

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO ESTADUAL DE PESQUISAS EM SENSORIAMENTO REMOTO E METEOROLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM SENSORIAMENTO REMOTO

BIANCA MATOS DE BARROS

**APRENDIZADO DE MÁQUINA APLICADO AO MAPEAMENTO DE POLUIÇÃO
PLÁSTICA MARINHA EM IMAGENS DO SENSOR MSI/SENTINEL-2**

PORTO ALEGRE

2023

BIANCA MATOS DE BARROS

**APRENDIZADO DE MÁQUINA APLICADO AO MAPEAMENTO DE POLUIÇÃO
PLÁSTICA MARINHA EM IMAGENS DO SENSOR MSI/SENTINEL-2**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Sensoriamento Remoto como requisito parcial para a obtenção do título de mestre em Sensoriamento Remoto e Geoprocessamento.

Orientador: Prof^a. Dr^a. Sílvia Beatriz Alves Rolim

Coorientador: Prof. Dr. Cristiano Lima Hackmann

PORTO ALEGRE

2023

CIP - Catalogação na Publicação

Barros, Bianca Matos de
Aprendizado de máquina aplicado ao mapeamento de
poluição plástica marinha em imagens do sensor
MSI/Sentinel-2 / Bianca Matos de Barros. -- 2023.
135 f.
Orientador: Sílvia Beatriz Alves Rolim.

Coorientador: Cristiano Lima Hackmann.

Dissertação (Mestrado) -- Universidade Federal do
Rio Grande do Sul, Centro Estadual de Pesquisas em
Sensoriamento Remoto e Meteorologia, Programa de
Pós-Graduação em Sensoriamento Remoto, Porto Alegre,
BR-RS, 2023.

1. Sensoriamento remoto. 2. Detritos marinhos. 3.
Classificação de imagens. 4. Redes neurais
artificiais. 5. Modelos de transferência radiativa. I.
Rolin, Sílvia Beatriz Alves, orient. II. Hackmann,
Cristiano Lima, coorient. III. Título.

Ao meu noivo Rafael, obrigada pelo amor de todos os dias. Aos meus pais, Rejane e Sidinei e a minha irmã Clara, obrigada por me desejaram sempre o melhor. A Deus, obrigada pela vida, pela fé e pela força para lutar.

AGRADECIMENTOS

À Universidade Federal do Rio Grande do Sul, por me oportunizar o desenvolvimento desta formação e a atuação como servidora pública.

Aos meus orientadores Sílvia e Cristiano, pela confiança e direcionamento desde o início deste trabalho e ao longo de toda a formação.

Aos colegas do meu grupo de pesquisa, Douglas e Mariana, sem a participação deles este trabalho não seria possível.

Aos meus colegas de trabalho, André, Carlos e Guilherme, pelo apoio e compreensão que me possibilitaram compatibilizar as minhas atividades de trabalho e formação acadêmica.

[...] entre os pobres mais abandonados e maltratados, conta-se a nossa terra oprimida e devastada [...]. Esquecemo-nos de que nós mesmos somos terra.

Papa Francisco, 2012
Carta Encíclica *Laudato Si*,
sobre o cuidado da Casa Comum

RESUMO

O lixo marinho causa um amplo espectro de impactos e representa uma ameaça crescente para o meio ambiente marinho e costeiro. O sensoriamento remoto pode atuar de forma integrada com técnicas tradicionais como operações de limpeza locais e pesquisas com redes de arrasto, oferecendo como vantagens a extensa área de cobertura e a observação frequente. Contudo, a pouca quantidade disponível de imagens com presença de poluição plástica confirmada, necessárias para validação de modelos, constitui uma limitação. Para sanar esta dificuldade, uma alternativa é o uso de modelos de transferência radiativa (MTRs) para geração de dados. Neste trabalho, utilizamos dados simulados por um MTR junto a dados adquiridos pelo sensor remoto Instrumento Multiespectral (MSI, do inglês *MultiSpectral Instrument*) da missão Sentinel-2 em aplicações com algoritmos de aprendizado de máquina, visando estudar o comportamento espectral da poluição marinha por plásticos e avaliar a aplicabilidade do MTR para esta área de pesquisa. Os resultados da classificação não supervisionada com o algoritmo KMeans demonstraram que o comportamento espectral dos poluentes é influenciado por fatores como tipo de polímero e percentual de cobertura do pixel. Os resultados da classificação supervisionada com uma Rede Neural Artificial (RNA) mostraram que o classificador treinado com dados sintéticos pode ser aplicável a dados reais em determinadas condições. Foram realizados 4 ciclos de 100 repetições de treino com dados simulados e teste com dados reais, usando 4 conjuntos de atributos de entrada diferentes. O melhor desempenho foi obtido com o conjunto de atributos composto exclusivamente pelos dados das bandas do sensor remoto, enquanto os demais conjuntos também incluíam dados de índices radiométricos. Com este conjunto de atributos, o modelo conseguiu atingir uma acurácia geral média de 80%, com média de 0,85 para pontuação F1 balanceada (ponderada pelo número de instâncias positivas de cada classe, visando atenuar o efeito do desbalanceamento).

Palavras-chave: Detritos marinhos. Sensoriamento remoto. Modelos de transferência radiativa. Classificação de imagens. Redes neurais artificiais.

ABSTRACT

Marine litter causes many impacts and is a growing threat to the marine and coastal environment. Remote sensing can work with traditional techniques such as local cleaning operations and trawl net surveys, offering advantages like extensive coverage and frequent observation. However, the small amount of available images with confirmed plastic pollution required for model validation constitutes a limitation. An alternative to healing this difficulty is using Radiative Transfer Models (RMTs) for data generation. In this work, we used RTM simulated data and Multispectral Instrument (MSI) data, from Sentinel-2 mission, in applications with machine learning algorithms, aiming to study the spectral behavior of marine plastic pollution and to evaluate the applicability of the MTR for this research area. The results of unsupervised classification with the KMeans algorithm evidenced that the spectral behavior of pollutants is influenced by factors such as polymer type and percentage of pixel coverage. The results of supervised classification with an Artificial Neural Network (ANN) evidenced that the classifier trained with synthetic data can apply to real data under certain conditions. Four cycles of 100 repetitions of training with simulated data and testing with real data were performed, using four sets of different input attributes. The best performance was reached using the set of attributes composed of remote sensor data, while the other sets also included data from radiometric indexes. With this set of attributes, the model was able to reach an average overall accuracy of 80%, with an average of 0.85 for a balanced F1-score (weighted by the number of positive instances of each class, aiming to mitigate the effect of imbalance).

Keywords: Marine debris. Remote sensing. Radiative transfer models. Image classification. Artificial neural networks.

LISTA DE ILUSTRAÇÕES

Figura 1 –	Trabalhos contendo as palavras “remote sensing” e “floating plastic” na base de dados Scopus publicados até 30 de junho de 2021.	22
Figura 2 –	Ilha Henderson, localizada no giro oceânico do Pacífico Sul, em 2015.	25
Figura 3 –	Representação dos giros oceânicos e das principais correntes superficiais.	26
Figura 4 –	Sistema desenvolvido pela organização The Ocean Cleanup para retirada de lixo da GPGP.	27
Figura 5 –	Riscos e impactos diretos do lixo marinho e dos plásticos. . . .	28
Figura 6 –	Comparação entre as bandas dos sensores a bordo dos satélites Landsat e Sentinel-2.	34
Figura 7 –	Assinaturas espectrais médias coletadas de alvos plásticos artificiais implantados no mar.	37
Figura 8 –	Coeficiente de absorção da água e sua influência na reflec- tância espectral.	38
Figura 9 –	Combinação dos índices FDI e NDVI para diferentes classes de cobertura da superfície em imagens com suspeita de poluição em águas naturais.	39
Figura 10 –	Assinaturas espectrais de polímeros coletados na natureza (A, B e C) e virgens (D).	40
Figura 11 –	Resoluções espaciais utilizadas em trabalhos de sensoria- mento remoto com aprendizado profundo.	46
Figura 12 –	Matriz de confusão para a classificação de um conjunto de dados contendo três classes	49
Figura 13 –	Espaço ROC com três classificadores.	52

Figura 14 –	Problemas direto (setas pretas) e inverso (setas amarelas) de sensoriamento remoto.	53
Figura 15 –	Cena DART contendo água, areia e plástico.	55
Figura 16 –	Matriz de células DART para o sistema Terra/Atmosfera.	56
Figura 17 –	Metodologia utilizada.	60
Figura 18 –	Alvos artificiais implantados na Grécia durante os PLPs 2019 (superior) e 2021 (inferior) em fotografias aéreas adquiridas por VANT.	65
Figura 19 –	Assinaturas espectrais do μ -NAPO (composto pela média de um conjunto de microplásticos coletados no Oceano Pacífico) e dos polímeros virgens PVC, PA 6.6, LDPE, PET e PP.	69
Figura 20 –	Quartis dos conjuntos de dados DART (simulados) e USGS (observados) após reamostragem.	74
Figura 21 –	Distribuição de classes nos conjunto de dados DART (simulados) e USGS (observados).	84
Figura 22 –	Distribuição de polímeros nos pixels contendo plástico nos conjuntos de dados DART (simulados) e USGS (observados).	84
Figura 23 –	Correlação entre atributos nos dados DART (simulados) e USGS (observados).	86
Figura 24 –	Fontes e datas de aquisição dos dados USGS (observados).	89
Figura 25 –	Média e desvio padrão nos conjuntos de dados DART (simulado) e USGS (observado).	91
Figura 26 –	Quartis nos conjuntos de dados DART (simulado) e USGS (observado)	92
Figura 27 –	Quartis das diferentes classes nos conjuntos de dados DART (simulado) e USGS (observado).	93
Figura 28 –	Assinaturas espectrais médias de cada classe por conjunto de dados e por percentual de cobertura.	94

Figura 29 –	Valores de FDI e NDVI agrupados por classe de cobertura da superfície nos conjuntos de dados DART (simulado) e USGS (observado).	95
Figura 30 –	Dispersão de valores de FDI e NDVI nos conjuntos de dados DART (simulado) e USGS (observado).	96
Figura 31 –	Box plot do plástico sobre a areia e a água na banda NIR1 nos dados DART (simulados) e USGS (observados), agrupado por percentual de cobertura.	97
Figura 32 –	Box plot do plástico sobre a areia e a água na banda RedEdge1 nos dados DART (simulados) e USGS (observados), agrupado por percentual de cobertura.	97
Figura 33 –	Dispersão dos valores de PI agrupados por classe de cobertura da superfície nos conjuntos de dados DART (simulado) e USGS (observado).	98
Figura 34 –	Importâncias de atributos para dados DART (simulados) na classificação com o algoritmo Random Forest.	99
Figura 35 –	Agrupamentos do algoritmo K-means para o conjunto de atributos B.	102
Figura 36 –	Distribuição de polímeros entre os agrupamentos do algoritmo K-means para o conjunto de atributos B.	103
Figura 37 –	Assinaturas espectrais médias por classe nos conjuntos de dados DART (simulado, parte superior) e USGS (observado, parte inferior).	104
Figura 38 –	Correlação de atributos para a classe madeira nos dados USGS (observados).	123
Figura 39 –	Correlação de atributos para a classe água nos dados DART (simulados) e USGS (observados).	124
Figura 40 –	Correlação de atributos para a classe plástico nos dados DART (simulados) e USGS (observados).	125

Figura 41 – Correlação de atributos para a classe areia nos dados DART (simulados) e para a classe costa nos dados USGS (observados).	126
Figura 42 – Gráfico de dispersão entre a banda Blue e o índice FDI nos dados DART (simulados) e USGS (observados).	127
Figura 43 – Gráfico de dispersão entre os índices AWEI e FDI nos dados DART (simulados) e USGS (observados).	128
Figura 44 – Gráfico de dispersão entre os índices AWEI e MNDWI nos dados DART (simulados) e USGS (observados).	128
Figura 45 – Gráfico de dispersão entre os índices MNDWI e NDWI nos dados DART (simulados) e USGS (observados).	129
Figura 46 – Gráfico de dispersão entre as bandas Blue e NIR2 nos dados DART (simulados) e USGS (observados).	130

LISTA DE TABELAS

Tabela 1 –	Análise da situação das imagens de 2018 e 2019.	66
Tabela 2 –	Análise da situação das imagens de 2021.	67
Tabela 3 –	Polímeros de interesse, siglas, características físicas e utilidades comuns.	70
Tabela 4 –	Bandas do MSI/Sentinel-2.	72
Tabela 5 –	Descrição dos conjuntos de dados simulados e observados. .	83
Tabela 6 –	Subconjuntos de atributos selecionados para classificação. .	99
Tabela 7 –	Médias das métricas de avaliação obtidas para os dados de teste (conjunto de imagens MSI/Sentinel-2 observadas). . .	106
Tabela 8 –	Erros e acertos do modelo agrupados por classe do pixel e ano de aquisição da imagem.	109
Tabela 9 –	Acurárias médias em amostras de plástico agrupados por percentual de cobertura do pixel para cada conjunto de atributos.	110
Tabela 10 –	Agrupamento dos dados DART (simulados) por classificação não supervisionada utilizando os conjuntos de atributos A e B.	131
Tabela 11 –	Agrupamento dos dados DART (simulados) por classificação não supervisionada utilizando os conjuntos de atributos C e D.	132
Tabela 12 –	Agrupamento dos dados USGS (observados) por classificação não supervisionada utilizando os conjuntos de atributos A e B.	133
Tabela 13 –	Agrupamento dos dados USGS (observados) por classificação não supervisionada utilizando os conjuntos de atributos C e D.	134
Tabela 14 –	Matriz de confusão média para o conjunto de atributos A. . .	135
Tabela 15 –	Matriz de confusão média para o conjunto de atributos B. . .	135
Tabela 16 –	Matriz de confusão média para o conjunto de atributos C. . .	135
Tabela 17 –	Matriz de confusão média para o conjunto de atributos D. . .	135

SUMÁRIO

1 INTRODUÇÃO	17
2 OBJETIVOS	21
3 DESENVOLVIMENTO	22
3.1 REFERENCIAL TEÓRICO	22
3.1.1 Poluição marinha por plásticos	23
3.1.2 Sensoriamento remoto e técnicas espetrorradiométricas	31
3.1.3 Experimentos de campo	34
3.1.4 Bancos de dados espectrais	38
3.1.5 Métodos de classificação de imagens	42
3.1.6 Dados para treinamento e validação de modelos	53
3.2 METODOLOGIA	60
3.2.1 Seleção de imagens observadas	61
3.2.2 Tratamento de imagens observadas	66
3.2.3 Seleção de polímeros	68
3.2.4 Simulação das cenas	69
3.2.5 Adição de índices	74
3.2.6 Limpeza de dados	76
3.2.7 Análise exploratória	76
3.2.8 Seleção de atributos	77
3.2.9 Classificação não supervisionada	77
3.2.10 Classificação supervisionada	79
3.3 RESULTADOS E DISCUSSÕES	83
3.3.1 Análise exploratória	83
3.3.2 Seleção de atributos	98
3.3.3 Classificação não supervisionada	100

3.3.4 Classificação supervisionada	105
4 CONCLUSÕES	111
5 FINANCIAMENTO	113
A Apêndice	123
A.1 Análise exploratória	123
A.2 Classificação	130

1 INTRODUÇÃO

Os oceanos são fundamentais para a manutenção da vida. Para a humanidade, eles fornecem alimento, petróleo, gás e meios de transporte, além de atividades culturais e de recreação. Para os ecossistemas em todo o planeta, os oceanos influenciam o clima, a distribuição de chuvas, secas, inundações e o desenvolvimento de tempestades, furacões e tufões (STEWART, 2008).

Atualmente, os ecossistemas marinhos e costeiros estão enfrentando mudanças naturais e antropogênicas, que incluem aquecimento, aumento do nível do mar, acidificação e perda de biodiversidade. A introdução de materiais sintéticos duráveis aumentou dramaticamente a poluição marinha. Estima-se que cerca de 60% a 90% dos detritos marinhos são compostos por diferentes produtos plásticos, de lenta decomposição. A entrada contínua nas águas naturais de grandes quantidades destes materiais resulta em seu acúmulo gradual no ambiente marinho e costeiro (BAKER et al., 2016; MAXIMENKO et al., 2019).

O amplo espectro de problemas gerados pela poluição marinha levou à inclusão deste tema na Agenda 2030 das Nações Unidas para o Desenvolvimento Sustentável. A redução da poluição marinha está presente nos Objetivos de Desenvolvimento Sustentável (ODS) números 6, 12 e 14 (UN, 2023a).

A questão do lixo plástico em águas naturais também é abordada por outras iniciativas internacionais, como o Programa das Nações Unidas para o Meio Ambiente (UN, 2023b), a Parceria Global sobre Poluição Plástica e Lixo Marinho (UN, 2023c), o Grupo de Especialistas sobre os Aspectos Científicos da Proteção do Meio Ambiente Marinho (IMO, 2023) e a Década do Oceano para promoção das ciências oceânicas com foco no desenvolvimento sustentável (UN, 2023d). No Brasil, o enfrentamento ao lixo marinho ainda é incipiente, mas já existem projetos como a Cátedra UNESCO para Sustentabilidade dos Oceanos (USP, 2023), o Plano Nacional de Combate ao Lixo no Mar da Marinha (BRASIL, 2019) e o Plano Estratégico de Monitoramento e Avaliação do Lixo no Mar do Estado de São Paulo (SP, 2021).

Considerando a ampla gama de fontes de lixo marinho, as medidas para resolver o problema precisam ser igualmente abrangentes em sua abordagem. Técnicas tradicionais como operações de limpeza e pesquisas com redes de arrasto são eficazes em quantificar e remover a poluição plástica de forma local, porém fornecem cobertura espacial e temporal limitadas e geralmente são demoradas, caras e até perigosas em alguns casos, colocando trabalhadores em situações inseguras devido a ventos fortes, solos escorregadios, chuva, ou exposição a substâncias químicas e resíduos hospitalares, dentre outros riscos. O sensoriamento remoto pode atuar de forma integrada com estas técnicas, oferecendo como vantagens a extensa área de cobertura e a observação frequente. A observação remota de detritos marinhos inclui o uso de plataformas como satélites, aeronaves e VANTs (Veículos Aéreos Não Tripulados), dentre outras, combinadas com uma grande variedade de sensores e amostradores que possibilitam diversos tipos de análises (MAXIMENKO et al., 2019).

Vários estudos em sensoriamento remoto têm realizado pesquisas na área de poluição marinha. Porém, em geral, são trabalhos recentes que apresentam resultados incipientes, com limitações relacionadas principalmente às resoluções dos sensores. Outra limitação significativa refere-se às abordagens para mitigar a baixa disponibilidade de imagens adquiridas remotamente com presença de plástico confirmada em campo, que possam ser usadas para treinamento e validação dos modelos de classificação de imagens. Para atenuar estas dificuldades, as pesquisas atuais optam por abordagens que incluem usar imagens com suspeita de plástico, sem confirmação *in situ* (BIERMANN et al., 2020), testar a aplicação de índices radiométricos (THEMISTOCLEOUS et al., 2020) e realizar sobrevoos com VANTs e câmeras aerotransportadas durante a passagem dos satélites, obtendo conjuntos de dados de múltiplos sensores simultaneamente (TOPOUZELIS; PAPAKONSTANTINOU; GARABA, 2019; TOPOUZELIS et al., 2020; PAPAGEORGIOU et al., 2022). São abordagens válidas, mas que frequentemente levam a limitações ou incertezas nos resultados obtidos.

Uma outra alternativa para sanar as dificuldades em obter dados de verdade de campo é o uso de modelos de transferência radiativa (MTRs). Estes modelos permitem

a criação de cenas simuladas nas mais variadas configurações, permitindo a testagem de diversos cenários hipotéticos com total controle dos elementos, o que também contribui para mitigar as limitações nas resoluções dos sensores. O uso de MTRs em uma aplicação pode, portanto, reduzir custos e aumentar as possibilidades de combinação de parâmetros, e já é uma realidade no sensoriamento remoto, especialmente para áreas de vegetação e agricultura (JACQUEMOUD et al., 2000; DUTHUIT et al., 2008; ALI et al., 2021).

A hipótese da detecção da poluição marinha por plásticos ser possível através do sensoriamento remoto é a questão de pesquisa central desta dissertação. Esta hipótese foi testada através do estudo do comportamento espectral das superfícies e do uso de dados simulados e observados com algoritmos de aprendizado de máquina.

Neste trabalho, o modelo de Transferência Radiativa Anisotrópica Discreta (DART, do inglês *Discrete Anisotropic Radiative Transfer*) foi utilizado para criar um conjunto de dados contendo cenas simuladas de areia e água sobrepostas por objetos plásticos. O conjunto simulado foi posteriormente utilizado por algoritmos de aprendizado de máquina, paralelamente a um conjunto de imagens observadas de alvos plásticos em praias, adquiridas pelo sensor Instrumento Multiespectral (MSI, do inglês *MultiSpectral Instrument*) da plataforma de satélites Sentinel-2.

Foram selecionados dois classificadores: o K-Means (não supervisionado) e a Rede Neural Artificial (RNA, supervisionada). Na classificação não-supervisionada, primeiro os dados simulados foram submetidos ao processo de agrupamento do algoritmo em diferentes configurações. Posteriormente, as mesmas configurações de classificação foram aplicadas à base de dados composta pelas imagens observadas. Os resultados obtidos em ambos os casos foram comparados, permitindo analisar características e padrões de associação entre elementos, bem como as diferenças de comportamento do algoritmo ao utilizar dados simulados e observados.

Na classificação supervisionada, o classificador foi inicialmente submetido a um processo de otimização de hiperparâmetros. Posteriormente, passou por quatro ciclos de 100 repetições de treino com os dados simulados e teste com os dados observa-

dos, usando um conjunto de atributos de entrada diferente em cada ciclo e extraíndo indicadores de acurácia. A avaliação de desempenho permitiu estimar a eficácia do modelo e avaliar a aplicabilidade do DART para treinamento de modelos supervisionados destinados à detecção de poluição marinha por plásticos em imagens observadas.

A dissertação está organizada nos capítulos Objetivos, Desenvolvimento e Conclusões. No capítulo Objetivos, são listados os objetivos gerais e específicos da dissertação. No capítulo Desenvolvimento, são apresentados o Referencial teórico, onde o atual estado da arte do tema deste trabalho é apresentado; Metodologia, onde o método utilizado nesta pesquisa é explicado em detalhes; e Resultados e discussões, onde os resultados obtidos são demonstrados e debatidos. No capítulo Conclusões, as contribuições da pesquisa são sintetizadas e trabalhos futuros são sugeridos.

2 OBJETIVOS

O objetivo geral do trabalho foi avaliar o potencial e as limitações da detecção de plásticos marinhos por imagens e métodos de sensoriamento remoto. Os objetivos específicos foram:

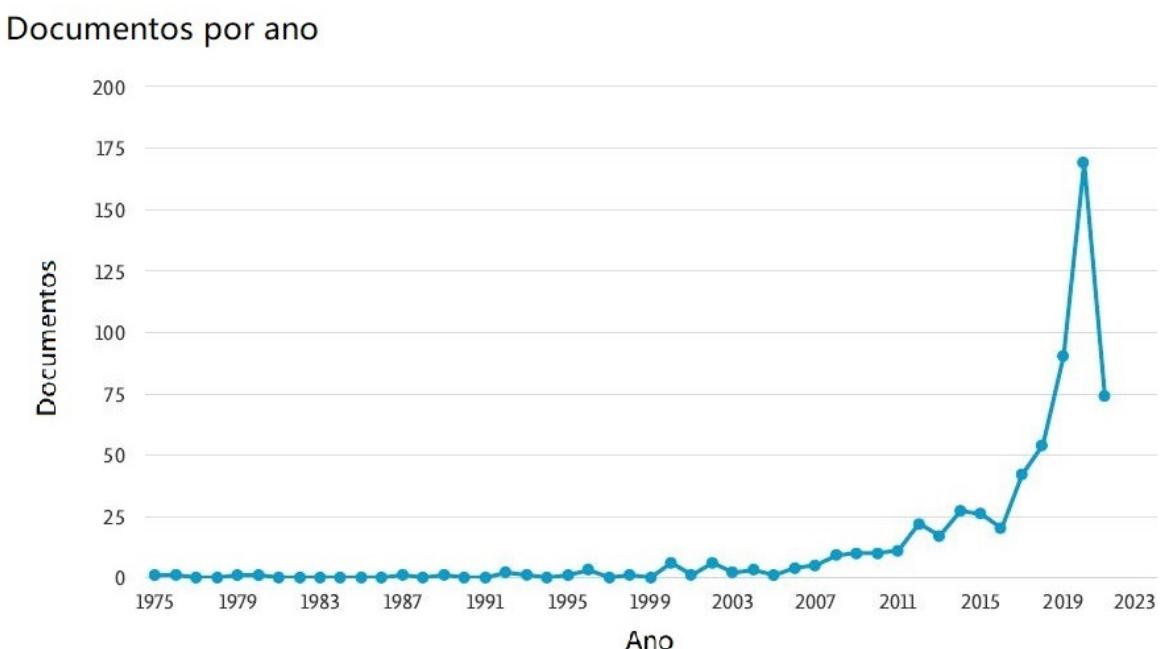
1. Estudar o comportamento espectral da poluição marinha;
2. Avaliar a aplicabilidade dos MTRs na detecção remota de detritos plásticos;
3. Avaliar a aplicabilidade de modelos de aprendizado de máquina treinados com imagens de sensoriamento remoto simuladas a imagens observadas.

3 DESENVOLVIMENTO

3.1 REFERENCIAL TEÓRICO

A aplicação de métodos de sensoriamento remoto para detecção de poluição plástica em águas naturais é relativamente recente. Na base de dados Scopus¹, uma busca por documentos contendo as palavras “remote sensing” e “floating plastic” em seu conteúdo revela uma quantidade relativamente pequena de trabalhos (203 encontrados em pesquisa realizada em 30 de junho de 2021), sendo quase sua totalidade publicada após 2010, conforme pode ser visto na Figura 1.

Figura 1 – Trabalhos contendo as palavras “remote sensing” e “floating plastic” na base de dados Scopus publicados até 30 de junho de 2021.



Fonte: Adaptação de relatório da base de dados Scopus.

Houve no máximo 1 trabalho por ano entre 1995 e 2010. A partir de 2011, o volume de publicações começou a apresentar tendência crescente, que segue até a atualidade. Em 2021, a quantidade exibida é menor, porém os dados contém apenas o

¹Disponível em: <<https://www.scopus.com/>>

primeiro semestre do ano, o que leva a uma expectativa de que a tendência crescente se mantenha.

O conteúdo relacionado a este tema de pesquisa pode ser agrupado em seis tópicos principais, apresentados nas próximas subseções: Poluição marinha por plásticos (Subseção 3.1.1), Sensoriamento remoto e técnicas espetrorradiométricas (Subseção 3.1.2), Experimentos de campo (Subseção 3.1.3), Bancos de dados espectrais (Subseção 3.1.4), Métodos de classificação de imagens (Subseção 3.1.5) e Dados para treinamento e validação de modelos (Subseção 3.1.6).

3.1.1 Poluição marinha por plásticos

Os detritos marinhos podem ser definidos como quaisquer materiais sólidos antropogênicos processados ou manufaturados que sejam descartados, perdidos ou deliberadamente abandonados no meio ambiente, incluindo os itens que são depositados no mar e na costa, ou são trazidos até eles por meio de rios, esgotos, águas pluviais, ondas ou vento (SHEVEALY; COURTNEY; PARKS, 2012).

O lixo marinho se origina de muitas fontes e viaja por longas distâncias com os ventos e as correntes oceânicas, causando um amplo espectro de impactos ambientais, econômicos, de segurança, de saúde e culturais, e representando uma ameaça vasta e crescente para o meio ambiente marinho e costeiro. Conforme já mencionado na Seção 1, estima-se que cerca de 60% a 90% dos detritos marinhos são compostos por diferentes produtos plásticos, de lenta decomposição (UNEP et al., 2005). Eles incluem desde pequenos itens, que escapam de nossos sistemas de gerenciamento de resíduos, até itens maiores, como artefatos de pesca abandonados e os contêineres com as mais variadas cargas que são perdidos no mar todos os anos (MAXIMENKO et al., 2019).

O plástico é leve e resistente e funde quando aquecido, permitindo a moldagem de diversos tipos de produtos, e pode ser classificado em dois grupos principais: os termorrígidos (fundidos apenas uma vez, não recicláveis) e os termoplásticos (podem ser fundidos várias vezes e remoldados, são reutilizáveis). Os termoplásticos compõem

grande parte dos plásticos de uso único e o setor de embalagens é o responsável pela parcela mais significativa do seu impacto ambiental (MAGRINI, 2012).

A contribuição do plástico para o desenvolvimento humano é inegável, pois ele torna equipamentos, vestuário e outros acessórios mais acessíveis para pessoas com menor poder aquisitivo, além de proporcionar armazenamento a baixo custo para alimentos e itens de higiene, ajudando a promover seu aproveitamento integral. Contudo, o uso intensivo, a produção linear e a alta durabilidade do plástico, unidos às dificuldades em implementar uma gestão de resíduos adequada, geraram um problema ambiental sem precedentes (MIHAI et al., 2022).

Desde sua invenção, o plástico substituiu o papel e o vidro nas embalagens, a madeira nos móveis e o metal nos carros, até tornar-se um material onipresente na atualidade, integrando os mais variados itens e gerando resíduos observados nos locais mais longínquos e desabitados do planeta, como no exemplo da ilha Henderson demonstrado na Figura 2, que mesmo sendo pequena, desabitada, e distante de centros urbanos, apresenta a maior densidade de detritos já relatada em qualquer lugar do mundo, chegando a 672 itens/m² (LAVERS; BOND, 2017). Os plásticos estão se tornando parte do registro fóssil da Terra e integrando as características da atual era geológica, o Antropoceno. Um novo habitat microbiano marinho foi designado como “plastisfera” (UNEP, 2016).

Os plásticos que entram nos oceanos podem flutuar ou afundar, dependendo de fatores como as características do detrito (peso, densidade, composição e forma) e a dinâmica local (correntes, ventos, turbidez da água e presença de outros elementos). Detritos mais densos que a água afundam imediatamente após entrarem no oceano, sendo possível observar leitos de mares próximos a grandes cidades, portos e fozes de rios densamente cobertos por garrafas PET inteiras e outros detritos de origem local. Outros tipos de detritos flutuam inicialmente, mas gradualmente perdem sua flutuabilidade devido à bioincrustação e degradação (MAXIMENKO et al., 2019; GALGANI; HANKE; MAES, 2015).

Figura 2 – Ilha Henderson, localizada no giro oceânico do Pacífico Sul, em 2015.

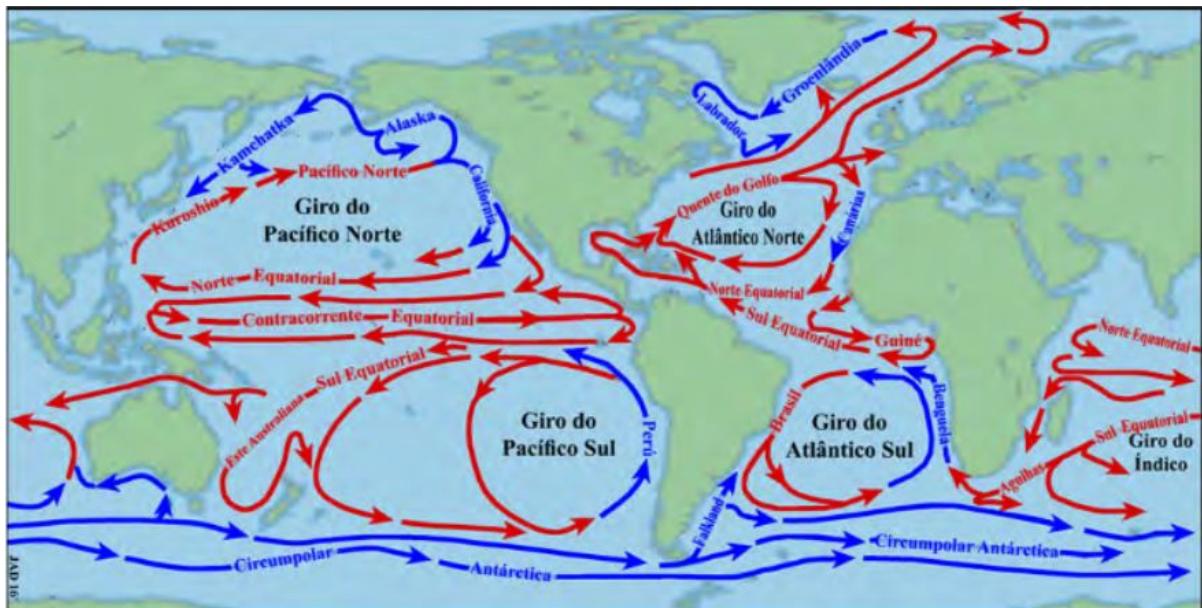


Fonte: <<https://tinyurl.com/ykzch5hs>>

Plásticos de baixa densidade podem flutuar por longos períodos na superfície dos corpos d’água, muitas vezes se acumulando em aglomerados que podem ser transportados por longas distâncias antes de afundar (CARLSON et al., 2017; KUBOTA, 1994). Parte deles entra nos giros oceânicos (Figura 3), que acabam se transformando em grandes sumidouros globais, e podem resultar em aglomerados de até vários quilômetros (KM) de tamanho, como a Grande Mancha de Lixo do Pacífico (GPGP, do inglês *Great Pacific Garbage Patch*), fotografado na Figura 4 (LEBRETON et al., 2018; LAW et al., 2010). Em algumas operações de limpeza nas regiões dos giros oceânicos, mais da metade dos plásticos encontrados flutuando haviam sido produzidos na década de 1990 ou anteriormente (UNEP, 2016).

A decomposição dos plásticos envolve uma combinação de vários processos, entre os quais desgaste mecânico, biodegradação e degradação foto e termo-oxidativa por radiação ultravioleta (UV). Os objetos maiores se fragmentam gradualmente em pe-

Figura 3 – Representação dos giros oceânicos e das principais correntes superficiais.

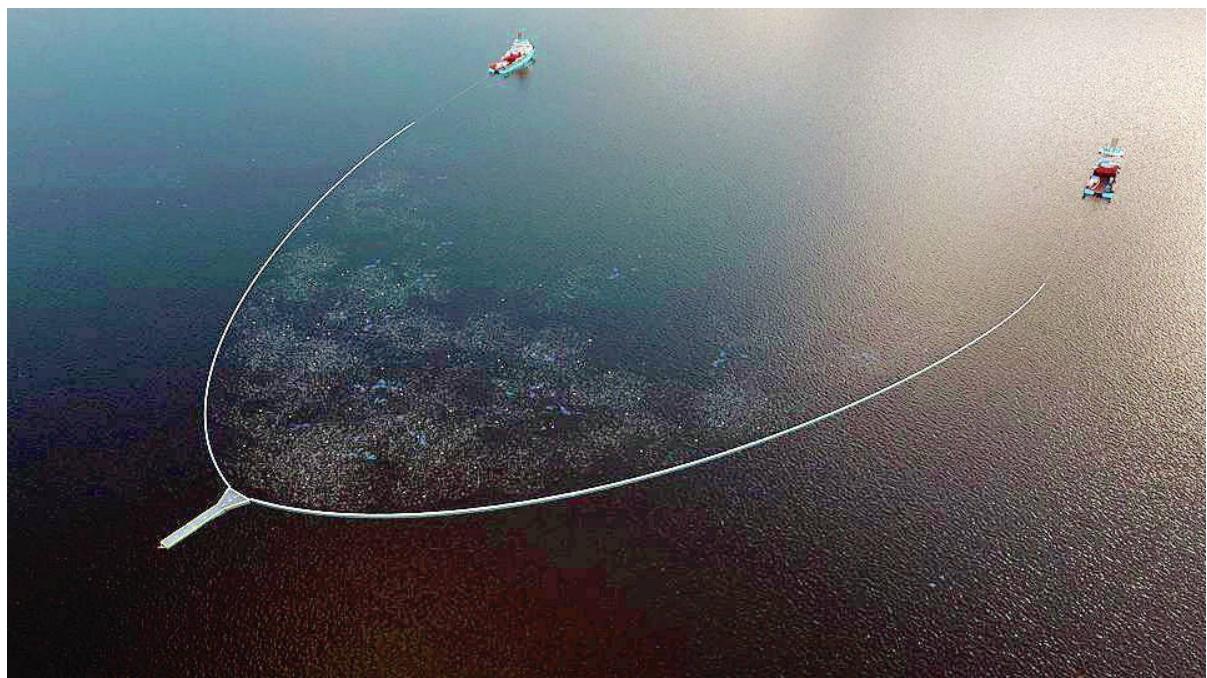


Fonte: DIAS (2018).

Nota: As correntes quentes estão representadas em vermelho e as frias em azul.

ças menores, que se tornam cada vez mais acessíveis a diferentes animais. Peixes, aves, tartarugas e outras espécies marinhas ingerem objetos dos mais variados tamanhos e formatos, incluindo microfibras, macroplásticos, microplásticos e nanoplásticos, podendo sofrer ferimentos, deformação, lacerção de tecidos internos, sufocamento, privação de oxigênio, fome e morte. Animais como os polvos bentônicos usam detritos como abrigo ou ninho, podendo sofrer contaminação pelos aditivos presentes nos materiais. Além disso, a “pesca fantasma” - assim chamada porque redes e armadilhas abandonadas podem continuar a capturar peixes e moluscos - fere animais e causa danos como abrasão e sufocamento aos recifes de corais. Os detritos podem, ainda, promover o transporte de espécies invasoras entre os mares e alterar o ciclo global do carbono por meio de seu efeito no plâncton e na produção primária nos ecossistemas, dentre uma série de outros impactos biológicos, econômicos e sociais, resumidos na Figura 5 (UNEP, 2016; MAES et al., 2021; COSTA et al., 2022; FERREIRA et al., 2022; FREITAS et al., 2022; JUSTINO et al., 2022; MONTEIRO et al., 2022; PÓVOA; ARAÚJO; SKINNER, 2022).

Figura 4 – Sistema desenvolvido pela organização The Ocean Cleanup para retirada de lixo da GPGP.

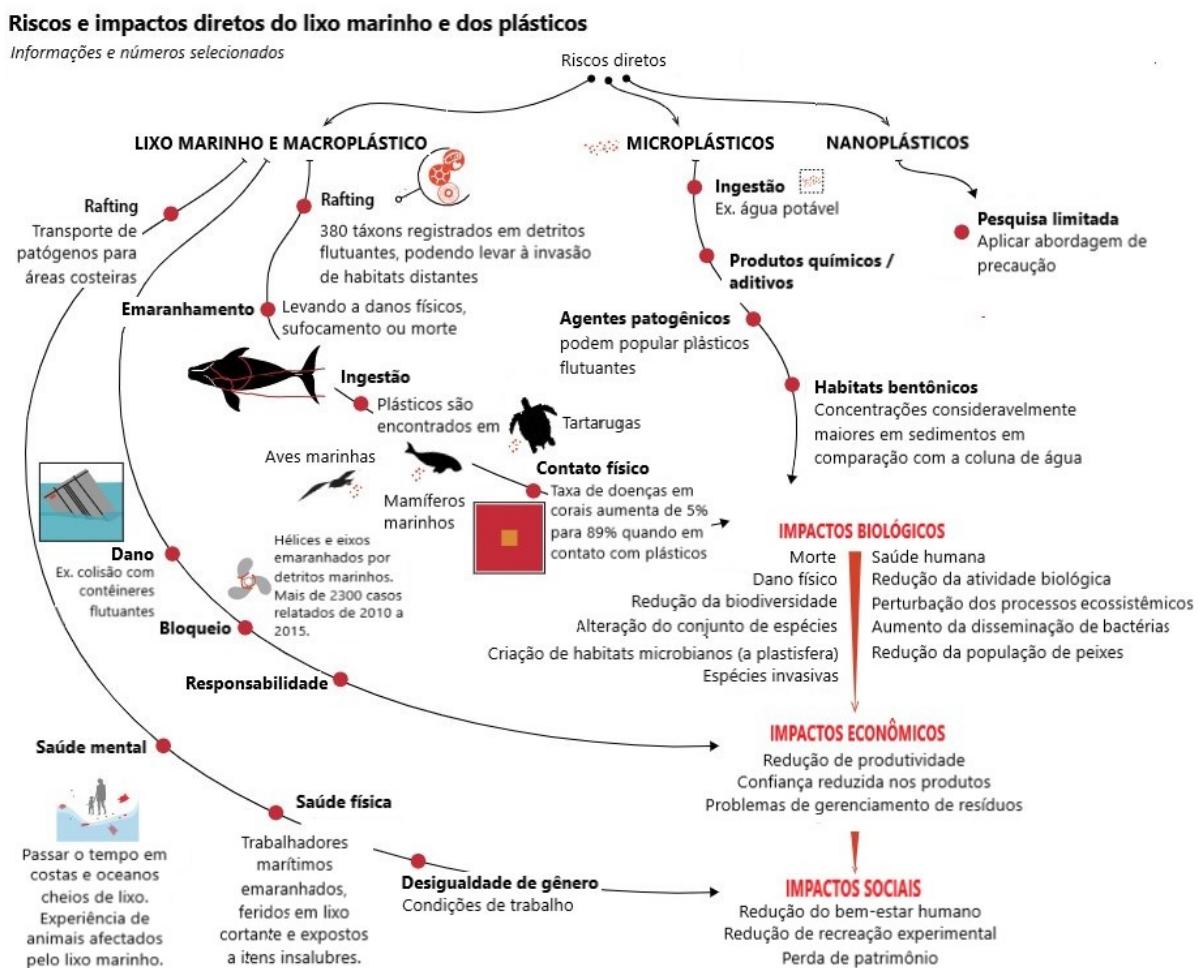


Fonte: <<https://tinyurl.com/ybykcshs>>

A predação accidental dos detritos plásticos pela biota marinha resulta em seu transporte através da cadeia alimentar por bioacumulação. Estes detritos também estão presentes na água e alimentos consumidos por seres humanos. Estimativas baseadas em aproximadamente 15% da ingestão calórica da população americana colocam o consumo anual de microplásticos por humanos variando de 39.000 a 52.000 partículas, dependendo da idade e do sexo. Essas estimativas aumentam para 74.000 e 121.000 quando a inalação é computada, junto a um adicional de 4.000 a 90.000 microplásticos quando se leva em consideração a ingestão das quantidades recomendadas de água através do consumo de água da torneira ou engarrafada, respectivamente (COX et al., 2019). Resíduos microplásticos também já foram encontrados em tecidos pulmonares humanos (AMATO-LOURENÇO et al., 2021).

Embora os efeitos do consumo de microplásticos na saúde humana sejam amplamente desconhecidos, possíveis consequências toxicológicas são sugeridas, como por exemplo inibição da proliferação celular e diminuição na atividade metabólica, de-

Figura 5 – Riscos e impactos diretos do lixo marinho e dos plásticos.



Fonte: Adaptação de MAES et al. (2021).

sestabilização e deformação das membranas celulares, além de outras alterações na morfologia das células. A exposição a longo prazo à inalação de microplásticos também é apontada como um risco para a saúde, pois o pulmão tende a atuar como um depósito permanente para as partículas. Como o plástico não é biodegradado por humanos, suas partículas tendem se acumular progressivamente no tecido pulmonar ao longo da vida (FLEURY; BAULIN, 2021; GOODMAN et al., 2021). Outros possíveis danos sugeridos incluem a liberação de monômeros constituintes, aditivos e toxinas por meio da absorção intestinal, podendo causar problemas que variam de estresse oxidativo a comportamento carcinogênico (COX et al., 2019; CAMPANALE et al., 2020; PRATA et al., 2020; TORRE, 2020; ZHANG et al., 2020; DANOPoulos et al., 2021).

A poluição marinha por plásticos também traz sérios riscos para a aquicultura, a pesca e as atividades econômicas relacionadas a elas: além do emaranhamento e dos danos físicos potencialmente reduzirem a eficiência e a produtividade, a pesca fantasma pode causar níveis significativos de mortalidade aos estoques comerciais que, em muitos casos, já estão sob pressão (MAES et al., 2021). O custo econômico estimado da poluição plástica, apenas para o capital natural marinho, é estimado entre US\$ 3.300,00 e US\$ 33.000,00 por tonelada de plástico ao ano (BEAUMONT et al., 2019).

Conforme apontado por Maximenko et al. (2019), são muitos os desafios enfrentados no processo de medição e compreensão dos detritos marinhos, dentre os quais:

- Variedade de tamanhos dos objetos (de dezenas de metros a mícrons);
- Variedade de formas dos objetos (pequenas esferas, filmes de embalagem, fibras e objetos compostos, dentre outros);
- Complexidade da composição química;
- Uso de protocolos e metodologias diferentes entre laboratórios e grupos de pesquisa, tornando as observações locais frequentemente difíceis de generalizar em um quadro global;
- Variação na acurácia dos dados coletados *in loco* (com dependência da habilidade do observador);
- Necessidade de registro de dados em ambientes de difícil acesso (como alto mar, por exemplo);
- Lacunas de conhecimento sobre as fontes, sumidouros, caminhos e processos de decaimento.

Estimativas globais da produção de plásticos em massa ao longo da história sugerem que, até o ano de 2017, cerca de 8300 milhões de toneladas métricas (TM)

de plásticos virgens foram fabricadas. Também indicam que, dos resíduos plásticos gerados até o ano de 2015, apenas aproximadamente 9% foram reciclados, enquanto 12% foram incinerados e 79% foram acumulados em aterros ou no ambiente natural (GEYER; JAMBECK; LAW, 2017). O volume de plásticos nos oceanos foi estimado entre 75 milhões e 199 milhões de TM (UNEP, 2016). Espera-se que o oceano contenha 1 TM de plástico para cada 3 TM de peixe até 2025, e em 2050, mais TM de plásticos do que de peixes (MACARTHUR; WAUGHRAY; STUCHTEY, 2016).

As estimativas de vazamento de lixo são muito limitadas, devido à baixa disponibilidade de dados consistentes, especialmente nos países do Sul Global (ALENCAR et al., 2022). Também há poucas informações disponíveis para a potencialmente significativa poluição por microplásticos via deposição atmosférica (LIU et al., 2019). Maximenko et al. (2019) demonstra que as estimativas da quantidade de microplásticos flutuando na superfície do mar chegam a no máximo 0,1% do total estimado de plástico adicionado ao oceano a partir de fontes terrestres, e mesmo após a inclusão de estimativas de outros tipos de detritos, sumidouros e reservatórios, o destino de pelo menos 90% dos detritos ainda não é contemplado pelos cálculos. As questões básicas sobre o lixo marinho ainda não foram totalmente respondidas e existem muitas lacunas e incertezas nas pesquisas sobre a dinâmica da poluição plástica no domínio marinho.

Conforme defendido por Emmerik, Vried e Peereboom (2022), embora a estratégia ideal para monitoramento de plástico em águas naturais seja altamente dependente do contexto local, há objetivos abrangentes importantes para direcionar o desenvolvimento e a implementação de sistemas de monitoramento da poluição em águas naturais e as pesquisas sobre a temática:

1. Apoio ao desenvolvimento de políticas voltadas para a redução da poluição;
2. Subsídio ao desenvolvimento de conhecimento de base;
3. Integração do monitoramento e da mitigação da poluição em ações de operação e manutenção, especialmente nas hidrovias e na infraestrutura hidráulica;
4. Estudo e implementação da melhor solução possível para cada contexto.

Este trabalho pretende contribuir com os objetivos 1 e 2, ao colaborar com o crescimento do conhecimento de base e oferecer uma possível alternativa de uso do sensoriamento remoto em políticas públicas de monitoramento da poluição.

3.1.2 Sensoriamento remoto e técnicas espetrorradiométricas

O sensoriamento remoto estuda a radiação eletromagnética proveniente da Terra e detectada por sensores orbitais. A Terra reflete parte desta radiação, proveniente principalmente do Sol, de volta para o espaço exterior, podendo também absorver a energia de comprimentos de ondas menores e então reemiti-la em comprimentos de onda maiores. A energia refletida ou reemitida pela Terra pode ser captada por sensores instalados em satélites com o intuito de analisar alvos da superfície da Terra (JENSEN et al., 2011).

O sensoriamento remoto em geral explora a interação dos alvos com o espectro eletromagnético em três domínios:

- Óptico, no qual o objeto de estudo é a proporção de radiação proveniente do Sol que é refletida pelos alvos, chamada de reflectância. É geralmente aplicado ao espectro visível (VIS) e ao infravermelho próximo (NIR, do inglês *Near Infrared*);
- Termal, onde mede-se a radiação infravermelha térmica emitida pelos objetos;
- Radar, onde verifica-se o retroespalhamento de uma fonte artificial de radiação.

Os sensores remotos possuem diferentes características, de forma que a determinação do sensor ideal depende da aplicação planejada, pois diferentes áreas de estudo apresentarão diferentes especificidades. Os sensores podem ser classificados de acordo com suas resoluções espacial, espectral, radiométrica e temporal. Conforme apresentado por Souza et al. (2007), a "resolução espacial ou geométrica refere-se ao campo de visada instantânea (IFOV, do inglês, *Instantaneous Field of View*) [...] e pode ser entendida como sendo a menor área do terreno que um sistema sensor é capaz de individualizar". O valor gravado em cada pixel de uma imagem adquirida por sensoriamento remoto é a média dos sinais dos diferentes objetos que ocorrem dentro

da respectiva área no terreno. Geralmente só objetos maiores do que a área do píxel podem ser identificados na imagem, de forma que o tamanho do píxel costuma ser a referência para resolução espacial.

Souza et al. (2007) conceitua a resolução espectral como “a capacidade de discriminação dos materiais da superfície da terra” sendo determinada pela “capacidade do sensor registrar a radiação em diferentes regiões do espectro”, ou seja, quanto maior o número e menor a largura das bandas espectrais, maior a capacidade discriminatória e, consequentemente, maior a resolução espectral. O autor também conceitua a resolução radiométrica como associada ao número de dígitos binários (bits) utilizados para armazenar os valores dos pixels, pois, quanto maior o número de bits, maior a faixa de valores que os pixels poderão assumir em uma imagem, e consequentemente, maior o detalhamento da superfície. Por fim, ele apresenta a definição de resolução temporal, que refere-se à frequência com que o sensor imageia uma determinada área, também referida como periodicidade, repetitividade ou tempo de revisita.

O sensoriamento remoto já é amplamente utilizado no monitoramento da superfície do oceano, oferecendo como vantagens seu potencial de alcançar áreas remotas e despovoadas e de preencher lacunas entre observações *in situ* esparsas, fornecendo cobertura global em longas escalas de tempo. Com relação à possível detecção de detritos marinhos, as observações espectrais fornecem uma riqueza de informações potencialmente aplicáveis a algoritmos para inferências relacionadas à abundância e à idade dos poluentes, aos tipos de polímeros e ao grau de intemperismo ou degradação, dentre outras questões relevantes (MAXIMENKO et al., 2019).

Contudo, não há missões de satélite específicas para a detecção de poluição plástica flutuante no ambiente marinho. A comunidade científica e as agências espaciais têm trabalhado na definição de especificações para as missões de satélite que sejam relevantes para a detecção e quantificação do lixo plástico marinho, mas a pesquisa na área até hoje só pode ser realizada utilizando dados de sensores projetados para outros tipos de aplicação.

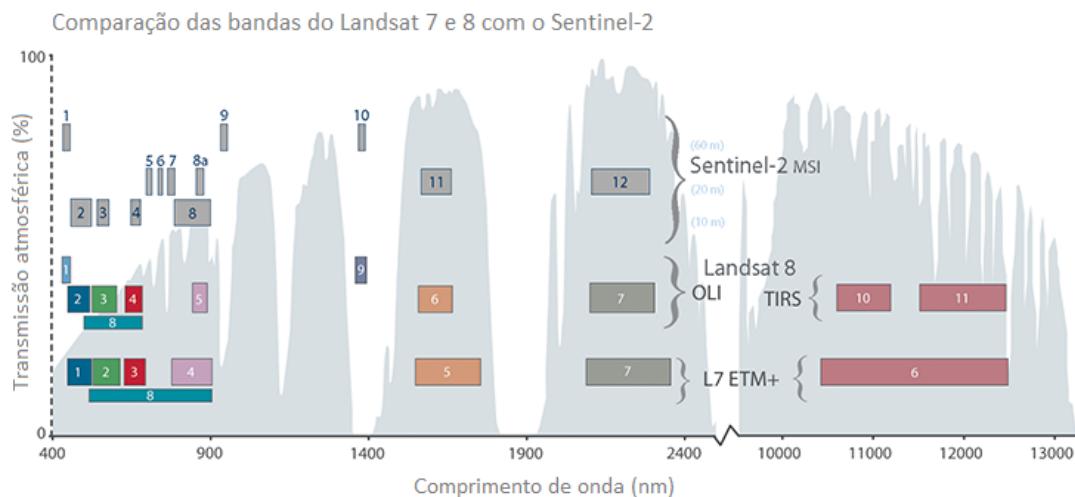
Conforme apontado por Biermann et al. (2020), as baixas resoluções temporal, espacial e espectral dos dados disponíveis até recentemente eram grandes limitadoras. Os exemplos citados pelos autores são missões muito conhecidas na comunidade do sensoriamento remoto e utilizadas nas mais variadas aplicações: o satélite Landsat 8, que fornece 9 bandas espectrais com uma resolução espacial de 30 m e temporal de 16 dias, e os satélites comerciais SkySat e RapidEye, que chegam a coletar imagens com resolução espacial submétrica, mas somente em 3 a 5 bandas espectrais. De acordo com a especificação de requisitos mínimos para sensoriamento remoto de poluição plástica apresentada por Martínez-Vicente et al. (2019), todos os exemplos citados estão aquém das especificações mínimas esperadas para viabilizar a detecção remota de poluição plástica em ambiente marinho.

O lançamento dos satélites Sentinel-2A e 2B pela Agência Espacial Europeia (ESA) em 2015 e 2017 trouxe uma evolução de recursos, com os sensores MSI colemando informações em 12 bandas com resolução espacial de até 10 m em tempos de revisita de 2 a 5 dias. Embora a missão tenha sido desenvolvida principalmente para serviços terrestres, a inclusão de águas costeiras globais em sua área de cobertura melhorou os resultados de aplicações de sensoriamento remoto em águas naturais, como as capacidades de discriminação da composição bentônica de recifes (HEDLEY et al., 2012) e de detecção e avaliação de mudanças nos componentes oceânicos e da criosfera (MALENOVSKÝ et al., 2012), melhorias estas atribuídas à estreiteza das bandas, à maior resolução espacial e à banda adicional em 443 nm, destinada à detecção de aerossol.

A hipótese de as resoluções terem melhorado o suficiente para a detecção de macroplásticos flutuantes a partir da órbita terrestre passou a ser considerada, pois, embora pedaços individuais de lixo marinho tendam a permanecer abaixo dos limites detectáveis na resolução do Sentinel-2, agregações de detritos podem tornar-se suficientemente grandes para viabilizar a detecção pelo MSI (BIERMANN et al., 2020; THEMISTOCLEOUS et al., 2020). A Figura 6 ilustra as diferenças entre as configurações de bandas do MSI/Sentinel-2 e alguns sensores da missão Landsat, lançada

antes da missão Sentinel. Os estudos descritos na Seção 3.1.3 investigam a detectabilidade de alvos plásticos através do MSI.

Figura 6 – Comparação entre as bandas dos sensores a bordo dos satélites Landsat e Sentinel-2.



Fonte: Adaptação de <<https://tinyurl.com/eyk7d7be>>.

3.1.3 Experimentos de campo

A primeira edição do Projeto de Lixo Plástico (PLP, do inglês *Plastic Litter Project*), realizada por Topouzelis, Papakonstantinou e Garaba (2019), avaliou o uso dos domínios óptico, termal e por radar na detecção de três alvos de plástico artificiais, compostos por garrafas, sacolas e redes de pesca, implantados em uma praia na Grécia durante a passagem do Sentinel-2. As imagens na região óptica ofereceram os melhores resultados: nas imagens termais, apenas as garrafas e os sacos plásticos foram detectados, sofrendo um provável efeito estufa devido à forma de agregação dos itens (diferente do que em geral se observa na natureza), enquanto nas imagens de radar, devido ao tipo de imageamento produzido e à condição de vento muito fraca, apenas o alvo das garrafas de plástico mostrou-se perceptível, enquanto as redes de pesca e os sacos plásticos apresentaram um sinal muito baixo, semelhante ao do mar.

O PLP passou a ter edições anuais com experimentos envolvendo a aquisição de dados remotos sobre alvos artificiais implantados em águas naturais, focando no sensoriamento remoto óptico a partir de 2019. Os resultados em geral indicam boas possibilidades de detecção dos alvos plásticos, especialmente no NIR, embora com grande dependência da proporção de plástico flutuante dentro dos pixels. Quanto maior o preenchimento do pixel na imagem da superfície da água, maior o sinal observado (TOPOUZELIS; PAPAKONSTANTINOU; GARABA, 2019; TOPOUZELIS et al., 2020; PAPAGEORGIOU et al., 2022).

Nesta dissertação, o sensoriamento remoto óptico foi selecionado como melhor abordagem disponível, e o MSI/Sentinel-2 como sensor mais adequado, seguindo na mesma diretriz do PLP e somando esforços para encontrar os melhores métodos aplicáveis aos conjuntos de dados adquiridos no projeto e posteriormente disponibilizados para a comunidade de pesquisa.

As características ópticas de diversos elementos frequentemente presentes nas cenas alvo foram resumidas por Biermann et al. (2020), em um estudo no qual imagens com suspeitas de agregação de macroplásticos em quatro países mostraram-se detectáveis nos dados do Sentinel-2. As características encontradas pelos autores foram:

- A água limpa é eficiente na absorção de luz no NIR e no infravermelho de ondas curtas (SWIR, do inglês *Shortwave Infrared*);
- Plásticos mostram um pico de reflectância no NIR;
- Algas marinhas refletem a luz nas bandas Green (560 nm) e Red (700–780 nm), mas podem absorver a luz na região do SWIR;
- A madeira mostra um pico de refletância no NIR e também reflete de forma relativamente forte na banda Red e no SWIR;
- A pedra-pomes é visivelmente brilhante em toda a faixa óptica, apresentando absorção apenas no NIR em aproximadamente 833 nm;

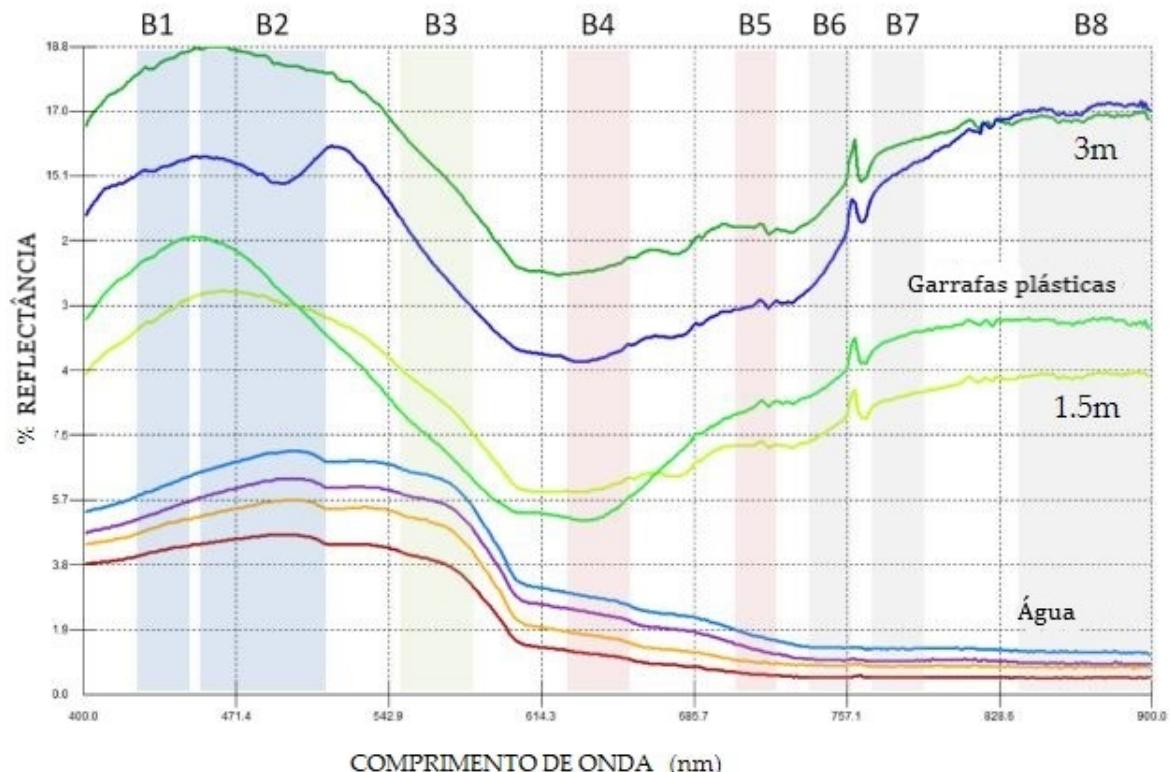
- A espuma das águas (provavelmente composta de detritos orgânicos em decomposição) mostra picos de reflectância mais altos nas bandas Green e Red e um pico menor no NIR.

Estes resultados são similares aos encontrados por Themistocleous et al. (2020), que realizou um experimento implantando alvos plásticos artificiais no mar do Chipre, compostos por garrafas plásticas agrupadas em conjuntos de 3 m × 10 m. As assinaturas espectrais da superfície da água e dos alvos de lixo plástico foram medidas com espectrorradiômetro, a 1,5 m e 3 m de altura, conforme demonstrado na Figura 7. As assinaturas espectrais foram identificadas e comparadas com as bandas do MSI, utilizando também um VANT com câmeras multiespectrais para fotografar o alvo de lixo plástico concomitantemente à passagem do satélite, para examinar diferentes comprimentos de onda em que o plástico poderia ser detectado na água do mar. Na região VIS, tanto os plásticos quanto a água demonstraram uma reflectância mais alta na banda Blue (B2, 496 nm), caindo nas bandas Green (B3, 560 nm) e Red (B4, 665 nm). Porém a refletância do plástico aumentou novamente no infravermelho, com um pico de refletância no NIR, onde a água absorve toda a radiação solar e não apresenta quase nenhuma reflectância, o que determina essa região como provável geradora de contraste entre o alvo e a água circundante.

Fica evidenciado que as regiões NIR e SWIR são chaves para a extração de informações espectrais relevantes. Também é importante considerar que materiais úmidos ou com qualquer grau de submersão também serão influenciados pelo coeficiente de absorção da água, demonstrado na Figura 8. A forte absorção da radiação infravermelha pela água constitui uma limitação significativa para a detecção de plásticos na água (GALGANI; HANKE; MAES, 2015).

Uma abordagem aplicável neste contexto é o uso de índices radiométricos. O índice NDVI, por exemplo, tende a ter valores altos para vegetação flutuante e valores baixos a negativos para água. Biermann et al. (2020), usando apenas NDVI chegaram a resultados nos quais os plásticos agrupados foram distinguíveis da água do mar, algas marinhas, materiais lenhosos, espuma do mar e pedra-pomes. Eles tam-

Figura 7 – Assinaturas espectrais médias coletadas de alvos plásticos artificiais implantados no mar.



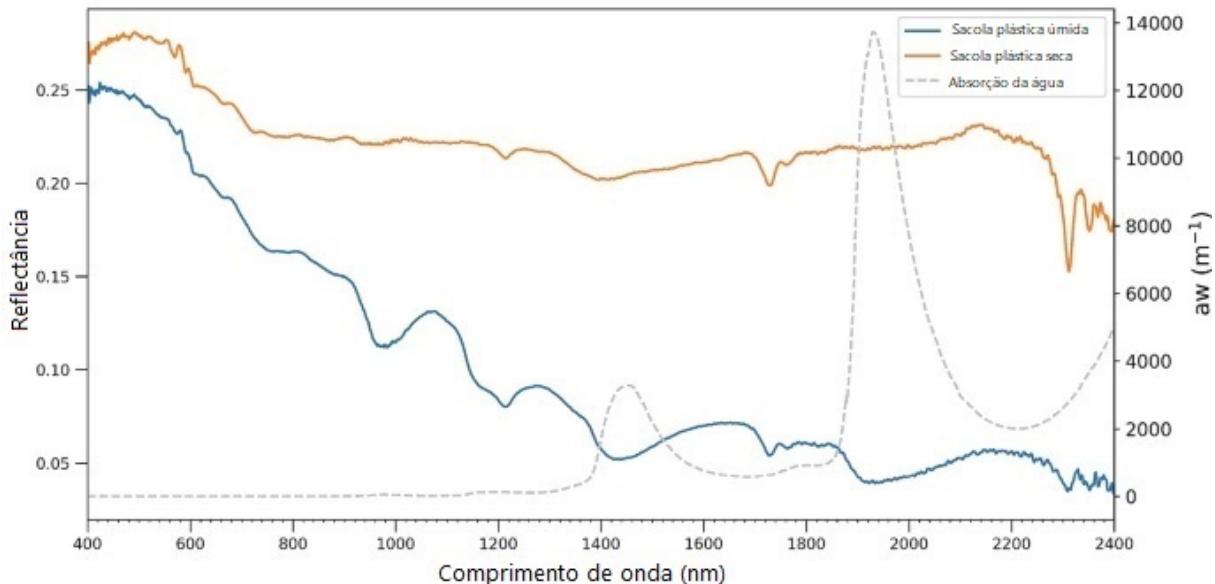
Fonte: Adaptação de THEMISTOCLEOUS et al. (2020).

Nota: As linhas vermelha, laranja, roxa e ciano indicam as assinaturas espectrais para a água, enquanto as linhas verde escuro, verde, azul e verde claro indicam as assinaturas espectrais para as garrafas de plástico na água em diferentes alturas (1,5 m e 3 m).

também propuseram o FDI (do inglês *Floating Debris Index*), um índice específico para identificação de plásticos testado nas imagens com suspeita de presença de detritos. Embora FDI e NDVI, isoladamente, só tenham distinguido algumas classes, quando examinados juntos, cada material ficou em um agrupamento distinto no gráfico, com total separabilidade (Figura 9). Contudo, o trabalho foi realizado com imagens sem confirmação *in situ* de poluição por plásticos, o que coloca um grau de incerteza nos resultados.

Themistocleous et al. (2020) também explorou o uso de diversos índices de bandas, incluindo a proposição do PI (*Plastic Index*), índice específico para detecção de macroplásticos, e obteve sucesso principalmente com índices utilizando as bandas

Figura 8 – Coeficiente de absorção da água e sua influência na reflectância espectral.



Fonte: Adaptação de MOSHTAGHI et al. (2021).

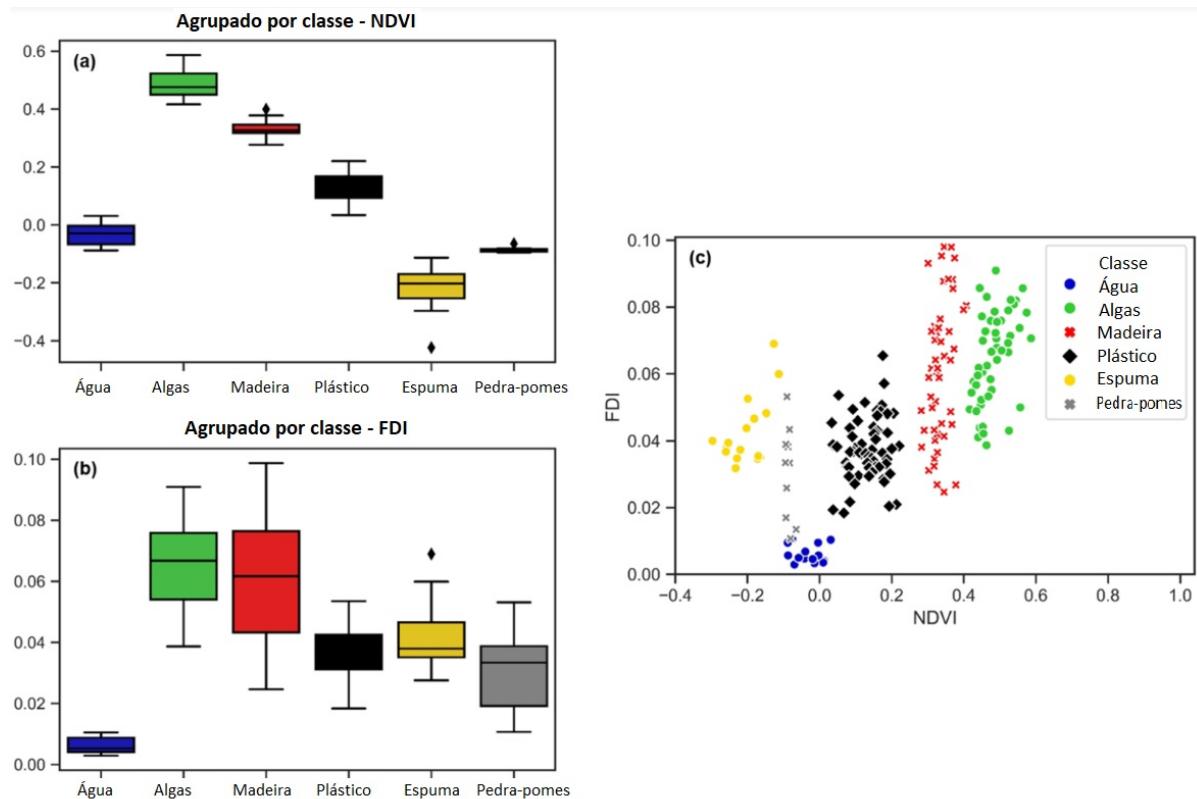
Red e NIR1 do MSI/Sentinel-2. Os índices aplicados nesses trabalhos foram selecionados para validação nesta dissertação.

3.1.4 Bancos de dados espectrais

Uma vez conhecido o comportamento espectral dos alvos, descritos na Seção 3.1.3, os dados adquiridos remotamente podem ser explorados através de composições coloridas, índices derivados de suas bandas espectrais, algoritmos computacionais e outras ferramentas. No caso dos MTRs, introduzidos na Seção 1, é necessário não somente conhecer as características espetrais gerais dos alvos, mas também obter suas assinaturas espetrais, que exibem os valores exatos de refletância do elemento em todo o espectro eletromagnético. Essa informação pode ser obtida com a utilização de um espectrorradiômetro, aparelho específico para este tipo de medição.

Os estudos de Acuña-Ruz et al. (2018) e Garaba e Dierssen (2018) forneceram bancos de dados espetrais de elementos relacionados à poluição marinha por plásticos. As assinaturas dos elementos estão ilustradas na Figura 10.

Figura 9 – Combinação dos índices FDI e NDVI para diferentes classes de cobertura da superfície em imagens com suspeita de poluição em águas naturais.

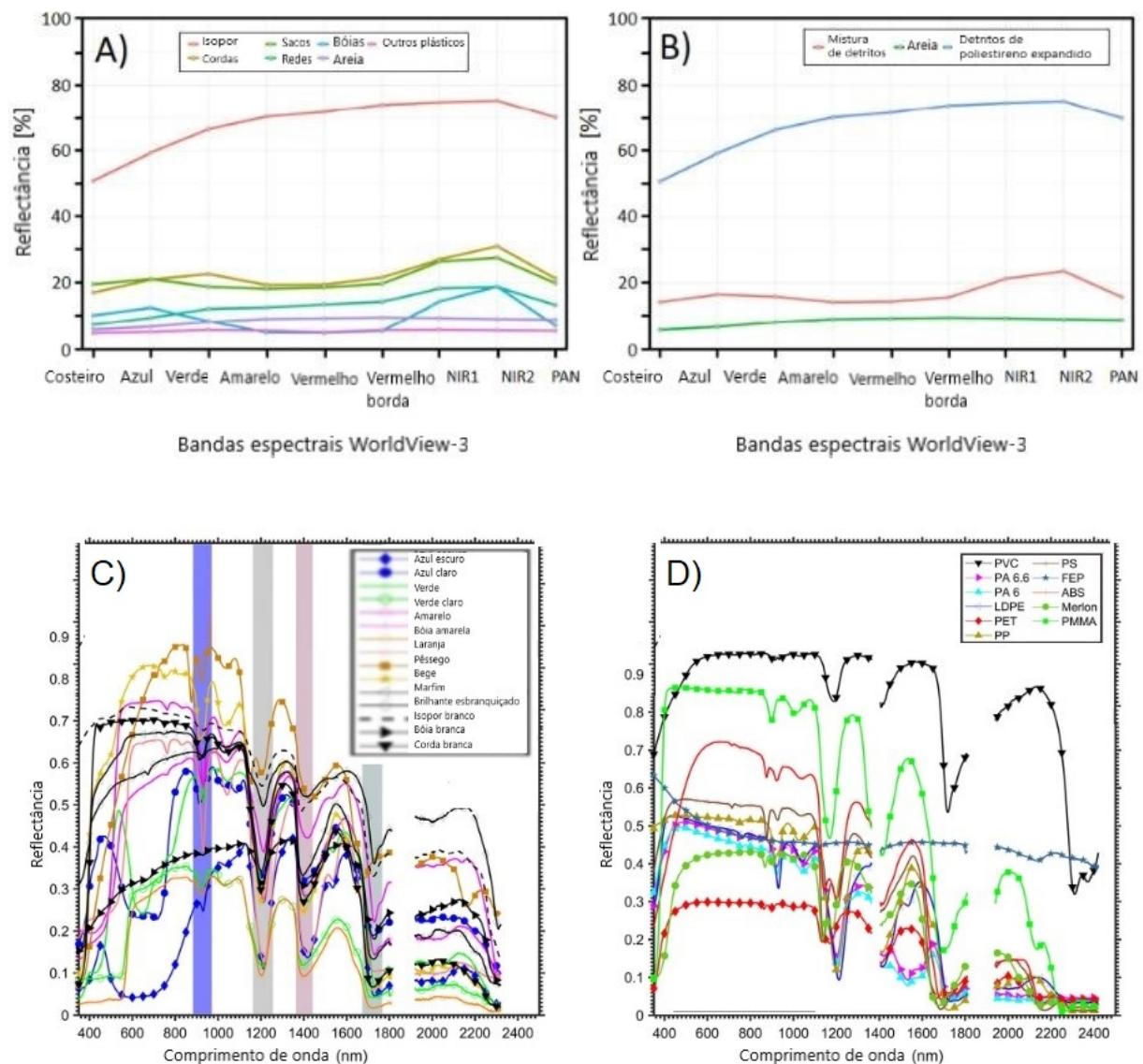


Fonte: Adaptação de BIERMANN et al. (2020).

Nota: Em (a) e (b) são exibidos os valores de NDVI e FDI, respectivamente, em gráficos agrupados por classe. Em (c), FDI e NDVI são combinados em um gráfico de dispersão bidimensional, também agrupado por classe.

Acuña-Ruz et al. (2018) estabeleceram um banco de dados de assinaturas espectrais a partir de amostras coletadas no litoral do Chile e analisaram sua detectabilidade em uma cena do sensor WorldView 3 (WV3). Foi possível identificar uma assinatura única para o poliestireno, bem como detectá-lo na cena do sensor, tendo em vista ele que possuía uma alta refletância e uma diferença espectral clara em relação aos demais materiais. Os outros componentes dos detritos mostraram baixa amplitude espectral e alta dispersão, e foram agregados em um sinal agrupado, sem identificação da composição. Apesar de ser desejável identificar as diferentes composições dos detritos, o uso de assinaturas mistas também pode ser útil: Topouzelis et al. (2020) conseguiu detectar alvos de tipos diferentes de plástico usando apenas

Figura 10 – Assinaturas espetrais de polímeros coletados na natureza (A, B e C) e virgens (D).



Fonte: Adaptação de ACUÑA-RUZ et al. (2018) (superior) e GARABA; DIERSSEN (2018) (inferior).

Nota: Na parte superior, gráfico adaptado das assinaturas espetrais apresentadas por Acuña-Ruz et al. (2018), sendo (A) assinaturas normalizadas em relação às bandas WV3 e (B) dados de entrada usados para classificação. Na parte inferior, gráfico adaptado das assinaturas espetrais apresentadas por Garaba, Dierssen (2018), sendo (C) refletância de macroplásticos coletados em ambiente natural, revelando as regiões de absorção (sombreados), e (D) refletância de polímeros virgens.

uma assinatura de tereftalato de polietileno (PET) em um cálculo de mistura espectral inversa, o que sugere que uma assinatura em massa pode ser usada para a detecção e quantificação de detritos marinhos em suspensão, uma grande vantagem perante às

dificuldades de construir e obter as bibliotecas espectrais necessárias, que demandam trabalho de campo com equipamento específico.

Já o estudo de Garaba e Dierssen (2018) apresentou uma caracterização óptica mais completa do ultravioleta ao SWIR (350–2500 nm) de polímeros coletados no Oceano Atlântico Norte e comparados com esferas plásticas virgens. Os resultados demonstraram de forma conclusiva características de absorção centradas em 931, 1215, 1417 e 1732 nm em macro e microplásticos secos e úmidos coletados no mar, constituindo uma assinatura espectral singular insensível ao tamanho, cor ou tipo de polímero. O estudo constatou que a refletância das partículas úmidas diminuiu em média $56 \pm 23\%$ em comparação com as partículas secas, com uma dependência espectral aumentando com o comprimento de onda de 12% no ultravioleta (UV) para quase 90% no SWIR, o que é compatível com a ação do coeficiente de absorção da água (Figura 8). Os macroplásticos amostrados no estudo foram apresentados separadamente; porém, para as amostras de microplásticos, foi definido um espectro médio, aqui referenciado como Micronapo (μ -NAPO).

Para estabelecer a biblioteca de referência espectral, onze tipos de plásticos virgens secos também foram medidos por Garaba e Dierssen (2018), sendo eles cloreto de polivinila (PVC), poliamida ou náilon (PA 6.6 e PA 6), polietileno de baixa densidade (LDPE), polietileno tereftalato (PET), polipropileno (PP), poliestireno (PS), etileno propileno teflon fluorado (FEP), terpolímero lustran 752 (ABS), merlon e polimetil metacrilato (PMMA). Os espectros destes plásticos virgens foram comparados com a medição dos macro e microplásticos coletados no mar na faixa de comprimento de onda onde as características únicas de absorção de plástico foram encontradas. A análise de similaridade espectral realizada pelos autores encontrou similaridades moderadas a muito fortes, embora não tenha sido possível combinar perfeitamente os espectros das amostras naturais com os dos polímeros brutos, tendo em vista que elas estavam sujeitas a intemperismo e degradação na natureza, além possuírem misturas de diferentes polímeros em sua composição.

O estudo também testou a viabilidade do uso das características detectadas usando imagens AVIRIS hiperespectrais, que possuem alta resolução espacial (7,1 m) e espectral (224 bandas), adquiridas sobre um aterro sanitário circundado por instalações industriais e de tratamento de água. As bandas próximas de 950 e 1400 nm foram descartadas por coincidirem com as bandas de absorção da água, e estarem sujeitas a interferências atmosféricas. As bandas 1215 e 1732 nm foram usadas para testar algoritmos para mapeamento automatizado, resultando no desenvolvimento de algoritmos de índice de hidrocarbonetos (HI1732 e HI1215), que foram capazes de identificar pixels com presença de plásticos. Posteriormente, Silva (2020) partiu deste trabalho para desenvolver o Índice de Hidrocarbonetos Simplificado (HIS), com objetivo de possibilitar a detecção da proporção de plásticos existentes em uma imagem a partir do valor HIS calculado para cada pixel.

A biblioteca espectral gerada no estudo foi disponibilizada na base de dados online Ecological Spectral Information System (GARABA; DIERSSEN, 2017; 2019a; 2019b). Os autores sugeriram o seu uso em conjunto com modelagem de transferência radiativa para determinar o potencial e as limitações para a detecção de plásticos marinhos usando sensoriamento remoto. Esse trabalho foi uma das principais referências para a proposta desta dissertação.

3.1.5 Métodos de classificação de imagens

Os métodos de sensoriamento remoto podem ser combinados com algoritmos de aprendizado de máquina. Diversos trabalhos buscam avaliar o uso destes algoritmos com imagens adquiridas por satélites ou VANTs para possíveis aplicações em identificação e quantificação automatizada de lixo plástico (ACUÑA-RUZ et al., 2018; MARTIN et al., 2018; BAK et al., 2019; KYLILI et al., 2019; GONÇALVES et al., 2020).

Dado um conjunto de observações - ou seja, os vetores com os valores dos pixels imageados em cada banda do sensor - o objetivo da classificação é atribuir um rótulo de classe único a cada observação. No caso dos plásticos em águas naturais, os rótulos atribuídos pelo classificador informariam as regiões onde há ou não presença

de plásticos, além de outras informações sobre a cobertura da superfície (BIOUCAS-DIAS et al., 2013).

A maior limitação dos modelos de aprendizado de máquina vem da necessidade de conjuntos de dados *in situ* para treinamento e validação, conforme mencionado na Seção 1. O número, qualidade e representatividade dos conjuntos de dados utilizados determinam a robustez do modelo, e as avaliações de acurácia e desempenho das previsões dependem de grandes volumes de dados, podendo levar a altos custos e diversos desafios. Conforme detalhado na Seção 3.1.6, uma alternativa é utilizar abordagens de base física, que permitem simular um grande conjunto de dados que pode ser utilizado para treinamento dos modelos, reduzindo assim a dependência de dados de campo. Outros desafios relevantes são a presença de pixels mistos nos dados, a necessidade de integrar informação espacial e espectral e a alta complexidade computacional de diversos classificadores (BIOUCAS-DIAS et al., 2013).

Os métodos de classificação em geral são supervisionados ou não supervisionados. A abordagem supervisionada atua de forma preditiva, usando informações prévias das quais o usuário dispõe para gerar um modelo capaz de identificar as classes de interesse, através de uma função que relate a classe (ou atributo alvo) do objeto aos seus atributos de entrada. Já a abordagem não supervisionada atua de forma descritiva, sem usar informações prévias para gerar o modelo, tendo como objetivo explorar ou descrever um conjunto de dados. Por isso, não utilizam informações sobre a classe do objeto, mas somente seus atributos de entrada (FACELI et al., 2021). No sensoriamento remoto, ela em geral busca similaridades entre as informações espetrais, agrupando os pixels de acordo com elas.

A pesquisa de poluição plástica em águas naturais por meio de dados remotamente adquiridos pode se beneficiar de ambas as abordagens. A abordagem supervisionada pode permitir o desenvolvimento de modelos visando o mapeamento da cobertura da superfície, e consequentemente, da presença de poluição plástica, enquanto a abordagem não supervisionada permite estudar os padrões de comportamento espectral dos elementos sobre a superfície.

Basu et al. (2021) usaram imagens MSI/Sentinel-2 dos alvos artificiais de plástico implantados nos PLPs e no estudo de Themistocleous et al. (2020), na Grécia e no Chipre, respectivamente, para avaliação de dois algoritmos de classificação não supervisionados, o K-means e o Fuzzy C-means (FCM), e também de dois algoritmos de classificação supervisionados, o Regressão Vetorial de Suporte (SVR, do inglês *Support Vector Regression*) e Fuzzy C-means Semi-supervisionado (SFM, do inglês *Semi-supervised Fuzzy C-means*). Das várias bandas disponíveis para o MSI/Sentinel-2, uma combinação de 6 bandas e dois índices foram selecionados para desenvolver os modelos, através da criação de três conjuntos de atributos: um contendo todos os atributos, e outros dois contendo subconjuntos dele.

A classificação supervisionada baseada em SVR atingiu uma acurácia na faixa de 96,9% a 98,4%, enquanto que para SFCM, FCM e K-means as faixas de acurácia alcançadas foram, respectivamente, 35,7% a 64,3%, 69,8% a 82,2% e, por fim, 69,8% para 81,4%. O desempenho dos modelos foi superior ao utilizar o maior conjunto de atributos, que continha todos os selecionados. O SVR não apontou plástico flutuante em nenhum pixel com água, enquanto os outros três algoritmos previram erroneamente a presença de plásticos em pixels onde ele não estava presente. A quantidade de dados disponível para o experimento era bastante limitada, tornando necessária a realização de análises adicionais com maior quantidade e variabilidade de amostras.

O desempenho dos métodos supervisionados Floresta Aleatória (RF, do inglês *Random Forest*), Análise Discriminante Linear (LDA, do inglês *Linear Discriminant Analysis*) e Máquina de Vetores de Suporte (SVM, do inglês *Support Vector Machine*) usando 3 kernels diferentes (Linear, Radial e Polinomial), foi testado no estudo Acuña-Ruz et al. (2018), usando dados WV3 e as assinaturas espectrais obtidas nas campanhas de campo, conforme descrito na Seção 3.1.2. As acuráncias obtidas com os diferentes modelos de classificação foram superiores a 75%. Para avaliar qual método apresentava um desempenho consistentemente melhor, 100 iterações da classificação foram feitas, com os dados sendo subamostrados aleatoriamente em cada caso.

SVM Radial obteve a melhor acurácia geral em média, estando com maior frequência próxima a 90%.

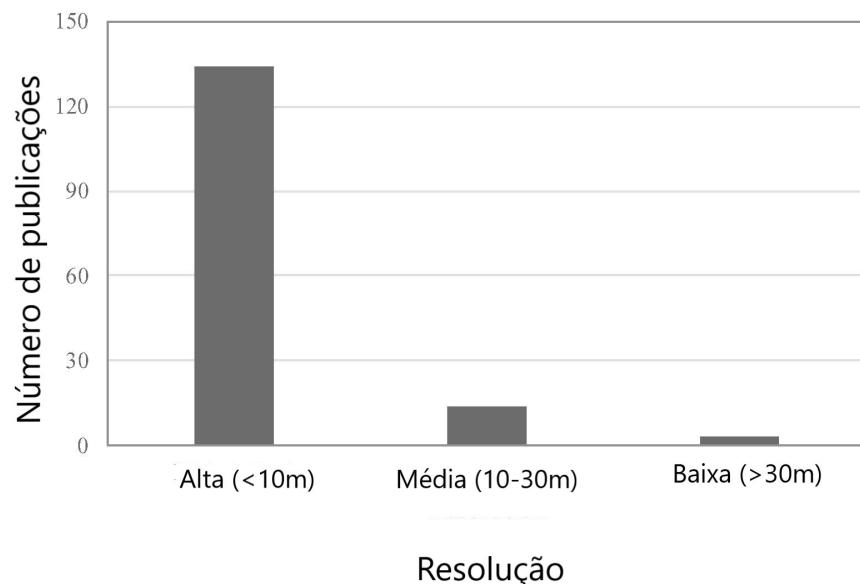
Outro experimento com classificação supervisionada foi desenvolvido por Wolf et al. (2020), comparando as técnicas de aprendizado de máquina exploradas por Acuña-Ruz et al. (2018) com redes neurais convolucionais (CNNs, do inglês *Convolutional Neural Networks*). Os métodos foram treinados com imagens de alta resolução espacial adquiridas por um VANT sobre rios e praias no Camboja. Os resultados apontaram, dentre o conjunto de métodos avaliados, que as CNNs produziram os resultados mais promissores, especialmente quando a classificação de tipos de plástico é desejada. O trabalho culminou no desenvolvimento do algoritmo APLASTIC-Q, baseado no método CNN com potencial de adaptação a aplicações para processamento de imagens de smartphones, câmeras portáteis, observatórios fixos e plataformas aéreas e espaciais tripuladas, e contendo as componentes PLD-CNN e PLQ-CNN, capazes de, respectivamente, identificar e quantificar objetos de lixo, obtendo alta acurácia.

As CNNs fazem parte das técnicas de aprendizado profundo. Elas são constituídas por uma série de camadas de processamento combinadas em rede e são capazes de aprender padrões hierárquicos espaciais. Seu uso no sensoriamento remoto tem crescido, porém costuma estar associado a altas resoluções espaciais, perdendo acurácia quando são usadas resoluções mais baixas (JAKOVLJEVIC; GOVEDARICA; ALVAREZ-TABOADA, 2020).

Conforme mostrado na Figura 11, o aprendizado profundo é usado com mais frequência no sensoriamento remoto com resoluções espaciais superiores a 10m. A maior parte dos estudos incluídos na meta-análise de Ma et al. (2019), que investigou o uso de aprendizado profundo em sensoriamento remoto, usavam resolução espacial superior a 2 m. A pouca disponibilidade de dados de referência adquiridos por sensores orbitais restringe o desenvolvimento das CNNs com este tipo de imageamento. É mais comum ver as CNNs e as técnicas de aprendizado profundo presentes em trabalhos com VANTs, que em geral possuem maior resolução espacial mas menor área de cobertura (ZHAO et al., 2015; ZHU et al., 2017; JAKOVLJEVIC; GOVEDARICA;

ALVAREZ-TABOADA, 2020; KENTSCH et al., 2020; LIU et al., 2020; NEZAMI et al., 2020; FENG et al., 2021).

Figura 11 – Resoluções espaciais utilizadas em trabalhos de sensoriamento remoto com aprendizado profundo.



Fonte: Adaptação de MA et al. (2019).

A tecnologia CNN se baseia nas Redes Neurais Artificiais (RNAs) e é responsável por uma expansão no seu uso. Os notáveis avanços em recursos computacionais nas últimas décadas também ampliaram muito o potencial de uso deste classificador, tendo em vista que sua abordagem de retropropagação no treinamento é computacionalmente intensiva. Paola, Schowengerdt (1995) já apontavam o potencial das RNAs se tornarem uma ferramenta padrão no sensoriamento remoto. Em seu estudo, eles compararam, para classificação de cobertura do solo, o uso de RNAs, que são não-paramétricas, com o método de máxima-verossimilhança, um classificador paramétrico muito usado em sensoriamento remoto. Ambos são classificadores supervisionados, e os dados levantados pelo artigo ajudam a entender as diferenças entre eles e também entre os classificadores paramétricos e não-paramétricos em geral. Os resultados sugerem que a abordagem da RNA (não-paramétrica) é mais robusta e acomoda mais facilmente rótulos heterogêneos, mesmo com conjuntos de treinamento menores, ao

passo que a máxima verossimilhança (paramétrica) mostra-se mais sensível à pureza das assinaturas de classe, gerando melhores resultados quando está restrita a classes espectralmente puras.

Conforme explicado por Oliveira (2014, pág. 251), “as RNAs são modelos matemáticos empíricos, com capacidade para reproduzir o funcionamento dos sistemas nervosos biológicos [...] através do treinamento e da resposta aos estímulos (variáveis de entrada)”. Segundo o mesmo autor, a composição de um neurônio artificial pode ser entendida como uma função de soma ponderada das suas entradas e uma função de ativação, normalmente não linear, que modifica o resultado desta soma, resultando em um sinal correspondente à resposta desejada. O alcance do resultado desejado depende do adequado ajuste dos pesos sinápticos entre os neurônios da RNA, por meio do processo de treinamento da rede.

O modelo matemático baseado no neurônio biológico foi proposto pela primeira vez por McCulloch e Pitts (1943), teorizando um método para obter o conjunto de constituintes computáveis da rede e caracterizar seu comportamento. Posteriormente, em um contexto histórico de proposição de diversos modelos representativos do cérebro, Rosenblatt (1958) projetou um sistema nervoso hipotético chamado Perceptron, com objetivo de ilustrar algumas das propriedades fundamentais de sistemas inteligentes em geral e exemplificar seu funcionamento. Widrow e Hoff (1960) criaram mais tarde a rede Adaline, que era linear, adaptativa e mais robusta que o Perceptron, capaz de aprender comparando as saídas desejadas com as saídas obtidas pelo modelo e gerar mudanças no design da rede.

Rumelhart, Hinton e Williams (1986) evoluíram o processo de aprendizado através da retropropagação do cálculo do erro pelas camadas da rede (do inglês *backpropagation*). O procedimento consiste em ajustar repetidamente os pesos das conexões entre os neurônios na rede de forma a minimizar os erros ou resíduos do vetor de saída desejado. O algoritmo retropropagativo tornou possível o treinamento de redes com múltiplas camadas - algo que não era viável nos modelos até aquele momento - e levou a uma nova onda de desenvolvimento das redes neurais, até então estagnadas.

Na literatura, as RNAs são frequentemente utilizadas em pesquisas na área de hidrologia (CAMPOLO; SOLDATI; ANDREUSSI, 2003; JAIN; SUDHEER; SRINIVASLU, 2004; BOWDEN; DANDY; MAIER, 2005; JAIN; KUMAR, 2007; LEAHY; KIELY; CORCORAN, 2008; OLIVEIRA; PEDROLLO; CASTRO, 2011). Elas também já obtiveram sucesso em determinar quais variáveis ambientais afetam a distribuição de lixo e em estimar as quantidades no fundo do Mediterrâneo central (FRANCESCHINI et al., 2019). A abordagem tradicional da RNA, utilizando um Perceptron multicamada com aprendizado por retropropagação, parece melhor adaptada que as CNNs à pesquisa de poluição marinha por sensoriamento remoto orbital, tendo em vista a necessidade de trabalhar com resoluções espaciais médias ou baixas e com pequenos conjuntos de dados com presença frequente de pixels mistos.

A aplicação de classificadores de aprendizado de máquina exige uma etapa de preparação de dados, onde as amostras são padronizadas e refinadas, de forma a serem corretamente lidas e interpretadas pelo classificador. A preparação de dados em geral inclui uma etapa de análise exploratória, na qual os dados passam por processos de caracterização e exploração, e uma etapa de pré-processamento, na qual os dados passam por processos de integração, seleção de atributos, amostragem, limpeza e transformação. A aplicabilidade de cada processo depende do estado do conjunto de dados e dos requisitos do classificador em relação a eles (FACELI et al., 2021).

Outra etapa importante ao usar modelos classificadores é a seleção de hiperparâmetros, que são parâmetros que não são aprendidos pelos classificadores, mas informados de forma direta pelo programador, e em geral definem configurações específicas do método. Eles podem ser definidos de forma arbitrária, mas é mais recomendável o uso de um processo de otimização, no qual um espaço amostral de valores para hiperparâmetros passa por uma busca exaustiva pela melhor pontuação de validação cruzada (PEDREGOSA et al., 2011).

A validação cruzada também é utilizada na avaliação dos modelos: nela, o conjunto de exemplos é dividido em r subconjuntos de tamanho aproximadamente igual, e os objetos de $r - 1$ partições são utilizados no treinamento de um preditor, posterior-

mente testado na partição restante. Esse processo é repetido r vezes, utilizando em cada ciclo uma partição diferente para teste. O desempenho final do modelo é dado pela média dos desempenhos observados sobre cada subconjunto de teste. O método também tem uma variação que inclui uma estratificação das amostras, na qual cada partição possui uma proporção de exemplos de cada classe semelhante à proporção contida no conjunto de dados total, algo especialmente relevante ao se trabalhar com dados desbalanceados (FACELI et al., 2021).

A medição do desempenho final do classificador pode ser realizado por meio de diversas métricas. Uma das mais utilizadas é a matriz de erro ou matriz de confusão, que ilustra o número de previsões corretas e incorretas em cada classe. Para determinado conjunto de dados, as linhas dessa matriz representam as classes verdadeiras, e as colunas, as classes preditas pelo classificador. Por meio do exame dessa matriz, tem-se medidas quantitativas de quais classes o algoritmo de aprendizado tem maior dificuldade em classificar. Na Figura 12, é apresentado um exemplo de matriz de confusão para um problema com três classes.

Figura 12 – Matriz de confusão para a classificação de um conjunto de dados contendo três classes

		Classe predita		
		1	2	3
Classe verdadeira	1	11	1	3
	2	1	4	0
	3	2	1	6

Fonte: FACELI et al., 2021.

Nota: No exemplo apresentado, 11 dos 15 exemplos da classe 1 foram corretamente classificados, um foi incorretamente classificado como pertencente à classe 2 e três foram preditos como pertencentes à classe 3. A matriz também informa que, dos 14 exemplos preditos pelo modelo como pertencentes à classe 1, 11 estavam corretamente classificados, enquanto um pertencia à classe 2 e dois pertenciam à classe 3. A mesma lógica aplica-se às demais classes exibidas na matriz.

Outras métricas calculam pontuações para o modelo com base em determinados critérios. Uma das mais difundidas é a Acurácia Geral (AG), também chamada de Acurácia Global ou Acurácia Total. Para um conjunto de dados classificados, ela pode ser calculada através da soma dos valores da diagonal principal da matriz de confusão dividida pela soma dos valores de todos os elementos da matriz. Para cada classe do conjunto individualmente, ela é calculada através da soma dos verdadeiros positivos (VP) e verdadeiros negativos (VN) divididos pelo total de amostras da classe (n), conforme descrito na Equação 1 (FACELI et al., 2021).

$$AG = \frac{VP + VN}{n} \quad (1)$$

Equação 1 – Acurácia Geral (AG), também chamada de Acurácia Global ou Acurácia Total

Para dados desbalanceados, a AG pode não ser uma boa métrica de avaliação de desempenho. Por exemplo: em um conjunto de dados composto por 95% de amostras de uma classe e 5% de outra classe, caso o classificador simplesmente aponte todas as amostras como pertencentes à primeira classe, ele obterá uma AG de 95%, mostrando-se insuficiente para a real avaliação do modelo, tendo em vista que, no exemplo acima, ele teria errado todos os exemplares de uma das classes.

Outras métricas muito difundidas envolvem os conceitos de Precisão (Prec, do inglês *Precision*) e Revocação (Rec, do inglês *Recall*). A Precisão é, para cada classe, a proporção de exemplos positivos classificados corretamente entre todos aqueles preditos como positivos pelo modelo (Equação 2) (MONARD; BARANAUSKAS, 2003; FACELI et al., 2021). Na análise de uma matriz de confusão, a Precisão pode ser associada aos conceitos de Acurácia do Usuário e Erro de Comissão, que calculam, para todo o conjunto de dados, os erros ao atribuir uma amostra a uma determinada classe quando ela na verdade pertence a outra classe (CONGALTON, 1991). São métricas associadas à exatidão e à confiabilidade do modelo para cada classe.

$$\text{Prec} = \frac{VP}{VP + FP} \quad (2)$$

Equação 2 – Precisão (Prec, do inglês *Precision*)

A Revocação é, para cada classe, a proporção de exemplos positivos classificados corretamente entre todos os exemplos positivos do conjunto de dados (Equação 3), também sendo chamada de Sensibilidade ou Taxa de Verdadeiros Positivos (TVP) (MONARD; BARANAUSKAS, 2003; FACELI et al., 2021). Na análise de uma matriz de confusão, a Revocação pode ser associada aos conceitos de Acurácia do Produtor e Erro de Omissão, que calculam, para todo o conjunto de dados, os erros cometidos quando o classificador deixa de atribuir as amostras às classes corretas (CONGALTON, 1991). São métricas associadas à sensibilidade e à completude do modelo para cada classe. Na pesquisa de poluição marinha por plástico, a revocação tem muita importância, pois fornece dados relacionados a capacidade do modelo detectar a presença do plástico em cenas marinhas ou costeiras.

$$\text{Rec} = \frac{VP}{VP + FN} \quad (3)$$

Equação 3 – Revocação (Rec, do inglês *Recall*)

Para problemas com múltiplas classes, as métricas consideram cada classe como positiva em relação ao conjunto das demais classes. Por exemplo, se a taxa que está sendo observada é a precisão, gera-se um valor de precisão para cada classe, no qual ela é considerada positiva, enquanto as demais são consideradas negativas (FACELI et al., 2021).

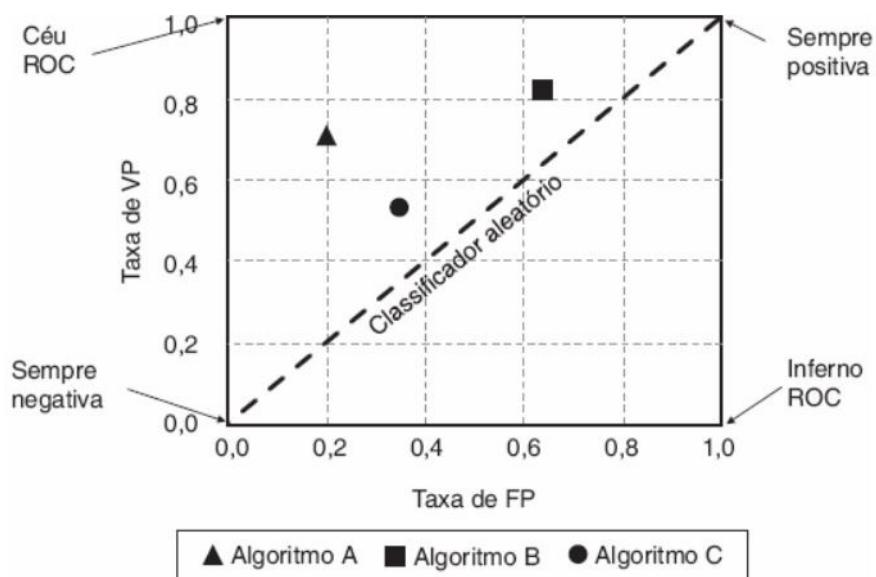
A Precisão e a Revocação fornecem informações complementares, de modo que, geralmente, não são discutidas isoladamente, mas são combinadas em uma única medida: a pontuação F1 (do inglês *F1-score*), a média harmônica ponderada das duas métricas (Equação 4). A pontuação F1 atinge seu melhor valor em 1 e pior em 0.

$$F1 = \frac{2 \times Prec \times Rec}{Prec + Rec} \quad (4)$$

Equação 4 – pontuação F1 (do inglês *F1-score*)

Uma outra alternativa para avaliar classificadores, aplicada a problemas binários (com apenas duas classes) são as curvas ROC (do inglês *Receiving Operating Characteristics*) (FAWCETT, 2006). O espaço ROC está ilustrado na Figura 13. Um classificador é considerado melhor que outro se seu ponto no espaço ROC encontra-se acima e à esquerda do ponto correspondente ao segundo classificador (PRATI, 2006). O procedimento mais comum para comparar o desempenho de diferentes algoritmos é gerar curvas no espaço ROC e usar a área abaixo da curva (AUC, do inglês *Area Under ROC Curve*) como medida. A AUC produz valores entre 0 e 1, sendo que os mais próximos de 1 são considerados melhores.

Figura 13 – Espaço ROC com três classificadores.



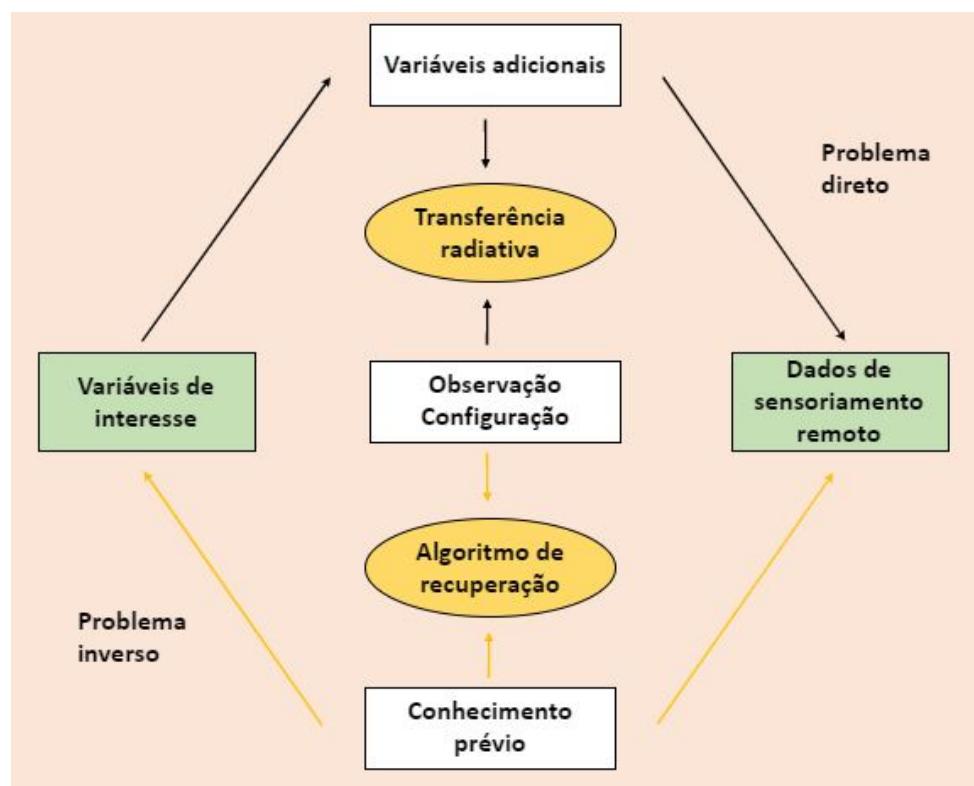
Fonte: FACELI et al. (2021).

3.1.6 Dados para treinamento e validação de modelos

A validação de dados em campo é fundamental para determinar a confiabilidade dos modelos em sensoriamento remoto. Para a pesquisa de poluição plástica usando o sensoriamento remoto, a pouca quantidade de cenas conhecidas nas quais há presença confirmada de plástico constitui um grande limitador.

O sistema de modelagem de sensoriamento remoto é ilustrado na Figura 14. Enquanto o problema inverso parte da radiação adquirida pelo sensor e pretende dar estimativas acuradas das variáveis de interesse, o problema direto envolve os MTRs, que simulam a radiação transmitida através da atmosfera para uma determinada cena em uma determinada configuração de observação. A resolução do problema de inversão implica o desenho de algoritmos que, partindo da radiação adquirida pelo sensor, possam dar estimativas acuradas das variáveis de interesse, “invertendo” o MTR.

Figura 14 – Problemas direto (setas pretas) e inverso (setas amarelas) de sensoramento remoto.



Fonte: Adaptação de BIOUCAS-DIAS et al. (2013).

Nesta dissertação, a abordagem selecionada foi a criação de um conjunto de dados simulado por meio de um MTR, baseando-se nas evidências apresentadas por Ali et al. (2021), cujo estudo envolveu o treinamento de algoritmos com simulações de um MTR e sua validação com medições de campo, examinando a previsibilidade de três características de biodiversidade em florestas temperadas mistas. O estudo concluiu que algoritmos de inversão podem produzir uma relação de acurácia de previsão semelhante para valores *in situ* e modelados.

O uso de MTRs também promove a exploração do comportamento radiativo e do balanço energético das superfícies, algo essencial para estudar o funcionamento da natureza com informação adquirida à distância. Na maioria dos casos as superfícies não são isotrópicas, o que torna a consideração de sua natureza tridimensional (3D) indispensável na construção dos modelos de transferência radiativa.

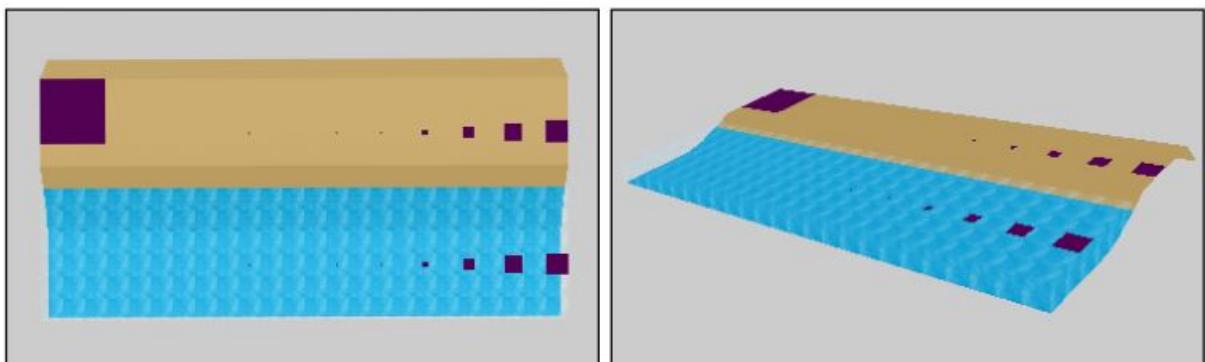
O DART, MTR selecionado para o trabalho, hoje é proposto como um dos modelos tridimensionais mais completos para simular a transferência radiativa e as observações de satélite das superfícies terrestres para qualquer configuração experimental e instrumental. Ele é usado em mais de 80 Centros Espaciais, Universidades e Centros de Pesquisa, e implementado através de uma ferramenta computacional que permite criar cenas sintéticas tridimensionais do sistema Terra-Atmosfera no plano do sensor para diferentes altitudes, da base ao topo da atmosfera. A Universidade de Toulouse III, na França, fornece licenças gratuitas do software para trabalhos científicos² (GASTELLU-ETCHEGORRY; GRAU; LAURET, 2012).

O DART é mais frequentemente utilizado para estudos envolvendo avaliação de áreas verdes (DUTHOIT et al., 2008; SEPULCRE-CANTO et al., 2009; JANOUTOVÁ et al., 2019; KÜKENBRINK et al., 2021). Contudo, ele permite a simulação de objetos com suas respectivas assinaturas espectrais, bem como das respostas de diferentes sensores, mostrando-se um candidato suficientemente robusto para utilização na pesquisa de poluição plástica em águas naturais. Ele também tem sido revisado e

²Disponível em: <<https://dart.omp.eu/>>

atualizado de forma a ampliar seu domínio de aplicação e melhorar sua acurácia. A Figura 15 ilustra um produto DART contendo água, areia e plástico.

Figura 15 – Cena DART contendo água, areia e plástico.



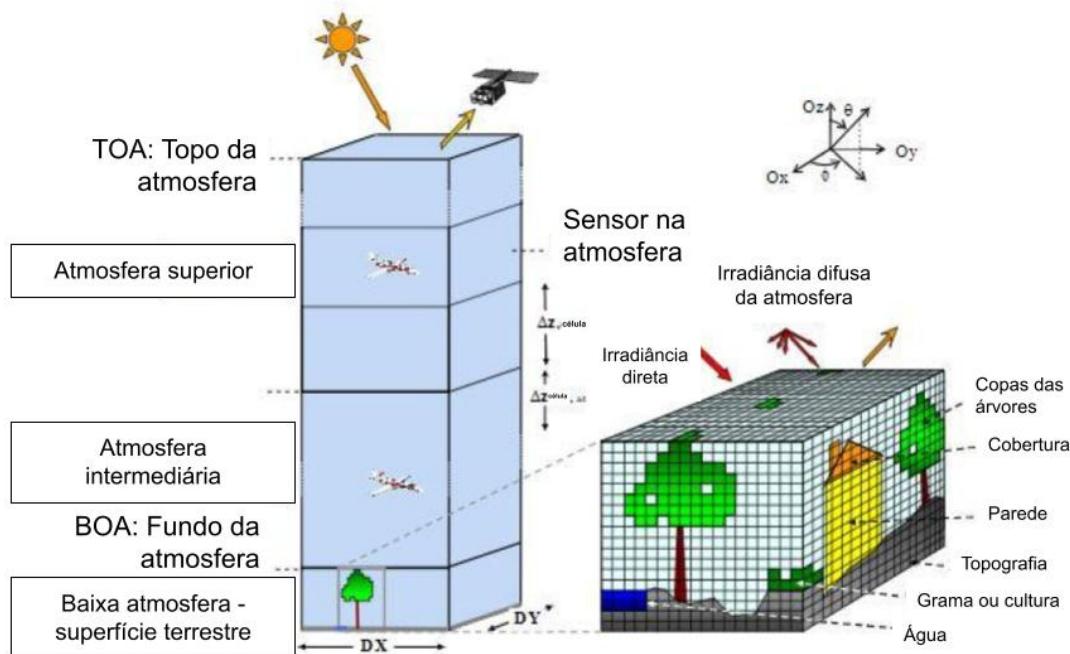
Fonte: Elaboração própria.

O DART atua nos domínios óptico e termal do sensoriamento remoto. A atmosfera pode ser simulada contendo gases, aerossóis e perfis de água, sem necessidade de modelo atmosférico, pois o modelo simula diretamente a transferência radiativa. As paisagens são simuladas como matrizes retangulares de células paralelepípedicas, conforme ilustrado na Figura 16.

Além de especificar os elementos da paisagem, também é necessário determinar o sensor a ser simulado. A Seção 3.2.1 detalha os estudos capazes de detectar poluição plástica em águas naturais via imageamento de satélite, determinando tanto quais são os conjuntos de imagens observadas disponíveis quanto qual o sensor mais indicado para as simulações, pois todos os estudos usaram a mesma família de missões: a Copernicus Sentinel, da União Européia. O sensor mais usado foi o MSI a bordo do Sentinel-2, que provê imageamento global em 13 bandas espectrais e fornece acesso aberto aos dados³, viabilizando seu uso em pesquisas sem dependência de orçamento (TOPOUZELIS; PAPAKONSTANTINOU; GARABA, 2019; BIERMANN et al., 2020; THEMISTOCLEOUS et al., 2020; TOPOUZELIS et al., 2020; PAPAGEORGIOU et al., 2022).

³Dados disponíveis para download em: <<https://scihub.copernicus.eu/>>

Figura 16 – Matriz de células DART para o sistema Terra/Atmosfera.



Fonte: Adaptação de GASTELLU-ETCHEGORRY; GRAU; LAURET (2012).

As diferentes resoluções espaciais das bandas do sensor - que variam entre 10, 20 ou 60 metros (m) - fazem com que bandas de diferentes resoluções apresentem diferentes números de pixels, que constituem as amostras básicas para a realização deste trabalho. Isso torna necessário uniformizar o número de amostras extraídos de cada banda. Conforme apresentado por Patel e Mistree (2013), isso pode ser feito através da interpolação, um processo de transferência de imagem de uma resolução para outra sem perda de qualidade. A interpolação pode ser realizada através de técnicas adaptativas, que levam em consideração características da imagem como valor de intensidade e informações sobre bordas e texturas, mas são mais intensivas computacionalmente, ou de técnicas não adaptativas, baseadas em manipulação direta dos pixels, seguindo um mesmo padrão para toda a imagem e exigindo um menor custo de cálculo. Neste trabalho, três métodos de interpolação não adaptativa muito difundidos na área do sensoriamento remoto foram utilizados:

- Interpolação do vizinho mais próximo: O pixel interpolado é substituído pelo pixel mais próximo. É um método simples de interpolação linear, fácil de implementar e que em geral oferece bons resultados quando a imagem tem pixels de alta resolução;
- Interpolação bilinear: Gera uma média ponderada dos 4 pixels mais próximos para calcular o valor interpolado. O resultado é uma imagem muito mais suave que a imagem original. Oferece um resultado melhor do que a interpolação do vizinho mais próximo e requer menos tempo computacional que a interpolação cúbica;
- Interpolação cúbica: Gera uma média ponderada de 16 pixels para calcular o valor interpolado. Esses pixels estão a diferentes distâncias do pixel desconhecido, e os pixels mais próximos recebem uma ponderação mais alta no cálculo. Tende a fornecer imagens mais nítidas que os dois métodos anteriores, mas requer maior tempo computacional.

As simulações no DART podem ser criadas como dados de reflectância de fundo de atmosfera, sem necessidade de correção atmosférica, diferentemente dos dados observados, que sempre necessitam deste tipo de correção. Alguns algoritmos de correção atmosférica também podem fazer a reamostragem de bandas, de forma a uniformizar a resolução espacial e o número de amostras por banda durante o processamento.

A qualidade da refletância da superfície é altamente dependente do método de correção atmosférica, das características da superfície e do projeto do sensor, existindo restrições práticas no nível de acurácia que pode ser alcançado (OKIN; GU, 2015). A correção atmosférica da água tem aspectos desafiadores: a maioria dos algoritmos voltados para cenas com cobertura de água foram desenvolvidos para sensores aplicados à cor do oceano, em satélites de resolução média ou baixa, limitando sua aplicação para águas mais complexas opticamente e levando ao uso de algoritmos de correção atmosférica originalmente desenvolvidos para a terra (WANG et al., 2019).

No caso da pesquisa de poluição plástica em águas naturais, a correção também pode remover o sinal dos plásticos ao confundi-los com espuma marinha ou aerossol (GARABA; DIERSSEN, 2018; TOPOUZELIS; PAPAKONSTANTINOU; GARABA, 2019).

Para produtos Sentinel, há dois algoritmos frequentemente usados: Sen2cor e Acolite. O Sen2Cor é disponibilizado pela Agência Espacial Europeia (ESA, do inglês *European Space Agency*) junto com a caixa de ferramentas Sentinel, software específico para tratamento de produtos desta família de missões. O Acolite⁴ foi desenvolvido por Vanhellemont e Ruddick (2016) para a correção atmosférica de imagens dos sensores Operational Land Imager (OLI) e MSI aplicadas a estudos marinhos e de águas interiores. Ele agrupa os algoritmos de correção atmosférica e softwares de processamento desenvolvidos pelo Instituto Real Belga de Ciências Naturais (RBINS, do inglês *Royal Belgium Institute of Natural Sciences*) para dados obtidos por satélites para aplicações em águas naturais (RBINS, 2022).

Topouzelis, Papakonstantinou, Garaba (2019) testaram ambos os algoritmos nas imagens marinhas com alvos artificiais de plástico, constatando que ambos foram capazes de realizar a correção atmosférica sem mascarar o sinal do plástico, embora os produtos Sen2Cor apresentassem uma refletância mais alta que os produtos Acolite. As diferenças percentuais não tendenciosas em diferentes bandas para cada alvo variaram entre 1% a 64%. Há estudos evidenciando que o Acolite pode oferecer bom desempenho em águas turvas (VANHELLEMONT; RUDDICK, 2021) e em águas interiores em determinadas condições (WANG et al., 2019). Porém, apesar destes resultados iniciais, outros estudos encontraram problemas com os algoritmos, optando por realizar a detecção do plástico na água usando os dados de refletância de topo de atmosfera, sem correção atmosférica, demonstrando a necessidade de mais estudos para validação (THEMISTOCLEOUS et al., 2020; WOLF et al., 2020).

O Acolite inclui dois algoritmos de correção atmosférica: o mais antigo *Exponential extrapolation* (EXP) e o padrão *Dark Spectrum Fitting* (DSF). O EXP assume a refletância da água como zero nas bandas SWIR, atribuindo qualquer sinal SWIR a

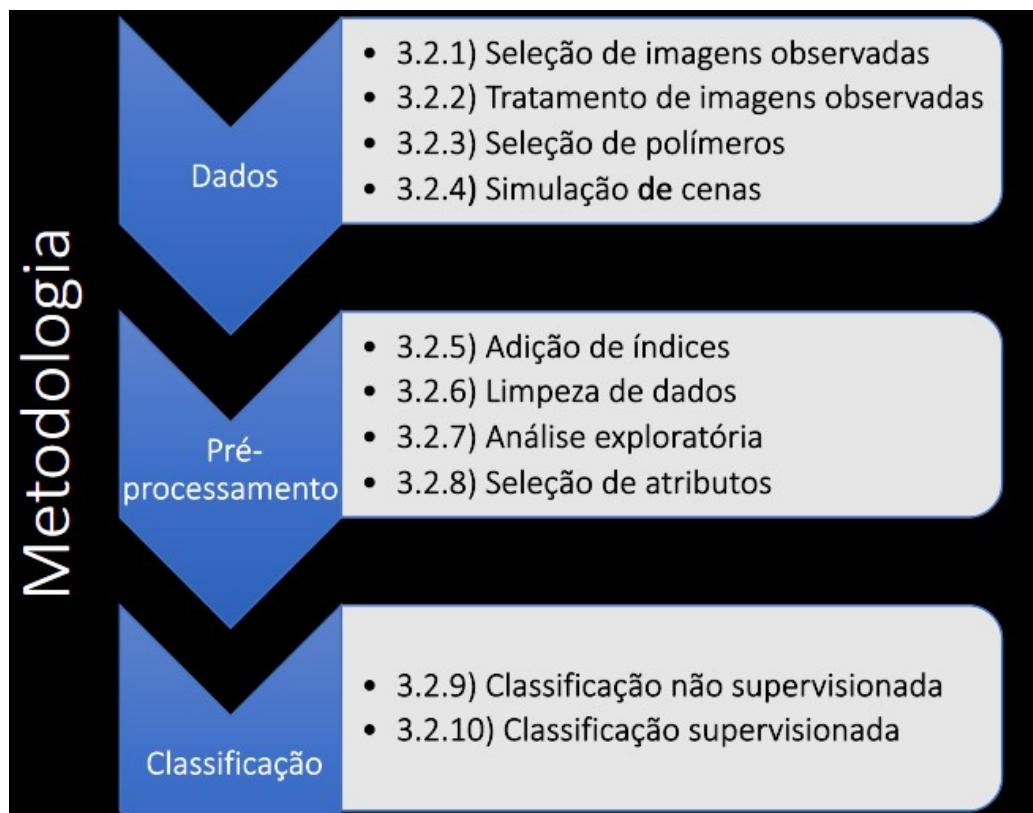
⁴O código-fonte mais recente em Python está disponível no GitHub <<https://github.com/acolite/acolite>>.

espalhamento Rayleigh ou aerossol. Já o DSF foi desenvolvido como uma proposta de correção atmosférica genérica e automatizada para imagens cobertas por água de sensores em escala métrica. Nele não há uma seleção de banda *a priori*, pois, para cada cena, a melhor banda para determinar a refletância é selecionada dinamicamente, com base em vários alvos escuros (VANHELLEMONT; RUDDICK, 2018). Além da correção atmosférica, o ACOLITE realiza o mascaramento de nuvens usando a banda de 1600 nm e a reamostragem das bandas para a melhor resolução espacial disponível. Para aplicações aquáticas, há evidências de melhor desempenho do DSF, que ao selecionar a banda mais apropriada automaticamente, evita a amplificação dos efeitos de brilho e adjacência na correção atmosférica (VANHELLEMONT, 2019).

3.2 METODOLOGIA

Nesta Seção, será apresentada a metodologia do presente trabalho. A Figura 17 mostra as principais fases e etapas que foram desenvolvidas ao longo da pesquisa. Na fase de Dados, foram realizados a seleção (Subseção 3.2.1) e o tratamento do banco de dados observado (Subseção 3.2.2), bem como a seleção dos polímeros para as simulações (Subseção 3.2.3) e a criação das cenas sintéticas (Subseção 3.2.4). Na fase de Pré-processamento, foram realizados a adição de índices (Subseção 3.2.5), a limpeza de dados (Subseção 3.2.6), a análise exploratória (Subseção 3.2.7) e a seleção de atributos (Subseção 3.2.8). Por fim, na fase de Classificação, foram feitas a classificação não supervisionada (Subseção 3.2.9) e a classificação supervisionada (Subseção 3.2.10).

Figura 17 – Metodologia utilizada.



Fonte: Elaboração própria.

3.2.1 Seleção de imagens observadas

Uma busca por imagens adquiridas por sensoriamento remoto orbital com confirmação *in situ* de presença de plástico em regiões costeiras foi realizada. Foram encontrados os experimentos abaixo.

1. Themistocleous et al., 2020: Neste trabalho, um aglomerado de lixo plástico medindo 3×10 m foi criado a partir de 1500 garrafas plásticas unidas por fio de nylon e emolduradas por tubos de PVC. O objeto foi instalado no mar, ao sul do Porto Velho de Limassol, no Chipre, durante a passagem do satélite Sentinel-2 no dia 15 de dezembro de 2018. Uma plataforma UAV com câmeras multiespectrais também fotografou o aglomerado, e um espectrorradiômetro foi usado para obter as assinaturas espectrais da superfície da água e do alvo de lixo plástico. A correção atmosférica foi testada usando o processador Sen2Cor, porém, ao aplicar a correção, houve absorção de informação referente às garrafas plásticas pelo algoritmo. Devido a isso, o trabalho foi realizado com as imagens sem correção atmosférica. O georreferenciamento do alvo plástico não foi fornecido pelos autores do estudo, de forma que sua localização só pode ser inferida por inspeção visual.
2. Topouzelis, Papakonstantinou, Garaba (2019) (*PLP 2018*): Neste estudo, um conjunto de três alvos artificiais flutuantes de plástico foi configurado nas águas perto da praia de Tsamakia, em Mitilene, na ilha grega de Lesvos. O material dos alvos consistia em garrafas PET, sacos plásticos de LDPE e redes de pesca em nylon. Cada alvo possuía 100 m^2 de área. O experimento foi conduzido em 07 de junho de 2018 durante a passagem do satélite Sentinel-2. Também foram adquiridas imagens com VANT, em resolução geoespacial mais alta que a do sensor do Sentinel-2, posteriormente utilizadas para melhorar o georreferenciamento das imagens de satélite e calcular sua cobertura percentual de plástico. Os efeitos dos algoritmos de correção atmosférica Sen2Cor e Acolite também foram avaliados. O georreferenciamento dos alvos plásticos não foi fornecido

pelos autores do estudo, de forma que sua localização só pode ser inferida por inspeção visual. Após este trabalho, experimentos similares passaram a ser realizados anualmente como parte do PLP, no âmbito do Marine Remote Sensing Group (MRSG), vinculado ao Departamento de Ciências Marinhas da Universidade do Egeu⁵, do qual parte dos autores do estudo são integrantes.

3. Topouzelis et al. (2020) (*PLP 2019*): Na sequência do trabalho descrito no item acima, este experimento foi realizado no mesmo local do ano anterior. Desta vez, um conjunto de alvos artificiais menores foi construído, em tamanhos de 1 x 5 m e 5 x 5 m, com conexões modulares. Estes tamanhos permitiram sua combinação em diferentes formatos, materiais de composição e percentuais de cobertura, com aquisição de dados remotos em configurações diversas e em várias datas ao longo do ano de 2019 (Figura 18, superior). Os materiais dos alvos incluíam garrafas transparentes (PET), sacos plásticos (Polietileno de baixa densidade, PEBD) e detritos naturais (juncos, *Arundo donax*). Os alvos foram georreferenciados com aparelhos de GPS, que possibilitavam obter sua localização exata. Foram adquiridas imagens do sensor MSI a bordo do satélite Sentinel-2 e, simultaneamente, com uma câmera com resolução geoespacial mais alta, aerotransportada via VANT, possibilitando a posterior sobreposição dos sinais GPS dos alvos às imagens georreferenciadas do satélite, para confirmar seu correto posicionamento. Cinco imagens de satélite Sentinel-2 capturaram o local de teste em boas condições climáticas, e os dados Nível-1C foram baixados e submetidos a correção atmosférica e mascaramento de nuvens usando Acolite v. 20190326.0. Os percentuais de cobertura plástica em cada pixel das imagens de satélite foram calculados com base nas fotografias adquiridas por VANT, por meio do software de segmentação eCognition Development 9.5.1. O comportamento espectral dos alvos plásticos foi investigado extraindo e analisando os valores dos pixels relevantes e comparando-os aos ambiente. Foram testadas as abordagens de separação espectral inversa e filtragem combinada nas imagens.

⁵Página do grupo: <<https://mrsg.aegean.gr/>>

Os resultados forneceram evidências de que, em condições adequadas, pixels com uma fração de abundância de PET de pelo menos 25% podem ser detectados com sucesso. Os dados do MSI/Sentinel-2, a geolocalização dos alvos e as porcentagens de presença de plástico para cada pixel foram fornecidas pelos autores, podendo ser obtidos na Página web do MRSG ou também no banco de dados Zenodo⁶, de forma que é possível localizar as imagens do experimento, mapear os pixels contendo os alvos de plástico, extrair as informações espectrais e relacioná-los com os respectivos percentuais de cobertura de cada classe.

4. Registro de experimento do *PLP 2020*: Ainda não havia artigo publicado sobre a edição de 2020 do PLP quando a busca foi realizada, porém o registro detalhado do experimento já estava disponível na Página do projeto⁷. Neste trabalho, alvos artificiais foram novamente implantados na Grécia, desta vez com formatos quadrados, circulares e cilíndricos de diferentes tamanhos. Foram realizadas medições de velocidade do vento e aquisições de imagens com VANT e pelos satélites Sentinel-2 e PRISMA. A maior parte dos alvos tinha tamanho inferior ao pixel do MSI, de forma que, sem os dados de geolocalização, não foi possível identificar os pixels contendo plástico por inspeção visual nas imagens.

5. Registro de experimento do *PLP 2021*: Ainda não havia artigo publicado sobre a edição de 2021 do PLP quando a busca foi realizada, porém o registro detalhado do experimento já estava disponível no site do projeto⁸. Neste trabalho, a implantação dos alvos aconteceu em uma nova localidade: uma pequena baía nas proximidades de Skala Loutron, na ilha grega de Lesvos. Dois grandes alvos de 28 m de diâmetro, sendo um coberto por malha de HDPE na cor branca e outro coberto por tábuas, foram ancorados na praia por um período de quatro meses (Figura 18, inferior), possibilitando múltiplas aquisições pelo Sentinel-2. Também foram adquiridas imagens de VANT para a produção de um mapa ortofotográfico da área de implantação. Adicionalmente, foram registradas a intensidade da luz

⁶Disponível em: <<https://zenodo.org/record/3752719#.Y0Rhh3bMKUk>>

⁷Disponível em: <<http://plp.aegean.gr/category/experiment-log-2020/>>

⁸Disponível em: <<http://plp.aegean.gr/category/experiment-log-2021/>>

incidente, a turbidez da água, a velocidade e a direção do vento, bem como foram coletados dados hiperespectrais da cena com um mini espectrorradiômetro. Foi possível observar o efeito de fenômenos como submersão e bioncrustação ao longo das aquisições. Este trabalho possui grande relevância, pois atende a um requisito levantado por Topouzelis et al. (2020): a construção de alvos artificiais com uma área efetiva grande o suficiente para garantir a cobertura total de pelo menos um pixel Sentinel-2 de 10×10 m, e se possível garantindo também a cobertura total de um pixel Sentinel-2 de 20×20 m, permitindo assim a análise adequada dos sinais SWIR e Red Edge, presentes em bandas com resolução espacial de 20 m. Dado o tamanho dos alvos, foi possível inferir sua localização por inspeção visual em um primeiro momento. Posteriormente, os mapas ortofotográficos em alta resolução geoespacial foram publicados no banco de dados Zenodo⁹, permitindo a confirmação da geolocalização dos alvos nas imagens. Embora seja possível identificar de forma acurada a localização dos alvos, os dados fornecidos pelos autores não incluíam os percentuais de cobertura plástica em cada pixel da imagem.

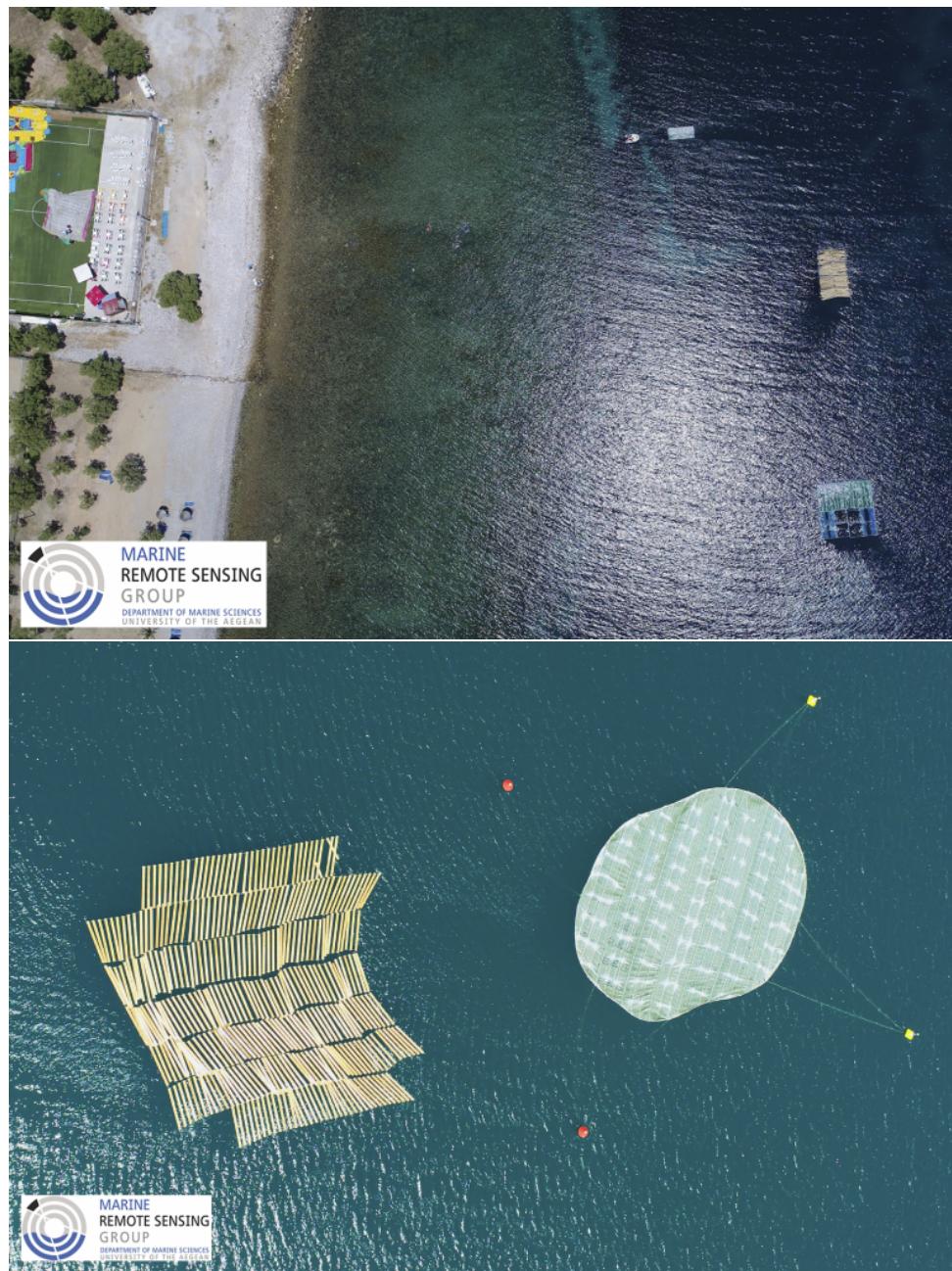
Os conjuntos de dados listados acima foram analisados na plataforma *Google Earth Engine*¹⁰ (GEE). Todas as imagens haviam sido obtidas pelo sensor MSI/Sentinel-2. A análise foi realizada usando os produtos nível 2A da plataforma. Somente os conjuntos 3 e 5 - referentes aos experimentos realizados no âmbito dos PLPs 2019 e 2021 - foram selecionados para o trabalho, por permitirem distinguir com clareza nas imagens os pixels onde os alvos artificiais estavam instalados. Os conjuntos de dados dos itens 1, 2 e 4 não foram considerados neste estudo a fim de evitar a introdução de dados incorretos ou com algum grau de incerteza na classificação.

Todas as cinco imagens do PLP 2019 foram selecionadas. As datas eram 18/04/2019, 03/05/2019, 18/05/2019, 28/05/2019 e 07/06/2019. Com relação ao PLP 2021, que possuía um conjunto maior de imagens, foram selecionadas cinco datas representativas do conjunto, que estivessem com boas condições climáticas e demonstrando as

⁹Disponível em: <<https://zenodo.org/record/7085112#.Y1ftq3bMK3B>>

¹⁰Disponível em: <<https://code.earthengine.google.com/>>

Figura 18 – Alvos artificiais implantados na Grécia durante os PLPs 2019 (superior) e 2021 (inferior) em fotografias aéreas adquiridas por VANT.



Fontes: <https://tinyurl.com/4yyr6s88> (superior); <https://tinyurl.com/ypmtsajb> (inferior).

transformações no alvo ao longo dos meses. As datas selecionadas foram 21/06/2021, 01/07/2021, 06/07/2021, 21/07/2021 e 25/08/2021.

Os produtos MSI/Sentinel-2 nível L1 referentes às imagens selecionadas foram obtidos através do portal do Serviço Geológico dos Estados Unidos¹¹ (USGS) ou do *Copernicus Open Access Hub*¹², conforme disponibilidade. O relatório do processo de seleção de imagens pode ser visto detalhadamente nas Tabelas 1 e 2.

Tabela 1 – Análise da situação das imagens de 2018 e 2019.

Data	Referência	Fonte	Selecionada?	Problemas encontrados
07/06/2018	PLP 2018	USGS	Não	Incerteza na localização dos alvos.
18/04/2019	PLP 2019	USGS	Sim	-
03/05/2019	PLP 2019	USGS	Sim	-
18/05/2019	PLP 2019	USGS	Sim	-
28/05/2019	PLP 2019	Copernicus Open Access Hub	Sim	-
07/06/2019	PLP 2019	USGS	Sim	-

Fonte: Elaboração própria.

3.2.2 Tratamento de imagens observadas

As imagens nível L1 selecionadas passaram por correção atmosférica com o processador Acolite¹³ usando o algoritmo DSF. O Acolite, além da correção atmosférica, fez também a reamostragem de todas as bandas da imagem para 10 m, melhor resolução disponível.

As regiões de interesse foram definidas para cada data, e as imagens foram recortadas no software SNAP¹⁴. Primeiramente, passaram por um recorte de 120 x 120 pixels sobre a região de interesse. Depois, foram extraídos dois subconjuntos de pixels do primeiro recorte: um contendo os alvos e a água circundante, em geral em um tamanho de 12x12 pixels (um pouco maior em algumas datas), e outro contendo um trecho da superfície terrestre, com tamanho de 10x10 pixels, em diferentes locais de extração para cada data. Os subconjuntos foram exportados para o formato TIFF.

Para as datas de 2019, a geolocalização dos alvos plásticos estava disponível para download no formato *shapefile* (SHP), conforme mencionado anteriormente, bem

¹¹Disponível em: <<https://earthexplorer.usgs.gov/>>

¹²Disponível em: <<https://scihub.copernicus.eu/dhus/>>

¹³Disponível em: <<https://github.com/acolite/acolite/releases/tag/20220222.0>>

¹⁴Disponível em: <<https://step.esa.int/main/download/snap-download/>>

Tabela 2 – Análise da situação das imagens de 2021.

Data	Referência	Fonte	Selecionada?	Status	Bioincrustação	Problemas encontrados
11/06/2021	PLP 2021	USGS	Não	Flutuante	Ausente	Instalação dos alvos ainda em andamento.
16/06/2021	PLP 2021	USGS	Não	Informação ausente	Informação ausente	Cobertura de nuvens.
21/06/2021	PLP 2021	USGS	Sim	Flutuante	Ausente	-
26/06/2021	PLP 2021	USGS	Não	Flutuante	Não	Similar às selecionadas (redundante).
01/07/2021	PLP 2021	USGS	Sim	Flutuante	Ausente	-
06/07/2021	PLP 2021	USGS	Sim	Flutuante	Baixa	-
11/07/2021	PLP 2021	USGS	Não	Flutuante	Média	Similar às selecionadas (redundante).
16/07/2021	PLP 2021	USGS	Não	Flutuante	Média	Similar às selecionadas (redundante).
21/07/2021	PLP 2021	USGS	Sim	Flutuante	Média / alta	-
26/07/2021	PLP 2021	USGS	Não	Flutuante	Média / alta	Similar às selecionadas (redundante).
31/07/2021	PLP 2021	USGS	Não	Flutuante	Alta	Similar às selecionadas (redundante).
05/08/2021	PLP 2021	USGS	Não	Flutuante	Alta	Similar às selecionadas (redundante).
10/08/2021	PLP 2021	USGS	Não	Flutuante	Alta	Similar às selecionadas (redundante).
15/08/2021	PLP 2021	Copernicus Open Access Hub	Não	Alta	Submerso	Similar às selecionadas (redundante).
20/08/2021	PLP 2021	USGS	Não	Submerso	Alta	Similar às selecionadas (redundante).
25/08/2021	PLP 2021	USGS	Sim	Parcialmente submerso	Alta	-
30/08/2021	PLP 2021	USGS	Não	Parcialmente submerso	Baixa	Objetos adicionais presentes na imagem.
04/09/2021	PLP 2021	USGS	Não	Flutuante misto	Baixa	Baixa visibilidade dos alvos.
09/09/2021	PLP 2021	Copernicus Open Access Hub	Não	Parcialmente submerso misto	Média	Baixa visibilidade dos alvos.
14/09/2021	PLP 2021	USGS	Não	Majoritariamente submerso misto	Média	Baixa visibilidade dos alvos.
19/09/2021	PLP 2021	USGS	Não	Majoritariamente submerso misto	Média	Baixa visibilidade dos alvos.
24/09/2021	PLP 2021	USGS	Não	Majoritariamente submerso	Média / alta	Baixa visibilidade dos alvos.
29/09/2021	PLP 2021	USGS	Não	Informação ausente	Informação ausente	Cobertura de nuvens.
04/10/2021	PLP 2021	USGS	Não	Majoritariamente submerso	Alta	Baixa visibilidade dos alvos.

Fonte: Elaboração própria.

Nota: Conforme categorização informada pelos autores do estudo, para a coluna "Status", considera-se "Flutuante" quando o alvo está no topo e acima da superfície; "Parcialmente submerso" quando o alvo está majoritariamente na superfície com algumas partes abaixo dela, entre 0.1 e 0.5 metro de profundidade; "Majoritariamente submerso" quando o alvo está abaixo da superfície em no máximo 1 metro de profundidade; "Submerso" quando a maior parte do alvo está em profundidade maior que 1 metro. A informação sobre alvo "misto" na coluna "Status" refere-se a ocasiões em que os pesquisadores sobrepuiseram os alvos de malha polimérica e madeira (nenhuma imagem com alvo misto foi selecionada para este trabalho). Para a coluna "Bioincrustação", também consideram-se os valores categorizados pelos autores do estudo (PAPAGEORGIOU et al., 2022).

como os respectivos percentuais de cobertura por pixel para cada alvo. Para as datas de 2021, em um primeiro momento a imagem do dia 01/07/2021 foi usada como refe-

rência para a criação de um SHP na plataforma GEE, marcando os pixels de plástico e madeira. O SHP foi usado na classificação dos pixels em todas as datas de 2021. Posteriormente, com a disponibilização das ortofotos georreferenciadas, a localização dos alvos foi corrigida. Não havia informações sobre percentuais de cobertura dos pixels pelos alvos para 2021, então eles foram estimados por inspeção visual em duas categorias: 100% (cobertura total do pixel) ou menor que 100% (cobertura parcial, com percentual desconhecido).

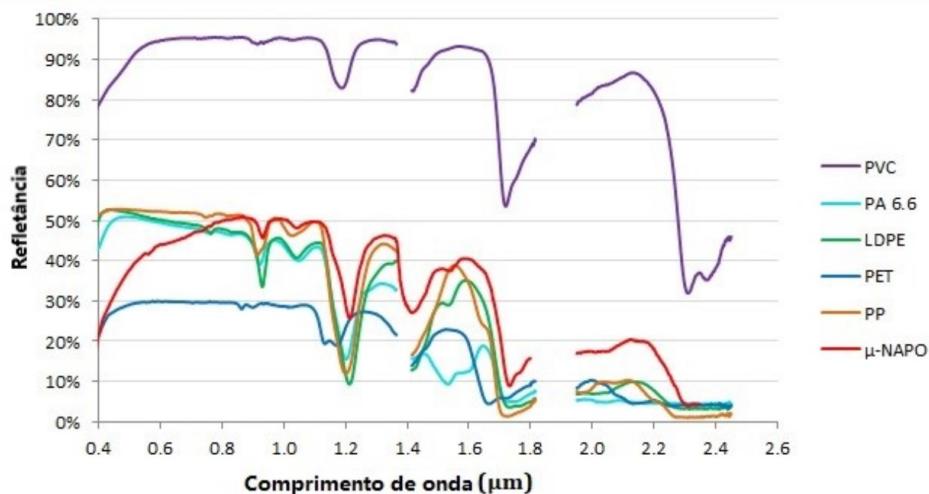
Os SHPs e as ortofotos foram sobrepostos aos respectivos arquivos TIFF das imagens de satélite no software QGIS¹⁵, permitindo o mapeamento dos pixels. As informações de mapeamento e os arquivos TIFF foram carregados em um programa Python, de forma que os dados relacionados à classificação e aos percentuais de cobertura dos pixels foram integrados aos dados de reflectância.

3.2.3 Seleção de polímeros

A seleção de um conjunto de polímeros com suas respectivas assinaturas especiais é necessária para a criação das imagens simuladas. Os polímeros neste trabalho foram selecionados a partir do trabalho de Garaba e Dierssen (2018), são eles: Poliamida 6 e 6.6 ou Nylon (PA 6 e PA 6.6), Policloreto de vinila (PVC), Polietieleno de baixa densidade (LDPE), PET e Polipropileno (PP). Além destes polímeros virgens, também foi selecionada uma assinatura média composta por diversos microplásticos coletados no Oceano Pacífico (μ -NAPO), referenciadas no mesmo estudo (GARABA; DIERSSEN, 2017; 2019a; 2019b). Estes polímeros foram escolhidos como uma amostra dos elementos mais frequentemente encontrados nos ambientes costeiros e marinhos. Suas assinaturas podem ser vistas na Figura 19. Percebe-se que as assinaturas possuem características singulares, com zonas de absorção centradas em 931, 1215, 1417 e 1732 nm. Na Tabela 3 estão listados os polímeros de interesse, suas siglas, características físicas e utilidades mais comuns.

¹⁵Disponível em:<https://qgis.org/pt_BR/site/>

Figura 19 – Assinaturas espetrais do μ -NAPO (composto pela média de um conjunto de microplásticos coletados no Oceano Pacífico) e dos polímeros virgens PVC, PA 6.6, LDPE, PET e PP.



Fonte: GARABA; DIERSSEN (2018).

3.2.4 Simulação das cenas

As simulações DART foram criadas através da ferramenta computacional específica do modelo. Foram criadas cenas tridimensionais de 1200 m x 1200 m, totalizando 720 km² de área cada. O sensor simulado foi o MSI/Sentinel-2, que compreende dois satélites idênticos - S2A e S2B - operando simultaneamente, faseados a 180° entre si, em uma órbita polar, sincronizada com o sol, a uma altitude média de 786 km. Os limites de cobertura do Sentinel-2 estão entre as latitudes 56° sul e 84° norte. A coleta de dados nesta região inclui toda a superfície terrestre e costeira, e também ilhas com mais de 100 km² de área ou com até 20 km de distância do litoral, bem como todas as ilhas da União Europeia, o Mar Mediterrâneo, todos os corpos d'água interiores e todos os mares fechados.

O sensor MSI funciona passivamente, coletando a luz solar refletida pela Terra. Seu design óptico permite um campo de visão de 290 KM. Os produtos Sentinel são classificados em diferentes níveis:

Tabela 3 – Polímeros de interesse, siglas, características físicas e utilidades comuns.

Polímero	Sigla	Características	Utilidades
Polietileno Tereftalato	PET (C-PET) (A-PET)	C-PET cristalino, opaco, pouco resistente a impactos, impermeável a gases, absorve água, resistente a produtos químicos, pouco resistente a bases, água quente, cetonas e halogenados. A-PET transparente, resistente a impactos, menos rígido, resiste a hidrocarbonetos halogenados.	Garrafas de bebidas, embalagem de alimentos, cosméticos e remédios, chapas, folhas e películas.
Polietieleno de baixa densidade	PEBD (LDPE)	Resistente à substâncias químicas, variadas características mecânicas, mantém propriedades mesmo quando submetido à temperaturas $\leq 60^{\circ}\text{C}$.	Isolantes de fios, sacos, garrafas, tampas, bolsas, revestimentos internos e externos caixas de leite, tubulações, mangueiras, telas de sombreamento.
Polipropileno	PP	Elevada resistência química a solventes, moldável, alta resistência mecânica, boa resistência ao impacto acima de 15°C , boa estabilidade térmica.	Brinquedos, copos plásticos, recipientes para alimentos, remédios e produtos químicos, tubos para cargas de canetas, carpetes, seringas, material hospitalar esterilizável, autopeças, cabos para ferramentas manuais.
Policloreto de Vinila	PVC	Conhecido como vinil. Essa resina consegue assumir tanto formas rígidas como flexíveis, graças aos aditivos. É leve e impermeável, é isolante térmico, elétrico e acústico. Boa resistência química, resistente ao fogo e às intempéries.	Filmes, laminados, soldados de calçados, brinquedos, embalagens para remédios (blister), mangueiras, tubulação para água e esgoto, revestimento de fios e cabos, cateteres e bolsas de soro.
Poliamida 6 e Poliamida 6.6 (Nylon)	PA 6 e PA 6.6	Possui alta resistência à fadiga, boa resistência ao impacto, alta temperatura de fusão, baixo coeficiente de atrito, resistência às intempéries, ótimas propriedades mecânicas e alta fluidez.	Engrenagens, peças automotivas, buchas, lacre e roupas.

Fonte: Elaboração própria.

- Level-0: Dados brutos compactados no formato Instrument Source Packet (ISP). Contém uma estrutura de metadados que descreve o produto e um conjunto consistente de ISPs. Este nível não é disponibilizado para os usuários.
- Level-1A: É obtido descompactando os dados da imagem bruta Level-0. Um modelo geométrico é desenvolvido, permitindo a localização de qualquer pixel da imagem. Este nível não é disponibilizado para os usuários.
- Level-1B: Fornece imagens corrigidas radiometricamente em valores de radiação no Topo da Atmosfera (TOA) e na geometria do sensor. Além disso, inclui

a geometria refinada que é usada para gerar o produto Nível-1C. Este nível é o mais baixo disponibilizado para os usuários.

- Level-1C: Resulta do uso de um Modelo Digital de Elevação (MDE) para projetar a imagem em coordenadas cartográficas. Medições radiométricas por pixel são fornecidas em reflectâncias do topo da atmosfera (TOA) com todos os parâmetros para transformá-los em radiâncias.
- Level-2A: Fornece imagens de reflectância de fundo de atmosfera (FOA) derivadas dos produtos Level-1C associados. Os produtos de nível 2A não são gerados sistematicamente, mas podem ser obtidos pelo usuário por meio da Caixa de ferramentas do Sentinel-2 usando como entrada o produto Level-1C associado.

A frequência de revisita de cada satélite individual é de 10 dias e a revisita da constelação combinada é de 5 dias. A resolução radiométrica do MSI é de 12 bits, permitindo que a imagem seja adquirida em uma faixa de 0 a 4095 valores de intensidade de luz. A resolução espacial varia de 10 a 60 m (Tabela 4):

- A resolução de 10 m abrange bandas localizadas nas regiões do VIS e NIR: Blue (B2, 496 nm), Green (B3, 560 nm), Red (B4, 665 nm) e NIR1 (B8, 835 nm, originalmente NIR e aqui referenciada como NIR1);
- A resolução de 20 m abrange bandas localizadas nas regiões do NIR e SWIR: RedEdge1 (B5, 704 nm), RedEdge2 (B6, 740 nm), RedEdge3 (B7, 782 nm), NIR2 (B8A, 865 nm, originalmente RedEdge4 e aqui referenciada como NIR2), SWIR1 (B11, 1614 nm) e SWIR2 (B12, 2202 nm);
- A resolução espacial de 60 m atua nas bandas B1 (Aerosols, 443 nm), B9 (Water vapor, 940 nm) e B10 (Cirrus, 1375 nm), estando uma na região VIS, uma na região NIR e uma na região SWIR (ESA, 2015).

Para unificar as nomenclaturas dos dados gerados no MTR DART e adquiridas pelo MSI/Sentinel-2, a banda B8, originalmente referenciada como NIR, é chamada

Tabela 4 – Bandas do MSI/Sentinel-2.

Banda	Região	Nome	Comprimento de onda central do S2A (nm)	Comprimento de onda central do S2B (nm)	Resolução (m)
B1	-	Aerosols	443,9	442,3	60
B2	VIS	Blue	496,6	492,1	10
B3	VIS	Green	560	559	10
B4	VIS	Red	664,5	665	10
B5	Red Edge	Red Edge 1	703,9	703,8	20
B6	Red Edge	Red Edge 2	740,2	739,1	20
B7	Red Edge	Red Edge 3	782,5	779,7	20
B8	NIR	NIR	835,1	833	10
B8A	NIR	Red Edge 4	864,8	864	20
B9	-	Water vapor	945	943,2	60
B10	-	Cirrus	1373,5	1376,9	60
B11	SWIR	SWIR 1	1613,7	1610,4	20
B12	SWIR	SWIR 2	2202,4	2185,7	20

Fonte: Adaptação de <<https://tinyurl.com/u8evfn7k>>.

neste trabalho de NIR1, enquanto a banda B8A, originalmente chamada de RedEdge4, aqui é referenciada como NIR2.

Os objetos plásticos foram criados com forma quadriculada e distribuídos na superfície de corpos d'água com dimensões de $100m^2$ (que cobre 100% do pixel do sensor MSI com resolução espacial de $10m$), $64m^2$ (80%), $36m^2$ (60%), $16m^2$ (40%), $4m^2$ (20%). As assinaturas espectrais descritas na Subseção 3.2.3 foram inseridas nestes objetos. As assinaturas espectrais da areia e da água foram fornecidas pelo próprio DART. Foram consideradas as bandas com resolução espacial de 10 e 20 m (Blue, Green, Red, RedEdge1, RedEdge2, RedEdge3, NIR1, NIR2, SWIR1, SWIR2).

Foi necessário criar uma simulação para cada uma das 5 proporções de cobertura com cada uma das 6 assinaturas espectrais de plástico selecionadas, totalizando 30 imagens sintéticas. Cada imagem foi simulada para cada uma das 10 bandas selecionadas, gerando 300 produtos de sensoriamento remoto simulado.

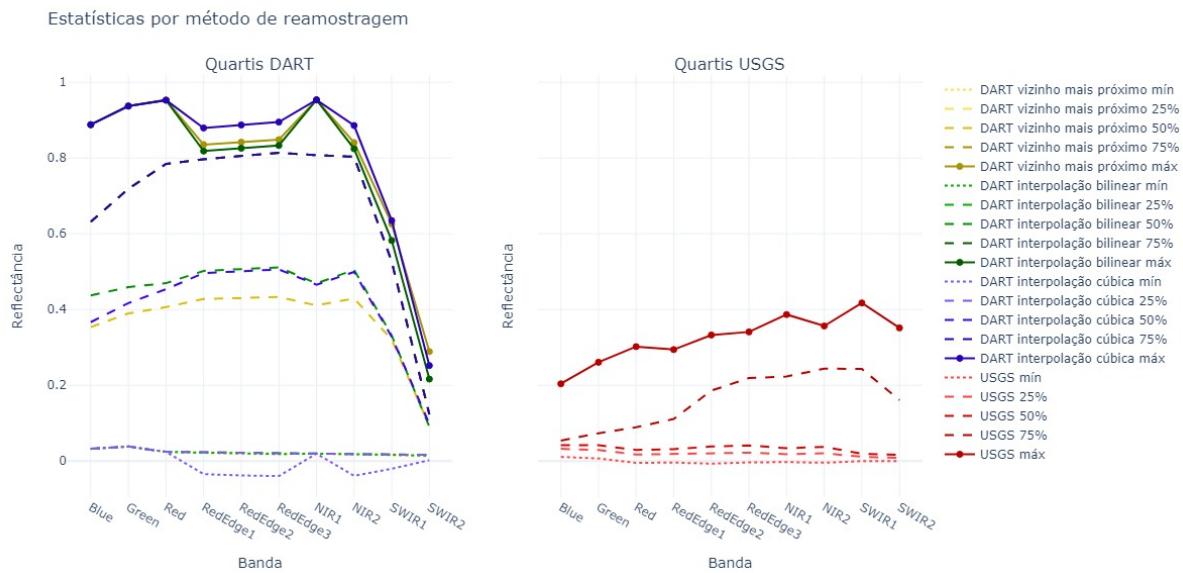
Para as bandas com resolução espacial de 20 m, cada cena possui 60 pixels x 60 pixels, totalizando 3600 pixels, dentre os quais 256 são de plástico, 1672 são de água e 1672 são de areia. Para as bandas com resolução espacial de 10 m, cada cena possui 120 pixels x 120 pixels, totalizando 14.400 pixels, dentre os quais 256 pixels são de plástico, 7.072 são de água e 7.072 são de areia. O tamanho e a localização dos objetos plásticos são iguais em ambas as resoluções espaciais.

As informações espaciais e espectrais das simulações foram exportadas em formato CSV e carregadas em um programa Python. A compilação das amostras em uma base de dados torna necessário adotar uma resolução espacial única; consequentemente, os pixels das bandas de 20 m foram reamostrados para a resolução de 10 m. Os métodos de reamostragem vizinho mais próximo, interpolação bilinear e interpolação cúbica foram testados, usando a classe *Resampling*¹⁶ que faz parte da biblioteca Python *Rasterio* (SEAN et al., 2013), com objetivo de adotar aquela que gerasse a distribuição que mais se aproximasse do conjunto de imagens observadas selecionadas na primeira etapa da metodologia, descrita na Seção 3.2.1.

As médias e desvio padrão dos conjuntos gerados usando os três métodos de reamostragem nas bandas de 20 m apresentaram diferenças ínfimas. Analisando os quartis dos conjuntos (Figura 20), percebe-se que os valores gerados pelos três métodos aplicados aos dados simulados são muito próximos, e em todos os casos possuem uma amplitude maior que os dados observados. A interpolação cúbica apresentou a maior amplitude entre os métodos, gerando inclusive valores negativos no limite inferior. O método selecionado foi a interpolação bilinear, por apresentar uma amplitude ligeiramente inferior aos outros dois métodos, ficando um pouco mais próxima da distribuição dos dados observados. O conjunto simulado reamostrado com a interpolação bilinear foi utilizado para todas as etapas posteriores. Além da reamostragem, os valores do conjunto de dados também foram arredondados para quatro casas decimais.

¹⁶Documentação disponível em: <<https://rasterio.readthedocs.io/en/latest/topics/resampling.html>>

Figura 20 – Quartis dos conjuntos de dados DART (simulados) e USGS (observados) após reamostragem.



Fonte: Elaboração própria.

Nota: À esquerda, quartis dos conjuntos de dados simulados após reamostragem das bandas de 20 m para 10 m pelos métodos vizinho mais próximo (amarelo), interpolação bilinear (verde) e interpolação cúbica (azul). À direita, quartis do conjunto de dados observado após reamostragem das bandas de 20 m para 10 m pelo algoritmo Acolite.

3.2.5 Adição de índices

Os índices radiométricos citados na Seção 3.1.2 foram calculados para os dados observados e simulados, e adicionados aos conjuntos como novos atributos. Primeiro, foram incluídos os índices NDWI (do inglês, *Normalized Difference Water Index*, Equação 5), WRI (do inglês, *Water Ratio Index*, Equação 6), NDVI (do inglês, *Normalized Difference Vegetation Index*, Equação 7), AWEI (do inglês, *Automated Water Extraction Index*, Equação 8), MNDWI (do inglês, *Modified Normalization Difference Water Index*, Equação 9), SR (do inglês, *Simple Ratio*, Equação 10), PI (do inglês, *Plastic Index*, Equação 11) e RNDVI (do inglês, *Reversed Normalized Difference Vegetation Index*, Equação 12). Estes índices foram avaliados por Themistocleous et al (2020), e foram incluídos aqui para validação.

$$\text{NDWI} = \frac{B3 - B8}{B3 + B8} \quad (5)$$

$$\text{WRI} = \frac{B3 + B4}{B8 + B12} \quad (6)$$

$$\text{NDVI} = \frac{B8 - B4}{B8 + B4} \quad (7)$$

$$\text{AWEI} = 4(B3 - B12) - 0.25B8 - 2.75B11 \quad (8)$$

$$\text{MNDWI} = \frac{B3 - B12}{B4 + B12} \quad (9)$$

$$\text{SR} = \frac{B8}{B4} \quad (10)$$

$$\text{PI} = \frac{B8}{B8 + B4} \quad (11)$$

$$\text{RNDVI} = \frac{B4 - B8}{B4 + B8} \quad (12)$$

Equações 5 a 12 – Índices NDWI, WRI, NDVI, AWEI, MNDWI, SR, PI e RNDVI, respectivamente (THEMISTOCLEOUS et al., 2020).

Além dos índices descritos pelas Equações 5 a 12, havia também o NDMI (do inglês, *Normalization Difference Moisture Index*), mas ele não foi incluído nas análises por apresentar uma formulação matemática idêntica à do índice NDWI no artigo. O índice FDI (do inglês *Floating Debris Index*, Equação 13, proposto por Biermann et al. (2020), também foi incluído nos conjuntos de dados.

$$\begin{aligned}
 FDI &= R_{rs,NIR} - R'_{rs,NIR} \\
 R'_{rs,NIR} &= R_{rs,RE2} + (R_{rs,SWIR1} - R_{rs,RE2}) \times \frac{(\lambda_{NIR} - \lambda_{RED})}{(\lambda_{SWIR1} - \lambda_{RED})} \times 10
 \end{aligned} \tag{13}$$

Equação 13 – Índice FDI (BIERMANN et al., 2020).

3.2.6 Limpeza de dados

Os conjuntos de dados observados e simulados originalmente não possuíam inconsistências, incompletudes, ruídos, redundâncias ou ausências perceptíveis. Contudo, o cálculo dos índices gerou divisões por zero em alguns casos. Foram identificadas 512 pixels de água com este problema no conjunto simulado e 20 no conjunto observado, sendo um de plástico e 19 de água. Como os quantitativos eram pequenos em relação aos totais, as amostras afetadas foram desconsideradas nas etapas subsequentes.

3.2.7 Análise exploratória

A análise exploratória de dados foi baseada principalmente em métodos visuais, com a geração de gráficos capazes de descrever os conjuntos de dados. Foram exploradas medidas de localidade, espalhamento e distribuição para ambos os conjuntos, e analisadas as relações entre conjuntos e atributos.

A escolha dos gráficos gerados foi guiada pelas seguintes perguntas:

- Qual a distribuição das classes nos conjuntos de dados?
- Como os conjuntos de dados observado e simulado se assemelham ou diferem entre si?
- Quais atributos/características são mais determinantes nas semelhanças e diferenças entre os conjuntos de dados observado e simulado?

Os conjuntos de dados observado e simulado também foram submetidos a um teste de Kolmogorov-Smirnov (DODGE, 2008), utilizado para verificar se as amostras de dois conjuntos pertencem a uma mesma distribuição.

3.2.8 Seleção de atributos

A primeira etapa da seleção de atributos foi feita usando o classificador Floresta Aleatória (RF, do inglês *Random Forest*), um estimador que gera classificações combinando técnicas de amostragem de dados com várias Árvores de Decisão, modelos que predizem variáveis alvo aprendendo regras de decisão simples inferidas a partir dos atributos de entrada. O RF foi escolhido para seleção de atributos porque, além da classificação, ele também gera uma pontuação de importância para cada atributo, relacionado a sua influência no resultado: quanto maior a pontuação, mais relevante o atributo. O algoritmo foi utilizado duas vezes, com os seguintes conjuntos de dados:

- Conjunto completo de dados simulados;
- Conjunto contendo somente as classes água e plástico dos dados simulados.

Os conjuntos de dados foram divididos em 75% para treinamento e 25% para teste. O valor de profundidade máxima selecionado para as árvores foi de 3 níveis, por ser o menor valor a alcançar a acurácia máxima na classificação dos dados simulados. O processo de treinamento e teste foi o mesmo para os dois conjuntos de dados. Ao final, os cálculos de importâncias dos atributos para cada um dos conjuntos foram colocados em gráficos.

As importâncias dos atributos foram comparadas com as matrizes de correlação deles. A correlação foi analisada tanto nos dados como um todo quanto classe a classe, com objetivo de selecionar conjuntos de atributos com alta pontuação atribuída pelo RF e também com baixa correlação entre si.

3.2.9 Classificação não supervisionada

O algoritmo de classificação não supervisionada utilizado foi o K-Means, implementado no Scikit-learn através da classe KMeans. Trata-se de um algoritmo que agrupa os dados tentando separar as amostras em n grupos de variância igual. Ele divide um conjunto de amostras em clusters disjuntos, cada um descrito pela média das

amostras contidas no agrupamento. As médias são comumente chamadas de “centróides” dos clusters, e o K-means busca escolher centróides que minimizem a inércia, ou critério de soma dos quadrados dentro do cluster (Equação 14). O algoritmo requer que o número de clusters seja especificado.

$$\sum_{i=0}^n \min_{\mu_j \in C} (\|x_i - \mu_j\|^2) \quad (14)$$

Após a inicialização, o K-means consiste em uma repetição de duas etapas: a primeira atribui cada amostra ao seu centróide mais próximo, e a segunda cria novos centróides tomando o valor médio de todas as amostras atribuídas a cada centróide anterior. A diferença entre os centróides antigos e os novos é calculada e o algoritmo repete essas duas últimas etapas até que esse valor seja menor que um determinado limite. Em outras palavras, ele se repete até que os centróides não se movam significativamente. O K-Means é altamente dependente da inicialização dos centróides. Como resultado, é comum que o cálculo seja feito várias vezes, com diferentes inicializações dos centróides (PEDREGOSA et al., 2011).

A classificação não supervisionada com K-Means foi testada com todos os conjuntos de atributos e incluindo todas as classes presentes nos dados, tanto para os conjuntos simulados quanto para os conjuntos observados. Em ambos os conjuntos, a classificação foi feita inicialmente com 3, 4 e 5 clusters, números próximos ao número de classes dos conjuntos, visando analisar as tendências de agrupamento. Todas as classificações foram realizadas com a mesma inicialização, com objetivo de avaliar as variações geradas pelas diferentes configurações iniciais com ambos os conjuntos.

A análise dos agrupamentos foi guiada pelas seguintes questões:

- Para um mesmo conjunto de atributos, o número de clusters mudou a classificação significativamente em algum caso?
- Para cada número de clusters, qual o conjunto de atributos classificou os dados com maior separabilidade?

- Quais as características dos pixels associados aos agrupamento maiores e menores? Há padrões de associação entre diferentes classes?

3.2.10 Classificação supervisionada

A etapa de classificação supervisionada começou com a seleção do classificador e dos respectivos hiperparâmetros. O classificador supervisionado escolhido foi a Rede Neural Artificial (RNA). A implementação foi realizada em Python usando a classe MLPClassifier do pacote Scikit-learn, que implementa um algoritmo *Multi-Layer Perceptron* (MLP) focado em problemas de classificação.

O MLP aprende uma função a partir de um treinamento que utiliza duas matrizes. Uma delas é bidimensional, contendo os atributos (colunas) e as amostras de treinamento (linhas), representadas como vetores dos atributos. A outra matriz é unidimensional, contendo os valores de destino (rótulos de classe) para as amostras de treinamento. Na versão implementada no MLPClassifier, o MLP possui uma camada de entrada formada por um conjunto de neurônios representando os atributos dos dados de treinamento. A camada de entrada é conectada a uma camada de saída, podendo haver uma ou mais camadas não lineares entre elas, chamadas de camadas ocultas. Cada neurônio na camada oculta transforma os valores da camada anterior com uma soma linear ponderada, seguida por uma função de ativação não linear.

O treinamento utiliza a retropropagação do cálculo de erro e do ajuste dos pesos, podendo optar, no caso do MLPClassifier, por um dos três algoritmos listados a seguir. O parâmetro alfa também é utilizado para regularização, penalizando pesos com grandes magnitudes, o que ajuda a evitar o sobreajuste (quando um modelo se ajusta muito bem ao conjunto de treinamento mas se mostra ineficaz para prever novos resultados).

- Gradiente Descendente Estocástico (SGD, do inglês *Stochastic Gradient Descent*): Método iterativo para otimizar uma função objetivo que atualiza os parâmetros usando o gradiente da função de perda em relação a um parâmetro que precisa de adaptação.

- Adam: Otimizador estocástico (como o SGD) capaz de ajustar automaticamente a atualização de parâmetros com base em estimativas adaptativas.
- L-BFGS: Otimizador da família quasi-Newton que usa uma estimativa da matriz Hessiana inversa para direcionar sua busca através do espaço variável.

A seleção dos melhores hiperparâmetros foi feita usando a classe GridSearchCV do Scikit-learn. Esta classe realiza uma busca exaustiva dentre um conjunto de valores de hiperparâmetros informados para um estimador utilizando validação cruzada, e retorna o modelo que obteve melhor pontuação para uma determinada métrica de avaliação. Na criação dos estratos para validação cruzada, foi utilizada a classe StratifiedKFold, que forma conjuntos estratificados, contendo aproximadamente o mesmo percentual de amostras da classe alvo que o conjunto completo. Somente os dados simulados foram utilizados nesta etapa (PEDREGOSA et al., 2011).

A arquitetura da rede foi construída com as configurações padrão do Scikit-learn, com um número de neurônios na camada de entrada correspondente ao número de atributos dos dados de treinamento e um único neurônio na camada de saída, com a função de ativação *logistic*. A busca pelos melhores hiperparâmetros foi realizada dentro do seguinte espaço de busca:

- Tamanho de camada oculta: (20), (30,30) e (50,50,50);
- Otimizador de pesos: lbfsgs, sgd e adam;
- Termo regularizador alpha: 0.00001, 0.0001, 0.001;
- Número máximo de iterações: 100, 250, 500;
- Função de ativação para as camadas ocultas: identity, logistic, tanh, relu.

A busca foi repetida para cada um dos conjuntos de atributos selecionados para as métricas AG, Prec, Rec e AUC (apresentadas na Seção 3.1.5). No Scikit-learn, há duas formas de computar a AG: a função *accuracy_score*, que retorna a proporção de previsões corretas para o total do conjunto de dados, e a função *balanced_accuracy_score*, aqui referenciado como Acurácia Geral Balanceada (AGB), na

qual cada amostra é ponderada de acordo com a prevalência inversa de sua classe verdadeira, de forma a evitar estimativas de desempenho infladas em conjuntos de dados desequilibrados. Ambas foram usadas na etapa de otimização de hiperparâmetros. Também foi utilizada a pontuação F_β , uma métrica que calcula a média harmônica ponderada entre Precisão e Revocação, atingindo seu valor ideal em 1 e seu pior valor em 0. Nesta função, o parâmetro beta determina o peso da Revocação na pontuação combinada: beta < 1 dá mais peso à Precisão, enquanto beta > 1 favorece a Revocação; beta = 0.5 atribui igual peso a ambas as métricas.

As métricas Prec, Rec, AUC e F_β foram calculadas, respectivamente, pelas funções `precision_score`, `recall_score`, `roc_auc_score` e `fbeta_score`. São funções que retornam valores globais, para todo o conjunto de dados; como estas métricas são calculadas por classe, e os conjuntos de dados possuem mais de uma classe, é necessário definir como será o cálculo da pontuação. Essa definição é feita através do parâmetro `average`, que pode assumir os seguintes valores:

- *binary*: Reporta apenas os resultados para a classe especificada em outro parâmetro, o `pos_label`. Aplicável somente quanto o atributo alvo é binário.
- *micro*: Calcula a métrica globalmente, contando o totais de verdadeiros e falsos positivos/negativos.
- *macro*: Calcula a métrica para cada classe e reporta sua média não ponderada, sem considerar o possível desbalanceamento de classes. A Revocação, quando pontuada com o valor *macro* atribuído ao parâmetro `average`, retorna o mesmo valor da AGB.
- *weighted*: Calcula a métrica para cada classe e reporta sua média ponderada pelo suporte (o número de instâncias verdadeiras para cada classe), considerando o desbalanceamento de classes no cálculo. A Revocação, quando pontuada com o valor *weighted* atribuído ao parâmetro `average`, retorna o mesmo valor da AG. A pontuação F1, quando calculada com o valor *weighted* atribuído ao pa-

râmetro *average*, pode não estar exatamente entre a Precisão e a Revocação, havendo uma certa distorção para levar o desbalanceamento em consideração.

- *samples*: Calcula a métrica para cada instância e encontra sua média (significativa apenas para classificação com múltiplas classes, quando difere de AG).

No caso da AUC, por se tratar de uma métrica destinada a conjuntos binários, além de configurar o parâmetro *average* como para as outras métricas, também é necessário definir, através do parâmetro *multi_class*, se a pontuação será computada para cada classe em comparação com as demais (valor *ovr*, sensível ao desbalanceamento de classes), ou como a AUC média de todas as combinações de pares possíveis entre as classes (valor *ovo*, insensível ao desbalanceamento de classes se *average* = *macro*). A busca exaustiva foi realizada com as métricas utilizando o parâmetro *average* configurado com os valores *macro* e *weighted*. AUC foi computada com o parâmetro *multi_class* configurado com o valor *ovr*.

Utilizando os hiperparâmetros com melhor desempenho na etapa de otimização, o modelo foi treinado com a totalidade dos dados simulados e testado com os dados observados. Da mesma forma que em Acuña-Ruz et al. (2018), o processo de amostragem, treino e teste foi repetido 100 vezes, com uma inicialização aleatória dos pesos da RNA diferente em cada uma delas. Em cada repetição, foram coletados os totais de erros e acertos por classe e percentual de cobertura, bem como as métricas AG, AGB, pontuação F1, pontuação $F\beta$, Prec e Rec. Para F1, $F\beta$, Prec e Rec o parâmetro *average* foi configurado com os valores *micro* e *weighted*.

As médias, os mínimos e os máximos atingidos por estas métricas foram analisados, bem como os padrões de acertos e erros por classe e percentual de cobertura do pixel. Por fim, foi realizada uma avaliação geral dos resultados obtidos, com intuito de verificar se os resultados indicam ser possível ou não utilizar classificadores treinados em ambientes simulados para aplicações em imagens observadas, além de inferir outras informações relevantes sobre os limites de detecção de plásticos marinhos por imagens e métodos de sensoriamento remoto.

3.3 RESULTADOS E DISCUSSÕES

3.3.1 Análise exploratória

O teste de Kolmogorov-Smirnov foi aplicado comparando os conjuntos de dados observado e simulado, atributo a atributo, tanto considerando todas as classes simultaneamente quanto de forma isolada. Em todos os casos, o valor-p apresentado foi próximo de zero, indicando que, em todas as condições testadas, os conjuntos de dados observado e simulado eram provenientes de distribuições diferentes. Isso configura uma violação de um dos pressupostos básicos do aprendizado de máquina, que determina que os conjuntos de treino, teste e validação precisam vir da mesma distribuição. A avaliação de desempenho dos classificadores (Subseções 3.3.3 e 3.3.4) foi realizada considerando que, neste cenário, a aplicabilidade do aprendizado de máquina pode ficar bastante limitada, tendo em vista que a função aprendida pelo modelo é baseada na distribuição estatística dos dados de treinamento e pode não ser generalizável para distribuições diferentes.

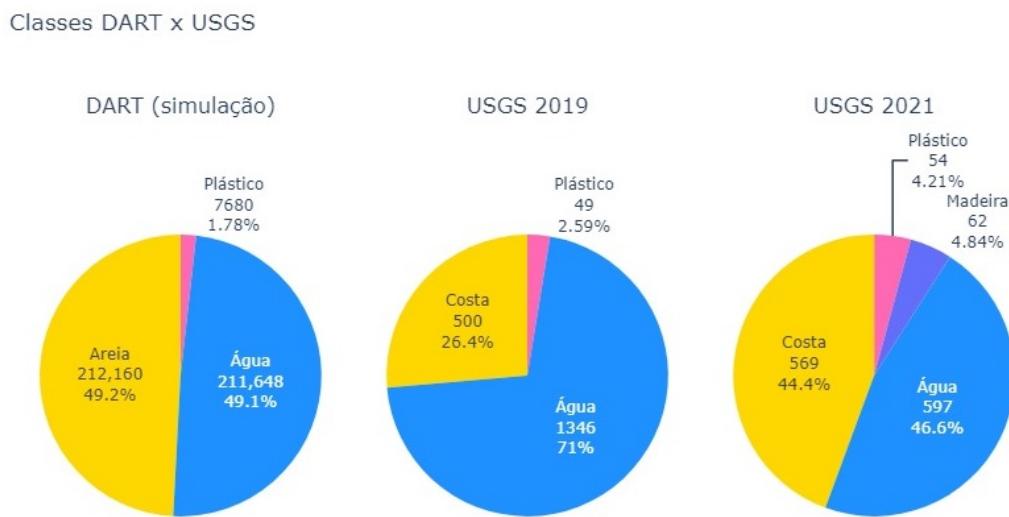
Os dados simulados contavam originalmente com 212.160 pixels de areia, 212.160 pixels de água e 7.680 pixels de plástico, totalizando 432.000 amostras. Após o pré-processamento, houve uma pequena redução no número de amostras de água, que passou a contar com 211.648 pixels. As demais classes mantiveram a quantidade original de amostras, como pode ser visto na Tabela 5 e na Figura 21. Dentre os pixels contendo plástico, todas as seis polímeros tinham o mesmo número de amostras (Figura 22).

Tabela 5 – Descrição dos conjuntos de dados simulados e observados.

Classe	Dados simulados (DART)	Dados observados (USGS)
Água	211.648	1.943
Areia	212.160	0
Costa	0	1.069
Plástico	7.680	103
Madeira	0	62
Total	431.488	3.177

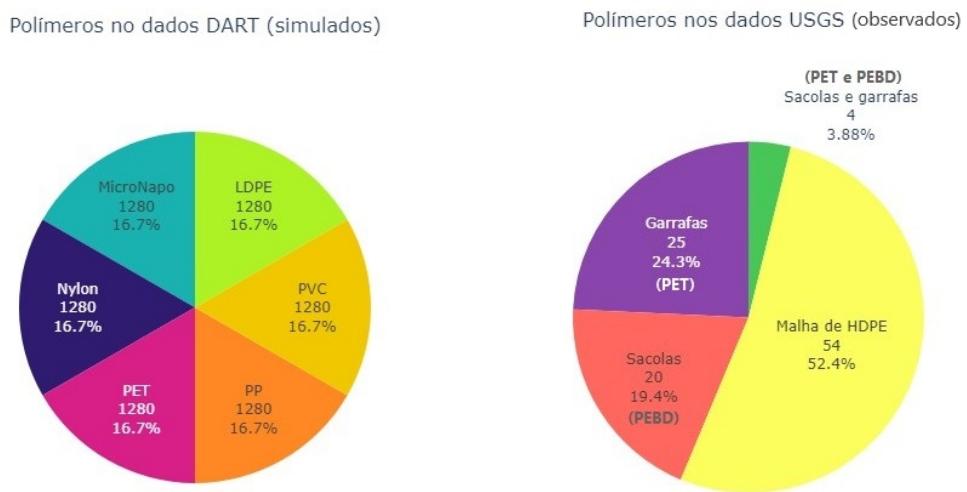
Fonte: Elaboração própria.

Figura 21 – Distribuição de classes nos conjunto de dados DART (simulados) e USGS (observados).



Fonte: Elaboração própria.

Figura 22 – Distribuição de polímeros nos pixels contendo plástico nos conjuntos de dados DART (simulados) e USGS (observados).



Fonte: Elaboração própria.

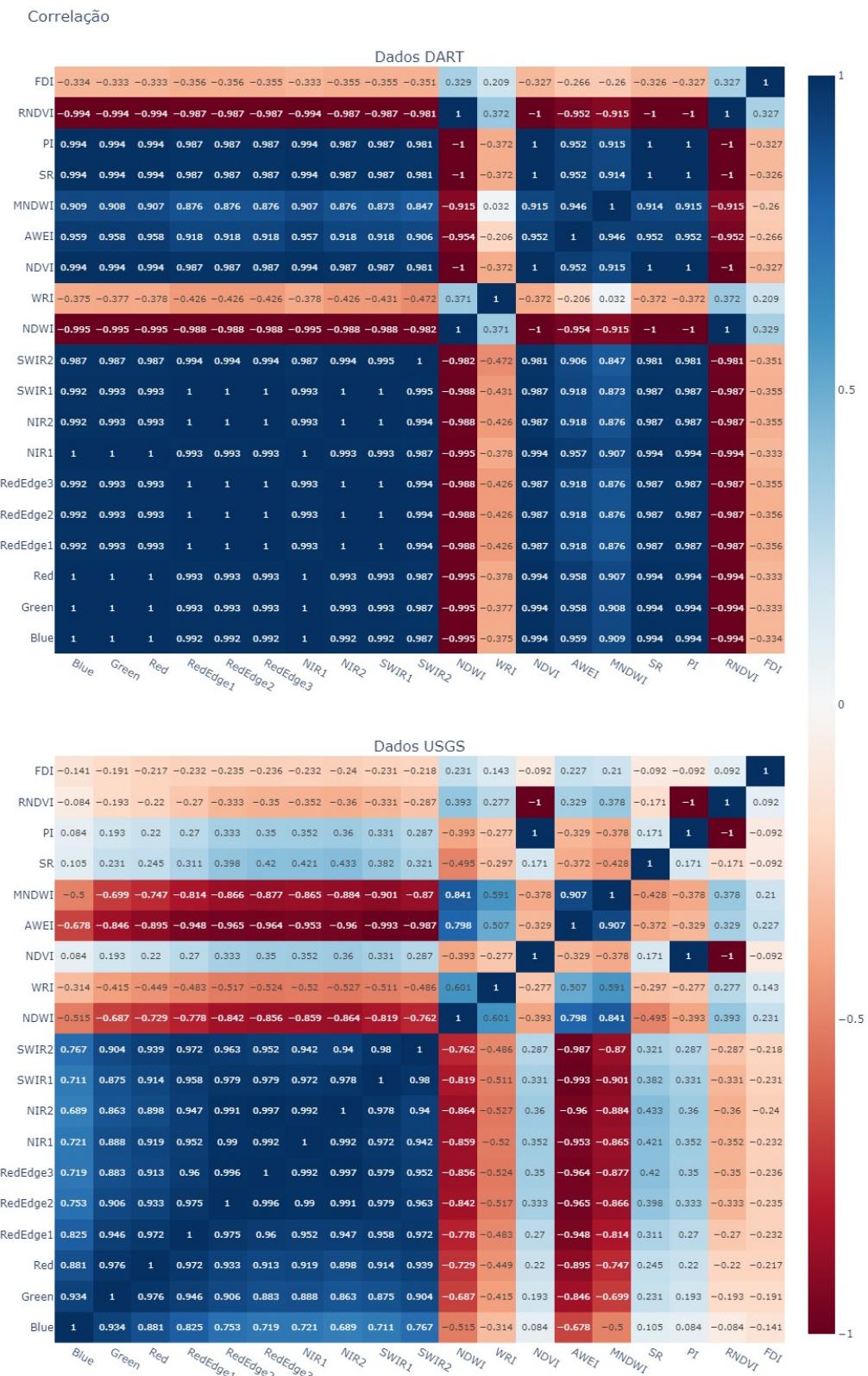
Os dados observados contavam originalmente com 3.197 amostras, das quais 3.177 foram consideradas após o pré-processamento. Conforme demonstrado na Tabela 5 e na Figura 21, o conjunto de dados observado conta com 1.943 pixels de água (1.346 em 2019 e 597 em 2021), 1.069 pixels extraídos da superfície terrestre da região costeira (500 em 2019 e 569 em 2021) e 165 pixels de alvos artificiais em suspensão na água, sendo 103 de plástico (49 em 2019 e 54 em 2021) e 62 de madeira. Dentre os pixels contendo plástico, havia uma predominância daqueles cobertos por malha de HDPE, derivados do PLP 2021, no qual todas as amostras continham somente este elemento. Dentre os demais, oriundos do PLP 2019, havia uma distribuição equilibrada de pixels contendo sacolas (PEBD) e garrafas (PET) (Figura 22).

Ambos os conjuntos de dados são altamente desbalanceados. O desbalanceamento das classes também é esperado em cenas naturais, majoritariamente compostas pela cobertura da superfície e contendo uma distribuição esparsa de detritos. A presença de diferentes resinas poliméricas nos pixels representando a poluição plástica é distribuída de forma equilibrada nos dados simulados mas desbalanceada nos dados observados. Cabe lembrar que, considerando que a aquisição de dados foi manipulada de forma artificial tanto no conjunto de dados simulado quanto no observado, também pode-se esperar maior variabilidade e desbalanceamento na distribuição destes elementos no meio ambiente.

Os dados simulados são altamente correlacionados, conforme demonstrado na Figura 23. Considerando o total do conjunto de dados, com todas as classes, os valores de correlação aproximam-se de 1 ou -1 para todas as bandas e quase todos os índices, sendo o WRI e o FDI as únicas exceções. Quando a correlação entre as bandas é observada separadamente para cada classe, os valores mudam. Os gráficos de correlação para cada classe podem ser vistos no Apêndice A.

Observando as três classes simuladas isoladamente, em todas elas há duas regiões com correlação tendendo a 1: uma contém as bandas do visível (Blue, Green e Red) e a banda NIR1 e a outra contém as bandas do Red Edge (RedEdge1, RedEdge2 e RedEdge3) e a banda NIR2. Já a correlação entre as bandas destas duas regiões

Figura 23 – Correlação entre atributos nos dados DART (simulados) e USGS (observados).



Fonte: Elaboração própria.

Nota: Os valores de correlação positivos são exibidos em azul e os negativos em vermelho. Cores mais escuras indicam maior proximidade dos valores extremos (-1 para vermelho e 1 para azul).

entre si é mais baixa, variando conforme a classe. O plástico possui a correlação maior entre as bandas destas duas regiões - 0.7 - enquanto a água e areia apresentam correlação tendendo a zero entre os dois grupos de bandas. Sendo assim, a correlação é baixa entre dois blocos de bandas, que podemos identificar como VISNIR1 (Blue, Green, Red e NIR1) e RENIR2 (RedEdge1, RedEdge2, RedEdge3 e NIR2).

Na região SWIR, a banda SWIR1 apresenta uma correlação com as demais bandas em um padrão muito similar à região do RENIR2 em todas as classes. Já o SWIR2, banda mais próxima dos maiores valores para o coeficiente de absorção da água, apresenta um padrão particular de correlação com as demais bandas. Para o plástico, a correlação é bastante alta com todas as demais bandas, variando entre 0.7 e 1. Já na areia e na água, o SWIR2 tem alta correlação com a Região RedEdge e as bandas NIR2 e SWIR1, enquanto tem um índice de correlação entre 0 e -0.1 com as bandas da região VIS e a banda NIR1.

Os índices radiométricos tiveram padrões de correlação variados, de forma que não foi possível encontrar um padrão discernível. Os resultados da análise de correlação entre bandas sugere que tanto novos índices radiométricos quanto a seleção de atributos poderiam explorar o contraste entre as regiões VISNIR1 e RENIR2 para diferenciação entre o comportamento espectral da água, da areia e do plástico.

Nos dados observados, a correlação entre as bandas também é bastante alta quando se considera o conjunto completo. A maior parte das bandas apresentam correlação superior a 0,8 entre si. A banda Blue apresenta correlações ligeiramente inferiores quando comparada às demais bandas. Já os índices demonstram correlações menores tanto entre si quanto em relação às bandas. Em especial, os índices FDI, RNDVI, PI, SR e NDVI possuem correlações bem baixas com a maior parte dos demais atributos. Quando se consideram as classes do conjunto separadamente, para água e plástico as correlações em geral são menores, mas se mantém o padrão de correlação mais alta entre as bandas e mais baixa tanto entre os índices em si quanto dos índices em relação às bandas. A classe madeira também segue esse padrão, com a diferença de apresentar baixos índices de correlação entre as bandas da região

SWIR e os demais atributos, e alguns graus de correlação mais altos entre alguns índices específicos. A classe costa apresenta correlação alta entre todos os atributos em geral, à exceção dos índices FDI e MNDWI. Alguns graus de correlação mais baixos são encontrados entre um e outro atributos específicos, como entre as bandas NIR1 e NIR2 e os índices NDVI e NDWI, mas em geral as correlações entre atributos são mais altas. Os gráficos de correlação para cada classe podem ser vistos no Apêndice A. As correlações mais baixas entre os índices e as bandas indica que pode haver potencial e relevância em seu uso para classificação das amostras, como novos atributos, adicionalmente às bandas espectrais (Seção 3.2.8).

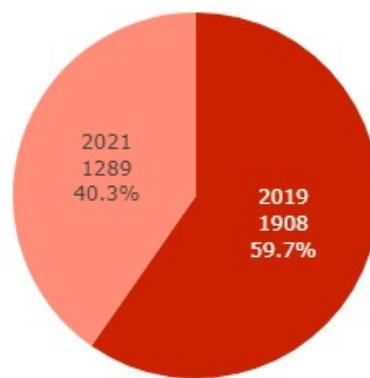
Todas as cenas simuladas estavam em igualdade de condições, como se os dados tivessem sido adquiridos no mesmo dia. Já os dados observados foram adquiridos em dias diferentes, estando sujeitos às variações oriundas das condições climáticas e ambientais e seus respectivos efeitos sobre os produtos finais do sensoriamento remoto. Como podemos observar na Figura 24, os dados observados são majoritariamente compostos pelas aquisições realizadas em 2019, havendo uma super-representação de duas datas: 03 e 18 de maio de 2019. Nessas datas, devido às diferentes configurações nos alvos plásticos, foi necessário definir uma região de interesse maior para capturar a totalidade da área.

Com relação às classes, a origem dos pixels das classes costa e plástico está dividida de forma equilibrada entre as duas edições do PLP, porém os da classe madeira só estavam presentes em 2021. Os pixels da classe água também tem uma super-representação do ano de 2019, como uma consequência direta da coleta de dados em áreas maiores dos dias 03 e 18 de maio daquele ano.

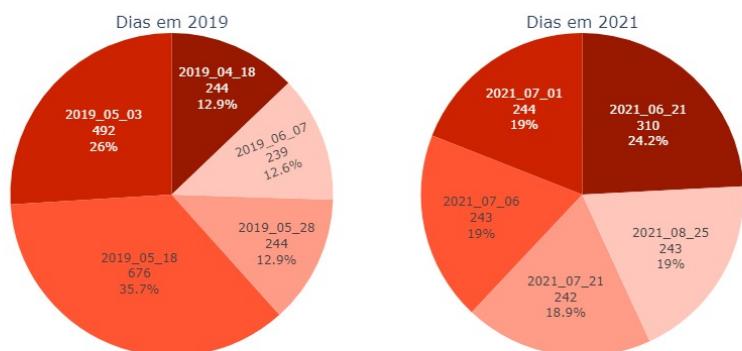
Os dados simulados apresentavam uma reflectância com média e desvio padrão mais altos que os dados observados, como pode ser observado na Figura 25. Quando somente as classes água e plástico foram consideradas - por serem as únicas presentes em ambos os conjuntos - a média do conjunto simulado ficou bem mais próxima da média do conjunto observado, embora o desvio padrão do conjunto simulado ainda tenha se mantido bem mais alto que o do conjunto observado, conforme demonstrado

Figura 24 – Fontes e datas de aquisição dos dados USGS (observados).

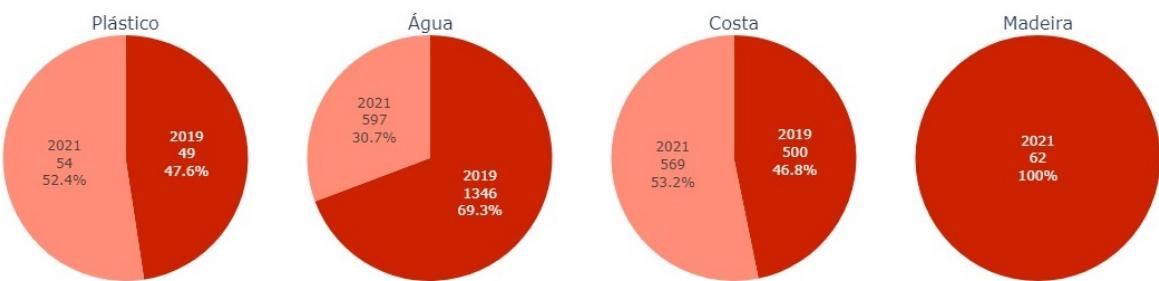
Fontes dos dados USGS



Datas de aquisição USGS



Fontes dos dados USGS - por classe



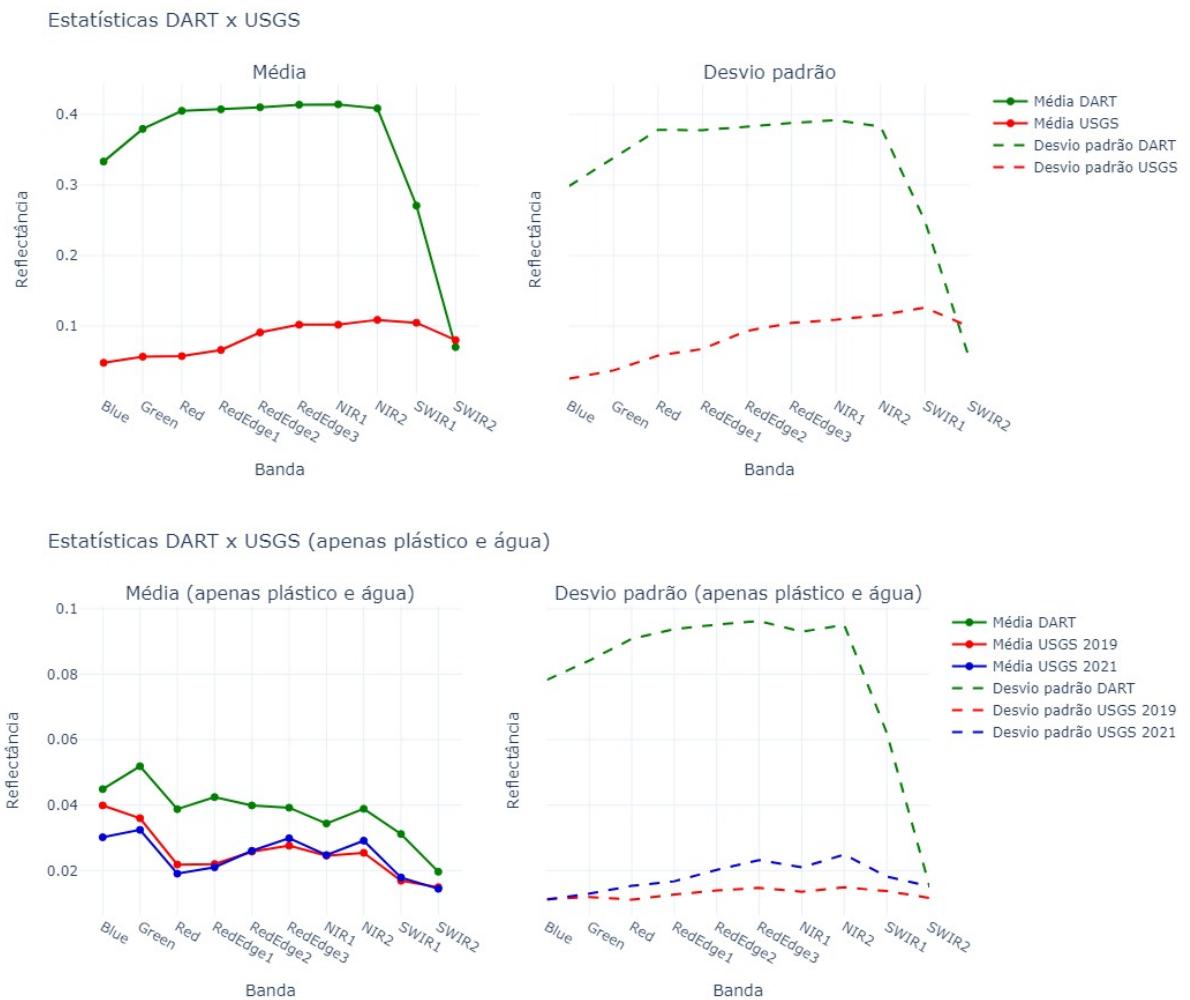
Fonte: Elaboração própria.

na Figura 25. O desvio padrão dos dados adquiridos em 2021 ficou ligeiramente mais próximo do desvio padrão do conjunto simulado quando comparado ao desvio padrão dos dados adquiridos em 2019.

Os quartis também demonstram maior amplitude para o conjunto de dados simulado, conforme demonstrado na Figura 26. Os dados simulados apresentaram uma maior diferença na variabilidade entre os substratos (água e areia) que os dados observados (água e costa). Nos dados observados, o conjunto de 2019 mostra mais diferenças entre o comportamento espectral das diferentes classes que os dados de 2021. Como pode ser visto na Figura 27, na comparação classe a classe entre os dados simulados e observados, o plástico observado apresentou uma variabilidade muito menor que o simulado. A costa apresentou um padrão de variação bem diferente da areia simulada, em uma faixa com valores mais baixos. A variabilidade da água nos dados observados ficou em uma faixa próxima a dos dados simulados. A madeira, presente apenas nos dados observados, variou em uma faixa próxima a da água.

As assinaturas espectrais médias por classe são exibidas na Figura 28. Nos dados simulados, exibidos na primeira coluna, percebe-se que as classes possuem assinaturas bastante distintas, à exceção da banda SWIR2. Nos conjuntos de dados observados, presentes nas duas colunas centrais da Figura 28, percebe-se que os dados de 2019, embora tenham uma assinatura característica para a classe costa, apresentam assinaturas indistinguíveis para água e plástico, que se sobrepõem uma à outra no gráfico. Já os dados de 2021 mostram uma assinatura minimamente distinta para cada uma das classes, possivelmente devido à presença de pixels com percentuais de cobertura plástica maiores neste conjunto de imagens. A quarta coluna da figura fortalece esta hipótese: ela exibe assinaturas espectrais médias da água e do plástico nos dados observados - considerando tanto o conjunto de 2019 quanto o de 2021 - porém incluindo somente os pixels com no mínimo 50% de cobertura plástica para composição da assinatura média de cada classe. Neste caso, percebe-se que as assinaturas mostram uma pequena separabilidade em todas as bandas (à exceção da banda SWIR2 novamente), diferentemente do que é exibido no conjunto de 2019.

Figura 25 – Média e desvio padrão nos conjuntos de dados DART (simulado) e USGS (observado).



Fonte: Elaboração própria.

Nota: Na parte superior, média e desvio padrão nos conjuntos de dados DART (simulado) e USGS (observado). Na parte inferior, média e desvio padrão apenas para as classes água e plástico nos conjuntos de dados DART e USGS.

As assinaturas espectrais médias confirmam as características de comportamento óptico descritas por Themistocleous et al. (2020) e Biermann et al. (2020): a água limpa é eficiente na absorção de luz, especialmente no NIR e no SWIR, ao passo que plástico e madeira mostraram picos de refletância no NIR. Adicionalmente, plástico e madeira também mostraram refletância mais alta na banda Red. Nos dados observados, também se confirma que tanto os plásticos quanto a água demonstraram uma refletância mais alta na banda Blue, caindo nas bandas Green e Red. Contudo, diferentemente do que foi apontado pelos autores, a madeira apresentou baixa refle-

Figura 26 – Quartis nos conjuntos de dados DART (simulado) e USGS (observado)



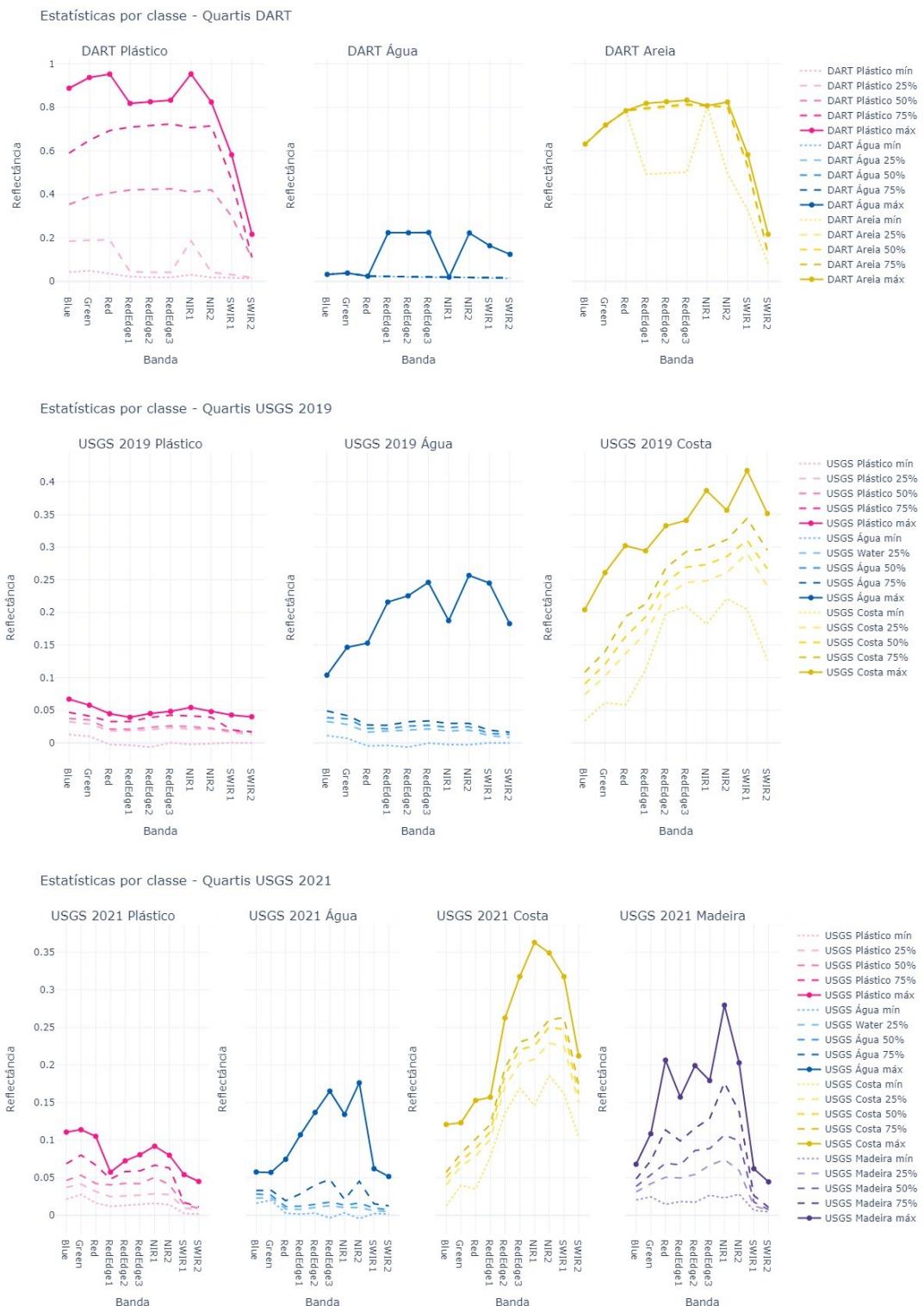
Fonte: Elaboração própria.

tância na região do SWIR, assim como o plástico, talvez como um efeito do coeficiente de absorção da água (Figura 8), tendo em vista que as amostras tinham algum grau de submersão em alguns casos (Tabela 2).

A hipótese levantada por Biermann et al. (2020) de que as classes de cobertura da superfície poderiam ser distinguíveis utilizando os índices FDI e NDVI individualmente em gráficos de *box plot* e combinados em gráficos de dispersão não se confirmou em nossos resultados. Como pode ser visto nas Figuras 29 e 30, os valores da água possuem muitos *outliers* e se sobrepõe aos valores das demais classes. Se os *outliers* forem desconsiderados nos *box plot*, pode-se perceber uma separabilidade entre água e areia/costa em ambos os conjuntos, mas ainda assim plástico e madeira ficam com valores muito próximos aos das classes do substrato.

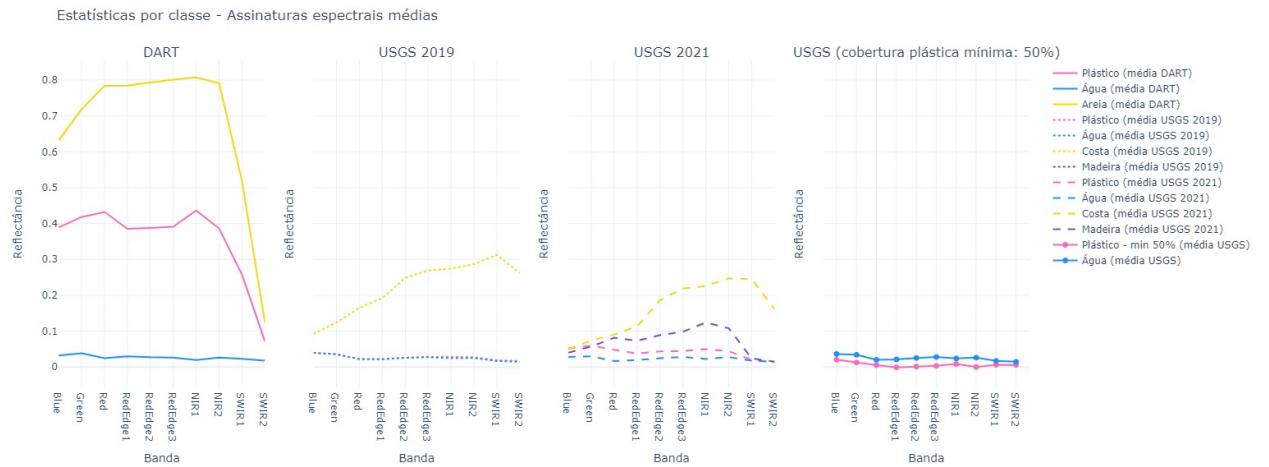
O plástico também foi analisado através de gráficos *box plot*, observando a distribuição dos valores para o substrato e para os diferentes percentuais de cobertura nos pixels contendo plástico, agrupados em cinco categorias: 20%, 40%, 60%, 80% e 100%. Nos dados simulados, cada agrupamento tinha exatamente o respectivo percentual de cobertura, enquanto nos dados observados os agrupamentos correspondiam ao limite superior de cada categoria. Por exemplo, nos gráficos USGS, o *box plot*

Figura 27 – Quartis das diferentes classes nos conjuntos de dados DART (simulado) e USGS (observado).



Fonte: Elaboração própria.

Figura 28 – Assinaturas espectrais médias de cada classe por conjunto de dados e por percentual de cobertura.



Fonte: Elaboração própria.

de plástico com 20% de cobertura contém pixels com cobertura entre 0% e 20%; o *box plot* de 40% contém pixels com cobertura entre 21% e 40%, e assim por diante.

Nos dados simulados, as bandas VIS e NIR1 mostraram padrões muito similares aos da Figura 31, onde é possível discernir o plástico do substrato, principalmente nas amostras com maior percentual de cobertura. Nas demais bandas, as amostras simuladas apresentam pouca variabilidade entre as classes e proporções de cobertura, que são bastante semelhantes entre si e ao mesmo tempo diferem do padrão observado nas bandas VIS (Figura 32).

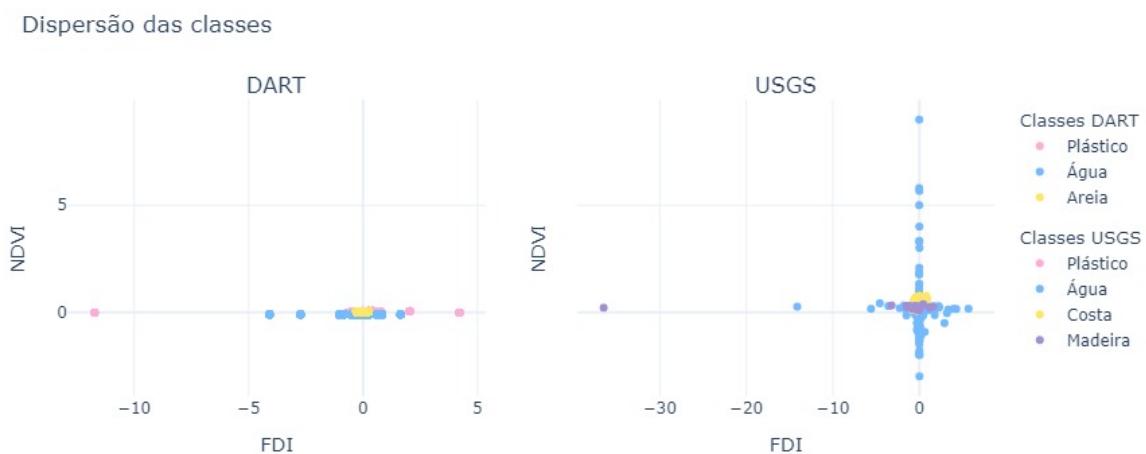
Nos dados observados, nenhuma das bandas demonstrou separabilidade espectral entre as classes. A impossibilidade de discernimento de um padrão por classe nos dados observados pode estar relacionado ao número limitado e desbalanceado das amostras. Além da predominância da água no conjunto de dados, poucos dentre os pixels de plástico possuem alta concentração do elemento; vários deles são majoritariamente compostos por água, com proporções de plástico abaixo de 50%. Os *box plot* dos índices também apresentaram padrões discerníveis no conjunto simulado em alguns casos, mas sem consistência nos dados observados.

Figura 29 – Valores de FDI e NDVI agrupados por classe de cobertura da superfície nos conjuntos de dados DART (simulado) e USGS (observado).



Fonte: Elaboração própria.

Figura 30 – Dispersão de valores de FDI e NDVI nos conjuntos de dados DART (simulado) e USGS (observado).



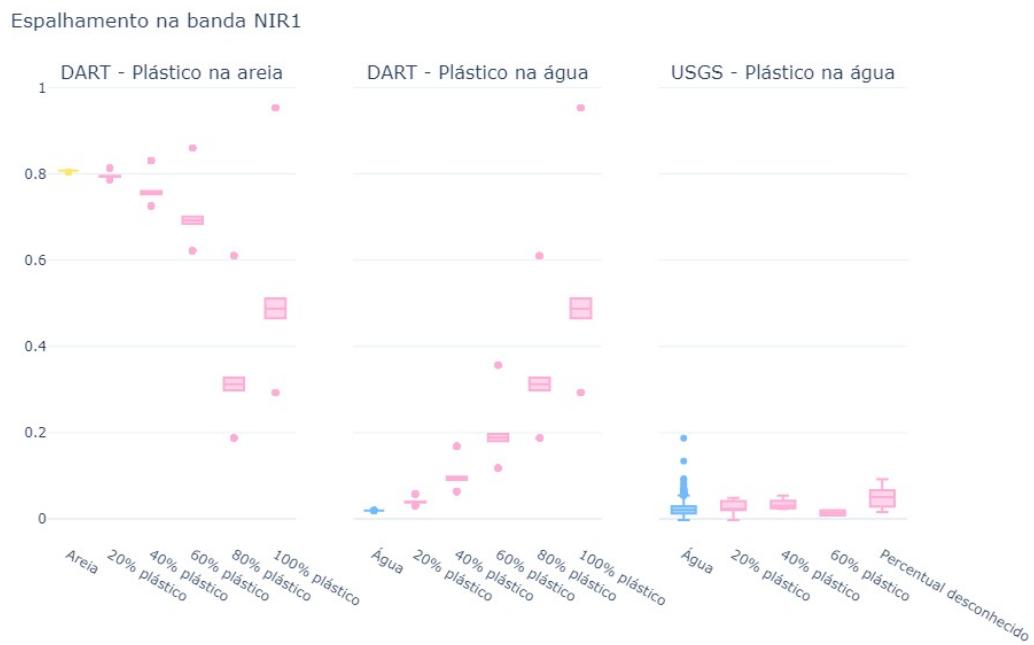
Fonte: Elaboração própria.

Nota: Os valores de FDI e NDVI foram combinados em um gráfico de dispersão bidimensional e agrupados por classe de cobertura, conforme legenda.

Os atributos também foram analisados em gráficos de dispersão, colocando cada atributo no eixo Y e os identificadores (IDs) dos pixels no eixo X, atribuindo diferentes cores para cada classe, de forma a analisar se em alguma banda ou algum dos índices era possível fazer a discriminação visual das classes em regiões de valores específicos para o referido atributo. Na maioria dos casos, em ambos os conjuntos, é possível diferenciar os substratos (água/areia ou água/costa) mas os pixels de plástico ou madeira aparecem sobrepostos ao substrato, sem uma faixa de valores que os discrimine. Isso contraria os achados de Themistocleous et al. (2020), que havia conseguido associar pixels de plástico a uma faixa de valores específica do índice PI. Aqui, o índice não demonstrou separabilidade, como ser visto na Figura 33.

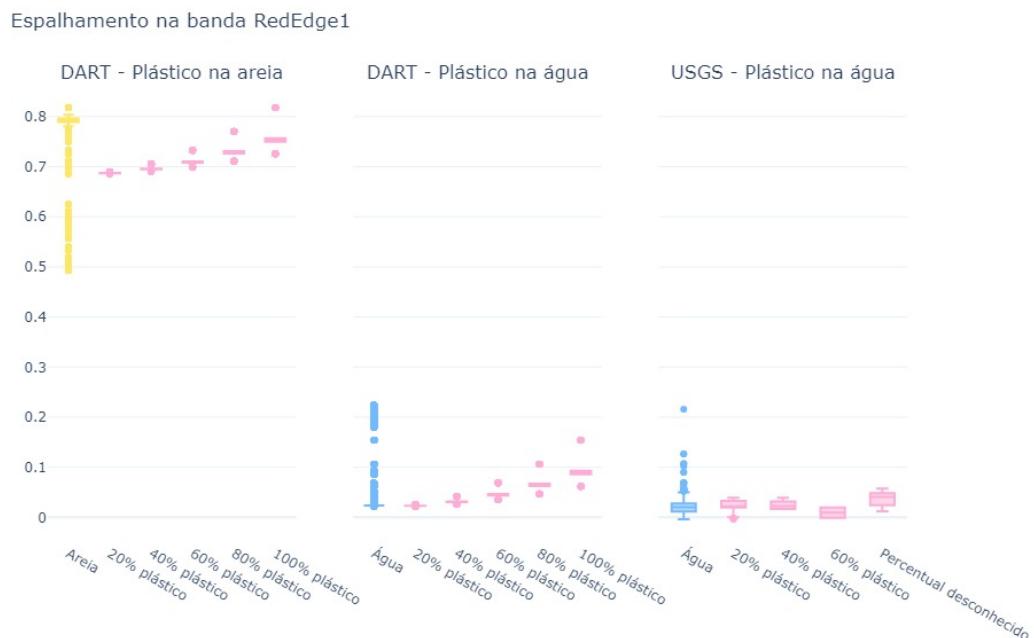
Por fim, foram gerados gráficos de dispersão bidimensionais para todos os atributos. Para cada atributo, o gráfico foi gerado colocando cada um dos demais atributos no eixo oposto, com cores específicas atribuídas a cada classe, visando observar se seria possível discernir as classes no espaço projetado. Em alguns casos, houve uma

Figura 31 – Box plot do plástico sobre a areia e a água na banda NIR1 nos dados DART (simulados) e USGS (observados), agrupado por percentual de cobertura.



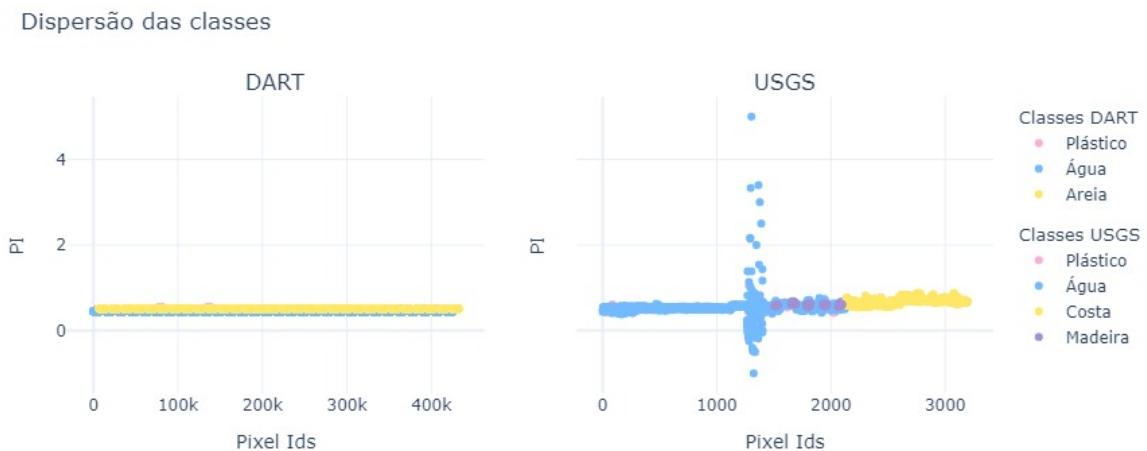
Fonte: Elaboração própria.

Figura 32 – Box plot do plástico sobre a areia e a água na banda RedEdge1 nos dados DART (simulados) e USGS (observados), agrupado por percentual de cobertura.



Fonte: Elaboração própria.

Figura 33 – Dispersão dos valores de PI agrupados por classe de cobertura da superfície nos conjuntos de dados DART (simulado) e USGS (observado).



Fonte: Elaboração própria.

pequena separabilidade entre as classes, mas não suficiente para destacar completamente uma classe da outra. A análise completa dos gráficos está no Apêndice A.

As técnicas de análise exploratória permitiram compreender melhor as características dos conjuntos de dados, e podem fornecer uma base de conhecimento para outros métodos em trabalhos futuros.

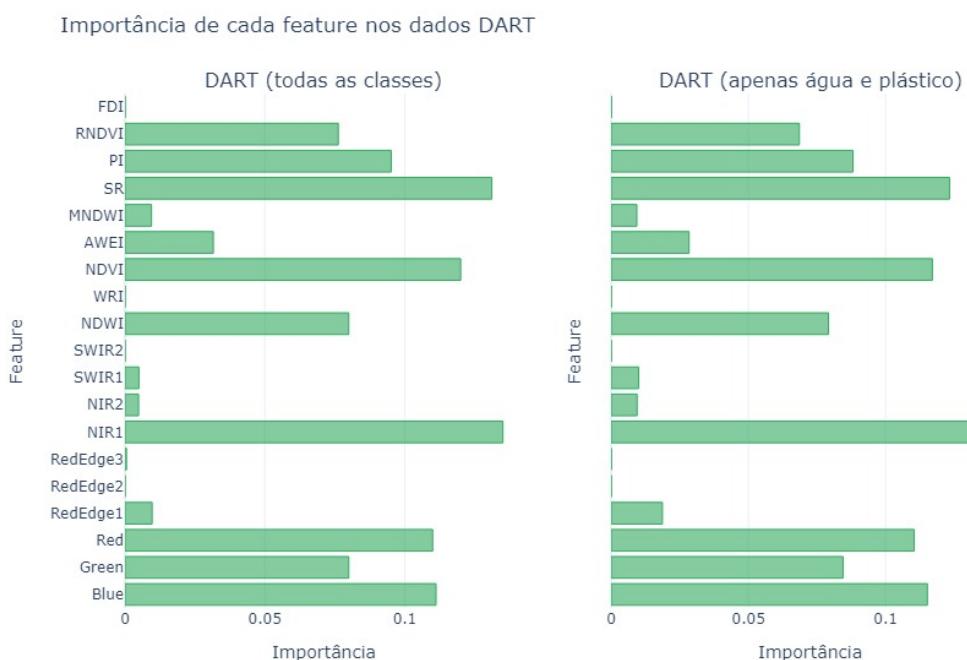
3.3.2 Seleção de atributos

As pontuações geradas pelo RF para as importâncias de atributos são exibidas na Figura 34. Não ocorreram grandes diferenças nas importâncias de atributos calculadas para os conjuntos contendo todas as classes ou somente água e plástico.

Na mesma linha de Basu et al. (2021), foram criados diferentes conjuntos de atributos para utilização no processo de classificação. Na seleção, além das pontuações de importância dos atributos geradas pelo RF, também foram considerados os valores de correlação entre eles. Os quatro conjuntos de atributos e seus critérios de seleção são apresentados na Tabela 6.

O cálculo das importâncias de atributos com RF foi realizado a partir das classificações geradas pelo algoritmo para cada conjunto de dados individualmente, dividindo

Figura 34 – Importâncias de atributos para dados DART (simulados) na classificação com o algoritmo Random Forest.



Fonte: Elaboração própria.

Tabela 6 – Subconjuntos de atributos selecionados para classificação.

Identificador	Atributos selecionados	Critérios de seleção
A	<i>Blue, Green, Red, RedEdge1, RedEdge2, RedEdge3, NIR1, NIR2, SWIR1, SWIR2, NDWI, WRI, NDVI, AWEI, MNDWI, SR, PI, RNDVI, FDI</i>	Todos os atributos disponíveis
B	<i>Blue, Green, Red, RedEdge1, RedEdge2, RedEdge3, NIR1, NIR2, SWIR1, SWIR2</i>	Todos as bandas espectrais disponíveis
C	<i>NDWI, WRI, NDVI, AWEI, MNDWI, SR, PI, RNDVI, FDI</i>	Todos os índices radiométricos disponíveis
D	<i>NIR1, SR, WRI, FDI</i>	Dois atributos com maior importância para o conjunto DART (<i>NIR1</i> e <i>SR</i>) junto aos dois atributos menos correlacionados com eles (<i>WRI</i> e <i>FDI</i>).

Fonte: Elaboração própria.

Nota: Na análise de correlação para o subconjunto D, foram considerados somente os atributos com valor de correlação entre - 0.75 e 0.75, tanto no total do conjunto quanto na análise classe a classe.

as amostras em subconjuntos de treino e teste. Neste cenário, o RF alcançou a acurácia máxima na classificação de dados simulados. Como se trata de um classificador supervisionado, da mesma forma que a RNA, escolhida na metodologia e explorada na Seção 3.2.10, também foi realizada uma tentativa treinando o RF com os dados simulados e testando-o com os dados observados. Nestas circunstâncias, o RF só alcançou 38% de acurácia, e por isso, seu uso ficou restrito à seleção de atributos.

3.3.3 Classificação não supervisionada

Os resultados da classificação não supervisionada com o algoritmo K-Means são apresentados de forma detalhada no Apêndice A. A análise destes resultados possibilitou avaliar quais aspectos parecem influenciar mais na formação dos agrupamentos, conforme descrito a seguir.

Na classificação não supervisionada com dados simulados, para todos os parâmetros testados, mais de 95% das amostras concentraram-se em apenas dois agrupamentos, sendo compostos majoritariamente por água em um dos casos e areia no outro, com cerca de metade dos pixels de plástico adicionada a cada um deles. Os demais agrupamentos gerados pelo algoritmo continham quantidades muito pequenas de água ou de areia misturadas a poucas amostras de plástico. Estes grupos menores concentraram mais frequentemente subgrupos de água do que de areia, e quando continham pixels de plástico, eram geralmente amostras com no mínimo 60% ou 80% de cobertura, demonstrando uma influência dos percentuais de cobertura na formação dos agrupamentos. Pixels com 20% a 60% de plástico em geral foram distribuídos de maneira uniforme entre os dois agrupamentos maiores de água ou areia, como se pertencessem a mesma classe do substrato onde estavam depositados.

Com dados observados, o algoritmo em geral formou dois agrupamentos somando mais de 99% do total de amostras, sendo um composto majoritariamente por água e outro composto majoritariamente por costa. A única exceção a esta tendência foi a classificação com o conjunto de atributos B, formado pelas bandas do sensor: essa configuração em geral concentrou entre 50% a 66% das amostras em um grupo,

com predominância de água, e distribuiu as demais amostras em percentuais razoavelmente equilibrados entre os demais agrupamentos. A Figura 35 demonstra um dos resultados obtidos com o conjunto de atributos B para ambos os conjuntos de dados.

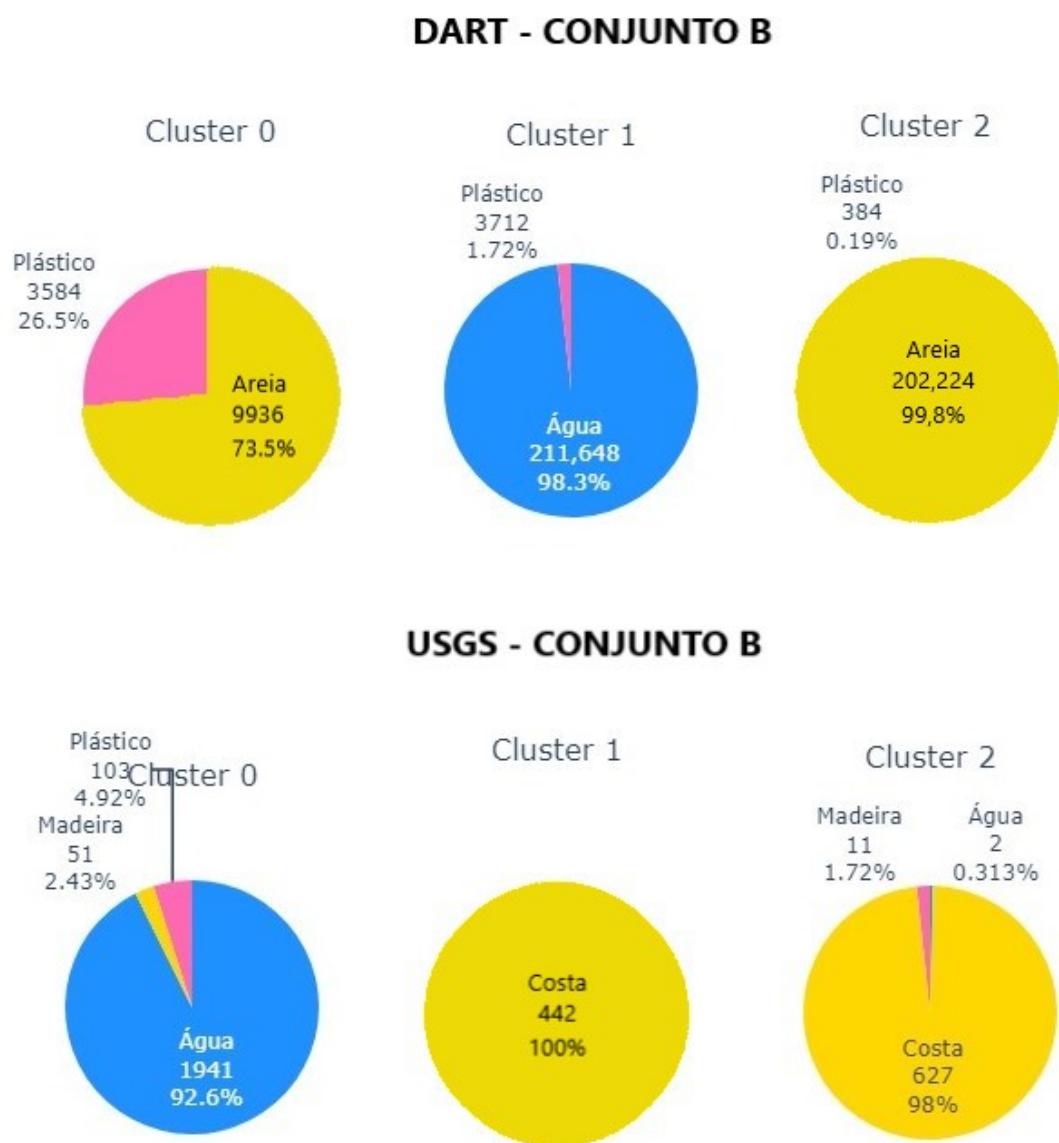
Com os dados simulados, os pixels de água e areia não foram colocados no mesmo agrupamento em nenhuma configuração. Já nos dados observados, água e costa eventualmente foram agrupados no mesmo cluster. É um resultado compatível com as características dos conjuntos de dados, tendo em vista que as simulações continham menos elementos e uma maior pureza espectral.

Nos dados simulados, foram observadas as seguintes tendências: PVC foi mais frequentemente associado à areia, enquanto PET foi mais frequentemente associado à água; o LDPE foi em geral agrupado à água ou em clusters exclusivos contendo poucos pixels, todos com 100% de cobertura, sendo o polímero com maior tendência a se separar dos demais elementos. A Figura 36 demonstra um exemplo de distribuição de polímeros para o conjunto de atributos B, para ambos os conjuntos de dados. As tendências confirmam as assinaturas médias de cada classe dos dados simulados exibidos na parte superior da Figura 37: dentre as assinaturas médias dos polímeros, PET apresenta a reflectância média mais baixa (e consequentemente mais próxima da assinatura da água) enquanto PVC apresenta a reflectância média mais alta (e consequentemente mais próxima da assinatura da areia). As diferenças entre as reflectâncias médias dos polímeros são maiores quando somente os pixels com 100% de cobertura são considerados.

Já nos dados observados, a totalidade de pixels com 100% de cobertura de malha de HDPE foi sempre adicionada a agrupamentos com predominância de água. Não houve outros padrões detectáveis de associação entre determinadas resinas ou percentuais de cobertura. Os grupos de água tendiam a agregar os pixels de madeira e plástico, e os grupos de costa apresentavam tendência ao isolamento, com menor frequência de mistura com amostras de outras classes.

Também foi verificada a distribuição das datas de aquisição das imagens entre os agrupamentos, verificando a hipótese de as condições dos alvos nas diferentes da-

Figura 35 – Agrupamentos do algoritmo K-means para o conjunto de atributos B.



Fonte: Elaboração própria.

Nota: Agrupamentos formados com o conjunto de atributos B e k=3 para dados DART (simulados) e USGS (observados). Os percentuais de amostras por classe se referem às proporções do cluster, não às do total do conjunto de dados (para ver as proporções de cada cluster em relação ao total, consulte o Apêndice A).

Figura 36 – Distribuição de polímeros entre os agrupamentos do algoritmo K-means para o conjunto de atributos B.



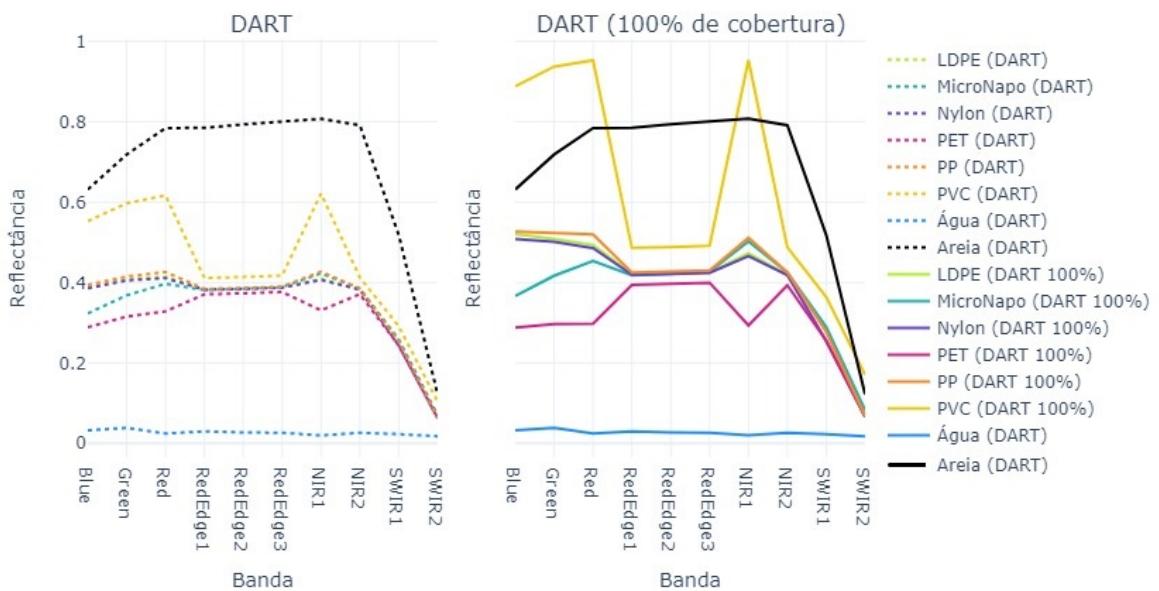
Fonte: Elaboração própria.

Nota: Agrupamentos formados com o conjunto de atributos B e k=3 para os dados DART (simulados, na parte superior) e USGS (observados, na parte inferior).

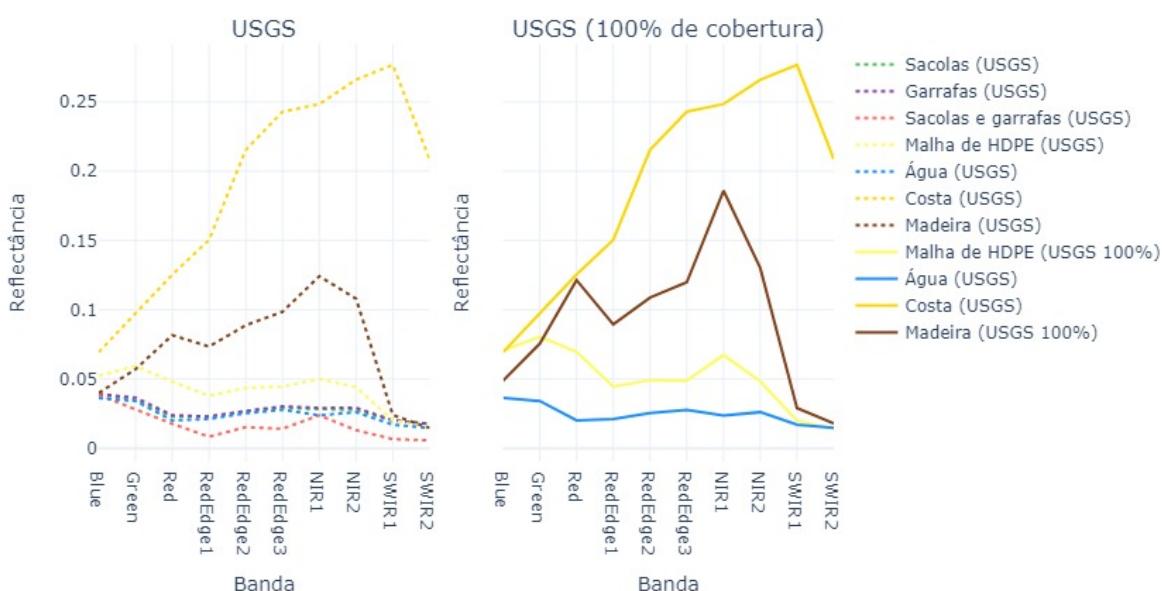
tas influenciar a formação dos agrupamentos. Contudo, o único padrão identificável de influência das datas de aquisição na formação dos agrupamentos foi a separação de imagens de 2019 e 2021 em alguns casos. Nenhuma configuração de classificação

Figura 37 – Assinaturas espectrais médias por classe nos conjuntos de dados DART (simulado, parte superior) e USGS (observado, parte inferior).

Estatísticas por elemento no DART - Assinaturas espectrais médias



Estatísticas por elemento no USGS - Assinaturas espectrais médias



Fonte: Elaboração própria.

com o KMeans separou diferentes datas de um mesmo ano em diferentes agrupamentos, de forma que não há evidência de impacto causado pelas diferenças nas condições dos alvos nas diferentes datas dos dois PLPs, como por exemplo as diferentes taxas de submersão e bioincrustação apontadas por Papageorgiou et al. (2022) nas imagens de 2021 (Tabela 2).

Respondendo às perguntas que guiaram a análise, o número de agrupamentos não afetou a classificação para nenhum dos conjuntos de atributos. Os padrões de associação entre os pixels foram influenciados pelo tipo de cobertura, no caso dos pixels contendo somente substrato, e pelo tipo de polímero e percentual de cobertura, no caso dos pixels contendo plástico ou madeira. Diferentes grupos de atributos e números de agrupamentos não geraram diferenças notáveis na separabilidade de classes.

3.3.4 Classificação supervisionada

Na etapa de otimização de hiperparâmetros, o melhor modelo encontrado apresentava os parâmetros abaixo. Utilizando os dados simulados para treino e teste, ele atingiu 100% de acurácia em todas as métricas para todos os conjuntos de atributos.

- Tamanho de camada oculta: (20);
- Otimizador de pesos: lbfgs;
- Termo regularizador alpha: 0.00001;
- Número máximo de iterações: 100;
- Função de ativação: identity.

A acurácia geral e a pontuação F1 atingiram 100% em todos os conjuntos de atributos para os dados de treinamento. Para os dados de teste, as médias das métricas nas 100 repetições de treino com dados simulados e teste com dados observados com cada um dos quatro conjuntos de atributos é apresentada na Tabela 7. Apenas a classificação realizada utilizando o conjunto de atributos B, formada pelas bandas do sensor, teve um desempenho razoável. Diferentemente de Basu et al. (2021), que

obteve o melhor desempenho com o conjunto de dados contendo todos atributos selecionados, aqui o melhor resultado foi obtido usando somente as bandas espectrais, sem utilizar os índices como variáveis de entrada. Na classificação não supervisionada, o conjunto B também havia gerado um modelo com comportamento diferente dos demais, gerando agrupamento com amostras distribuídas de forma mais equilibrada. As diferenças expressivas nas acurácia com dados de treino e teste podem ser consequências de um possível sobreajuste do modelo aos dados de treinamento, mas também podem ser causadas pelas diferenças entre as distribuições estatísticas dos dois conjuntos de dados.

As pontuações alta e baixa para Precisão ponderada e geral, respectivamente, indicam que o modelo não costuma classificar pixels de plástico como água, mas costuma classificar pixels de água como plástico. É possível chegar a esta conclusão, pois, como o plástico tem uma proporção muito pequena do conjunto de dados, o fato de a pontuação subir quando é ponderada pelo suporte demonstra que a pontuação da métrica está muito alta para a classe minoritária. Para a Revocação, a pontuação também sobe quando ponderada pelo suporte, embora a diferença entre as médias seja menor. Uma boa pontuação de Revocação para o plástico é muito positiva, pois é um indicativo de sensibilidade do modelo para a classe alvo.

Tabela 7 – Médias das métricas de avaliação obtidas para os dados de teste (conjunto de imagens MSI/Sentinel-2 observadas).

Métrica	Conjunto A	Conjunto B	Conjunto C	Conjunto D
Acurácia geral	0,359	0,797	0,170	0,148
Acurácia geral balanceada	0,594	0,666	0,530	0,508
F1 (macro)	0,307	0,546	0,165	0,141
F1 (weighted)	0,476	0,849	0,222	0,180
$F\beta$ (macro)	0,390	0,544	0,244	0,201
$F\beta$ (weighted)	0,671	0,894	0,403	0,325
Precisão (macro)	0,521	0,551	0,513	0,505
Precisão (weighted)	0,931	0,927	0,926	0,913
Revocação (macro)	0,594	0,666	0,530	0,508
Revocação (weighted)	0,359	0,797	0,170	0,148

Fonte: Elaboração própria.

Nota: Métricas obtidas em 100 iterações de treino com dados simulados e teste com dados observados para cada conjuntos de atributos, com uma inicialização aleatória do classificador diferente em cada repetição. Apenas as classes água e plástico foram consideradas. Os melhores resultados para cada métrica foram destacados em negrito.

A Tabela 8 detalha o padrão de erros e acertos por classe e ano de aquisição das imagens, para cada conjunto de atributos. Destacados em negrito estão os resultados mais acurados obtidos. Todos os conjuntos tiveram mau desempenho ao classificar a água das imagens de 2019, e apenas o conjunto de atributos C obteve acurácia médias razoáveis ao classificar o plástico nas imagens de 2019. Nos dados de 2021, apenas o conjunto B teve um bom desempenho com a classificação da água, mas todos os conjuntos de atributos alcançaram boas acurácia para o plástico. O classificador atribuiu a maior parte das amostras à classe minoritária - o plástico - um comportamento diferente do esperado para um conjunto de dados tão desbalanceado.

O bom desempenho dos classificadores com os pixels contendo plástico nas imagens de 2021 parece estar relacionado ao maior percentual de cobertura plástica neste conjunto, conforme resultados apresentados na Tabela 9: todos os conjuntos de atributos apresentam ótimas acurácia para os pixels com 100% de cobertura plástica. Os pixels totalmente cobertos por plástico eram oriundos do PLP 2021, que utilizou apenas um tipo de polímero para construção do alvo. Isso significa que, além de percentuais de cobertura maiores devido ao tamanho do alvo, a assinatura espectral do polímero também era homogênea, enquanto no PLP 2019 diversos polímeros com diferentes assinaturas faziam parte da composição do alvo. A maior concentração de uma assinatura que também é mais homogênea pode ter sido um fator determinante para os resultados melhores.

Para os pixels com percentuais de cobertura menores, os conjuntos A, C e D obtiveram altas taxas de acerto, contudo, observando as métricas da Tabela 7, percebe-se que os modelos que utilizaram estes atributos classificaram a maior parte dos pixels das imagens como plástico, de forma que os acertos com os percentuais de cobertura mais baixos de plástico não podem ser atribuídos à sensibilidade do modelo. Já o conjunto de atributos B confundiu frequentemente os pixels com baixo percentual de cobertura plástica com água, mas acertou quase a totalidade dos pixels com 100% de cobertura plástica.

As taxas médias de acerto do modelo gerado pelo conjunto de atributos B foram de 98,2% e 93,6% nos dias 21/07/2021 e 25/08/2021, respectivamente. Para a maior parte das demais datas, o classificador obteve acurácia média similar ou inferiores. Nestas datas, os alvos artificiais apresentavam submersão ou bioincrustação. Logo, o fato de ter obtido boas acurácia indica que estes fatores não influenciaram negativamente a detectabilidade do material pelo classificador.

Os resultados encontrados são promissores, pois confirmam o potencial dos MTRs e do aprendizado de máquina na detecção remota de poluição marinha e costeira por plásticos. Os MTRs, sendo ferramentas que oferecem total controle dos cenários simulados, podem explorar as características dos mais variados sensores, poluentes e superfícies, enquanto os modelos de aprendizado de máquina podem ser refinados com estratégias de adaptação às características da base de dados.

A qualidade dos dados é uma questão central para gerar simulações e modelos acurados, e a disponibilização de conjuntos novos e mais diversos tem o potencial de impulsionar soluções baseadas em MTRs e aprendizado de máquina para detecção de poluição marinha e costeira por plásticos.

Tabela 8 – Erros e acertos do modelo agrupados por classe do pixel e ano de aquisição da imagem.

Ano	Classe	Total	Conjunto de atributos	Média de acertos	Variação de acertos (mínimo - máximo)
2019	Água	1346	A	460 (34,2%)	341 - 626 (25,3% - 46,5%)
2019	Água	1346	B	1064,8 (79,1%)	682 - 1184 (50,7% - 88%)
2019	Água	1346	C	174,87 (13%)	139 - 218 (10,3% - 16,2%)
2019	Água	1346	D	173,27 (12,9%)	96 - 672 (7,1% - 49,9%)
2019	Plástico	49	A	34,13 (69,7 %)	23 - 39 (46,9% - 79,6%)
2019	Plástico	49	B	15,32 (31,3%)	11 - 22 (22,4% - 44,9%)
2019	Plástico	49	C	44,68 (91,2%)	44 - 45 (89,8% - 91,8%)
2019	Plástico	49	D	42,49 (86,7%)	16 - 47 (32,7% - 95,9%)
2021	Água	597	A	185,39 (31,1 %)	128 - 268 (21,4% - 44,9%)
2021	Água	597	B	512,5 (85,8%)	466 - 570 (78,1% - 95,5%)
2021	Água	597	C	77,81 (13%)	44 - 122 (7,4% - 20,4%)
2021	Água	597	D	36,26 (6,1%)	14 - 209 (2,3% - 35%)
2021	Plástico	54	A	53,97 (99,9 %)	53 - 54 (98,1% - 100%)
2021	Plástico	54	B	38,34 (71%)	32 - 43 (59,3% - 79,6%)
2021	Plástico	54	C	51,06 (94,6%)	47 - 53 (87% - 98,1%)
2021	Plástico	54	D	51,03 (94,5%)	42 - 54 (77,8% - 100%)

Fonte: Elaboração própria.

Nota: Destacados em negrito estão os resultados mais acurados obtidos para ambas as classes de um mesmo conjunto.

Tabela 9 – Acurárias médias em amostras de plástico agrupados por percentual de cobertura do pixel para cada conjunto de atributos.

Percentual de cobertura	Total	Conjunto de atributos	Média de acertos	Variação de acertos (mínimo - máximo)
<=25%	39	A	24,63 (63,2%)	15 - 29 (38,5% - 74,4%)
<=25%	39	B	11,75 (30,1%)	8 - 17 (20,5% - 43,6%)
<=25%	39	C	34,68 (88,9%)	34 - 35 (87,2% - 89,7%)
<=25%	39	D	32,86 (84,3%)	12 - 37 (30,8% - 94,9%)
26% - 50%	9	A	8,5 (94,4%)	7 - 9 (77,8% - 100%)
26% - 50%	9	B	3,57 (39,7%)	3 - 5 (33,3% - 55,6%)
26% - 50%	9	C	9 (100%)	9 (100%)
26% - 50%	9	D	8,64 (96,0%)	4 - 9 (44,4% - 100%)
51% - 99%	1	A	1 (100%)	1 (100%)
51% - 99%	1	B	0 (0%)	0 (0%)
51% - 99%	1	C	1 (100%)	1 (100%)
51% - 99%	1	D	0,99 (99%)	0 - 1 (0% - 100%)
100%	15	A	15 (100%)	15 (100%)
100%	15	B	14,63 (97,5%)	12 - 15 (80% - 100%)
100%	15	C	13,79 (91,9%)	13 - 14 (86,7% - 93,3%)
100%	15	D	14,11 (94,1%)	12 - 15 (80% - 100%)
Desconhecido	39	A	38,97 (99,9 %)	38 - 39 (97,4% - 100%)
Desconhecido	39	B	23,71 (60,8 %)	20 - 28 (51,8% - 71,8%)
Desconhecido	39	C	37,27 (95,6 %)	34 - 39 (87,2% - 100%)
Desconhecido	39	D	36,92 (94,7 %)	30 - 39 (76,9% - 100%)

Fonte: Elaboração própria.

4 CONCLUSÕES

A metodologia desenvolvida atendeu aos objetivos do trabalho. Foi possível avaliar o potencial e as limitações da detecção de plásticos marinhos por imagens e métodos de sensoriamento remoto: há evidências de que a detecção é viável, embora ainda existam limitações relacionadas principalmente às resoluções dos sensores e aos bancos de dados espectrais disponíveis. Embora tenham ocorrido avanços, a detecção ainda precisa ser realizada em nível de subpixel, dependendo de fatores como pureza espectral e percentual de cobertura do alvo.

Os resultados da análise exploratória e da classificação não supervisionada permitem estudar o comportamento espectral da poluição marinha, demonstrando similaridades, contrastes e padrões de associação. Tanto na classificação supervisionada quanto na não supervisionada, a diferenciação entre os alvos de plástico ou madeira e o substrato tende a ocorrer somente a partir de percentuais mais altos de cobertura do pixel. Em ambos os casos, o conjunto de atributos B, formado somente pelas bandas do sensor, teve um desempenho de destaque, enquanto o uso dos índices como variáveis de entrada não apresentou bons resultados.

Os resultados da classificação supervisionada demonstraram que os MTRs são aplicáveis à detecção remota de detritos plásticos, e que os classificadores supervisionados treinados com dados simulados gerados por MTRs podem ser aplicados a dados observados em determinadas condições. No entanto, há uma dependência da pureza espectral e do percentual de cobertura plástica dos pixels, sendo necessário buscar outras estratégias para lidar com casos de mistura espectral, ruído e necessidade de detecção em nível de subpixel. Possíveis abordagens incluem modelos de separação espectral, técnicas de super-resolução de imagens, metodologias específicas para tratamento do desbalanceamento dos dados e criação de simulações com maior variedade de parâmetros, utilizando, quando possível, bancos de dados espectrais com maior diversidade de elementos e condições, de forma a aproximar as cenas sintéticas das situações encontradas no meio ambiente.

Melhorias nos conjuntos de dados disponíveis podem influenciar positivamente o uso dos modelos de aprendizado de máquina. São necessários novos bancos de dados espectrais para criação de novas simulações, com maior variabilidade de elementos e condições, permitindo incluir fatores como submersão e bioncrustação nas cenas sintéticas. Também é importante a criação de novos conjuntos de dados observados de alta qualidade com validação *in situ*, para análises mais robustas dos fatores que influenciam a detectabilidade da poluição marinha.

Atingir uma acurácia razoável em uma das configurações testadas é um grande ganho, tendo em vista que o modelo conseguiu aprender informações sobre o domínio a partir de dados simulados que pertenciam a uma distribuição diferente da dos dados observados, utilizados no teste. Mesmo sem o cumprimento deste pressuposto básico do aprendizado de máquina, o modelo conseguiu, em determinadas condições, extrair relações entre as variáveis presentes em ambos os conjuntos, o que sugere que estas relações são suficientemente fortes para serem detectáveis mesmo em conjuntos de distribuições diferentes.

5 FINANCIAMENTO

Não houve financiamento para a realização deste trabalho.

REFERÊNCIAS

- ACUÑA-RUZ, T. et al. Anthropogenic marine debris over beaches: Spectral characterization for remote sensing applications. *Remote Sensing of Environment*, Elsevier, v. 217, p. 309–322, 2018.
- ALENCAR, M. V. et al. How far are we from robust estimates of plastic litter leakage to the environment? *Journal of Environmental Management*, Elsevier, v. 323, p. 116195, 2022.
- ALI, A. M. et al. Machine learning methods' performance in radiative transfer model inversion to retrieve plant traits from sentinel-2 data of a mixed mountain forest. *International Journal of Digital Earth*, Taylor & Francis, v. 14, n. 1, p. 106–120, 2021.
- AMATO-LOURENÇO, L. F. et al. Presence of airborne microplastics in human lung tissue. *Journal of Hazardous Materials*, Elsevier, p. 126124, 2021.
- BAK, S. et al. Detection and monitoring of beach litter using uav image and deep neural network. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 2019.
- BAKER, E. et al. *Marine litter vital graphics*. [S.I.]: Grid-Arendal, 2016. Disponível em: <https://gridarendal-website-live.s3.amazonaws.com/production/documents/s_document/11/original/MarineLitterVG.pdf?1488455779>. Acesso em 18 de março de 2023.
- BASU, B. et al. Development of novel classification algorithms for detection of floating plastic debris in coastal waterbodies using multispectral sentinel-2 remote sensing imagery. *Remote Sensing*, MDPI, v. 13, n. 8, p. 1598, 2021.
- BEAUMONT, N. J. et al. Global ecological, social and economic impacts of marine plastic. *Marine pollution bulletin*, Elsevier, v. 142, p. 189–195, 2019.
- BIERMANN, L. et al. Finding plastic patches in coastal waters using optical satellite data. *Scientific reports*, Nature Publishing Group, v. 10, n. 1, p. 1–10, 2020.
- BIOUCAS-DIAS, J. M. et al. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and remote sensing magazine*, IEEE, v. 1, n. 2, p. 6–36, 2013.
- BOWDEN, G. J.; DANDY, G. C.; MAIER, H. R. Input determination for neural network models in water resources applications. part 1—background and methodology. *Journal of Hydrology*, Elsevier, v. 301, n. 1-4, p. 75–92, 2005.
- BRASIL. Ministério do Meio Ambiente. *Plano Nacional de Combate ao Lixo no Mar*. 2019. Disponível em: <<https://www.marinha.mil.br/dellaguna/sites/www.marinha.mil.br.dellaguna/files/Downloads/Plano-de-Combate-ao-Lixo-no-Mar-20.03-FINAL.pdf>>. Acesso em: 18 de março de 2023.

- CAMPANALE, C. et al. A detailed review study on potential effects of microplastics and additives of concern on human health. *International journal of environmental research and public health*, Multidisciplinary Digital Publishing Institute, v. 17, n. 4, p. 1212, 2020.
- CAMPOLO, M.; SOLDATI, A.; ANDREUSSI, P. Artificial neural network approach to flood forecasting in the river arno. *Hydrological Sciences Journal*, Taylor & Francis, v. 48, n. 3, p. 381–398, 2003.
- CARLSON, D. F. et al. Combining litter observations with a regional ocean model to identify sources and sinks of floating debris in a semi-enclosed basin: the adriatic sea. *Frontiers in Marine Science*, Frontiers Media SA, v. 4, p. 78, 2017.
- CONGALTON, R. G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*, Elsevier, v. 37, n. 1, p. 35–46, 1991.
- COSTA, I. D. da et al. Microplastics in fishes in amazon riverine beaches: Influence of feeding mode and distance to urban settlements. *Science of The Total Environment*, Elsevier, p. 160934, 2022.
- COX, K. D. et al. Human consumption of microplastics. *Environmental science & technology*, ACS Publications, v. 53, n. 12, p. 7068–7074, 2019.
- DANOPOULOS, E. et al. A rapid review and meta-regression analyses of the toxicological impacts of microplastic exposure in human cells. *Journal of Hazardous Materials*, Elsevier, p. 127861, 2021.
- DIAS, J. *Todo o mundo é composto de mudança": Considerações sobre o clima ea sua história. II-Factores astronómicos*. 2018. Disponível em: <https://www.researchgate.net/publication/301553846_Todo_o_mundo_e_composto_de_mudanca_Con sideracoes_sobre_o_clima_e_a_sua_historia_II_-O_Sistema_Climatico_Terrestre>.
- DODGE, Y. *The concise encyclopedia of statistics*. [S.I.]: Springer Science & Business Media, 2008.
- DUTHUIT, S. et al. Assessing the effects of the clumping phenomenon on brdf of a maize crop based on 3d numerical scenes using dart model. *Agricultural and forest meteorology*, Elsevier, v. 148, n. 8-9, p. 1341–1352, 2008.
- EMMERIK, T. van; VRIEND, P.; PEEREBOOM, E. C. Roadmap for long-term macroplastic monitoring in rivers. *Frontiers in Environmental Science*, Frontiers, v. 9, p. 716, 2022.
- ESA, E. S. A. Sentinel-2 user handbook. *ESA Standard Document*, 2015. Disponível em: <https://sentinel.esa.int/documents/247904/685211/Sentinel-2_User_Handbook.pdf/8869acdf-fd84-43ec-ae8c-3e80a436a16c?t=1438278087000>.
- FACELI, K. et al. *Inteligência Artificial: Uma abordagem de aprendizado de máquina*. 2. ed. [S.I.]: Grupo Gen-LTC, 2021.
- FAWCETT, T. An introduction to roc analysis. *Pattern recognition letters*, Elsevier, v. 27, n. 8, p. 861–874, 2006.

- FENG, Q. et al. Mapping of plastic greenhouses and mulching films from very high resolution remote sensing imagery based on a dilated and non-local convolutional neural network. *International Journal of Applied Earth Observation and Geoinformation*, Elsevier, v. 102, p. 102441, 2021.
- FERREIRA, G. V. et al. Plastic in the inferno: Microplastic contamination in deep-sea cephalopods (*vampyroteuthis infernalis* and *abralia veranyi*) from the southwestern atlantic. *Marine Pollution Bulletin*, Elsevier, v. 174, p. 113309, 2022.
- FLEURY, J.-B.; BAULIN, V. A. Microplastics destabilize lipid membranes by mechanical stretching. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 118, n. 31, 2021.
- FRANCESCHINI, S. et al. Rummaging through the bin: modelling marine litter distribution using artificial neural networks. *Marine pollution bulletin*, Elsevier, v. 149, p. 110580, 2019.
- FREITAS, T. B. et al. In an octopus's garden in the shade: Underwater image analysis of litter use by benthic octopuses. *Marine Pollution Bulletin*, Elsevier, v. 175, p. 113339, 2022.
- GALGANI, F.; HANKE, G.; MAES, T. Global distribution, composition and abundance of marine litter. *Marine anthropogenic litter*, Springer International Publishing, p. 29–56, 2015.
- GARABA, S. P.; DIERSSEN, H. M. *Spectral reference library of 11 types of virgin plastic pellets common in marine plastic debris*. 2017. <<https://tinyurl.com/y7de3cup>>. (Accessed on 08/07/2021).
- GARABA, S. P.; DIERSSEN, H. M. An airborne remote sensing case study of synthetic hydrocarbon detection using short wave infrared absorption features identified from marine-harvested macro-and microplastics. *Remote sensing of environment*, v. 205, p. 224 – 235, 2018. ISSN 0034-4257. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0034425717305722>>.
- GARABA, S. P.; DIERSSEN, H. M. *Spectral reflectance of dry and wet marine-harvested microplastics from Kamilo Point, Pacific Ocean*. 2019. <<https://tinyurl.com/2amnd89u>>. (Accessed on 08/07/2021).
- GARABA, S. P.; DIERSSEN, H. M. *Spectral reflectance of washed ashore macroplastics*. 2019. <<https://ecosis.org/package/spectral-reflectance-of-washed-ashore-macroplastics>>. (Accessed on 08/07/2021).
- GASTELLU-ETCHEGORRY, J.-P.; GRAU, E.; LAURET, N. Dart: A 3d model for remote sensing images and radiative budget of earth surfaces. *Modeling and simulation in Engineering*, p. ISBN-978, 2012.
- GEYER, R.; JAMBECK, J. R.; LAW, K. L. Production, use, and fate of all plastics ever made. *Science advances*, American Association for the Advancement of Science, v. 3, n. 7, p. e1700782, 2017.

GONÇALVES, G. et al. Mapping marine litter using uas on a beach-dune system: a multidisciplinary approach. *Science of the Total Environment*, Elsevier, v. 706, p. 135742, 2020.

GOODMAN, K. E. et al. Exposure of human lung cells to polystyrene microplastics significantly retards cell proliferation and triggers morphological changes. *Chemical Research in Toxicology*, ACS Publications, v. 34, n. 4, p. 1069–1081, 2021.

HEDLEY, J. et al. Capability of the sentinel 2 mission for tropical coral reef mapping and coral bleaching detection. *Remote Sensing of Environment*, Elsevier, v. 120, p. 145–155, 2012.

IMO. *International Maritime Organization. GESAMP - Group of Experts on the Scientific Aspects of Marine Environmental Protection*. 2023. Disponível em: <<http://www.gesamp.org/>>. Acesso em: 18 de março de 2023.

JACQUEMOUD, S. et al. Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode. *Remote Sensing of Environment*, Elsevier, v. 74, n. 3, p. 471–481, 2000.

JAIN, A.; KUMAR, A. M. Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing*, Elsevier, v. 7, n. 2, p. 585–592, 2007.

JAIN, A.; SUDHEER, K.; SRINIVASULU, S. Identification of physical processes inherent in artificial neural network rainfall runoff models. *Hydrological processes*, Wiley Online Library, v. 18, n. 3, p. 571–581, 2004.

JAKOVLJEVIC, G.; GOVEDARICA, M.; ALVAREZ-TABOADA, F. A deep learning model for automatic plastic mapping using unmanned aerial vehicle (uav) data. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 12, n. 9, p. 1515, 2020.

JANOUTOVÁ, R. et al. Influence of 3d spruce tree representation on accuracy of airborne and satellite forest reflectance simulated in dart. *Forests*, Multidisciplinary Digital Publishing Institute, v. 10, n. 3, p. 292, 2019.

JENSEN, J. R. et al. *Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres*. [S.I.]: Parêntese Editora, 2011.

JUSTINO, A. K. et al. The role of mesopelagic fishes as microplastics vectors across the deep-sea layers from the southwestern tropical atlantic. *Environmental Pollution*, Elsevier, v. 300, p. 118988, 2022.

KENTSCH, S. et al. Computer vision and deep learning techniques for the analysis of drone-acquired forest images, a transfer learning study. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 12, n. 8, p. 1287, 2020.

KUBOTA, M. A mechanism for the accumulation of floating marine debris north of hawaii. *Journal of Physical Oceanography*, v. 24, n. 5, p. 1059–1064, 1994.

KÜKENBRINK, D. et al. Modelling of three-dimensional, diurnal light extinction in two contrasting forests. *Agricultural and Forest Meteorology*, Elsevier, v. 296, p. 108230, 2021.

- KYLILI, K. et al. Identifying floating plastic marine debris using a deep learning approach. *Environmental Science and Pollution Research*, Springer, v. 26, p. 17091–17099, 2019.
- LAVERS, J. L.; BOND, A. L. Exceptional and rapid accumulation of anthropogenic debris on one of the world's most remote and pristine islands. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 114, n. 23, p. 6052–6055, 2017.
- LAW, K. L. et al. Plastic accumulation in the north atlantic subtropical gyre. *Science*, American Association for the Advancement of Science, v. 329, n. 5996, p. 1185–1188, 2010.
- LEAHY, P.; KIELY, G.; CORCORAN, G. Structural optimisation and input selection of an artificial neural network for river level prediction. *Journal of hydrology*, Elsevier, v. 355, n. 1-4, p. 192–201, 2008.
- LEBRETON, L. et al. Evidence that the great pacific garbage patch is rapidly accumulating plastic. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 1–15, 2018.
- LIU, J. et al. Urban green plastic cover mapping based on vhr remote sensing images and a deep semi-supervised learning framework. *ISPRS International Journal of Geo-Information*, Multidisciplinary Digital Publishing Institute, v. 9, n. 9, p. 527, 2020.
- LIU, K. et al. Consistent transport of terrestrial microplastics to the ocean through atmosphere. *Environmental science & technology*, ACS Publications, v. 53, n. 18, p. 10612–10619, 2019.
- MA, L. et al. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS journal of photogrammetry and remote sensing*, Elsevier, v. 152, p. 166–177, 2019.
- MACARTHUR, D. E.; WAUGHRAY, D.; STUCHTEY, M. R. The new plastics economy, rethinking the future of plastics. In: ELLEN MACARTHUR FOUNDATION AND MCKINSEY & COMPANY LONDON, UK. *World Economic Forum*. 2016. Disponível em: <<https://www.ellenmacarthurfoundation.org/publications/the-new-plastics-economy-rethinking-the-future-of-plastics>>.
- MAES, T. et al. *From pollution to solution: A global assessment of marine litter and plastic pollution*. [S.I.]: United Nations Environment Programme, 2021.
- MAGRINI, A. *Impactos ambientais causados pelos plásticos: uma discussão abrangente sobre os mitos e os dados científicos*. [S.I.]: Editora E-papers, 2012.
- MALENOVSKÝ, Z. et al. Sentinels for science: Potential of sentinel-1,-2, and-3 missions for scientific observations of ocean, cryosphere, and land. *Remote Sensing of environment*, Elsevier, v. 120, p. 91–101, 2012.
- MARTIN, C. et al. Use of unmanned aerial vehicles for efficient beach litter monitoring. *Marine pollution bulletin*, Elsevier, v. 131, p. 662–673, 2018.

- MARTÍNEZ-VICENTE, V. et al. Measuring marine plastic debris from space: Initial assessment of observation requirements. *Remote Sensing*, MDPI, v. 11, n. 20, p. 2443, 2019.
- MAXIMENKO, N. et al. Toward the integrated marine debris observing system. *Frontiers in marine science*, v. 6, 2019.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, p. 115–133, 1943.
- MIHAI, F.-C. et al. Plastic pollution in marine and freshwater environments: abundance, sources, and mitigation. In: *Emerging Contaminants in the Environment*. [S.I.]: Elsevier, 2022. p. 241–274.
- MONARD, M. C.; BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. *Sistemas inteligentes-Fundamentos e aplicações*, v. 1, n. 1, p. 32, 2003.
- MONTEIRO, R. et al. Glove: The global plastic ingestion initiative for a cleaner world. *Marine Pollution Bulletin*, Elsevier, v. 185, p. 114244, 2022.
- MOSHTAGHI, M. et al. Spectral reflectance of marine macroplastics in the vnir and swir measured in a controlled environment. *Scientific Reports*, Nature Publishing Group, v. 11, n. 1, p. 1–12, 2021.
- NEZAMI, S. et al. Tree species classification of drone hyperspectral and rgb imagery with deep learning convolutional neural networks. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 12, n. 7, p. 1070, 2020.
- OKIN, G. S.; GU, J. The impact of atmospheric conditions and instrument noise on atmospheric correction and spectral mixture analysis of multispectral imagery. *Remote Sensing of Environment*, Elsevier, v. 164, p. 130–141, 2015.
- OLIVEIRA, G. G. d.; PEDROLLO, O. C.; CASTRO, N. M. dos R. Metodologia de análise de sensibilidade e exclusão de variáveis de entrada em simulação hidrológica por redes. 2011.
- OLIVEIRA, G. G. d.; PEDROLLO, O. C.; CASTRO, N. M. dos R. O desempenho das redes neurais artificiais (rnas) para simulação hidrológica mensal. *Revista Brasileira de Recursos Hídricos*, v. 19, n. 2, p. 251–265, 2014.
- PAOLA, J. D.; SCHOWENGERDT, R. A. A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and remote sensing*, IEEE, v. 33, n. 4, p. 981–996, 1995.
- PAPAGEORGIOU, D. et al. Sentinel-2 detection of floating marine litter targets with partial spectral unmixing and spectral comparison with other floating materials (plastic litter project 2021). *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 14, n. 23, p. 5997, 2022.
- PATEL, V.; MISTREE, K. A review on different image interpolation techniques for image enhancement. *International Journal of Emerging Technology and Advanced Engineering*, v. 3, n. 12, p. 129–133, 2013.

- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.
- PÓVOA, A. A.; ARAÚJO, F. V. de; SKINNER, L. F. Macroorganisms fouled in marine anthropogenic litter (rafting) around a tropical bay in the southwest atlantic. *Marine Pollution Bulletin*, Elsevier, v. 175, p. 113347, 2022.
- PRATA, J. C. et al. Environmental exposure to microplastics: An overview on possible human health effects. *Science of the Total Environment*, Elsevier, v. 702, p. 134455, 2020.
- PRATI, R. C. *Novas abordagens em aprendizado de máquina para a geração de regras, classes desbalanceadas e ordenação de casos*. Tese (Doutorado) — Universidade de São Paulo, 2006.
- RBINS. Acolite python user manual. Royal Belgium Institute of Natural Sciences, 2022. Disponível em: <https://github.com/acolite/acolite/releases/download/20220222.0/acolite__manual__20220222.pdf>.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, American Psychological Association, v. 65, n. 6, p. 386, 1958.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group UK London, v. 323, n. 6088, p. 533–536, 1986.
- SEAN, G. et al. *Rasterio: geospatial raster I/O for Python programmers*. 2013. Disponível em: <<https://github.com/mapbox/rasterio>>.
- SEPULCRE-CANTO, G. et al. Discriminating irrigated and rainfed olive orchards with thermal aster imagery and dart 3d simulation. *Agricultural and Forest Meteorology*, Elsevier, v. 149, n. 6-7, p. 962–975, 2009.
- SHEVEALY, S.; COURTNEY, K.; PARKS, J. E. The honolulu strategy: a global framework for prevention and management of marine debris. 2012.
- SILVA, J. A. D. d. *Proposta de um índice para detecção de macro e microplásticos em zonas costeiras através de sensoriamento remoto hiperespectral*. 2020. <<https://lume.ufrgs.br/handle/10183/209920>>. Dissertação (Mestrado), UFRGS (Universidade Federal do Rio Grande do Sul), Porto Alegre, Brazil.
- SOUZA, C. C. d. et al. Técnicas de sensoriamento remoto como subsídios aos estudos de florestas implantadas no brasil-uma revisão bibliográfica. *Ciência Florestal*, SciELO Brasil, v. 17, p. 409–417, 2007.
- SP. *Secretaria de Infraestrutura e Meio Ambiente do Estado de São Paulo. Plano Estratégico de Monitoramento e Avaliação do Lixo no Mar do Estado de São Paulo*. 2021. Disponível em: <<https://www.pemalm.com/o-plano>>. Acesso em: 18 de março de 2023.
- STEWART, R. H. *Introduction to physical oceanography*. [S.I.]: Robert H. Stewart, 2008.

- THEMISTOCLEOUS, K. et al. Investigating detection of floating plastic litter from space using sentinel-2 imagery. *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 12, n. 16, p. 2648, 2020.
- TOPOUZELIS, K. et al. Remote sensing of sea surface artificial floating plastic targets with sentinel-2 and unmanned aerial systems (plastic litter project 2019). *Remote Sensing*, Multidisciplinary Digital Publishing Institute, v. 12, n. 12, p. 2013, 2020.
- TOPOUZELIS, K.; PAPAKONSTANTINOU, A.; GARABA, S. P. Detection of floating plastics from satellite and unmanned aerial systems (plastic litter project 2018). *International Journal of Applied Earth Observation and Geoinformation*, Elsevier, v. 79, p. 175–183, 2019.
- TORRE, G. E. De-la. Microplastics: an emerging threat to food security and human health. *Journal of food science and technology*, Springer, v. 57, n. 5, p. 1601–1608, 2020.
- UN. *United Nations. Objetivos de Desenvolvimento Sustentável | As Nações Unidas no Brasil*. 2023a. Disponível em: <<https://brasil.un.org/pt-br/sdgs>>. Acesso em: 18 de março de 2023.
- UN. *United Nations. PNUMA - Programa das Nações Unidas para o Meio Ambiente | As Nações Unidas no Brasil*. 2023b. Disponível em: <<https://www.unep.org/pt-br/sobre-onu-meio-ambiente>>. Acesso em: 18 de março de 2023.
- UN. *United Nations. GPML - Global Partnership On Plastic Pollution and Marine Litter*. 2023c. Disponível em: <<https://www.gpmarinelitter.org/>>. Acesso em: 18 de março de 2023.
- UN. *United Nations. Ocean Decade – The Science We Need For The Ocean We Want*. 2023d. Disponível em: <<https://oceandecade.org/>>. Acesso em: 18 de março de 2023.
- UNEP et al. Marine litter: An analytical overview. 2005. Disponível em: <<https://wedocs.unep.org/handle/20.500.11822/8348?show=full>>.
- UNEP, U. N. E. P. *Marine Plastic Debris and Microplastics: Global Lessons and Research to Inspire Action and Guide Policy Change*. 2016. Disponível em: <<https://wedocs.unep.org/20.500.11822/7720>>.
- USP. *Universidade de São Paulo | Cátedra UNESCO para Sustentabilidade dos Oceanos*. 2023. Disponível em: <<https://catedraoceano.iea.usp.br/decadadooceano/>>. Acesso em: 18 de março de 2023.
- VANHELLEMONT, Q. Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the landsat and sentinel-2 archives. *Remote Sensing of Environment*, Elsevier, v. 225, p. 175–192, 2019.
- VANHELLEMONT, Q.; RUDDICK, K. Acolite for sentinel-2: Aquatic applications of msi imagery. In: *Proceedings of the 2016 ESA Living Planet Symposium, Prague, Czech Republic*. [S.I.: s.n.], 2016. p. 9–13.

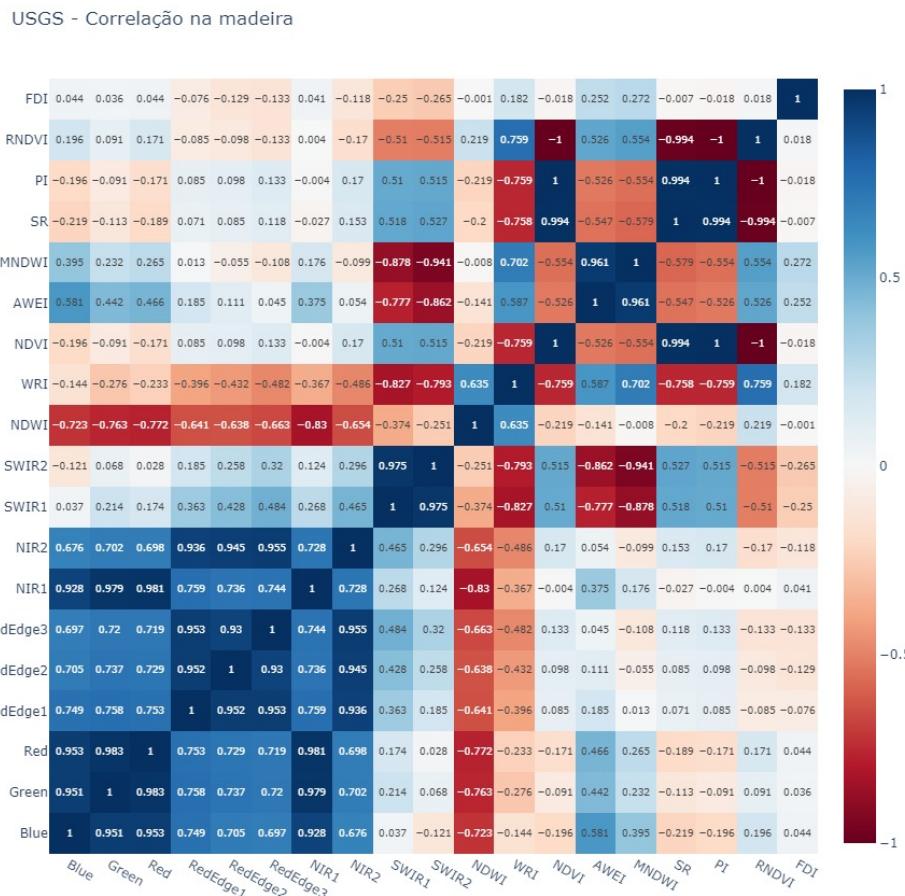
- VANHELLEMONT, Q.; RUDDICK, K. Atmospheric correction of metre-scale optical satellite data for inland and coastal water applications. *Remote sensing of environment*, Elsevier, v. 216, p. 586–597, 2018.
- VANHELLEMONT, Q.; RUDDICK, K. Atmospheric correction of sentinel-3/olci data for mapping of suspended particulate matter and chlorophyll-a concentration in belgian turbid coastal waters. *Remote Sensing of Environment*, Elsevier, v. 256, p. 112284, 2021.
- WANG, D. et al. The assessment of landsat-8 oli atmospheric correction algorithms for inland waters. *Remote Sensing*, MDPI, v. 11, n. 2, p. 169, 2019.
- WIDROW, B.; HOFF, M. E. *Adaptive switching circuits*. [S.I.], 1960.
- WOLF, M. et al. Machine learning for aquatic plastic litter detection, classification and quantification (aplastic-q). *Environmental Research Letters*, IOP Publishing, v. 15, n. 11, p. 114042, 2020.
- ZHANG, Q. et al. A review of microplastics in table salt, drinking water, and air: direct human exposure. *Environmental science & technology*, ACS Publications, v. 54, n. 7, p. 3740–3751, 2020.
- ZHAO, W. et al. On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery. *International Journal of Remote Sensing*, Taylor & Francis, v. 36, n. 13, p. 3368–3379, 2015.
- ZHU, X. X. et al. Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, IEEE, v. 5, n. 4, p. 8–36, 2017.

A Apêndice

A.1 Análise exploratória

Os dados complementares da análise exploratória são apresentados aqui. As correlações por classe dos conjuntos de dados são apresentados nas Figuras 38, 39, 40 e 41.

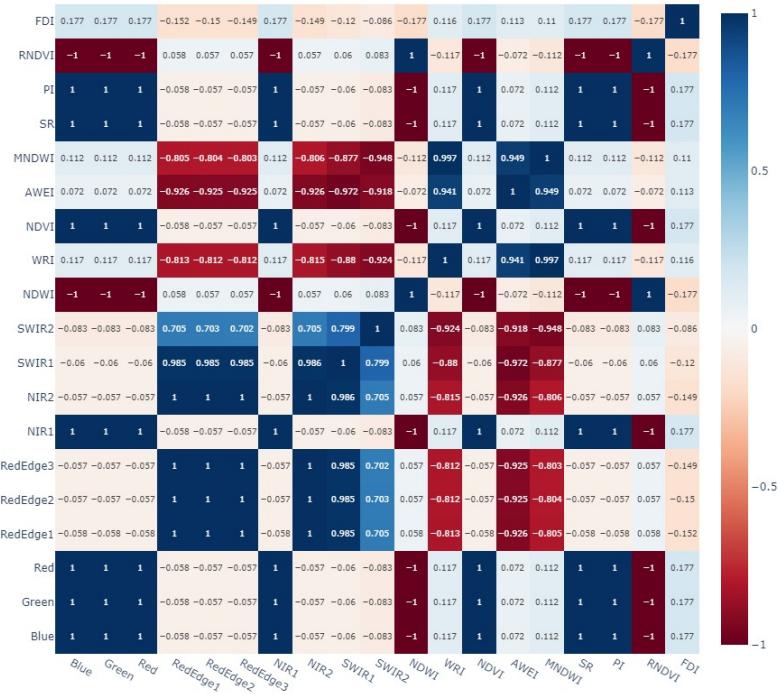
Figura 38 – Correlação de atributos para a classe madeira nos dados USGS (observados).



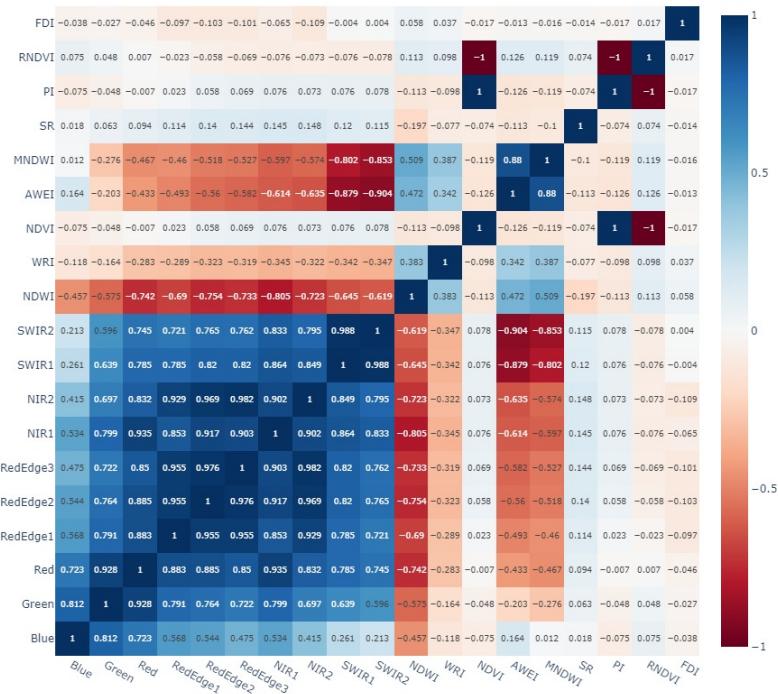
Fonte: Elaboração própria.

Figura 39 – Correlação de atributos para a classe água nos dados DART (simulados) e USGS (observados).

DART - Correlação na água



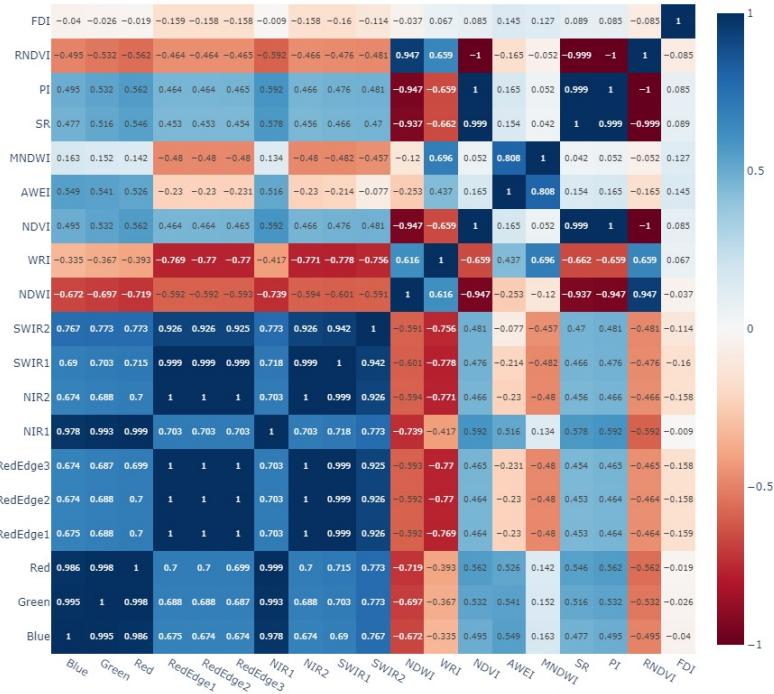
USGS - Correlação na água



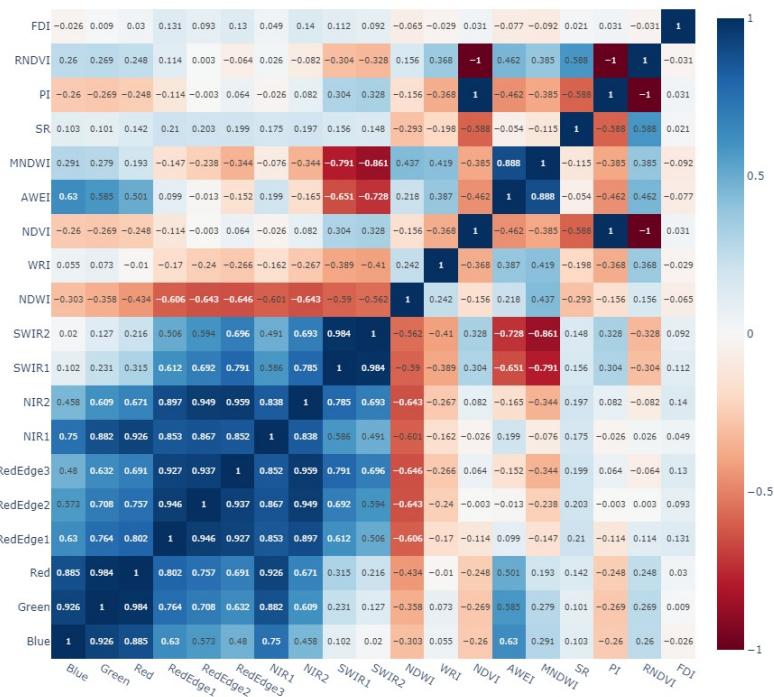
Fonte: Elaboração própria.

Figura 40 – Correlação de atributos para a classe plástico nos dados DART (simulados) e USGS (observados).

DART - Correlação no plástico



USGS - Correlação no plástico



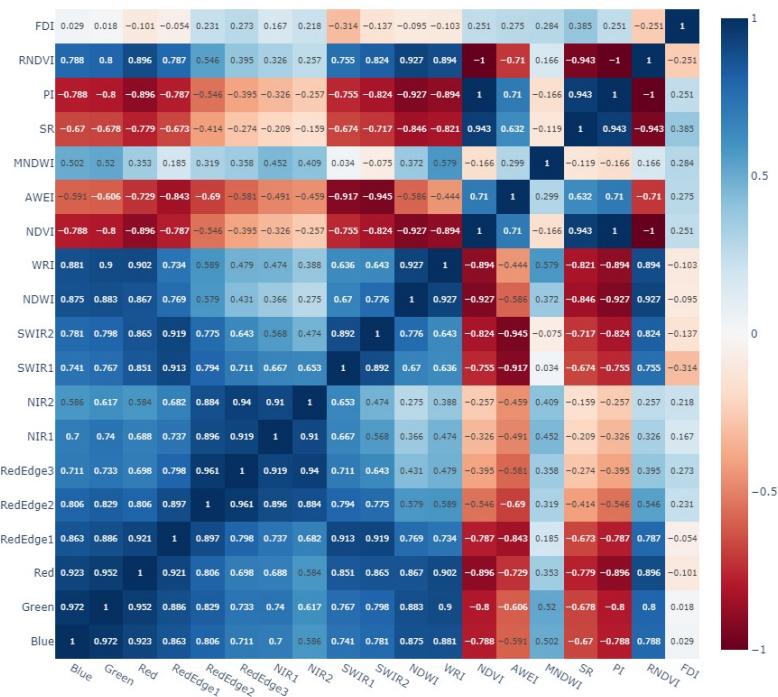
Fonte: Elaboração própria.

Figura 41 – Correlação de atributos para a classe areia nos dados DART (simulados) e para a classe costa nos dados USGS (observados).

DART - Correlação na areia



USGS - Correlação na costa

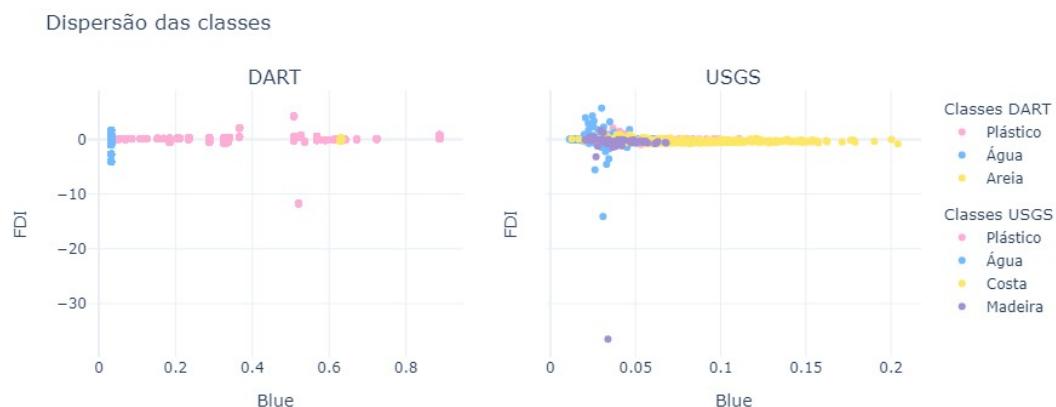


Fonte: Elaboração própria.

Nos gráficos de dispersão bidimensionais, os padrões observados foram:

- Nenhuma separabilidade (Figura 30): Não há separabilidade alguma entre as classes em nenhum dos conjuntos.
- Separabilidade parcial apenas nos dados simulados (Figura 42): As amostras de água e areia ficam separadas nos dados simulados, porém não é possível discernir as amostras de plástico das do substrato. Nos dados observados, não há separabilidade alguma entre as classes.
- Separabilidade parcial nos dados observados e simulados (Figura 43): As amostras de água e areia ficam separadas, porém as de madeira e plástico aparecem misturadas ao substrato tanto nos dados simulados quanto nos observados.
- Separabilidade total nos dados simulados e parcial nos dados observados (Figura 44): As amostras de cada classe ficam quase completamente separadas no gráfico referente ao conjunto de dados simulados, mas apenas as classes do substrato (água e costa) ficam separados no gráfico dos dados observados, enquanto plástico e madeira aparecem misturados a eles.

Figura 42 – Gráfico de dispersão entre a banda Blue e o índice FDI nos dados DART (simulados) e USGS (observados).



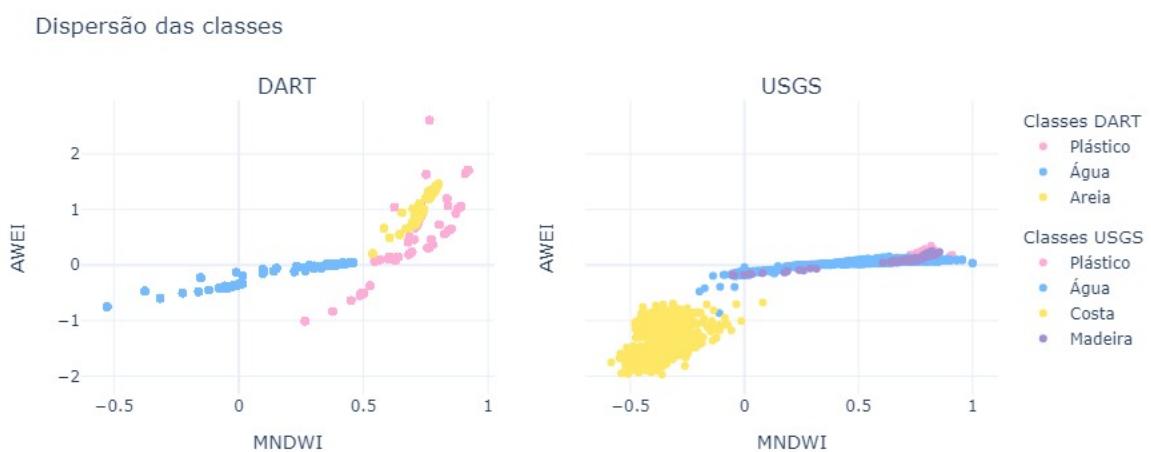
Fonte: Elaboração própria.

Figura 43 – Gráfico de dispersão entre os índices AWEI e FDI nos dados DART (simulados) e USGS (observados).



Fonte: Elaboração própria.

Figura 44 – Gráfico de dispersão entre os índices AWEI e MNDWI nos dados DART (simulados) e USGS (observados).



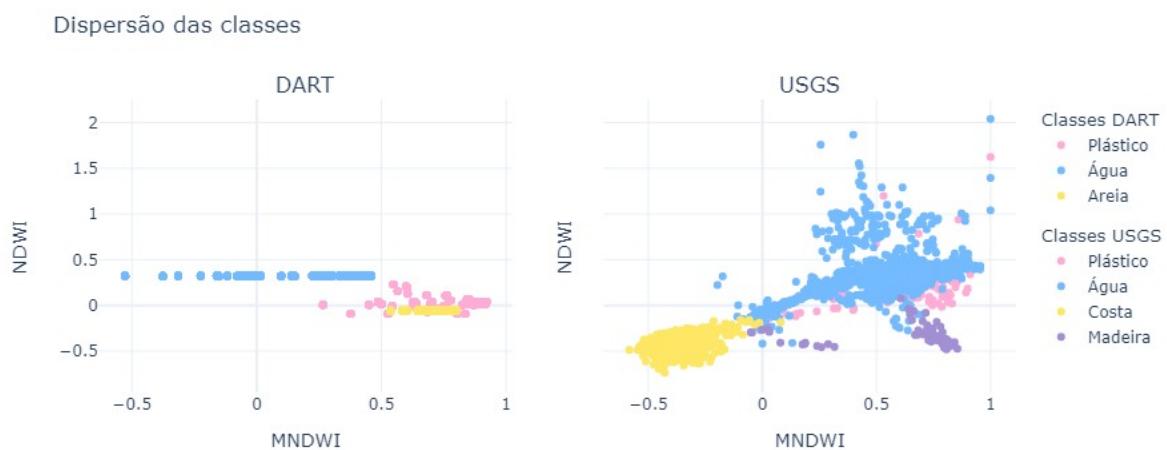
Fonte: Elaboração própria.

O ideal seria ter encontrado uma separabilidade total tanto nos dados observados quanto nos simulados, mas essa condição não foi observada em nenhum dos casos. Contudo, alguns atributos apresentaram alguma separabilidade nos dados observados, aproximando-se da condição ideal de visualização. Foram eles:

- O índice NDWI tem alguma separabilidade entre as classes no conjunto de dados observado quando colocado em dispersão com o atributo MNDWI no eixo oposto (Figura 45);
- As bandas Blue e Green demonstram alguma separabilidade entre as classes no conjunto de dados observado quando colocadas no gráfico de dispersão com os atributos NIR1, NIR2, RedEdge1, RedEdge2 ou RedEdge3 no eixo oposto (Figura 46).

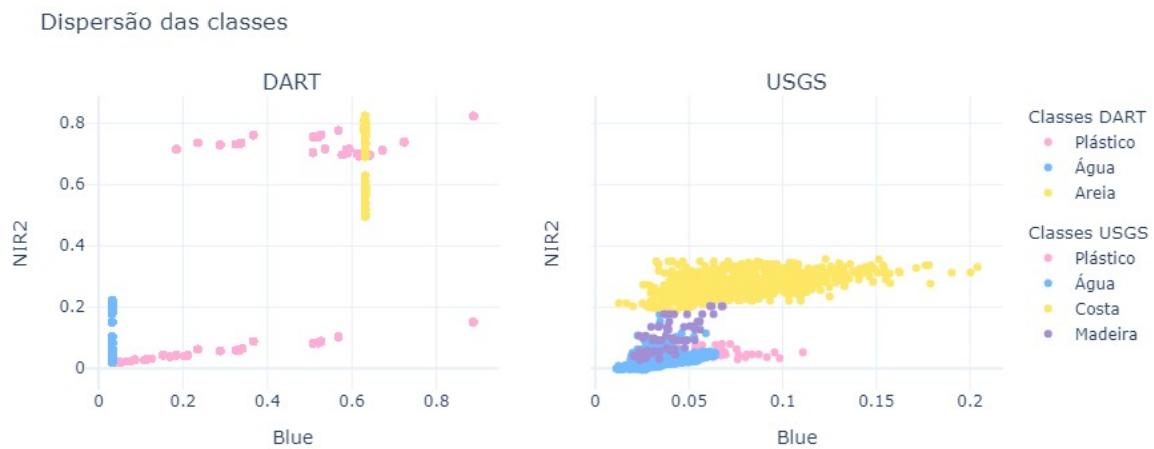
Essas possíveis combinações de atributos podem ser exploradas para validação em outros conjuntos de dados, bem como através de outras técnicas de análise de dispersão.

Figura 45 – Gráfico de dispersão entre os índices MNDWI e NDWI nos dados DART (simulados) e USGS (observados).



Fonte: Elaboração própria.

Figura 46 – Gráfico de dispersão entre as bandas Blue e NIR2 nos dados DART (simulados) e USGS (observados).



Fonte: Elaboração própria.

A.2 Classificação

Os resultados detalhados da classificação não supervisionada são apresentados nas Tabelas 10, 11, 12 e 13.

Os resultados detalhados da classificação supervisionada são apresentadas nas Tabelas 14, 15, 16 e 17, que trazem as médias das matrizes de confusão obtidas nas 100 iterações de treino com dados simulados e teste com dados observados para cada um dos quatro conjuntos de atributos.

Tabela 10 – Agrupamento dos dados DART (simulados) por classificação não supervisionada utilizando os conjuntos de atributos A e B.

Conjunto atributos	Dados	Número de clusters	Classes cluster 0	Classes cluster 1	Classes cluster 2	Classes cluster 3	Classes cluster 4
A	DART	3	210.624 (98,4%) Água e 3.328 (1,56%) Plástico [49,6% total]	212.160 (98%) Areia e 4.224 (1,95%) Plástico [50,1% total]	1.024 (88,9%) Água e 128 (11,1%) Plástico [0,03% total]	-	-
A	DART	4	202.672 (98,4%) Água e 3.200 (1,55%) Plástico [47,8% total]	212.160 (98,3%) Areia e 3.712 (1,72%) Plástico [50,1% total]	1.024 (88,9%) Água e 128 (11,1%) Plástico [0,2% to- tal]	7.952 (92,6%) Água e 640 (7,45%) Plástico [1,9% to- tal]	-
A	DART	5	212.160 (98,3%) Areia e 3.712 (1,72%) Plástico [50,1% total]	202.672 (98,4%) Água e 3.200 (1,55%) Plástico [47,8% total]	1.024 (100%) Água [0,02% total]	128 (100%) Plástico [<0,1% total]	7.952 (92,6%) Água e 640 (7,45%) Plástico [1,9% to- tal]
B	DART	3	9.936 (73,5%) Areia e 3.584 (26,5%) Plástico [3% total]	211.648 (98,3%) Água e 3.712 (1,72%) Plástico [50% total]	202.224 (99,8%) Areia e 384 (0,19%) Plástico [47% total]	-	-
B	DART	4	202.224 (99,8%) Areia e 384 (0,19%) Plástico [47% total]	211.648 (98,9%) Água e 2.304 (1,08%) Plástico [49,6% total]	1.536 (100%) Plástico [0,4% total]	9.936 (74,2%) Areia e 3.456 (25,8%) Plástico [3% total]	-
B	DART	5	201.520 (99,8%) Areia e 384 (0,19%) Plástico [46,8% total]	211.648 (98,9%) Água e 2.304 (1,08%) Plástico [49,6% total]	10.640 (83%) Areia e 2.176 (17%) Plástico [3% total]	1280 (100%) Plástico [0,3% total]	1.536 (100%) Plástico [0,3% total]

Fonte: Elaboração própria.

Tabela 11 – Agrupamento dos dados DART (simulados) por classificação não supervisionada utilizando os conjuntos de atributos C e D.

Conjunto atributos	Dados	Número de clusters	Classes cluster 0	Classes cluster 1	Classes cluster 2	Classes cluster 3	Classes cluster 4
C	DART	3	210.624 (98,7%) Água e 2.688 (1,26%) Plástico [49,4% total]	212.160 (97,8%) Areia e 4.864 (2,24%) Plástico [50,3% total]	1.024 (88,9%) Água e 128 (11,1%) Plástico [0,03% total]	-	-
C	DART	4	210.624 (98,7%) Água e 2.688 (1,26%) Plástico [49,4% total]	212.160 (97,8%) Areia e 4.864 (2,24%) Plástico [50,3% total]	1.024 (100%) Água [0,3% total]	128 (100%) Plástico [<0,1% total]	-
C	DART	5	212.160 (97,8%) Areia e 4.864 (2,24%) Plástico [50,3% total]	202.160 (98,9%) Água e 2.304 (1,13%) Plástico [47,4% total]	8.464 (95,7%) Água e 384 (4,34%) Plástico [2% do total]	128 (100%) Plástico [<0,1% total]	1.024 (100%) Água [0,02% total]
D	DART	3	210.624 (98%) Água e 4.224 (1,97%) Plástico [49,8% total]	212.160 (98,5%) Areia e 3.328 (1,54%) Plástico [49,9% total]	1.024 (88,9%) Água e 128 (11,1%) Plástico [0,3% total]	-	-
D	DART	4	210.624 (98%) Água e 4.224 (1,97%) Plástico [49,8% total]	212.160 (98,5%) Areia e 3.328 (1,54%) Plástico [49,9% total]	1.024 (100%) Água [0,2% total]	128 (100%) Plástico [<0,1% total]	-
D	DART	5	212.160 (98,5%) Areia e 3.328 (1,54%) Plástico [49,9% total]	209.600 (98,1%) Água e 3.968 (1,86%) Plástico [49,5% total]	128 (100%) Plástico [<0,1% total]	1.024 (100%) Água [0,2% total]	1.024 (80%) Água e 256 (20%) Plástico [0,3% total]

Fonte: Elaboração própria.

Tabela 12 – Agrupamento dos dados USGS (observados) por classificação não supervisionada utilizando os conjuntos de atributos A e B.

Conjunto atributos	Dados	Número de clusters	Classes cluster 0	Classes cluster 1	Classes cluster 2	Classes cluster 3	Classes cluster 4
A	USGS	3	1.918 (92,5%) Água, 100 (4,82%) Plástico e 55 Madeira (2,65%) [65,2% total]	1.069 (97%) Costa, 24 (2,18%) Água, 7 (0,635%) Madeira e 2 (0,181%) Plástico [34,7% total]	1 (50%) Água e 1 (50%) Plástico [<0,1% total]	-	-
A	USGS	4	1.919 (92,5%) Água, 100 (4,82%) Plástico e 55 (2,65%) Ma- deira [65,3% total]	1.069 (97,2%) Costa, 22 (2%) Água, 7 (0,64%) Madeira e 2 (0,18%) Plástico [34,6% total]	1 (50%) Água e 1 (50%) Plástico [<0,1% total]	1 (100%) Água [<0,1% total]	-
A	USGS	5	1.069 (97,3%) Costa, 22 (2%) Água, 6 (0,546%) Madeira e 2 (0,18%) Plástico [34,6% total]	1.919 (92,5%) Água, 100 (4,82%) Plástico e 55 (2,65%) Ma- deira [65,3% total]	1 (50%) Água e 1 (50%) Plástico [<0,1% total]	1 (100%) Madeira [<0,1% total]	1 (100%) Água [<0,1% total]
B	USGS	3	1.941 (92,6%) Água, 103 (4,92%) Plástico e 51 (2,43%) Ma- deira [66% total]	442 (100%) Costa [14% total]	627 (98%) Costa, 11 (1,72%) Madeira e 2 (0,313%) Água [20% total]	-	-
B	USGS	4	1.519 (96,2%) Água, 54 (3,42%) Plástico e 6 (0,38%) Ma- deira [49,7% total]	629 (98,9%) Costa, 6 (0,943%) Madeira e 1 (0,157%) Água [20% total]	440 (100%) Costa [13,9% total]	423 (81%) Água, 50 (9,58%) Madeira e 49 (9,39%) Plástico [16,4% total]	-
B	USGS	5	1.518 (96,2%) Água, 54 (3,42%) Plástico e 6 (0,38%) Ma- deira [49,7% total]	302 (99,7%) Costa e 1 (0,33%) Água [09,5% total]	535 (98,7%) Costa e 7 (1,29%) Madeira [17,1% total]	424 (81,2%) Água, 49 (9,39%) Plástico e 49 (9,39%) Madeira [16,4% total]	232 (100%) Costa [07,3% total]

Fonte: Elaboração própria.

Tabela 13 – Agrupamento dos dados USGS (observados) por classificação não supervisionada utilizando os conjuntos de atributos C e D.

Conjunto atributos	Dados	Número de clusters	Classes cluster 0	Classes cluster 1	Classes cluster 2	Classes cluster 3	Classes cluster 4
C	USGS	3	1.068 (96,3%) Costa, 31 (2,8%) Água, 8 (0,721%) Madeira e 2 (0,18%) Plástico [34,9% total]	1.911 (92,5%) Água, 100 (4,84% Plástico), 54 (2,61%) Madeira e 1 (0,05%) Costa [65% total]	1 (50%) Plástico e 1 (50%) Água [<0,1% total]	-	-
C	USGS	4	1.906 (92,5%) Água, 100 (4,85%) Plástico e 54 (2,62%) Ma- deira [64,8% total]	1.069 (96%) Costa, 35 (3,14%) Água, 8 (0,718%) Madeira e 2 (0,18%) Plástico [35,1% total]	1 (50%) Plástico e 1 (50%) Água [<0,1% total]	1 (100%) Água [<0,1% total]	-
C	USGS	5	1.069 (95%) Costa, 47 (4,2%) Água, 7 (0,6%) Madeira e 2 (0,2%) Plástico [35,4% total]	1.894 (92,5%) Água, 100 (4,88%) Plástico e 54 (2,64%) Ma- deira [64,5% total]	1 (50%) Plástico e 1 (50%) Água [<0,1% total]	1 (100%) Madeira	1 (100%) Água [<0,1% total]
D	USGS	3	1.069 (77,1%) Costa, 258 (18,6%) Água, 44 (3,17%) Madeira e 16 (1,15%) Plástico [43,7% total]	1.684 (94,2%) Água, 86 (4,8%) Plás- tico e 18 (1%) Madeira [56,3% total]	1 (50%) Plástico e 1 (50%) Água [<0,1% total]	-	-
D	USGS	4	1.679 (94,2%) Água, 85 (4,8%) Plás- tico e 18 (1%) Madeira [56,1% total]	1.069 (76,8%) Costa, 262 (18,8%) Água, 44 (3,2%) Ma- deira e 17 (1,2%) Plás- tico [43,8% total]	1 (50%) Plástico e 1 (50%) Água [<0,1% total]	1 (100%) Água [<0,1% total]	-
D	USGS	5	1.677 (94,3%) Água, 85 (4,7%) Plás- tico e 17 (1%) Madeira [56% total]	1.069 (76,7%) Costa, 264 (18,9%) Água, 44 (3,2%) Ma- deira e 17 (1,2%) Plás- tico [43,9% total]	1 (50%) Plástico e 1 (50%) Água [<0,1% total]	1 (100%) Água [<0,1% total]	1 (100%) Madeira [<0,1% total]

Fonte: Elaboração própria.

Tabela 14 – Matriz de confusão média para o conjunto de atributos A.

	Água	Plástico
Água	646,13	1296,87
Plástico	14,9	88,1

Fonte: Elaboração própria.

Tabela 15 – Matriz de confusão média para o conjunto de atributos B.

	Água	Plástico
Água	1577,3	365,7
Plástico	49,34	53,66

Fonte: Elaboração própria.

Tabela 16 – Matriz de confusão média para o conjunto de atributos C.

	Água	Plástico
Água	252,68	1690,32
Plástico	7,26	95,74

Fonte: Elaboração própria.

Tabela 17 – Matriz de confusão média para o conjunto de atributos D.

	Água	Plástico
Água	209,53	1733,47
Plástico	9,48	93,52

Fonte: Elaboração própria.