

Mapeamento de áreas de pastagens do Pontal do Paranapanema para estimativa da produção de etanol

Mapping of Pontal do Paranapanema pasture areas to estimate ethanol production

Mapeo de áreas de pastos en Pontal do Paranapanema para estimar la producción de etanol

Recebido: 25/05/2022 | Revisado: 12/06/2022 | Aceito: 13/06/2022 | Publicado: 14/06/2022

Vitor Siqueira Ferreira

ORCID: <https://orcid.org/0000-0003-1993-9467>
Universidade Estadual Paulista, Brasil
E-mail: siqueirafferreira.victor@gmail.com

Andréia Fátima Zanette

ORCID: <https://orcid.org/0000-0003-2770-2582>
Universidade Estadual Paulista, Brasil
E-mail: andrea.zanette@unesp.br

Claudia Gonçalves de Azevedo

ORCID: <https://orcid.org/0000-0002-0203-3776>
Universidade Estadual Paulista, Brasil
E-mail: claudia.azevedo@unesp.br

Leticia Sabo Boschi

ORCID: <https://orcid.org/0000-0003-4596-1377>
Universidade Estadual Paulista, Brasil
E-mail: leticia.sabo@unesp.br

Resumo

A crescente preocupação com as mudanças climáticas bem como uma legislação ambiental cada vez mais rigorosa e a degradação ambiental impulsionam o consumo de biocombustíveis visando reduzir emissões poluentes. O etanol obtido a partir da cana de açúcar tem sido amplamente utilizado uma vez que é uma fonte de energia renovável e produz baixas emissões de poluentes. No Brasil, o Proálcool incentivou a expansão de lavouras de cana de açúcar pelo estado de São Paulo e a região do Pontal do Paranapanema está presenciando uma crescente substituição de áreas de pastagem por esta cultura. Nesse contexto, o presente trabalho tem como objetivo estimar a produção de etanol para uma pequena região do Pontal do Paranapanema a partir da substituição das áreas de pastagens por cultura de cana-de-açúcar. Para tanto, foi gerado um mapa de uso e cobertura do solo da região através da classificação supervisionada de uma imagem do satélite Sentinel 2A no Google Earth Engine por meio do algoritmo Random Forest e dos índices NDVI, NDBI, MNDWI e RS. A classificação apresentou uma acurácia excelente, com valores de Índice Kappa de 99,7% e acurácia global de 99,78%. As pastagens ocuparam 48,08% da área de estudo, com 196.416,71 hectares e ao substituir as pastagens por cana de açúcar a área apresenta um potencial de produção entre 1.079.741.938,32 e 1.214.709.680,61 litros de etanol. Desse modo, a área mostra um enorme potencial de produção de etanol sendo atrativa para investimentos no setor, favorecendo a economia e a população dos municípios da região.

Palavras-chave: Mapeamento do uso e cobertura do solo; Google Earth Engine; Biocombustíveis.

Abstract

The growing concern about climate change as well as increasingly stringent environmental legislation and the environmental degradation drive the consumption of biofuels in order to reduce polluting emissions. Ethanol obtained from sugar cane has been widely used as it is a renewable energy source and produces low pollutant emissions. In Brazil, the Proálcool encourages the expansion of sugarcane flavors throughout the state of São Paulo and the region of Pontal do Paranapanema is showing a growing replacement of pasture areas by this crop. In this context, the present work aims to estimate the production of ethanol for a small region of Pontal do Paranapanema from the replacement of pasture areas by sugarcane crops. For this, a land use and land cover map of the region was generated through the supervised classification of an image from the Sentinel 2A satellite in Google Earth Engine using the Random Forest algorithm and the NDVI, NDBI, MNDWI and RS indices. The classification had excellent accuracy, with Kappa Index values of 99.7% and overall accuracy of 99.78%. Pastures occupy 48.08% of the study area, with 196,416.71 hectares and when replacing pastures for sugar cane, the area presents a production potential between 1,079,741,938.32 and 1,214,709,680.61 liters of ethanol. Thus, the shows an enormous potential for ethanol production, being attractive for investments in the sector, favoring the economy and the population of the municipalities in the region.

Keywords: Mapping of land use and land cover; Google Earth Engine; Biofuels.

Resumen

La creciente preocupación por el cambio climático así como una legislación ambiental cada vez más estricta y el deterioro ambiental impulsan el consumo de biocombustibles con el fin de reducir las emisiones contaminantes. El etanol obtenido de la caña de azúcar ha sido ampliamente utilizado ya que es una fuente de energía renovable y produce bajas emisiones contaminantes. En Brasil, el Proálcool fomentó la expansión de los cultivos de la caña de azúcar en todo el estado de São Paulo y la región de Pontal do Paranapanema está mostrando una creciente sustitución de áreas de pastos por este cultivo. En este contexto, el presente trabajo tiene como objetivo estimar la producción de etanol para una pequeña región del Pontal do Paranapanema a partir de la sustitución de áreas de pastos por cultivos de caña de azúcar. Para ello se elaboró un mapa de uso y cobertura del suelo de la región mediante la clasificación supervisada de una imagen del satélite Sentinel 2A en Google Earth Engine a través del algoritmo Random Forest los índices NDVI, NDBI, MNDWI y RS. La clasificación presentó una precisión excelente, con valores del índice Kappa del 99,7 % y una precisión general del 99,78 %. Los potreros ocuparon el 48,08% del área de estudio, con 196.416,71 hectáreas y al sustituir los pastos por caña de azúcar, el área presenta un potencial de producción entre 1.079.741.938,32 y 1.214.709.680,61 litros de etanol. Por lo tanto, el área muestra un enorme potencial para la producción de etanol, siendo atractiva para inversiones en el sector, favoreciendo la economía y la población de los municipios de la región.

Palabras clave: Mapeo del uso y cobertura del suelo; Google Earth Engine; Biocombustibles.

1. Introdução

Os principais combustíveis que compõem a matriz energética mundial são os combustíveis fósseis, obtidos do carvão mineral, gás natural e derivados do petróleo. Ao longo das últimas décadas o uso desses combustíveis tem aumentado a preocupação global devido à alta emissão de gases poluentes e a degradação ambiental gerada na extração dos mesmos. Além disso, existe uma preocupação mundial com relação às mudanças climáticas causadas pela interferência humana no meio ambiente, como o aumento de temperatura, que em grande parte é ocasionado pelo efeito estufa, e o aumento expressivo da temperatura global, causados majoritariamente pelo uso de combustíveis fósseis como carvão mineral, petróleo e o gás natural. Nesse contexto, estão sendo discutidas e estudadas diferentes maneiras para contribuir com a redução da emissão de gases poluentes (The Economist, 2021).

Com o enrijecimento das políticas de proteção ambiental visando reduzir as emissões de gases poluentes nas últimas décadas tem ocorrido um aumento no incentivo, bem como na busca e desenvolvimento por combustíveis menos poluentes. Exemplo desta preocupação é a Conferência das Partes da Convenção-Quadro das Nações Unidas de Mudanças Climáticas realizada em 2021 e que reuniu líderes mundiais com o objetivo de discutir problemas globais atuais, com enfoque na diminuição em 45% das emissões globais de CO₂ até 2030. A partir deste encontro, estima-se que os biocombustíveis receberão mais incentivos e investimentos para substituir o petróleo como principal combustível nas frotas de veículos no mundo (The Economist, 2021).

De acordo com a Embrapa (2020), o etanol produzido a partir do processo de destilação do caldo de cana-de-açúcar é uma alternativa promissora dentro do cenário ambiental mundial, uma vez que é uma fonte renovável e em termos de emissão de gases poluentes, quando comparado à gasolina, se torna um excelente combustível devido sua baixa emissão. A substituição do combustível obtido a partir do refino do óleo bruto por um biocombustível e o uso de veículos movidos a etanol permite reduzir 73% das emissões de CO₂ na atmosfera, quando comparada à emissão dos veículos a combustão (Grofalo et al., 2020).

Historicamente, a cana de açúcar sempre esteve presente na agricultura do Brasil e tem papel relevante na economia brasileira. Inicialmente a cana de açúcar era cultivada para a produção de açúcar e o país se destacou entre os maiores produtores de açúcar para exportação, promovendo assim, a expansão de fazendas por todo o país (Grofalo et al., 2020). Durante a década de 30, já havia produção do etanol no Brasil, entretanto, sua participação na matriz energética brasileira era pouco expressiva, haja vista que o etanol anidro produzido era destinado para o uso como aditivo da gasolina.

Com a crise do petróleo na década de 70, os países começaram a investir em novas tecnologias para o desenvolvimento de combustíveis alternativos. No Brasil, o governo passou a incentivar a produção de etanol bem como a produção de veículos movidos a etanol, com o objetivo de diminuir a dependência do petróleo importado, que representava aproximadamente 80% do

petróleo consumido no país. Desse modo, em 1978 surgiram no mercado os primeiros automóveis movidos 100% a etanol e durante o programa Proálcool houve uma mudança na frota nacional. Em 1979, 0,45% dos veículos eram movidos a etanol, e em 1986, a participação do etanol hidratado como combustível principal passou a ser de 76,1% (Kohlhepp, 2010).

A estabilidade do preço do barril do petróleo no mercado a partir de 1990 levou a um aumento do uso de motores a combustão na frota nacional, entretanto, a partir de 2003 com o advento da tecnologia *Flex Fuel Vehicles* nos veículos nacionais houve a necessidade de se manter alta a produção do etanol (Costa, 2019).

Atualmente a produção de açúcar e etanol continuam em expansão no Brasil, de modo que o país é o maior produtor mundial de cana-de-açúcar, produzindo aproximadamente 578,73 milhões de toneladas destinados à produção de 18,69 bilhões de litros de etanol hidratado, 11,19 bilhões de litros de etanol anidro e 34,94 milhões de toneladas de açúcar na safra 2021/22 (Ministério da agricultura, pecuária e abastecimento, 2022). O Estado de São Paulo lidera a produção no país, sendo responsável pela produção de 298,5 milhões de toneladas destinados à produção de 11,94 bilhões de litros de etanol total, ou seja, anidro e hidratado, e 21,41 milhões de toneladas de açúcar, representando 61% da produção nacional, na safra 2021/22 (CONAB, 2022).

As empresas sucroalcooleiras estão expandindo ou migrando para regiões mais interioranas do estado de São Paulo devido a incentivos fiscais para criação de empregos e industrialização de regiões majoritariamente agrárias. Devido à crescente industrialização dos principais polos sucroalcooleiros como Campinas e Ribeirão Preto, a indústria do etanol passou a realizar uma expansão para o oeste do estado (Andrade et al., 2009).

Historicamente, o Pontal do Paranapanema possui uma característica predominante de atividades de agropecuária de corte, pois na década de 40 ocorreu uma expansão da indústria de agropecuária na região, aumentando assim o desmatamento. De modo que a vegetação nativa foi substituída por fazendas de corte. Cabe ressaltar que, durante as primeiras décadas de expansão agrícola nos municípios houve uma expansão das regiões de cultura, principalmente de algodão, mandioca e amendoim (Bento, 2020).

Entre os anos de 2002 e 2012 a paisagem da região do Pontal do Paranapanema sofreu mudanças expressivas em virtude da disseminação das áreas canavieiras. As incorporações e implantação de novas agroindústrias, bem como o crescimento das culturas da cana-de-açúcar foram os principais responsáveis por estas modificações. Nos municípios onde as terras eram destinadas para pastagens/pecuária ou para outras lavouras, as mesmas passaram a ser utilizadas para a produção da cultura canavieira (Barreto & Thomaz Jr, 2012). Atualmente, a região tem presenciado uma expansão da ocupação do solo por plantações de cana-de-açúcar, uma vez que o bioma do cerrado é um ambiente adaptável ao plantio, favorecendo assim a migração de empresas sucroalcooleiras para a região (Bento, 2020).

Desse modo, o presente trabalho tem como objetivo estimar a produção de etanol de cana-de-açúcar para uma pequena região do Pontal do Paranapanema a partir de áreas de pastagens que poderiam ser substituídas por cultura de cana-de-açúcar. Para tanto, será produzido um mapa de uso e cobertura do solo da região a partir da classificação supervisionada de uma imagem do satélite Sentinel 2A utilizando o *Google Earth Engine* (GEE) através do algoritmo *Random Forest* (RF) e dos índices NDVI, NDBI, MNDWI e RS.

2. Metodologia

2.1 Área de Estudo

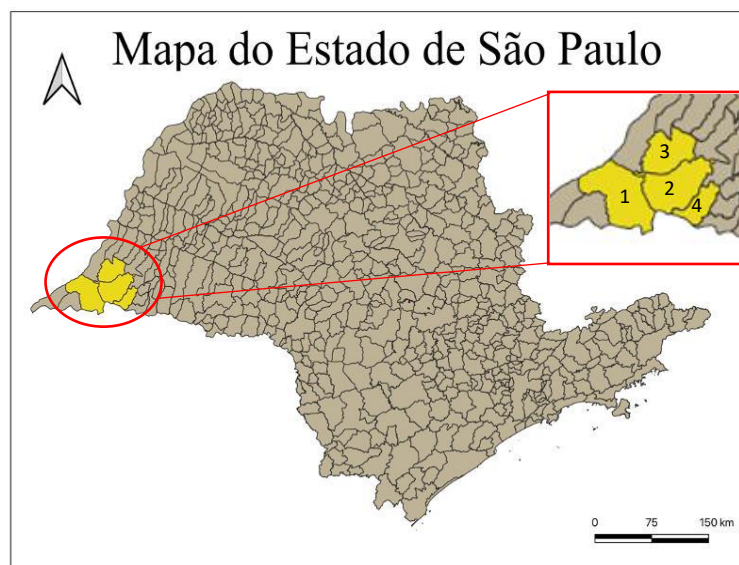
A metodologia abordada no presente trabalho tem caráter exploratório - descritivo bem como quantitativo, e consiste em realizar um estudo de caso dos municípios de Teodoro Sampaio, Mirante do Paranapanema, Sandovalina e Marabá Paulista, que pertencem à região do Pontal do Paranapanema e possuem usinas sucroalcooleiras, a fim de estimar a produção de etanol de cana-de-açúcar para as regiões atualmente destinadas a pastagens.

A escolha desta área para o presente estudo deve-se ao fato de que tal região está sofrendo uma expansão canavieira intensa bem como o interesse das indústrias da região em reduzir as distâncias e transporte da matéria-prima para as fábricas (Rezende, 2011). Essa expansão profunda e intensificada pela rapidez da adesão de investidores na região do Oeste Paulista trouxeram a necessidade de um plantio em grande escala que é atraído por diversos fatores. Dentre eles pode-se citar: relevo propício à mecanização da produção; tipo de solo propício à produção de uma cana-de-açúcar; o valor do metro quadrado da região é baixo quando comparado com outras regiões do estado; a logística favorável que a região possui, com rodovias (SP-270/Raposo Tavares), ferrovias e hidrovias (o potencial navegável de ambos os rios Paraná e Paranapanema).

Além dos aspectos físicos, algumas medidas políticas incentivam, ainda que indiretamente, a expansão da produção de cana de açúcar no Pontal, inclusive medidas legislativas como a Lei Estadual 11.241 / 02 que proíbe as queimadas nos canaviais, levando à mecanização, favorecendo a região a adotar a expansão e ser tornar alvo de investimentos (Barreto e Thomaz Jr, 2012).

A região selecionada para a avaliação compreende quatro municípios da macrorregião do Pontal do Paranapanema no Oeste Paulista. Os municípios escolhidos foram Teodoro Sampaio, Marabá Paulista, Mirante do Paranapanema e Sandovalina, cuja população é de 46.956 habitantes (IBGE, 2020). Uma vez que os quatro municípios já possuem usinas sucroalcooleiras instaladas, há um interesse em avaliar o potencial dessas regiões. A Figura 1 apresenta a localização dos municípios selecionados.

Figura 1. Mapa do Estado de São Paulo com destaque para os municípios analisados, sendo (1) Teodoro Sampaio, (2) Mirante do Paranapanema, (3) Marabá Paulista, (4) Sandovalina.



Fonte: Adaptado pelos autores.

De acordo com a classificação Köppen e Geiger, a região do Pontal do Paranapanema é classificada como clima tropical com estação seca (Aw), com temperatura média anual de 21,9 °C e média de precipitação de 1.370mm durante o ano. A vegetação predominante é do tipo Cerrado com resquícios de Mata Atlântica e zonas de transição entre elas (Peel et al., 2007).

A data utilizada para a aquisição das imagens foi 26 de dezembro de 2020, haja vista que neste período a região encontra-se entre a safra e entressafra do cultivo de cana-de-açúcar e foi a única data que atendeu a exigência máxima de 10% de nuvens estipulada para o trabalho.

2.2 Indústrias na Região

A Tabela 1 apresenta as quatro usinas de açúcar e etanol localizadas na região do Pontal do Paranapanema, bem como os municípios onde estão instaladas e os grupos investidores.

Tabela 1. Usinas dos municípios que compõem a área de estudos.

Nome da Usina	Município	Grupo Investidor Responsável
Alcídia	Teodoro Sampaio – SP	Atvos (Antiga Odebrecht)
Conquista do Pontal	Mirante do Paranapanema -SP	Atvos (Antiga Odebrecht)
Paranapanema II	Sandovalina – SP	Umoe BioEnergy
Usina Decasa	Marabá Paulista – SP	Grupo Olival Tenório

Fonte: Elaborado pelos autores.

A Destilaria Alcídia S/A, localizada em Teodoro Sampaio – SP e a Usina Conquista do Pontal S/A, localizada no município de Mirante do Paranapanema – SP, estão sob a responsabilidade da empresa Atvos pertencente ao Grupo Novonor (antigo Grupo Odebrecht Agroindustrial). A Destilaria Alcídia S/A foi inaugurada durante o período do Governo Militar do Presidente Ernesto Geisel, sendo fruto da primeira fase do Programa Proálcool com início da produção do etanol no ano de 1978 e a Usina Conquista do Pontal S/A foi fundada em 2004 e pertence ao grupo de usinas que vieram para suprir a expansão recente de etanol devido à popularização dos veículos *Flex Fuel*. Ambas têm enfoque na produção de açúcar VHP (açúcar *Very High Polarization*), energia elétrica e etanol de cana hidratado e possuem capacidade máxima de moagem de 4 milhões de toneladas de cana por ano (Atvos, 2021).

A Destilaria Paranapanema II, localizada no município de Sandovalina – SP, sob direção da Umoe BioEnergy pertencente ao Grupo norueguês Umoe AS foi fundada em 2006 visando suprir a demanda de produção de etanol hidratado e anidro. A Umoe possui 100% de colheita mecanizada em todas as suas áreas cultiváveis que compreendem 44 mil hectares de área. A planta de Sandovalina possui uma capacidade de moagem de 2,8 milhões de toneladas de cana por ano (Umoe Bioenergy, 2021).

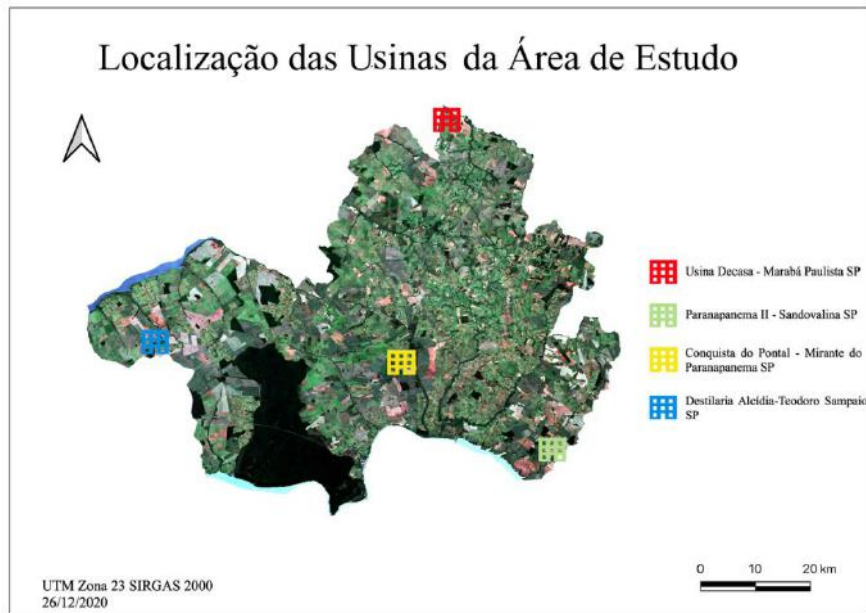
A Destilaria Decasa Açúcar e Álcool, localizada no município de Marabá Paulista, pertencente atualmente ao Grupo Olival Tenório, foi inaugurada em 1982 durante os investimentos do Projeto Governamental do Proálcool sob controle do Grupo Jacinto e Associados. A planta possui capacidade de moagem de 1,6 milhões de toneladas de cana por ano (NovaCana, 2021). A Figura 2 apresenta a localização das usinas acima citadas.

2.3 Softwares Utilizados

Para o processamento das imagens do satélite Sentinel 2A foram utilizados dois softwares: o *Google Earth Engine* (GEE) para o processamento dos dados e o Quantum GIS (QGIS) para o tratamento dos dados. Desse modo, a principal ferramenta utilizada para o processamento das imagens foi o *Google Earth Engine* e o *software* QGIS foi utilizado para a elaboração dos mapas temáticos e geração das análises numéricas.

O *Google Earth Engine* (GEE) é uma plataforma em nuvem desenvolvida em 2016 pela Google com a finalidade de auxiliar em processos de análises ambientais em grande escala que permite a realização de atividades de geoprocessamento através da seção *Google Earth Engine Code Editor*. Tal ferramenta permite o uso de algoritmos em *JavaScript* ou em *Python* para a implementação de análises geoespaciais em um ambiente em nuvem, facilitando assim processamentos em larga escala uma vez que não é necessário efetuar *download* das imagens de satélite. O GEE disponibiliza, em seu catálogo o acesso direto a imagens dos sensores Sentinel 1 e 2; Landsat 5, 7 e 8; Aster, Modis e NOAA (Google Inc, 2021).

Figura 2. Mapa com a localização das usinas.



Fonte: Elaborado pelos autores.

O Quantum GIS (QGIS) é um *software* gratuito desenvolvido pela *QGIS Development Team* que é um SIG (Sistema de Informação Geográfica) projetado pela *Open Source Geospatial Foundation* (OSGeo). Tal ferramenta permite realizar análises de dados espaciais com um sistema em código aberto programado em C++ e *Python* bem como edição e análise de dados georreferenciados. O QGIS tem como característica um amplo desenvolvimento de uma arquitetura de *plug-ins* que podem ser baixados para executar análises específicas bem como a liberdade de escolha de *download* de sensores presentes no mercado, com a necessidade de fazer o *download* das bandas para efetuar o processamento (Qgis Development Team, 2021).

2.4 Imagens Sentinel 2

O programa Copernicus foi desenvolvido pela Agência Espacial Européia (ESA – *European Spatial Agency*) em parceria com a Organização Europeia para Exploração de Satélites Meteorológicos (EUMETSAT - *European Organisation for the Exploitation of Meteorological Satellites*), Centro Europeu de Previsões Meteorológicas de Médio Prazo, Agências da UE e Mercator Océan, sendo estabelecido em 2014 pela regulamentação (EU) N° 377/2014.

O programa foi desenvolvido com o objetivo de atingir uma capacidade de observação global da Terra com alta qualidade, em larga escala e de forma precisa. O acesso às imagens é gratuito e permite que sejam utilizadas para estudos de mitigação ambiental, impactos de mudanças climáticas, estudos no campo da geociência abordando os campos de monitoramento atmosférico, estudo do uso e cobertura do solo, ecossistema marinho, mudanças climáticas, manejo de situações emergenciais e segurança para organizações europeias e não europeias.

O programa Copernicus conta com 2 satélites, o Sentinel 2A e o 2B na mesma órbita, localizados a uma altitude de 786 km, cobrindo 290 km de largura e com uma diferença de fase de 180°. Devido à geometria de operação dos satélites é possível uma revisão da coleta de imagens a cada 5 dias (ESA, 2021).

O satélite Sentinel 2 desempenha a função da aquisição sistemática de imagens de alta resolução espacial com 10m, 20m a 60m de densidade de pixel. Dispõe de um sensor MSI (*MultiSpectral Instrument*) com 13 bandas multiespectrais dispostas entre os espectros visível (VIS), infravermelho de ondas curtas (SWIR) e infravermelho próximo (NIR) (ESA, 2021).

No presente trabalho foram utilizadas imagens provindas do satélite Sentinel 2A por possuir uma resolução espacial de 10m x 10m no pixel, considerada uma boa resolução, além de ser disponibilizada gratuitamente (<https://earthexplorer.usgs.gov/>).

A Tabela 2 apresenta as características das bandas presentes no Sentinel 2A. É importante destacar que o presente trabalho utiliza de imagens providas de satélites e não de veículos aero tripulados (drones), pois esses veículos possuem uma abrangência de área bem menor quando comparado com as imagens providas de sensores. Devido a essa baixa abrangência de área o uso de drones se torna inviável para análises de áreas extensas.

Tabela 2. Características das bandas do Sentinel 2A.

Banda	Nome da Banda	Comprimento de onda central (nm)	Resolução Espacial (m)
B01	Coastal Aerosol	443	60
B02	Blue	490	10
B03	Green	560	10
B04	Red	665	10
B05	Vegetation Red Edge	705	20
B06	Vegetation Red Edge	740	20
B07	Vegetation Red Edge	783	20
B08	Near Infrared	842	10
B08A	Near Infrared Narrow	865	20
B09	Water Vapour	945	60
B10	Shortwave Infrared Cirrus	1380	60
B11	Shortwave Infrared	1610	20
B12	Shortwave Infrared	2190	20

Fonte: Adaptada de ESA (2021).

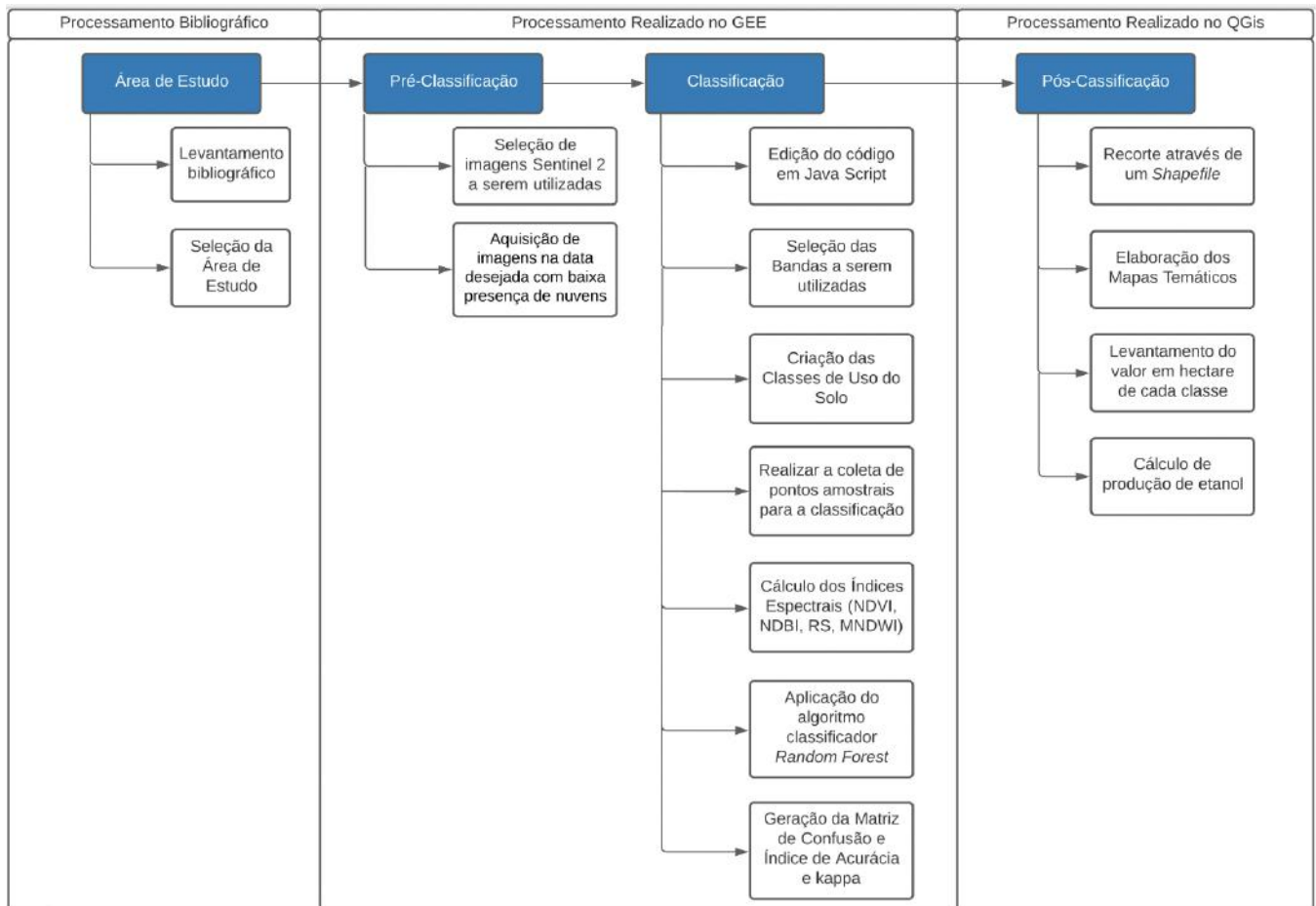
2.5 Método utilizado

Conforme mencionado, os resultados obtidos no presente trabalho foram gerados através da combinação do uso de duas ferramentas, o GEE e o QGis. Os pontos favoráveis desta combinação foram: alto poder de processamento do GEE, o fato de não haver necessidade de *download* das imagens e de uma máquina de alto desempenho, pois os processamentos são realizados em uma nuvem, bem como a facilidade do uso das ferramentas e o tempo de processamento reduzido.

A utilização do GEE e do QGis para produzir os mapas e os resultados numéricos permite manter a qualidade do processamento, bem como reduzir o tempo de desenvolvimento. De acordo com Cho et al. (2021), Vale (2019) e Alencar et al. (2020), o desempenho da ferramenta da Google para a classificação de uso e cobertura do solo se mostrou eficiente e apresentou resultados excelentes de classificação. Reforçando assim, que a metodologia definida para o processo de classificação é eficiente.

O processo metodológico desenvolvido no presente trabalho é descrito pelo fluxograma da Figura 3.

Figura 3. Fluxograma de atividades realizadas na metodologia.



Fonte: Elaborado pelos autores.

2.5.1 Pré-Processamento dos Dados

Na etapa de pré-processamento foi definida a área de interesse a partir de um polígono desenhado na imagem e, na sequência, é selecionado o satélite a partir do qual serão extraídas as imagens multiespectrais. Foram utilizadas imagens do satélite Sentinel 2A uma vez que possuem uma resolução espacial de 10m, considerada adequada para o objetivo do trabalho.

Em seguida, as imagens foram selecionadas compreendendo o período de 01/12/2020 a 31/12/2020 e foi definida a porcentagem máxima de nuvem em 10%. Optou-se pelo mês de dezembro em virtude da baixa presença de áreas que foram aradas.

A aquisição das imagens foi feita através da aba principal na plataforma em nuvem na aba Code Editor do GEE no link <https://code.earthengine.google.com/>.

2.5.2 Processamento

Após o pré-processamento foram selecionadas as bandas, sendo elas: B2, B3, B4, B5, B6, B7, B8, B11, B12 e após a escolha das bandas foram definidas as seis classes a serem mapeadas. A Tabela 3 descreve cada uma das classes.

Após a definição das classes foi realizada a coleta das amostras de cada uma delas, necessárias para a etapa da classificação supervisionada. As amostras foram adquiridas a partir de polígonos referenciando as informações contidas naquele pixel ou grupo de pixels que representa um tipo específico de cobertura terrestre (Lillesand & Kiefer, 1995). A Figura 4 apresenta a coleta de pontos amostrais realizada no GEE.

Tabela 3. Classes do mapa de uso e cobertura do solo.

Classe	Definição
Corpo d'água	Área de águas superficiais, onde se destacam rios, lagos, lagoas e oceanos.
Mata Nativa	Área de vegetação densa, não alterada, compondo a vegetação nativa da região.
Área Urbana	Área com alta interferência humana de impermeabilização de solo.
Cultura	Áreas com presença de vegetação, desde monocultura ou cultura por ciclos, vegetação de porte médio e com simetria de espaçamento entre uma planta e outra.
Pastagem	Áreas com presença de vegetação rasteira ou regiões que ocorreu a retirada da vegetação original para o desenvolvimento da atividade pastoril.
Solo Exposto	Área onde não há a presença de vegetação.

Fonte: Elaborado pelos autores.

Figura 4. Recorte da área de estudo com os polígonos da coleta de pontos amostrais.



Fonte: Elaborado pelos autores.

2.5.3 Índices Espectrais Utilizados

Após a criação dos polígonos amostrais de cada uma das classes foram calculados os índices espectrais selecionados para compor e melhorar o desempenho da classificação. Os índices utilizados foram NDVI (Índice de vegetação por diferença normalizada), NDBI (Índice de Construção por Diferença Normalizada), MNDWI (Índice da diferença normalizada da água modificado) e a RS (Razão simples). Os índices promovem uma maior acurácia, desse modo suas respostas serão utilizadas pelo algoritmo no momento da separação dos pixels de cada classe.

NDVI

O Índice de Vegetação por Diferença Normalizada (*NDVI - Normalized Difference Vegetation Index*) foi desenvolvido por Rouse et al. (1974) e aprimorado posteriormente por Campbell e Wynne (2011) e tem a finalidade de avaliar a densidade da cobertura vegetal levando em consideração a resposta espectral da vegetação com relação a absorção e reflectância dentro do

espectro. Esse índice analisa a condição da vegetação, avaliando a intensidade da presença de clorofila associada ao vigor vegetativo na região. O NDVI é bastante utilizado em averiguação de desmatamento, monitoramento de safras, expansão ou encolhimento de áreas vegetativas e é calculado através da diferença entre a reflectância das bandas do infravermelho próximo (ρ_{nir}) e do vermelho (ρ_{red}), dividido pela soma das duas reflectâncias utilizadas (Equação 1).

$$NDVI = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \quad (1)$$

No Sentinel 2, a banda do infravermelho próximo (ρ_{nir}) é equivalente à banda 08 e a banda do vermelho (ρ_{red}) equivale a banda 04. A formulação resulta em uma resposta numérica entre -1 e 1. Quanto mais próximo de 1 maior o vigor vegetativo no local representado pelo pixel e valores mais próximos de -1 compreendem áreas com uma menor ou nenhuma incidência de matéria composta por clorofila.

Razão Simples

O Índice de Vegetação da Razão Simples (*RS – Simple Ratio*) desenvolvido inicialmente por Birth e McVey (1968) é um índice físico que permite avaliar, identificar e realçar, através da análise de dados de reflectância, regiões com a presença de vegetação e exibe boa relação com características das florestas. A RS se assemelha muito com a análise do NDVI, pois também leva em consideração a resposta de absorção da vegetação na faixa do visível. Entretanto, a análise RS é uma avaliação mais simplificada. A RS é um índice calculado pela divisão da banda vermelha (ρ_{red}), representada pela banda 04 no Sentinel 2, pela banda do infravermelho próximo (ρ_{nir}), representada pela banda 08 no Sentinel 2. A razão simples é calculada pela Equação 2:

$$RS = \frac{\rho_{red}}{\rho_{nir}} \quad (2)$$

MNDWI

O índice MNDWI (*Modified Normalized Difference Water Index*) é fundamentado pela modificação do índice de diferença normalizada de água (NDWI – *Normalized Difference Water Index*). Tal índice foi proposto por McFeeters (1996) sendo desenvolvido para aplicação de estudos ambientais a fim de identificar cursos d'água, visando maximizar a reflectância de áreas com a presença de água. Esse índice tem como principais aplicações a avaliação e manejo ambiental de recursos hídricos, realizando estudos de cursos de rios, monitoramento de áreas inundadas e gestão de recursos hídricos. O NDWI é dado pela relação matemática entre as bandas verde (ρ_{green}) e a banda infravermelha próxima (ρ_{nir}), no Sentinel 2 a banda verde (ρ_{green}) representado pela banda 03 e a infravermelha próxima (ρ_{nir}) representado pela banda 08. O NDWI é calculado pela Equação 3:

$$NDWI = \frac{\rho_{green} - \rho_{nir}}{\rho_{green} + \rho_{nir}} \quad (3)$$

O MNDWI foi desenvolvido para suprir a deficiência do índice NDWI proposto por Xu (2006), uma vez que este apresenta limitações devido a ruídos provenientes de áreas construídas que interferem na resposta espectral dos pixels. Desse modo, o MNDWI tem uma aplicação no monitoramento de corpos hídricos que possam ter interferências devido a áreas impermeabilizadas. O cálculo do MNDWI difere do NDWI, uma vez que considera a relação entre a banda verde (ρ_{green}) e a infravermelho médio (ρ_{Swir1}), sendo determinado a partir da Equação 4:

$$MNDWI = \frac{\rho_{green} - \rho_{Swir1}}{\rho_{green} + \rho_{Swir1}} \quad (4)$$

No Sentinel 2 o infravermelho médio (ρ_{Swir1}) equivale à banda 11 e a verde (ρ_{green}) equivale à banda 03.

NDBI

O Índice de Construção por Diferença Normalizada (NDBI - *Normalized Difference Built-up Index*) tem como base a abordagem dos estudos de Zha et al. (2003). O NDBI possui como principal finalidade identificar áreas construídas e impermeabilizadas através da resposta que o pixel fornece durante a análise, pois solos impermeabilizados de áreas urbanas tem a predominância de materiais que absorvem a irradiância solar (asfalto, concreto, metal, entre outros), diferindo assim, da resposta de áreas com presença de vegetação. O NDBI é determinado através de equações aritméticas da diferença entre a reflectância das bandas do infravermelho médio (ρ_{Swir1}) e do infravermelho próximo (ρ_{nir}) e, dividido pela soma das duas reflectâncias utilizadas, conforme a Equação 5:

$$NDBI = \frac{\rho_{Swir1} - \rho_{nir}}{\rho_{Swir1} + \rho_{nir}} \quad (5)$$

No Sentinel 2, o infravermelho médio (ρ_{Swir1}) equivale a banda 11 e o infravermelho próximo (ρ_{nir}) é equivalente a banda 08. O resultado expresso pelo índice produz uma imagem com valores de -1 a 1. Espera-se que o valor mais alto indique os pixels incluídos na área construída e o valor negativo indica os pixels presentes na área não construída.

2.5.4 Aplicação do Algoritmo Random Forest

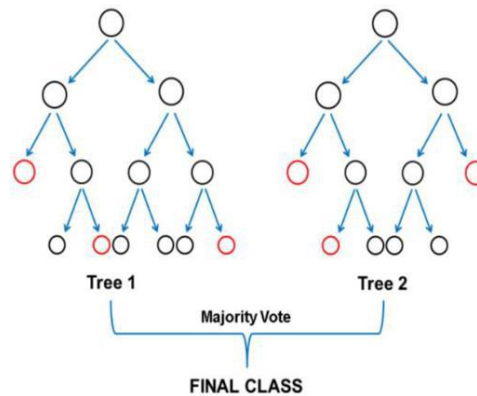
O código fonte desenvolvido em *JavaScript* foi disponibilizado por Vale (2019) e alterado para atender aos objetivos do presente trabalho. O código foi desenvolvido com o algoritmo Random Forest e foi utilizado no GEE para a classificação supervisionada da área de estudos.

O Random Forest (RF) é um algoritmo de classificação supervisionada, baseado no conjunto de árvores de decisão gerados de uma forma randômica a partir do treinamento com as classes amostrais coletadas no processo de classificação. Constituído pela reamostragem aleatória do treinamento inicial, o RF funciona pela classificação de um vetor com uma característica específica, submetendo o vetor à árvore da floresta de decisão, sendo o algoritmo que toma uma decisão do vetor com a melhor resposta através de votação. Na Figura 5 é apresentado o fluxograma de operação do *Random Forest* na tomada de decisão (Vale, 2019).

Como algoritmo para classificação, o *Random Forest* tem se mostrado uma excelente ferramenta quando comparado a outros modelos devido à rapidez na etapa de treinamento e à minimização do efeito de ajustamento do modelo apenas para os dados de treinamento. Uma desvantagem do RF é a necessidade de uma enorme quantidade de dados de treinamento, o que gera possíveis lacunas de dados e uma possível lentidão no processamento devido a necessidade da utilização de diversas árvores de decisão para se chegar em um resultado mais coeso com o esperado (Breiman, 2001).

Utilizando-se o GEE esta desvantagem do RF é minimizada uma vez que o processamento é realizado em nuvem. A utilização desse algoritmo no script alinhado com os índices espectrais (NDVI, NDBI, MNDWI, entre outros) torna ainda mais eficiente a classificação, trazendo uma acurácia maior para o processamento e dados mais condizentes com a verdade terrestre da região (Belgiu & Dragut, 2016; Schott, 2019).

Figura 5. Fluxograma de operação do *Random Forest*.



Fonte: Vale, (2019).

Segundo Cho et al. (2021), dentre diversos classificadores, esse algoritmo se destaca na detecção de vegetação e possui um desempenho considerado excelente para ser utilizado como ferramenta de classificação, tendo como vantagens um bom resultado de acurácia e desempenho ágil de processamento quando combinado a uma ferramenta como o GEE, mostrando resultados consistentes com a verdade terrestre da região estudada.

Após as etapas de seleção de área e definição de índices foi realizada a aplicação do algoritmo e o treinamento do classificador foi realizado com 70% dos pixels selecionados na criação de classes amostrais. Na sequência foram calculados os índices para verificação da qualidade da classificação.

2.5.5 Geração do Índice Kappa, Acurácia Global e Matriz de Confusão

A veracidade e desempenho da classificação das imagens e do mapa do uso e cobertura do solo produzido foram avaliados a partir dos índices Kappa e a Acurácia Global. A Matriz de Confusão descreve as classes amostrais e a quantidade de pixels que podem interferir em uma boa resposta devido à confusão gerada por pixels semelhantes.

Matriz de Confusão

A Matriz de Confusão ou Matriz de Erro é uma das melhores técnicas para avaliar a acurácia de uma classificação. A matriz de confusão (Tabela 4) é constituída pela formação de matriz quadrada contendo números arranjados em linhas e colunas, que apresentam o número de pixels atribuídos a uma classe específica em relação à verdadeira classe à qual o pixel pertence. As linhas representam as classes definidas através do processo de classificação e as colunas os dados de referência. Os números de pixels da diagonal representam os pixels que foram classificados de forma correta (Congalton & Green, 2008).

Tabela 4. Exemplificação da Matriz de Confusão.

	Mata Nativa	Área urbana	Cultura	Pastagem	Solo exposto
Mata Nativa	X ₁₁	X ₁₂	X ₁₃	X ₁₄	X ₁₅
Área urbana	X ₂₁	X ₂₂	X ₂₃	X ₂₄	X ₂₅
Cultura	X ₃₁	X ₃₂	X ₃₃	X ₃₄	X ₃₅
Pastagem	X ₄₁	X ₄₂	X ₄₃	X ₄₄	X ₄₅
Solo exposto	X ₅₁	X ₅₂	X ₅₃	X ₅₄	X ₅₅

Fonte: Adaptada de Congalton e Green (2008).

Ademais, as informações marginais das linhas e colunas permite realizar uma distinção entre erros de omissão e de inclusão. Erros de omissão são aqueles onde os pixels pertencentes a uma classe de interesse são alocados a outras classes e os erros de inclusão ocorrem quando os pixels de outras classes são associados à classe de interesse (Congalton & Green, 2008).

A partir da matriz de confusão pode-se avaliar quais classes amostrais apresentaram mais confusão e obter algumas medidas descritivas, tais como Acurácia Global e estimativa do Índice Kappa.

Índice Kappa

O Índice Kappa é desenvolvido como medida de avaliação estatística da relação de exatidão entre os dados de referência e os dados de classificados, fornecendo um valor numérico da diferença entre esses dados. Esse índice é indicado para avaliar a legitimidade de uma classificação, avaliando as interações realizadas e utiliza como base os dados da matriz de confusão para estimar a exatidão do processo. Proposto por Cohen (1960) a validação do cálculo desse índice pode ser expressa pela Equação 6:

$$K = \frac{\sum_{i=1}^k n_{ii} - \sum_{i=1}^k n_{i+} n_{+i}}{n^2 - \sum_{i=1}^k n_{i+} n_{+i}} \quad (6)$$

Na qual o k refere-se o número de linhas e colunas presentes na matriz, n_{ii} é o número de observações na coluna i e linha i, n se refere ao número total de observações e $n_{i+} n_{+i}$ são os valores totais da linha e coluna i (Cohen, 1960).

De acordo com Landis e Koch (1997) o valor de resposta do Índice Kappa é atribuído à qualidade da classificação variando de Péssima a Excelente, sendo tabelada por valores atribuindo uma associação qualitativa de processo. A Tabela 5 apresenta a divisão e classificação.

Tabela 5. Associação de Qualidade de Classificação com Valores de Kappa.

Valor do Índice Kappa	Desempenho
0,00	Péssima
0,01 a 0,20	Ruim
0,21 a 0,40	Razoável
0,41 a 0,60	Boa
0,61 a 0,80	Muito Boa
0,81 a 1,00	Excelente

Fonte: Adaptada de Landis e Koch (1997).

Acurácia Global

A Acurácia Global (AG) ou Exatidão Global é uma medida de análise de confiabilidade de classificação, sendo uma medida simples relacionada aos valores presentes na diagonal principal da matriz de confusão e relacionada com o total de pontos amostrados na classificação.

De acordo com Congalton e Green (2008), a Acurácia Global (AG) ou Exatidão Global é calculada através da divisão da somatória da diagonal principal da matriz de confusão n_{ii} , pelo número de amostras totais coletadas como descrita pela Equação 7:

$$AG = \sum_{i=1}^K \frac{n_{ii}}{n} \quad (7)$$

A acurácia pode ser dividida em dois tipos específicos para a avaliação dos resultados, a Acurácia do Usuário (AU) e a Acurácia do Produtor (AP). A AU se refere à divisão entre o número de pixels classificados corretamente em cada classe e o total de pixels classificados na classe (total da linha). A AP é a divisão do número de pixels corretamente classificados em cada classe pelo número de pixels do conjunto de dados de treinamento usado nas classes (total da coluna) (Congalton & Green, 2008). As Equações 8 e 9 são utilizadas para determinar as duas acurácias:

$$AP = \frac{n_{ii}}{n_{+j}} \quad (8)$$

$$AU = \frac{n_{ii}}{n_{j+}} \quad (9)$$

Após a determinação dos índices foi realizado o tratamento dos dados no QGIS a fim de obter a espacialização dos resultados obtidos no GEE.

2.5.6 Tratamento dos dados no QGIS

Após a realização do processamento utilizando o GEE, os resultados foram exportados para o Google Drive e baixados para a elaboração dos mapas temáticos no software QGIS. O objetivo dessa etapa é apresentar os resultados obtidos na classificação supervisionada no GEE na forma de um mapa a fim de melhorar a visualização dos resultados.

Após baixar o arquivo Geotiff, gerado no GEE, foi aplicada uma máscara shapefile dos municípios escolhidos para desenvolver o trabalho. Através desta máscara é feito o recorte da área e a elaboração do mapa temático para uma melhor visualização dos resultados.

Após a elaboração do mapa de uso e ocupação do solo foi realizada a estimativa das áreas de cada uma das classes e, a partir das áreas estimadas, foi calculada a produção de etanol.

2.5.7 Estimativa de Área das Classes e Cálculo da Produção do Etanol

Realizado o recorte da área aplicando o shapefile, dentro das ferramentas do QGIS utilizou-se a ferramenta de processamento (processing r.report) para gerar um relatório estatístico de área de modo a levantar os valores em hectare de cada classe presente na classificação. Em seguida, os dados são disponibilizados e os resultados obtidos são inseridos em uma tabela.

A partir dos valores em hectare para cada uma das classes é possível realizar a estimativa do potencial de produção de biocombustível. A estimativa tem como objetivo avaliar o cenário de substituição das áreas de pastagem da região por cana-de-açúcar. Devido a expansão crescente na região do Oeste Paulista, esse cenário de substituição vem se tornando uma realidade.

De acordo com a CONAB (2020) a produtividade média nacional para a safra 2020/21 foi de 76,35 ton/ha, ou seja, um hectare de plantação de cana equivale a 76,35 toneladas de cana-de-açúcar.

De posse dessa informação é possível estimar o valor de produção, em tonelada de cana, substituindo toda a área de pastagem da região pela cultura. O valor encontrado na classificação foi aplicado nesse cálculo para assim dar continuidade aos cálculos de estimativa de produção.

Mundialmente, o etanol é majoritariamente produzido a partir do processo de fermentação. De modo geral, a fermentação alcoólica é um processo químico de transformação de açúcares fermentescíveis (glicose e frutose) em uma cadeia alcoólica como o etanol através da ação de leveduras que facilitam o processo fermentativo. Esse processo pode ocorrer como

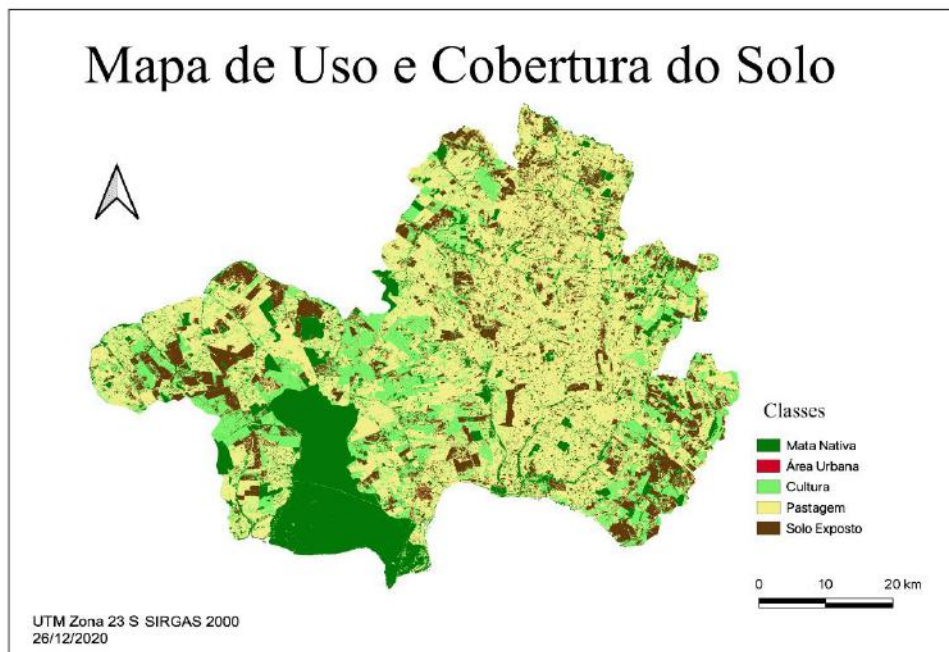
uma reação anaeróbica, sem a presença de oxigênio. A primeira etapa é o processo de glicólise, onde fosfatos são incorporados pelas moléculas de glicose facilitando a sua quebra em 2 moléculas de ácido pirúvico. A segunda etapa consiste na fermentação alcoólica realizada na presença das leveduras, nessa reação o ácido pirúvico (C3) é descarboxilado gerando o acetaldeído. Através do NADH ocorre a redução do acetaldeído, gerando etanol (C₂H₆O) e dióxido de carbono (CO₂) (Góes-Favoni et al., 2018).

3. Resultados e Discussão

A partir do desenvolvimento metodológico foi possível gerar uma representação visual do uso e cobertura do solo da área estudada juntamente com a construção de um cenário de estimativa do potencial de produção de etanol da região estudada.

A Figura 6 apresenta o mapa de uso e cobertura do solo da área de estudos resultante da classificação supervisionada produzida no GEE, através do algoritmo *Random Forest* e das bandas B2, B3, B4, B5, B6, B7, B8, B11, B12 do satélite Sentinel 2 para a data de 26 de dezembro de 2020. Na legenda são apresentadas as classes que compõem o mapa.

Figura 6. Mapa de uso e cobertura do solo.



Fonte: Elaborado pelos autores.

A partir do mapa gerado verifica-se uma maior concentração das áreas de cultura na porção sul e à leste da área de estudos. De acordo com Thomaz Jr. (2009), esse fato se deve à expansão da monocultura de cana-de-açúcar na região em função da implantação de usinas canavieiras em vários municípios, intensificando o cultivo.

Observa-se que a região possui grande presença de áreas de pastagem. Como descrito por Bento (2020), a região possui uma indústria agropecuária forte, mas que vem sendo substituída pelo plantio de cana devido a sua maior rentabilidade e incentivos municipais. Diante disso, futuramente a região pode experimentar o aumento da classe de cultura em seu território.

A fim de validar o mapa gerado pela classificação das imagens Sentinel 2 foi produzida a matriz de confusão (Tabela 6). A matriz de confusão mostra em sua diagonal o número de pixels de verdade terrestre corretamente rotulados pelo classificador e, fora da diagonal, o número de pixels que foi classificado erroneamente para cada classe.

Tabela 6. Matriz de Confusão.

	Mata Nativa	Área Urbana	Cultura	Pastagem	Solo Exposto	Total	Erro de Inclusão (%)
Mata Nativa	74.880	0	6	1	0	74.887	0,01
Área Urbana	0	629	3	8	105	745	15,57
Cultura	3	1	27.832	259	22	28.117	1,01
Pastagem	3	2	140	29.520	14	29.679	0,54
Solo Exposto	0	14	13	4	28.021	28.052	0,11
Total	74.886	646	27.994	29.792	28.162	161.480	
Erro de Omissão (%)	0,01	2,63	0,58	0,91	0,50		

Fonte: Elaborado pelos autores.

Analisando a matriz de confusão verifica-se que as classes que apresentaram maior confusão foram as de Cultura e Pastagem. Essa confusão se dá pela proximidade da resposta espectral dessas classes entre si, mas os valores de confusão representam uma interferência baixa no restante da classificação, representando 0,24% de pixels na classificação.

As classes de Área Urbana e Solo Exposto também foram as classes que apresentaram maior confusão, também em função das respostas espectrais serem semelhantes, essa confusão gera uma baixíssima interferência no resultado, representando 0,07% dos pixels na classificação. As demais confusões apresentaram valores bem inferiores aos apresentados por essas classes.

Com base nos valores que estão fora da diagonal principal da matriz de confusão foram determinados os erros de omissão e inclusão. Os erros de omissão foram de: 0,01%, 15,57%, 1,01%, 0,54% e 0,1%, respectivamente para as classes Mata Nativa, Área Urbana, Cultura, Pastagem e Solo Exposto. Os erros de inclusão encontrados foram: 0,01%, 2,63%, 0,58%, 0,91% e 0,5% para as classes Mata Nativa, Área Urbana, Cultura, Pastagem e Solo Exposto, respectivamente.

Tanto o erro de omissão quanto o de inclusão foram bastante baixos, com exceção da área urbana que apresentou um erro de inclusão de 15,57%, em função de parte dos pixels da classe de solo exposto (105 pixels) terem sido classificados como área urbana.

A classificação supervisionada mostrou resultados bastante satisfatórios, ou seja, condizentes com a verdade terrestre da região, o que pode ser verificado através do índice Kappa e da Acurácia Global, determinados com base na matriz de confusão. Ambos têm o objetivo de representar o grau de confiabilidade dos dados classificados, de modo que quanto mais próximo de 1 ou 100%, maior concordância os dados classificados têm com os dados de referência.

Os valores de Índice de Acurácia Global e Índice Kappa foram de 99,78% e 99,70%, respectivamente (Tabela 7). O valor do Índice Kappa, segundo Landis e Koch (1997), foi considerado “excelente”.

Em virtude dos valores altos obtidos é possível afirmar que a GEE atuou como um excelente classificador. O índice Kappa com 99,7% de precisão demonstra que houve poucas confusões de pixel e a AG mostrou valores considerados excelentes de classificação. A partir dos dados da matriz de confusão foram determinados os valores de Acurácia do Produtor e Acurácia do Usuário, apresentados na Tabela 8.

Tabela 7. Índices Kappa e Acurácia Global.

Índice	
Acurácia	0,9977905647154292
Kappa	0,9969595919883522

Fonte: Elaborado pelos autores.

Tabela 8. Valores de Acurácia do Produtor (ferramenta).

Acurácia Produtor	
Mata Nativa	0,9999065258322539
Área Urbana	0,8442953020134228
Cultura	0,989863784761888
Pastagem	0,9946426766400486
Solo Exposto	0,9988593020354329

Fonte: Elaborado pelos autores.

Os valores de acurácia para cada classe, com exceção da área urbana, foram acima de 0,97. A classe Mata Nativa foi a mais bem identificada, com valores de acurácia iguais a 0,999, tanto para a acurácia do produtor quanto do usuário. A acurácia do produtor demonstra a veracidade da resposta da ferramenta, neste caso o GEE. Segundo os dados apresentados, a classificação atingiu valores de alta excelência e a classe de área urbana é a única que se encontra inferior a 90%. Esse valor de 84,43% deve-se, muito provavelmente, à confusão de pixels presente na matriz. Essa confusão se deve à presença de cidades de pequeno porte na região do Pontal, o que impossibilita a coleta alta de pontos amostrais e a presença de estradas não pavimentadas que possuem resposta espectral semelhante ao solo exposto. Pequenas cidades/áreas urbanas tendem a apresentar um valor menor de acurácia devido a ser uma classe muito mista (tendo a presença desde vegetação até áreas de solo exposto em uma área pequena).

A Tabela 9 apresenta os valores de Acurácia do Usuário, que apresentou valores elevados de resposta, legitimando uma excelente classificação. Os valores elevados de acurácia devem-se à boa coleta de pontos amostrais feita pelo operador.

Tabela 9. Valores de Acurácia do Usuário.

Acurácia Produtor	
Mata Nativa	0,9999198782148866
Área Urbana	0,9736842105263158
Cultura	0,9942130456526399
Pastagem	0,9908700322234156
Solo Exposto	0,9949932533200767

Fonte: Elaborado pelos autores.

Após a análise da qualidade do mapeamento, na etapa de pós-processamento, foi realizada no QGis a determinação das áreas, em hectare, de cada uma das classes do mapa de uso e cobertura do solo (Tabela 10).

Tabela 10. Valores em Hectare Encontrados.

Classes	Valor Área (hectare)	Valor Área (%)
Mata Nativa	68.015,19	16,65
Área Urbana	696,71	0,17
Cultura	77.227,65	18,90
Pastagem	196.416,71	48,08
Solo Exposto	66.160,98	16,20
Total	408.517,24	100

Fonte: Elaborado pelos autores.

Ao quantificar as áreas em função do tipo de uso e cobertura (Tabela 9), verifica-se que a classe pastagem ocupou 48,08% da área de estudo. A classe cultura foi a segunda maior em extensão territorial e, na sequência, as áreas de mata nativa e solo exposto.

A partir dos valores de áreas calculados foi possível realizar uma análise da estimativa da substituição da pastagem por plantações de cana-de-açúcar com a finalidade da produção do etanol. De posse do valor encontrado de 196.416,71 hectares de área destinada a pastagem e das estimativas de eficiência e perdas propostas pela literatura, uma vez que 1 hectare de plantio de cana-de-açúcar possui potencial de produção de 76,35 toneladas de cana (CONAB, 2020) tem-se que a região apresenta um potencial de produção de 14.996.415,81 de toneladas de cana de açúcar, podendo variar de acordo com a produtividade da área plantada.

No processo de produção de etanol ocorrem diversas perdas inerentes ao processo, e tais perdas devem levadas em consideração para os cálculos de estimação dos Açúcares Redutores Totais Fermentescíveis, ART (Delgado et al. 2019; Raízen Energia, 2021). Atualmente as usinas utilizam tecnologias e metodologias modernas, o que permite que as usinas mais bem gerenciadas operem com perdas industriais totais entre 8% a 10% (RPAnews Cana&Indústria, 2020).

Considerando o cenário onde as perdas correspondem a 10%, tem-se que serão perdidas 1.499.641,58 toneladas durante o processo de produção. Desse modo, serão utilizadas para a produção do etanol 13.496.774,23 toneladas.

Atualmente uma tonelada de cana produz entre 80 e 90 litros de etanol (Delgado, 2019), de modo que a área de estudo composta por 4 municípios do Pontal do Paranapanema possui um potencial futuro de produção que varia entre 1.079.741.938,32 e 1.214.709.680,61 litros de etanol para ser utilizado no setor de transportes do país ao se substituir as áreas de pastagens determinadas no mapa de uso e cobertura do solo por cultura de cana-de-açúcar.

4. Conclusão

A partir dos resultados obtidos com a classificação de uso e cobertura do solo no Pontal do Paranapanema verifica-se que a região possui um enorme potencial futuro de produção de biocombustíveis, podendo atrair investimentos de indústrias interessadas na expansão sucroalcooleira na região do Oeste Paulista. Embora o potencial agrícola da região ainda seja pouco explorado, a área possui solo e relevo propícios para a expansão de lavouras.

Com relação à ferramenta utilizada, o *Google Earth Engine* se mostrou uma excelente alternativa, apresentando como principais vantagens a eficiência de processos, a agilidade de processamento e a execução de todas as análises na nuvem, evitando assim a necessidade de *download* de bancos de dados de imagens pesados. Tais vantagens contribuíram para a obtenção de bons resultados, como os valores do índice Kappa e da Acurácia Global que apresentaram valores de precisão de 99,78% e 99,70%, respectivamente. De acordo com a literatura, tais valores legitimam a classificação como excelente, comprovando que essa

ferramenta vinculada com a metodologia aplicada é bastante eficaz para geração de resultados coerentes com a verdade terrestre. Essa eficiência está atrelada ao uso do algoritmo *Random Forest* que, alinhado com os índices espectrais, trouxe uma boa resposta de classificação.

A partir do estudo realizado estima-se que a região tem um potencial de produção entre 1.079.741.938,32 e 1.214.709.680,61 litros de etanol. Em virtude do elevado potencial e dos incentivos mundiais para a utilização de biocombustíveis, nos próximos anos a região pode sofrer mudanças na utilização do solo, pois a substituição de terras destinadas à pecuária por lavouras vem se tornando algo recorrente nos últimos anos, principalmente e em função do maior retorno financeiro que o arrendamento para produção de cana-de-açúcar oferece para os proprietários de terra.

De modo geral verifica-se que a região do Pontal do Paranapanema analisada apresenta potencial econômico viável para a produção de etanol. A região encontra-se pouco explorada e o presente estudo apresenta resultados positivos de que a região é atrativa para investimentos no setor, favorecendo assim a economia e a população da região.

Apesar do etanol ser uma fonte de energia renovável, sustentável e limpa, cabe ressaltar que a expansão do cultivo da cana-de-açúcar pode causar danos ao meio ambiente, dentre os quais, a contaminação de rios, lagos e solos em virtude do uso excessivo de fertilizantes e agrotóxicos e da aplicação não controlada de vinhaça ou torta de filtro, bem como, a perda na qualidade do solo, alterando assim, a fotossíntese das plantas e a biodiversidade.

Referências

- Alencar, A., Shimbo, J. Z., Lenti, F., Marques, C. B., Zimbres, B., Rosa, M., Arruda, V., Castro, I., Ribeiro, J. P. F. M., Varella, V., Alencar, I., Piontekowski, V., Ribeiro, V., Bustamante, M. C., Sano, E. E., & Barroso, M. (2020). Mapping Three Decades of Changes in the Brazilian Savanna Native Vegetation Using Landsat Data Processed in the Google Earth Engine Platform. *Mdpi - Multidisciplinary Digital Publishing Institute*, Basel, Switzerland, 12(6), 1-23.
- Andrade, E. T., Carvalho, S. R. G., & Souza, L. F. (2009). Programa do proálcool e o etanol no Brasil. *Engevista*, 11(2), 127-136.
- Atvos (Brasil) (2021). Grupo Novonor (emp.). Sobre a Empresa.
- Barreto, M. J., & Thomaz Jr, A. (2012). O cenário do agronegócio canavieiro na região do Pontal do Paranapanema-SP. In: XXI *Encontro Nacional de Geografia Agrária*, UFU – MG.
- Belgiu, M., & Drăgut, L. (2016). Random Forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, 24-31.
- Bento, F. S. (2020). Os impactos da expansão do agronegócio canavieiro para as relações campo-cidade no Pontal do Paranapanema (SP). *Revista Geografia em Atos* (Geo Atos online). 03(18), 181-205. 10.35416/geoatos.v3i18.7219.
- Breiman, L. (2001). *Random Forests*. *Machine Learning*, 45, 05-32.
- Campbell, J.B., & Wynne, R.H. (2011). *Introduction to Remote Sensing*. (4a ed.).
- Cho, D. F., Schwaída, S. F., Cicerelli, R. E., Almeida, T., Ramos, A. P. M., & Sano, E. E. (2021). Desempenho do Algoritmo de Classificação de Imagens Random Forest para Mapeamento do Uso e Cobertura do Solo no Cerrado Brasileiro. *Anuário do Instituto de Geociências*, Rio de Janeiro, 44.
- Cohen, J. A. (1960). Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(01), 37-46.
- CONAB - Companhia Nacional de Abastecimento (Brasil). (2020). *Acompanhamento da safra brasileira de cana-de-açúcar – Safra 2020-21*, Brasília, DF, v. 7, n. 1, agosto 2020.
- CONAB - Companhia Nacional de Abastecimento (Brasil). (2022). *Acompanhamento da safra brasileira de cana-de-açúcar – Safra 2022-23*, Brasília, DF, v. 9, n. 1, abril 2022.
- Congalton, R. G., & Green, K. (2008). *Assessing the accuracy of remotely sensed data: principles and practices*. CRC press.
- Costa, J. A. (2019). *A indústria sucroalcooleira: uma análise da competitividade do setor no período 2003 a 2018*. Monografia (TCC) - Curso de Ciências Econômicas, Universidade Federal do Maranhão, São Luís.
- Delgado, A. A., Cesar, M. A. A., & Silva, F. C. (2019). *Elementos de Tecnologia e Engenharia da Produção do Açúcar, Etanol e Energia*. Fealq. 984 p.
- ESA – European Spatial Agency (Paris França). (2021). *Satellite Constellation*. https://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2/Satellite_constellation.
- Goés-Favoni, S., Monteiro, A. C.C., Dorta, C., Crippa, M. G., & Shigematsu, E. (2018). Fermentação alcoólica na produção de etanol e os fatores determinantes do rendimento. *Revista Ibero-Americana de Ciências Ambientais*, 9(4), 285-296.

Google Inc. (2021). *Earth engine data*. <https://developers.google.com/earth-engine/datasets>

Grofolo, D. F. T., Packer, A. P. C., Ramos, N. P., Kondo, V. Y., Folegatti, M. I. S., & Cabral, O. M. R. (2020). Dinâmica do cultivo de cana-de-açúcar no Brasil – 1990 a 2018. *Empresa Brasileira de Pesquisa Agropecuária Embrapa Meio Ambiente e Ministério da Agricultura, Pecuária e Abastecimento*.

IBGE - Instituto Brasileiro de Geografia e Estatística - Rio de Janeiro - RJ. (2021). Instituto Nacional de Estatística. Pesquisa do Censo 2020. <https://www.ibge.gov.br/ibge/est%C3%A9stica/popula%C3%A7%C3%A3o/censo2020>.

Kohlhepp, G. (2010). Análise da situação da produção de etanol e biodiesel no Brasil. *Estudos Avançados*, 24, 223-253. <https://doi.org/10.1590/S0103-40142010000100017>.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(01), 159-174.

Lillesand, T. M., & Kiefer, R. W. (1995). *Remote sensing, and image interpretation*. New York: John Wiley e Sons Inc., 708p.

Ministério da agricultura, pecuária e abastecimento. (2022). *Produção brasileira de cana-de-açúcar, açúcar e etanol*. https://www.gov.br/agricultura/pt-br/assuntos/sustentabilidade/agroenergia/arquivos-producao/001PRODUOBRASILEIRADECANADEACARACAREETANOL_11052022.pdf.

NovaCana. (2021). *Mapa das usinas de etanol e açúcar do Brasil*. https://www.novacana.com/usinas_brasil/mapa.

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.*, 11, 1633-1644.

Qgis Development Team (Estados Unidos das Américas). (2021). *Open-Source Geospatial Foundation*. *Quantum GIS*. <https://www.qgis.org/en/site/>.

Raízen Energia (2021). *Relatório Anual 2020/2021*. <https://www.raizen.com.br/relatorioanual/2021/pdf/raizen-rs2021-pt.pdf>.

RPAnews Cana&Indústria. (2020). *Como identificar, gerenciar e reduzir perdas industriais*. <https://revistarpanews.com.br/como-identificar-gerenciar-e-reduzir-perdas-industriais/>.

Rezende, W. S. (2011). *O setor sucroalcooleiro no pontal do paranapanema: uma análise a partir do licenciamento ambiental*. 128 f. TCC (Graduação) - Curso de Engenharia Ambiental, Faculdade de Ciências e Tecnologia, Universidade Estadual Paulista Júlio de Mesquita Filho - Unesp, Presidente Prudente.

Rouse, J. H. (1974). Monitoring Vegetation Systems in the Great Plains with ERTS. In: Proceedings, *Third Earth Resources Technology Satellite-1 Symposium, Greenbelt*, 3010-3017.

Schott, M. (2019). Random Forest Algorithm for Machine Learning: Part 4 of a Series on Introductory Machine Learning Algorithms.

The Economist. (2021). Was COP26 in Glasgow a success? The Economist. Glasgow Scotland, 109-110.

Thomaz Júnior, A. (2009). *Dinâmica Geográfica do Trabalho no Século XXI*. (Limites Explicativos, Autocrítica e Desafios Teóricos). p.941, Livro Docência - Universidade Estadual Paulista. Presidente Prudente.

Vale, J. R. B. (2019). *Análise da dinâmica do uso e cobertura da terra nas áreas desflorestadas do estado do Pará por meio da plataforma google earth engine*. 69 f. TCC (Graduação) - Curso de Engenharia Cartográfica e de Agrimensura, Universidade Federal Rural da Amazônia, Belém – PA.

Umoe Bioenergy (Brasil). (2021). Umoe As. Sobre a Empresa.

Xu, H. (2006). Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14), 3025-3033.

Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3), 583-594.