEE, E. (Editors). A Study of Global Sand Seas. Washington D.C.: U.S. Geological Survey Paper, p. 253 - 304. 1052.
- Breiman, L. 1984. Classification and Regression Trees. Belmont, California, U.S.A.: Wadsworth Publishing Company, 1984. (Statistics/Probability Series). 15-18.
- Breiman, L. 2001. Random Forests. *Machine Learning* 45, 5–32.
- Calderoni, L., Ferrara, M., Franco, A., Maio, D. 2015. Indoor localization in a hospital environment using random Forest classifiers. *Expert Syst. Appl.* 42 (1), 125–134.
- Campbell, J. B. Introduction to Remote Sensing. New York: The Guilford Press, 1996.
- Castro, C., Zúñiga, A., Pattillo, C. 2012. Geomorfología y geopatrimonio del Mar de Dunas de Atacama, Copiapó (27° S), Chile. *Revista de Geografía Norte Grande*, 53: 123-136.
- Cartwright, M. L. 1956. Integral Functions, Cambridge Tracts in Math and Math. Phy., v. 44, Cambridge Univ. Press.
- Ciren. 2007. Estudio Agrológico Valle del Copiapó y Valle del Huasco, III Región. Descripciónes de suelos, Materiales y símbolos. (Pub. Centro

- de Información de Recursos Naturales de Chile - CIREN N°135).
- Ciren. 2013. Caracterización de Humedales Altoandinos para una gestión sustentable de las actividades productivas del sector norte del país". Informe Recurso Suelo III Atacama. Centro de Información de Recursos Naturales de Chile.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, P. 37 – 46.
- Congalton, R. G., Green, K. 2001. Accuracy assessment and validation of remotely sensed and other spatial information. *International Journal of Wildland Fire*, 2001, 10, 321–328.
- Congalton, R. G., Green, K. 2009. Assessing the accuracy of remotely sensed data: principles and applications. Lewis Publishers, Boca Raton.
- Coira B., Davidson J., Mpodozis C., Ramos V. 1982. Tectonic and magmatic evolution of the Andes of northern Argentina and Chile. *Earth-Sci. Rev.* 18: 303– 332.
- Couto, P. 2003. Assessing the accuracy of spatial simulation models. *Ecological Modelling*, 167, 181–198. doi:10.1016/S0304-3800(03)00176-5.
- Chen, W., Li, X., Wang, Y., Chen, G., Liu, S. 2014. Forested landslide detection using LiDAR data and the random forest algorithm: a case study of the three gorges, China. *Remote Sens. Environ.* 152, 291–301.
- Drummond, T., Goedtel, A., Cunha, R., Fontes, W. 2007. Aplicação do algoritmo Random Forest como classificador de padrões de falhas em rolamentos de motores de indução. XIII Simpósio Brasileiro de Automação Inteligente Porto Alegre – RS, 1o – 4 de outubro de 2017.
- Dunai T., González G., Juez-Larre J. 2005. Oligocene–Miocene age of aridity in the Atacama Desert revealed by exposure dating of erosion-sensitive landforms. *Geology* 33: 321–324. DOI: 10.1130/G21184.1.
- Everitt, W. N. 1968. Some positive definite differential operators. *London Math. Soc.* 43 465 – 473.
- Fryberger, S., Dean, G. 1979. Dune forms and wind regime. In: Mckee, E. (Editors). *A Study of Global Sand Seas*. Washington D.C.: U. S. Geological Survey Paper, p. 137- 170.
- Ganem, K. 2017. Classificação da Cobertura do Solo na Caatinga a partir de imagens Do Landsat-8 e da ferramenta Google Earth Engine: uma comparação entre dados com e sem correção atmosférica. Dissertação de Mestrado. Univ. de Brasília.
- Godoy, E., Marquardt, C., Blanco, N. 2003. Carta Caldera, Región de Atacama. Santiago de Chile: Servicio Nacional de Geología y Minería, Carta Geológica de Chile, Serie Geología Básica, N° 76.
- Gómez, N. 2014. Climatología Urbana De Copiapó Como Ciudad Localizada En Un Medio Ambiente Árido. Memoria para optar al Título Profesional de Geógrafa da Universidad de Chile.
- Gorelick, N., Hancher, M., Dixon, M. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, n. n 2016.
- Griem. W. Las Dunas de Atacama. Atacama virtual. 2009. Disponível em: <https://www.geovirtual2.cl/Museovirtual/tur210.htm>. Acesso em: 5 nov. 2020.
- Gunter, B.; Ries, J. L.; Srinivas Bettadpur, S.J. 2006. A simulation study of the errors of omission and commission for GRACE RL01 gravity fields. 80: 341 – 351.
- Haggard, E. A. 1958. *Intraclass correlation and the análisis of variance*. New York: Dryden Press.
- Hasan, M.A.M., Nasser, M., Pal, B., Ahmad, S. 2014. Support vector machine and Random forest modeling for intrusion detection system (IDS). *J. Intell. Learn. Syst. Appl.* 6 (1), 45.
- INE, 2012. Resultados del XVIII Censo Nacional de Población y VII de Vivienda 2012. Instituto Nacional de Estadísticas, Chile. [online] Disponível:[http://www.ineatacama.cl/archivos/files/pdf/Otras%20publicaciones/Minuta\\_Censo\\_2012.pdf](http://www.ineatacama.cl/archivos/files/pdf/Otras%20publicaciones/Minuta_Censo_2012.pdf). Acesso: 23 nov. 2020.
- Juliá, C, Montecinos. S & Maldonado. A. 2008. Libro rojo de la flora nativa y de los sitios prioritarios para su conservación: Región de Atacama. Capítulo 3: características climáticas de la Región de Atacama. Ediciones Universidad de la Serena, p 25-42.
- Lam, D. K., Rimmel, T. K.; Drezner, T. D. 2012. Tracking Desertification in California Using Remote Sensing: A Sand Dune Encroachment Approach. *Remote Sensing*. 2011, 3, p. 1-13.
- Laity, J. 2008. *Deserts and Desert Environment*. 3a ed. Wiley-Blackwell.
- Luebert, F. Y Plissock, P. 2017. Sinopsis bioclimática y vegetacional de Chile. Santiago de Chile: Editorial Universitaria, Segunda Edición.
- Mendonça, F. A. 1990. A evolução sócio - econômica do norte novíssimo do Paraná (PR) e os impactos ambientais: desertificação? 323 f. Dissertação (Mestrado) – Universidade de São Paulo.

- Meneses, O. R.; Almeida, T. 2012. Introdução ao Processamento de Imagens de Sensoriamento Remoto. Universidade de Brasília (UnB). Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). 191p.
- Novo; E.L.M. 1989. Sensoriamento remoto: princípios e aplicações. São Paulo: Edgar Blucher. p.1-8.
- Oncken O., Hindle D., Kley J., Elger K., Victor P. 2006. Deformation of the central Andean upper plate system: Facts, fiction, and constraints for plateau models, in Oncken, O., et al., eds., The Andes: Active subduction orogeny. *Frontiers in Earth Sciences*: London, Springer Verlag, 3–28.
- Paskoff, R.; Cuitiño, L.; Manríquez, H. 2003. Origen de las arenas dunares de la Región de Copiapó, Desierto de Atacama, Chile. *Revista Geológica de Chile*, Vol. 30, N° 2, 355-361.
- Pliscoff, P.; Zanetta, N.; Hepp, J.; Machuca, J. 2017. Efectos sobre la flora y vegetación del evento de precipitación extremo de agosto 2015 en Alto Patache, Desierto de Atacama, Chile. *Revista de Geografía Norte Grande*, 68: 91-103.
- Poelking, E. L.; Lauermann, A.; Dalmolin, R.S.D. 2007. Imagens CBERS na geração de NDVI no estudo da dinâmica da vegetação em período de estresse hídrico. *Anais do XII Simpósio Brasileiro de Sensoriamento Remoto*. Florianópolis – SC, p. 4145-4150.
- Ponzoni, F. J., Shimabukuro, Y. E. 2007. Sensoriamento Remoto no Estudo da Vegetação. São José dos Campos, SP. Editora Parêntese.
- Queiroz, R. B., Rodrigues, A. G., Gómez, A. T. 2004. Redes Neurais: Um comparativo com Máxima Verossimilhança Gaussiana na Classificação de Imagens CBERS 1. In: *Workshop De Tecnologia Da Informação Aplicada Ao Meio Ambiente*, Itajaí. *Anais... Itajaí: CBComp*, 2004. p. 746-749.
- Rosa, M. R., Brancalion, P., Crouzeilles, R., Tambosi, L., Piffer, P., Lenti, F., Hirota, M., Santiamand, E., Metzger, J. P. Hidden destruction of older forests threatens Brazil's Atlantic Forest and challenges restoration programs. *Science Advances* 20 Jan 2021. Vol 7, P. 1.
- Rube, R., Kotteck, M. 2010. Observed and projected climate shifts 1901–2100 depicted by world maps of the Koppen-Geiger climate classification. *Meteorologische Zeitschrift*, Vol. 19, No. 2, 135-141.
- Sánchez, L; Aguilar. G; Börgel, R; Comte, D; Ortega, F; Regard, V. 2015. Estudio de la interrelación entre procesos aluvionales y la inundación de Tierra Amarilla asociado a las lluvias torrenciales de marzo de 2015. XIV Congreso Geológico Chileno.
- Santos, A. M., Galvêncio, J. D., Moura, M. S. B. 2008. Aplicação do índice de vegetação por diferença normalizada (NDVI) para identificação da cobertura fitogeográfica da Bacia Hidrográfica do Rio Goiana – PE. II Simpósio de Geografia Física do NORDESTE, João Pessoa - PB, 2008. *Anais... João Pessoa, PB*, p.1-9.
- Sernageomin, 2003. Mapa Geológico de Chile: Version Digital. Base Geológica escala 1:1.000.000. *Publicacion Geologica Digital*, No. 4, 2003 CD-ROM, versión 1.0. Servicio Geológico de Chile.
- Silva, A. B.; Lobão, J. S. B.; Sano, E. E. 2017. Geographical Simultaneity: A New Index to Validate Results Obtained From Digital Image Processing Of Thematic Maps. *Revista Brasileira de Cartografia*, n. 69/2, p. 361 – 373.
- Silva, A. 2018. *Análise Quantitativa Espacial. Conceitos e fundamentos*. Curitiba, PR, p. 151 – 172.
- Simões, F. L. 2014. *Arqueologia da paisagem nas Dunas Holocênicas: O estudo de caso do sítio Cardoso (Lagoa Redonda, Pirambu, Se)*. Dissertação (Mestrado) – Universidade Federal de Sergipe.
- Stehman S.V. 2006. Design, analysis, and inference for studies comparing thematic accuracy of classified remotely sensed data: A special case of map comparison. *Journal of Geographical Systems*, 8, pp. 209-226.
- Tsoar, H.; Arens, S. 2003. Mobilização e Estabilização de dunas em climas úmidos e secos. *Mercator - Revista de Geografia da UFC*, ano 02, número 03.
- Vale, J. R., Costa, J., Santos, J., Silva. 2018. *Análise Comparativa de Métodos de Classificação Supervisionada Aplicada ao Mapeamento da Cobertura do solo no município de Medicilândia, Pará*.
- Venturieri, A. 2007. *Introdução às técnicas de Sensoriamento Remoto*. Belém: UFPA.
- Youssef, A.M., Pourghasemi, H.R., Pourtaghi, Z.S., Al-Katheeri, M.M. 2015. Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia. *Landslides* 1–18.