



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
Programa de Pós-Graduação em Engenharia Ambiental
Campus Londrina e Apucarana



PÂMELA ROBERTA FRANCISQUETTI VENTURIN

**APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA E
IMAGENS MULTIESPECTRAIS PARA A ESTIMATIVA DA
CONCENTRAÇÃO DE SÓLIDOS SUSPENSOS TOTAIS NO LAGO
IGAPÓ, NA CIDADE DE LONDRINA (PR)**

LONDRINA

2022

PÂMELA ROBERTA FRANCISQUETTI VENTURIN

**APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA E
IMAGENS MULTIESPECTRAIS PARA A ESTIMATIVA DA
CONCENTRAÇÃO DE SÓLIDOS SUSPENSOS TOTAIS NO LAGO
IGAPÓ, NA CIDADE DE LONDRINA (PR)**

Applying machine learning techniques and multispectral images to
estimate the total suspended solids concentration in Igapó Lake in
Londrina (PR)

Dissertação apresentada como requisito para
obtenção do título de Mestre no Programa de Pós-
graduação em Engenharia Ambiental da
Universidade Tecnológica Federal do Paraná
(UTFPR).

Orientador: Prof. Dr. Jorge Alberto Martins.

Coorientador: Prof. Dr. Alessandro Botelho Bovo.

LONDRINA

2022



[4.0 Internacional](https://creativecommons.org/licenses/by-nc/4.0/)

Esta licença permite compartilhamento do trabalho, mesmo para fins comerciais, sem a possibilidade de alterá-lo, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.



**Ministério da Educação
Universidade Tecnológica Federal do Paraná
Campus Londrina**



PAMELA ROBERTA FRANCISQUETTI VENTURIN

APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA E IMAGENS MULTIESPECTRAIS PARA A ESTIMATIVA DA CONCENTRAÇÃO DE SÓLIDOS SUSPENSOS TOTAIS NO LAGO IGAPÓ, NA CIDADE DE LONDRINA

Trabalho de pesquisa de mestrado apresentado como requisito para obtenção do título de Mestra Em Engenharia Ambiental da Universidade Tecnológica Federal do Paraná (UTFPR). Área de concentração: Engenharia Ambiental.

Data de aprovação: 15 de Agosto de 2022

Dr. Jorge Alberto Martins, Doutorado - Universidade Tecnológica Federal do Paraná

Dra. Joseane Debora Peruco Theodoro, Doutorado - Universidade Tecnológica Federal do Paraná

Dra. Nariane Marselhe Ribeiro Bernardo, Doutorado - Universidade Estadual Paulista - Unesp

Thiago Gentil Ramires, - Universidade Tecnológica Federal do Paraná

Documento gerado pelo Sistema Acadêmico da UTFPR a partir dos dados da Ata de Defesa em 15/08/2022.

Dedico este trabalho aos meus pais Vilma e Cidinho e as minhas duas estrelinhas
Vó Alzira e Vó Jana que hoje estariam orgulhosas.

AGRADECIMENTOS

A realização desta pesquisa seria impossível se não houvesse a colaboração e o apoio de diversas pessoas, que de diferentes formas contribuíram para a excelência deste estudo.

Agradeço primeiramente a Deus por ter me dado forças para não desistir do meu objetivo, em meio a tantas tribulações. Aos meus pais Vilma e Cidinho e ao meu irmão Juninho, sou eternamente grata por me incentivarem a nunca desistir dos meus objetivos, dando todo o amparo necessário. Agradeço ao meu namorado Renan, por não me deixar sozinha e sempre apoiar minhas decisões, buscando me ajudar naquilo que fosse preciso. Agradeço também aos meus familiares e amigos que estiveram ao meu lado durante estes anos.

Agradeço aos professores Jorge Alberto Martins e Alessandro Botelho Bovo pela oportunidade de realizar esta pesquisa, pela orientação, pela paciência e pela motivação garantindo o pleno desenvolvimento deste trabalho.

Agradeço os professores Edson, Leila e Alessandra por disponibilizarem o laboratório e os materiais necessários para coleta e análise de água.

Agradeço imensamente aos amigos do Laboratório EAE por terem auxiliado em todas as etapas da campanha de campo, bem como nas análises laboratoriais não medindo esforços para fazer aquilo que fosse necessário para a realização desta pesquisa, o meu muito obrigada à: Thaís, Anderson, Priscila, Ronaldo, Adriano, Audrey, Luiza, Juliana e professor Jorge.

Gostaria de agradecer também aos membros da banca, Dra. Joseane, Dr. Thiago e Dra. Nariane, por terem aceito o convite e pelas contribuições.

Por fim, agradeço a Fundação Araucária pela bolsa de fomento à pesquisa.

RESUMO

VENTURIN, Pâmela Roberta Francisquetti. **Aplicação de técnicas de aprendizado de máquina e imagens multiespectrais para a estimativa da concentração de sólidos suspensos totais no Lago Igapó, na cidade de Londrina (PR)**. 2022. 96 f. Dissertação (Mestrado) - Curso de Programa de Pós-Graduação em Engenharia Ambiental, Universidade Tecnológica Federal do Paraná (UTFPR), Londrina, 2022.

A drenagem urbana de águas pluviais representa um dos grandes desafios da área ambiental em todo o mundo. Um dos efeitos indesejados e imediatos do escoamento superficial das águas da chuva é o transporte de sedimentos e de poluentes para dentro dos corpos hídricos. Os Sólidos Suspensos Totais (SST) representam um dos parâmetros críticos e de maior interesse no monitoramento ambiental e na implementação de Soluções Baseadas na Natureza. Dessa forma, este trabalho se propõe a avaliar uma alternativa às medidas *in situ*, através da predição da concentração de SST de forma automatizada, a partir do uso de técnicas de Aprendizado de Máquina (AM), combinadas com imagens multiespectrais de alta resolução espacial capturadas por Veículo Aéreo não Tripulado (VANT) do Lago Igapó, na cidade de Londrina (PR). A campanha de medidas considerou 40 pontos de coleta, com alguns parâmetros analisados *in situ* com o auxílio de uma sonda multiparamétrica (potencial hidrogeniônico (pH), Condutividade elétrica (CE), Temperatura, Porcentagem de oxigênio dissolvido (OD), Concentração de OD, Salinidade; Pressão e Sólidos Dissolvidos Totais (SDT)) e as concentrações de SST medidas em laboratório. Simultaneamente ao processo de coleta das amostras de água, foram capturadas imagens multiespectrais nas bandas *Green, Red, Red Edge* e *Near Infrared* com o auxílio de um VANT. As técnicas de AM utilizadas foram Floresta Aleatória (FA), Regressão Linear (RL), Regressão de Ridge, Regressão de Lasso e Regressão ElasticNet, que foram aplicadas em diferentes condições (número original de amostragens, número reduzido pela remoção de outliers e número ampliado por interpolação). A capacidade de predição dos modelos de AM testados foram avaliados pelos valores do Erro quadrático médio (EQM) e pelo coeficiente de determinação R-quadrado (R^2). Os parâmetros mensurados, foram comparados com os padrões de qualidade estabelecidos na Resolução do Conselho Nacional De Meio Ambiente (CONAMA) 357/2005, para água doce classe 1, assim, foi possível afirmar que os valores encontrados nesta pesquisa, satisfazem as exigências estabelecidas nesta resolução. Os modelos de AM que apresentaram melhores capacidade de predição foram: FA com R^2 de 62,30% e EQM de 5,48; RL com R^2 de 16,53% e EQM de 18,62; Regressão Ridge, de Lasso e de Elasticnet com R^2 de 16,37% e EQM de 18,69. Todos os melhores resultados foram encontrados ao ser utilizado os valores de SST interpolados. Esta metodologia, trata-se de uma ferramenta promissora e com grande chance de substituir as medidas *in situ* no futuro. A ampliação do número de pontos de amostragem, considerando campanhas realizadas sob diferentes cargas de sedimentos, pode melhorar significativamente a capacidade de predição dos modelos testados.

Palavras-chave: VANT; análises *in situ*; capacidade de predição; parâmetros de qualidade.

ABSTRACT

VENTURIN, Pâmela Roberta Francisquetti. **Applying machine learning techniques and multispectral images to estimate the total suspended solids concentration in Igapó Lake in Londrina (PR)**. 2022. 96 f. Dissertation (Master) - Postgraduate Program course in Environmental Engineering, Federal Technological University of Paraná (UTFPR), Londrina, 2022.

Urban rainwater drainage represents one of the most significant challenges in the environmental area around the world. One of the undesired and immediate effects of the surface runoff of rainwater is the transport of sediments and all sorts of pollutants into water bodies. Total Suspended Solids (TSS) represent one of the critical parameters of most significant interest in environmental monitoring and the implementation of Nature-Based Solutions. Thus, this work proposes to evaluate an alternative to in situ measurements through the prediction of TSS concentration in an automated way, using Machine Learning (ML) techniques combined with high spatial resolution multispectral images captured by Unmanned Aerial Vehicle (UAV) from Lake Igapó, in the city of Londrina (PR). The measurement campaign considered 40 collection points, with some parameters analyzed in situ with the aid of a multiparametric probe (Hydrogenionic potential (pH), Electrical Conductivity (EC), Temperature, OD Percentage, OD Concentration, Salinity; Pressure, and Total Dissolved Solids (TDS)) and TSS concentrations measured in the laboratory. Simultaneously with collecting the water samples, multispectral images were captured in the Green, Red, Red Edge, and Near-Infrared bands with the aid of a UAV. The ML techniques used were Random Forest (RF), Linear Regression (LR), Ridge Regression, Lasso Regression, and ElasticNet Regression, which were applied under different conditions (the original number of samples, reduced number by removing outliers, and increased number by interpolation). The ML models' predictive capacity was evaluated using the Mean Square Error (EQM) and the R-square (R^2) determination coefficient. The measured parameters were compared with the quality standards established in the Resolution of the National Environmental Council (CONAMA) 357/2005, for class 1 fresh water, thus, it was possible to affirm that the values found in this research, satisfy the requirements established in this resolution. The AM models that presented the best prediction capacity were: FA with R^2 of 62.30% and EQM of 5.48; RL with R^2 of 16.53% and EQM of 18.62; Ridge, Lasso and Elasticnet regression with R^2 of 16.37% and EQM of 18.69. All the best results were found when using interpolated SST values. This methodology is a promising tool with a great chance of replacing in situ measurements in the future. Expanding the number of sampling points, considering campaigns carried out under different sediment loads, can significantly improve the prediction capacity of the models tested.

Keywords: UAV; in situ analyses; prediction ability; quality parameters.

LISTA DE ILUSTRAÇÕES

Figura 1: Espectro eletromagnético destacando a resposta espectral da vegetação, solo e água.	19
Figura 2: VANT de asa fixa (a) e multi-rotor (b).	22
Figura 3: VANT e câmeras RGB integrada e Multiespectral acoplada.	22
Figura 4: Mapa de Localização da Bacia Hidrográfica do Ribeirão Cambé e Lago Igapó.	26
Figura 5: Sub-bacia Hidrográfica Ribeirão Cambé, lagos Igapó I, II, III e IV e seus afluentes.	26
Figura 6: Barragem Lago Igapó (a) em fase de construção e (b) atualmente.	27
Figura 7: Vista panorâmica do Lago Igapó I em 1960.	28
Figura 8: Vista panorâmica do Lago Igapó I em 2021.	28
Figura 9: Fluxograma do processo metodológico.	35
Figura 10: Pontos amostrais de coleta de água no Lago Igapó I e II.	38
Figura 11: Materiais utilizados na coleta de água e medição in situ.	39
Figura 12: Turbidímetro digital PoliControl AP 2000®.	40
Figura 13: Esquema de determinação de SST e equipamentos utilizados.	41
Figura 14: Membrana de fibra de vidro GF/C 47mm x 1,2µm da marca FilterPro.	42
Figura 15: VANT Dji Phantom 4 Pro® e câmera integrada.	43
Figura 16: Câmera multiespectral Parrot Sequoia® que foi acoplada no VANT.	44
Figura 17: Correlação entre as bandas e o SST com 40 pontos.	57
Figura 18: Análise exploratória dos dados com 40 pontos.	57
Figura 19: Correlação entre as bandas e o SST com 38 pontos.	58
Figura 20: Análise exploratória dos dados com 38 pontos.	59
Figura 21: Mapa resultante da interpolação IDW para os SST.	62
Figura 22: Correlação entre as bandas e o SST com 16.966 pontos.	63
Figura 23: Análise exploratória dos dados com 16.966 pontos.	63
Figura 24: Correlação entre as bandas e o SST com 15.368 pontos (Tukey).	64
Figura 25: Correlação entre as bandas e o SST com 16.273 pontos (Z-Score).	65
Figura 26: Análise exploratória dos dados com 15.368 pontos (Tukey).	65
Figura 27: Análise exploratória dos dados com 16.273 pontos (Z-Score).	66

LISTA DE TABELAS

Tabela 1: Divisão do espectro eletromagnético.....	17
Tabela 2: Precipitação no período de 15 dias que antecederam a data da coleta.....	36
Tabela 3: Localização dos pontos amostrais.....	37
Tabela 4: Valores máximo, mínimo, médio e o desvio padrão dos parâmetros analisados.....	55

LISTA DE QUADROS

Quadro 1: Câmeras multiespectrais utilizadas em VANT e suas características.	23
Quadro 2: Características das bandas da câmera multiespectral Parrot Sequoia ®.	44
Quadro 3: Tabela de dados criada com os valores de pixel e SST (obtidos em laboratório) para cada ponto amostral.	46
Quadro 4: Tabela de dados criada com os valores de pixel e SST (obtidos em laboratório) para os pontos interpolados.	48
Quadro 5: Tabela de dados com as características e a variável.	50
Quadro 6: Resultados dos parâmetros analisados para cada ponto amostral.	54
Quadro 7: Resultados de R^2 e EQM para cada teste utilizando 40 pontos, o valor de pixel das bandas G, R, RE e NIR e o valor do SST.	60
Quadro 8: Resultados de R^2 e EQM para cada teste utilizando 38 pontos, o valor de pixel das bandas G, R, RE e NIR e o valor do SST.	60
Quadro 9: Resultados de R^2 e EQM para cada teste utilizando 38 pontos, o valor do pixel das bandas G e R e o SST.	61
Quadro 10: Resultados de R^2 e EQM para cada teste utilizando 16.966 pontos, o valor de pixel das bandas G, R, RE e NIR e o valor do SST.	68
Quadro 11: Resultados de R^2 e EQM para cada teste utilizando 15.368 pontos, o valor de pixel das bandas G, R, RE e NIR e o valor do SST.	68
Quadro 12: Resultados de R^2 e EQM para cada teste utilizando 16.273 pontos, o valor de pixel das bandas G, R, RE e NIR e o valor do SST.	69
Quadro 13: Resultados de R^2 e EQM para cada teste utilizando 16.966 pontos, o valor do pixel das bandas G e R e o SST.	70
Quadro 14: Resultados de R^2 e EQM para cada teste utilizando 15.368 pontos, o valor do pixel das bandas G e R e o SST.	71
Quadro 15: Resultados de R^2 e EQM para cada teste utilizando 16.273 pontos, o valor do pixel das bandas G e R e o SST.	71
Quadro 16: Melhores resultados encontrados de R^2 e EQM para cada modelo de AM, em ordem decrescente.	72

LISTA DE SIGLAS E ABREVIATURAS

AGUASPARANÁ	Instituto das Águas do Paraná
AM	Aprendizado de Máquina
AR	Análises de Regressão
ARO	Avaliação de Risco Operacional
B	<i>Blue</i>
Bits	<i>Binary digits</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CBERS	<i>China-Brazil Earth Resources Satellite</i>
Cfa	Clima Temperado Húmido com verão quente
CE	Condutividade Elétrica
cm	Centímetro
CMOS	<i>Complementary Metal Oxide Semiconductor</i>
DECEA	Departamento de Controle do Espaço Aéreo
DQO	Demanda Química de Oxigênio
EQM	Erro quadrático médio
FA	Floresta Aleatória
G	<i>Green</i>
g	Gramas
GLONASS	<i>Global Navigational Satellite System</i>
GNSS	<i>Global Navigation Satellite System</i>
GPS	<i>Global Positioning System</i>
IA	Inteligência Artificial
IAP	Instituto Ambiental do Paraná
IAPAR	Instituto Agrônômico do Paraná
IAT	Instituto de Água e Terra
IBGE	Instituto Brasileiro de Geografia e Estatística
IDW	<i>Inverse Distance Weighting</i>
INPE	Instituto Nacional de Pesquisas Espaciais
IP67	A prova de poeira totalmente vedado
IP68	Protegido contra imersão contínua em água
ITCG	Instituto de Terras, Cartografia e Geologia do Paraná
Kg	Quilograma
km	Quilômetro
km/h	Quilômetro por hora
km ²	Quilômetro quadrado
LANDSAT	<i>Land Remote Sensing Satellite</i>
LASSO	<i>Least Absolute Shrinkage and Selection Operator</i>
M	Metro
mAh	Miliampere-hora
mg/L	Miligramas por Litro
mg/L de OD	Miligramas por Litro de OD
mL	Mililitros

mm	Milímetro
Mpx	Megapixel
MSI	<i>MultiSpectral Imager</i>
NBS	Soluções Baseadas na Natureza
NIR	<i>Near Infrared</i>
nm	Nanômetro
NTU	Unidade Nefelométrica de Turbidez
°C	Graus Celsius
OD	Oxigênio Dissolvido
OLI	<i>Operational Land Imager</i>
P1	Tara do cadinho
P2	Cadinho com amostra após filtração e secagem
pH	Potencial hidrogeniônico
PNAD	Pesquisa Nacional por Amostra de Domicílios
%OD	Porcentagem de oxigênio dissolvido
PR	Paraná
PSI	Libras por Polegadas Quadradas (PSI),
PSU	Unidade de Salinidade Prática
R	<i>Red</i>
R ²	R-squared
RE	<i>Red Edge</i>
RL	Regressão Linear
RLM	Regressão Linear Múltipla
RGB	<i>Red, Green and Blue</i>
RNA	Rede Neural Artificial
RNAs	Redes Neurais Artificiais
SABESP	Companhia de Saneamento Básico do Estado de São Paulo
SBN	Soluções Baseadas na Natureza
SDT	Sólidos Dissolvidos Totais
SS	Sólidos Suspensos
SST	Sólidos Suspensos Totais
Tmax	Temperatura máxima
Tmed	Temperatura média
Tmin	Temperatura mínima
TSS	<i>Total Suspended Solid</i>
UAV	<i>Unmanned Aerial Vehicle</i>
Ur	Umidade relativa
UTFPR	Universidade Tecnológica Federal do Paraná
VANT	Veículo Aéreo Não Tripulado
VANTs	Veículos Aéreos Não Tripulados
Vol.amostra	Volume da amostra
µS/cm	Microsiemens por Centímetro
µm	Micrômetro

SUMÁRIO

1 INTRODUÇÃO	12
2 OBJETIVOS	16
2.1 Objetivo geral	16
2.2 Objetivos específicos	16
3 REFERENCIAL TEÓRICO	17
3.1 Sensoriamento remoto	17
3.2 Veículo aéreo não tripulado	21
3.2.1 Tipos de VANT e sensores.....	21
3.2.2 Uso na gestão ambiental de recursos hídricos.....	24
3.3 O Lago Igapó	25
3.4 Aprendizado de máquina	29
3.4.1 Floresta Aleatória	31
3.4.2 Regressão Linear	31
3.4.3 Regressão de <i>Ridge</i> , Regressão de Lasso e Regressão <i>ElasticNet</i>	32
4 METODOLOGIA	34
4.1 Área de estudo e campanha de campo	36
4.2 Amostragem de água	38
4.3 Aquisição das imagens aéreas com VANT	42
4.4 Análise de dados	44
4.4.1 Processamento das imagens aéreas	45

4.4.2 Interpolação IDW.....	47
4.4.3 Manipulação de dados	49
5 RESULTADOS E DISCUSSÃO	53
5.1 Resultado da amostragem de água	53
5.2 Resultado da aplicação de técnicas de AM	56
5.2.1 Utilizando os pontos medidos em campanha de campo	56
5.2.2 Utilizando os pontos interpolados e as bandas G, R, RE e NIR.....	62
5.2.3 Utilizando os pontos interpolados e as bandas G e R.....	70
5.3 Discussão dos resultados	72
6 CONCLUSÃO	75
REFERÊNCIAS	78
ANEXO 1 – Ofício autorização de voo	84
ANEXO 2 – Avaliação de Risco Operacional para Operação com RPA em Londrina-PR	86

1 INTRODUÇÃO

Esta pesquisa faz parte de um consórcio internacional intitulado “*Research-based Assessment of Integrated approaches to Nature-based SOLUTIONS (RainSolutions)*” e conta com a participação de Suécia, Noruega, Estônia, Irlanda, Holanda, Espanha, África do Sul e Brasil, cada país contribuindo com um caso de estudo. O consórcio tem por objetivo desenvolver metodologias para a aplicação de Soluções Baseadas na Natureza (SBN) para a restauração e reabilitação de sistemas de recursos hídricos em áreas urbanas. Cada país participante do consórcio contribui com um caso de estudo, e o sistema de lagos Igapó, na cidade de Londrina, é um dos casos de estudo. O sistema de lagos Igapó é tema de seis dissertações de mestrado do Programa de Pós-Graduação em Engenharia Ambiental da Universidade Tecnológica Federal do Paraná - campus Apucarana/Londrina, cada uma contribuindo em um tema específico. Juntas, essas dissertações devem contribuir na proposta de SBN para mitigar os impactos causados pela expansão urbana acelerada no Lago Igapó e seu entorno, em particular aqueles relacionados à drenagem urbana e o assoreamento/poluição do mesmo.

Os problemas relacionados à drenagem urbana e assoreamento do Lago Igapó estão intrinsecamente relacionados às grandes transformações ocorridas no Brasil ao longo de sua história. Em meados das décadas de 1970 e 1980, o Brasil passou por intenso processo de êxodo rural, onde trabalhadores oriundos do campo saíram para as cidades em busca de novas oportunidades de trabalho, principalmente devido ao início do processo de mecanização agrícola (WAGNER; WARD, 1980). De acordo com os dados da Pesquisa Nacional por Amostra de Domicílios (PNAD) realizada em 2015, constatou-se que 84,72% da população brasileira vive em áreas urbanas (IBGE, 2020b).

A expansão e o adensamento urbano acelerados ocasionaram modificações rápidas e intensas no meio físico, impactando, sobretudo os processos ambientais. Por exemplo, a diminuição das áreas com coberturas naturais impactou diretamente na taxa de infiltração, consequentemente impedindo que o solo desempenhe suas funções hidrológicas eficientemente (TUCCI; BERTONI, 2003; ALMEIDA; NETO, 2015; NUNES, 2015).

Devido à rapidez como aconteceu o processo de expansão e adensamento urbano, não houve tempo suficiente para a condução de estudos rigorosos que

permitissem prever possíveis impactos no meio físico. O problema se torna mais acentuado no momento atual, em que a ocorrência de eventos naturais extremos tem se tornado mais frequente.

Paralelamente, a impermeabilização do solo impede a infiltração de águas pluviais, tornando-se em grande parte responsável pelo aumento do escoamento superficial. Essa lâmina d'água, que escoar superficialmente, acaba carreando materiais orgânicos e inorgânicos para as áreas mais baixas do relevo, que alcançam, geralmente em um corpo hídrico. Sendo assim, a carga de poluentes e sedimentos nos corpos hídricos aumenta significativamente diante desse cenário impactado pelas atividades humanas, causando alterações nos parâmetros de qualidade ambiental, bem como ampliação das áreas de inundação e ocorrência de enchentes (SOUTO *et al.*, 2019; TUCCI; BERTONI, 2003). Deste modo, os corpos hídricos receptores, necessitam de ações de monitoramento constantes, a fim de preservá-los e evitar desastres ambientais em casos de eventos extremos.

O monitoramento da qualidade da água dos corpos hídricos é de suma importância para a sua preservação. O monitoramento convencional é realizado através de medições *in situ* e análises laboratoriais. Segundo Yigit Avdan *et al.* (2019), medir e monitorar a qualidade da água de corpos hídricos *in situ* pode tornar-se um processo oneroso e com elevados custos financeiros, que estão relacionados com as dificuldades de acessar determinados locais, com a necessidade de aquisições de equipamentos analíticos para medições *in situ* e laboratoriais. Adiciona-se a isso a necessidade de pessoal qualificado e todo impasse inerente à sua contratação e permanência no serviço de monitoramento. Por esses motivos, métodos alternativos devem ser explorados.

Uma alternativa aos métodos tradicionais é a utilização de técnicas de sensoriamento remoto, seja através de imagens de satélite ou produzidas através de sensores acoplados em Veículos Aéreos Não Tripulados (VANTs), mais popularmente conhecidos como Drones.

O sensoriamento remoto torna-se uma ferramenta alternativa ao oferecer dados de forma mais segura, acessível, rápida e mais econômica para o monitoramento ambiental de grandes áreas. Considere-se ainda que boa parte dos dados de satélites estão disponibilizados gratuitamente (VILELA, 2010). A aplicação dessa ferramenta é possível devido à presença de componentes opticamente ativos na água, que através de um sistema sensor resulta em diferentes padrões de

absorção e retroespalhamento da energia incidente, capaz de distinguir cada componente (GUIMARÃES *et al.*, 2019). Essa resposta espectral diferenciada pode ser interpretada como um indicador das diferenças nas propriedades físico-químicas da coluna d'água.

Técnicas de sensoriamento remoto para estudos de monitoramento de qualidade da água têm sido frequentemente utilizadas. Considera-se como monitoramento, toda observação repetitiva e com frequência definida, de um local ou um fenômeno na qual deseja-se obter informações sobre a dinâmica de seu comportamento (NOVO, 2007). No entanto, para o monitoramento de pequenos lagos, represas ou rios, a resolução espacial das imagens obtidas por satélites, com média ou baixa resolução espacial (10 a 30 m), pode não ser adequada para se obter informações detalhadas e precisas desses corpos hídricos em tais imagens (YIGIT AVDAN *et al.*, 2019; KUPSSINSKÜ *et al.*, 2020).

Tendo em vista a impraticabilidade em utilizar imagens de satélite em pequenos corpos hídricos, faz-se necessário aplicar um método que se assemelha às imagens de satélite, mas que garanta uma alta resolução espacial. Neste caso, os sensores remotos considerados de alta resolução espacial são os que conseguem capturar objetos à aproximadamente 50 cm (INPE, 2021). Os VANTs representam uma tecnologia com potencial de monitorar pequenos rios, lagos e reservatórios, sendo capaz de coletar dados de sensoriamento remoto com alta resolução, rapidez e custos relativamente menores, como desejável para o monitoramento. Estudos com VANTs para monitoramento da qualidade da água de pequenos corpos hídricos estão começando a ser difundidos na gestão ambiental de recursos hídricos. De acordo com McDonald (2019), dados multiespectrais e hiperespectrais de imagens demonstram ser bons indicadores de turbidez, sólidos em suspensão e clorofila de corpos hídricos.

Em estudo realizado por Ehmann *et al.* (2018), foi avaliada a aplicação de Veículo Aéreo não tripulado (VANT) equipado com câmeras multiespectrais para monitoramento de turbidez e vazão em um riacho no centro de Nova York. Já Woodget *et al.* (2017) utilizaram VANT para realizar uma classificação do tamanho do substrato fluvial em um trecho do rio San Pedro (Chile). Su e Chou (2015) utilizaram sensores embarcados em VANT que capturam imagens em *red* (R), *green* (G), *blue* (B) e *near infrared* (NIR) para mapear o estado trófico de um pequeno reservatório em Taiwan. Guimarães *et al.* (2019) utilizaram análises de regressão (AR) e redes neurais artificiais (RNAs) para prever sólidos suspensos (SS) em um lago no Rio Grande do

Sul, a partir de imagens de VANT. Kupssinskü *et al.* (2020) utilizaram imagens do satélite e imagens multiespectrais capturadas por VANT e técnicas de AM a fim de propor um método para predição de SST e Clorofila-a em um lago artificial.

O objeto de estudo desta pesquisa são os lagos Igapó I e II, que fazem parte do sistema de Lago Igapó composto por 4 lagos (Lago Igapó I, Lago Igapó II, Lago Igapó III e Lago Igapó IV) localizado no município de Londrina – PR. O sistema de Lago Igapó foi criado ao longo do leito do Ribeirão Cambé, com a finalidade de solucionar problemas de drenagem e criar uma área de lazer para a população. Pista de caminhada, parques e academia ao ar livre, quadras de vôlei, campo de futebol e infraestrutura para a prática de esportes náuticos estão entre as atrações do Igapó, tornando-se o principal espaço público do Município, atraindo londrinenses e turistas (SILVA, 2017). No entanto, com o passar dos anos, houve aumento da urbanização ao redor do lago, que tem impactado significativamente no ciclo hidrológico da bacia, aumentando o escoamento superficial e causando eventos de enchentes, alagamentos, inundações bem como assoreamento em alguns pontos do Lago Igapó (VACARIO; MACHADO, 2017; VACARIO; MACHADO, 2019).

Considerando os problemas de drenagem que afetam a bacia do Igapó, e tendo em vista os diversos usos de VANT para o monitoramento de corpos hídricos, este projeto avaliou o desempenho de um método alternativo às medições *in situ* e ao uso de imagens de satélite, com base em imagens capturadas por VANT, com o intuito de prever quais trechos dos lagos Igapó I e II recebem maior carga de sedimentos. Mais especificamente, propõe-se desenvolver um método para a estimativa da concentração de SST, utilizando técnicas de Aprendizado de Máquina (AM) (*Machine Learning*), aplicadas nas imagens multiespectrais de alta resolução espacial capturadas por VANT. Pretende-se assim, identificar os locais com maior carga de SST ao longo dos lagos Igapó I e II, e auxiliar os gestores na tomada de decisões mais eficazes na busca por SBN para os problemas relacionados à drenagem urbana no sistema de lagos Igapó.

2 OBJETIVOS

2.1 Objetivo geral

Estimar a concentração de SST a partir de técnicas de AM e imagens multiespectrais de alta resolução espacial capturadas por VANT, a fim de identificar os locais com maior carga de SST ao longo dos lagos Igapó I e II.

2.2 Objetivos específicos

Para atingir o objetivo geral deste projeto, os seguintes objetivos específicos foram atendidos:

- a) Mapear as áreas de interesse a partir de imagens multiespectrais de alta resolução espacial;
- b) Caracterizar limnologicamente, através de análises laboratoriais, as amostras de água dos lagos Igapó I e II;
- c) Obter dados de referência para validação de modelos de AM;
- d) Validar os modelos de AM para estabelecimento de ferramentas de monitoramento;
- e) Diagnosticar os gradientes de concentração de SST nos lagos Igapó I e II.

3 REFERENCIAL TEÓRICO

3.1 Sensoriamento remoto

O sensoriamento remoto é uma ferramenta difundida em escala global pela sua capacidade de levantamento e monitoramento de recursos terrestres de forma rápida e prática. Trata-se de uma tecnologia capaz de obter imagens da superfície terrestre sem que haja o contato direto entre o sensor e o objeto. Isto é possível devido às propriedades da radiação eletromagnética, que consegue percorrer o espaço entre o sensor e o objeto, sendo o sol a fonte mais potente de radiação eletromagnética (MENESES; ALMEIDA, 2012). Qualquer corpo com temperatura maior que 0 Kelvin (zero grau absoluto) emite energia eletromagnética, cuja frequência ou comprimento de onda depende dessa temperatura.

Portanto, a radiação eletromagnética possui diferentes comprimentos de onda, que podem ser absorvida, refletida ou transmitida pelos objetos. Em geral a radiação eletromagnética é subdividida em distintos intervalos de comprimento de onda. De acordo com seu uso e aplicações, esta subdivisão é denominada de espectro eletromagnético (MENESES; ALMEIDA, 2012) (Tabela 1).

Tabela 1: Divisão do espectro eletromagnético.

INTERVALO ESPECTRAL	COMPRIMENTO DE ONDA
Raios cósmicos	0,01 x 10 ⁻¹⁰ m
Raios gama	0,01 – 0,1 x 10 ⁻¹⁰ m
Raios X	0,1 – 10 x 10 ⁻¹⁰ m
Ultravioleta	100 nm – 0,38 µm
Visível*	0,38 – 0,76 µm
Infravermelho próximo*	0,76 – 1,2 µm
Infravermelho de ondas curtas*	1,2 – 3,0 µm
Infravermelho médio*	3,0 – 5,0 µm
Infravermelho termal*	5,0 – 1 mm
Micro-ondas*	1 mm – 100 cm
Rádio	1 m – 10 km

Legenda: (*) Os intervalos mais utilizados no sensoriamento remoto.

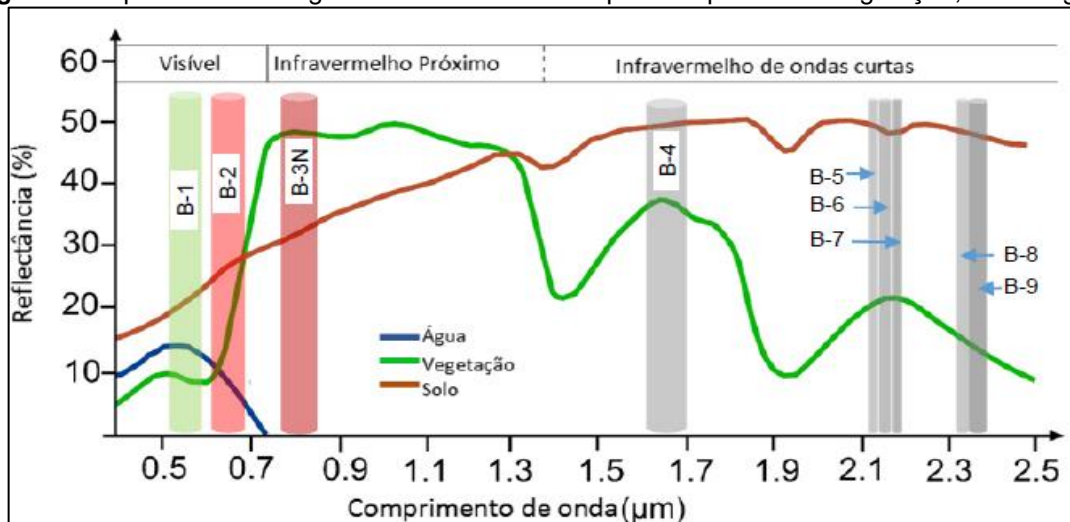
Fonte: Adaptado de Meneses e Almeida (2012).

O espectro eletromagnético possui comprimentos de onda que variam desde os raios cósmicos, que são muito curtos, até ondas de rádio, com comprimentos de onda compreendidos numa ampla faixa e podendo chegar a centenas de km.

Nem todas as bandas espectrais são passíveis de uso pelos sensores instalados em VANTs, aeronaves ou satélites. Em geral, o intervalo utilizado por sensores dessa natureza está entre a banda do visível e micro-ondas (MENESES; ALMEIDA, 2012). O espectro visível é muito utilizado no sensoriamento remoto e corresponde ao conjunto de energias que o olho humano é capaz de perceber gerando uma sensação de cor, que é produzida pela luz e está associada aos intervalos espectrais do violeta, azul, verde, amarelo, laranja e vermelho (MORAES, 2002). Outra banda muito utilizada em sensoriamento remoto é o Infravermelho, que é subdividido entre Infravermelho próximo (0,7 a 1,3 μm), médio (1,3 a 6 μm) e distante (6 a 1000 μm) (MORAES, 2002).

Um objeto é capaz de absorver, refletir ou transmitir a radiação eletromagnética incidente. A capacidade do objeto de absorver energia é denominada de absorvância, enquanto a capacidade de refletir é indicada pela refletância e, do mesmo modo, a capacidade de transmitir energia radiante é denominada de transmitância. O somatório de todos esses parâmetros é expresso por valores variando entre 0 e 1 (FERREIRA; PEREIRA FILHO, 2009). Deste modo, é importante conhecer o comportamento espectral dos objetos, para que seja possível escolher a região do espectro ideal para determinada aplicação (MORAES, 2002). Por exemplo, a água pura em estado líquido, apresenta comportamento espectral de baixa refletância (da ordem de 10%) para a faixa de 0,38 e 0,7 μm e absorção máxima para comprimentos de onda acima de 0,7 μm (faixa do visível e infravermelho próximo) (Figura 1). No entanto, o comportamento espectral dos corpos hídricos é alterado de acordo com os materiais dissolvidos e em suspensão presentes e que interferem no processo de absorção e espalhamento (MORAES, 2002). Os principais responsáveis pelo espalhamento de energia radiante na água são as partículas suspensas e os detritos orgânicos e inorgânicos. Para a absorção de radiação os principais responsáveis são os pigmentos fotossintéticos, detritos orgânicos e matéria orgânica (FERREIRA; PEREIRA FILHO, 2009).

Figura 1: Espectro eletromagnético destacando a resposta espectral da vegetação, solo e água.



Fonte: Novais (2017).

Para um melhor entendimento da interação da radiação com os corpos hídricos e o uso dessa informação em sensoriamento remoto, deve-se ter em vista alguns aspectos relativos à resolução das imagens obtidas pelos sensores. Sabe-se que a resolução de uma imagem em sensoriamento remoto é subdividida em quatro: Espacial, Espectral, Radiométrica e Temporal (MENESES; ALMEIDA, 2012).

A resolução espacial determina o tamanho do menor objeto que pode ser identificado em uma imagem. Sensores capazes de cobrir milhares de quilômetros quadrados, como os da série LANDSAT e CBERS, são considerados de resoluções espaciais baixas (20 a 30 metros) e são utilizados em estudos que não necessitam de detalhamento dos alvos. Já os sensores que recobrem áreas pequenas, tais como aqueles embarcados nos satélites *Ikonos* e *Orbview*, possuem altas resoluções espaciais (menor que 1 metro), pois são capazes de mostrar detalhes da área de estudo (MENESES; ALMEIDA, 2012).

O termo resolução espectral envolve o número de bandas, a largura em comprimento de onda das bandas, e as posições das bandas no espectro eletromagnético. Um sensor é considerado com melhor resolução espectral à medida que aumenta o seu número de bandas em diferentes regiões espectrais e com larguras de comprimentos de onda estreitos (MENESES; ALMEIDA, 2012).

A resolução radiométrica é a medida de intensidade de radiância da área de cada pixel medida pelo detector. Quanto maior a capacidade do detector de medir as diferenças de intensidade de radiância, maior será a resolução radiométrica, ou seja, maior será a qualidade visual da imagem. A resolução radiométrica é determinada em

bits, quanto maior o número de bits, melhor a resolução da imagem (MENESES; ALMEIDA, 2012).

A resolução temporal se refere à frequência com que o sensor revisita uma determinada área (MENESES; ALMEIDA, 2012), que depende de parâmetros orbitais definidos pelo projeto do satélite no qual o sensor está embarcado. Por exemplo, um satélite do programa LANDSAT só consegue produzir imagem de um mesmo local a cada 16 dias. A resolução temporal é especialmente importante para regiões com intensa cobertura de nuvens. Uma baixa resolução temporal pode reduzir consideravelmente as chances de se conseguir uma imagem de qualidade dessas regiões. Para suprir essa limitação, alguns programas trabalham com projetos envolvendo constelações de satélites, o que aumenta a frequência de visitas sobre uma dada região. O *RapidEye* e *Pleiades* são exemplos desses programas.

Diversos estudos ao redor do mundo estão sendo desenvolvidos utilizando imagens de satélite para monitoramento de corpos hídricos. Liu e Wang (2019) desenvolveram um modelo para avaliação da turbidez em reservatórios em Taiwan, utilizando imagens do satélite LANDSAT 8 OLI. Mushtaq e Lala (2017) fizeram uso de imagens de satélite do LANDSAT 8 OLI para estimativa remota de parâmetros da qualidade da água (SST, SDT), OD, demanda química de oxigênio (DQO), turbidez, etc.) em um corpo hídrico na Índia. Marquez *et al.* (2018) utilizaram imagens de satélite do LANDSAT 8 OLI para avaliar a qualidade e a profundidade de um grande reservatório na Colômbia (11.647 hectares). Dezordi *et al.* (2019) analisaram a turbidez superficial do reservatório da Usina Hidrelétrica de Itaipú no Paraná, a partir de imagens do satélite LANDSAT 8 OLI. Pizani *et al.* (2020) estimaram a qualidade da água do reservatório da Usina Hidrelétrica de Três Marias em Minas Gerais, utilizando imagens dos satélites LANDSAT 8 OLI e Sentinel-2 MSI.

Portanto, isso demonstra a frequência de utilização do sensoriamento remoto em estudos que visam monitorar parâmetros de qualidade da água. No entanto, sensoriamento remoto não se aplica apenas a imagens de satélite. As fotografias aéreas também estão inclusas nesta definição (MENESES; ALMEIDA, 2012). Conforme citado anteriormente, segundo Yigit Avdan *et al.* (2019), para o monitoramento de pequenos corpos hídricos (lagos, represas, córregos, dentre outros), as resoluções espaciais média ou baixa (10 a 30 m) podem não ser adequadas. Por este motivo, imagens aéreas obtidas por câmeras acopladas em VANT

apresentam uma tecnologia com potencial de monitorar pequenos corpos hídricos, sendo capaz de coletar dados de sensoriamento remoto com alta resolução, rapidez e custos relativamente menores.

3.2 Veículo aéreo não tripulado

Nos últimos anos houve um crescente emprego de VANT em diferentes áreas do conhecimento, tais como, monitoramento de tráfego em rodovias, segurança pública, mapeamento topográfico de imóveis, mapeamento de uso do solo, agricultura de precisão, monitoramento ambiental e monitoramento da qualidade de corpos hídricos. Aeronave não tripulada, veículo aéreo não tripulado e drones possuem o mesmo significado no Brasil (DECEA, 2020).

Em pesquisa realizada pelo autor no Portal de Periódicos CAPES em setembro de 2020, utilizando o termo “VANT *or* UAV” como filtro, no período de 2000 a 2020, foram encontrados 105.730 artigos publicados. No período de 1980 a 2000, foram encontradas 17.071 publicações, com base no mesmo termo de busca, mostrando o crescimento no número de publicações e, conseqüentemente, indicando a crescente demanda por VANT nos últimos 20 anos.

Segundo McDonald (2019), a utilização de VANT pode resolver muitos desafios da gestão de águas pluviais urbanas, pois pode coletar dados de sensoriamento remoto rapidamente, com resoluções espaciais e temporais incomparáveis às imagens de satélites. Deste modo, o uso de VANT pode desempenhar papel importante no gerenciamento e monitoramento de corpos hídricos.

Com o aumento do uso desta tecnologia, houve um grande aumento de fabricantes de VANT, câmeras, sensores e *softwares* de voos e processamento de dados (MCDONALD, 2019).

3.2.1 Tipos de VANT e sensores

Os tipos de VANT podem ser amplamente subdivididos em duas categorias: asa fixa e multi-rotor (Figura2). O VANT de asa fixa recobre maior área e suporta maior carga útil, porém não pode pairar sobre o local e necessita de uma grande área livre

para levantar voo e pousar, o que pode ser uma dificuldade na área urbana (MCDONALD, 2019). Já o modelo multi-rotor supera essas dificuldades por possuir sistemas de hélice que lhe permite pairar no ar, decolar e pousar de forma perpendicular ao solo. (MCDONALD, 2019).

Figura 2: VANT de asa fixa (a) e multi-rotor (b).



Fonte: (a) Su e Chou (2015); (b) Autoria própria (2021).

Em estudo realizado por Su e Chou (2015) foi utilizado um VANT de asa fixa, com o objetivo de mapear o estado trófico de um pequeno reservatório em Taiwan. Já Veronez *et al.* (2018) utilizaram um sensor de baixo custo acoplado em VANT multi-rotor para monitorar a correlação entre SST e matéria orgânica dissolvida em um pequeno lago no Rio Grande do Sul. Portanto, para um sobrevoo eficiente que seja capaz de atingir os objetivos desejados, é necessário um estudo prévio do local e planejamento de voo, para definir o melhor modelo a ser utilizado.

Faz-se necessário também determinar a melhor câmera ou sensor que serão utilizados, pois existem diferentes modelos sensíveis a diferentes faixas do espectro e capazes de reconhecer diferentes materiais. Existem as câmeras RGB (*Red Green Blue*) (câmera fotográfica), NIR, multiespectrais e hiperespectrais (FRANCO JUNIOR, 2019) (Figura 3). As câmeras multiespectrais, hiperespectrais e radiométricas estão sendo utilizadas em VANT em várias aplicações de monitoramento ambiental (MCDONALD, 2019).

Figura 3: VANT e câmeras RGB integrada e Multiespectral acoplada.



Legenda: (a) Camera Multiespectral *Parrot Sequoia*[®]; (b) Câmera RGB do VANT *Dji Phantom 4 Pro*[®].
Fonte: Autoria Própria (2021).

No mercado estão disponíveis modelos de câmeras multiespectrais que capturam imagens na faixa do visível e algumas bandas além da faixa de luz visível, normalmente no espectro infravermelho. As diferenças entre elas estão geralmente no intervalo espectral (395 - 1000 nm). No Quadro 1 são apresentados alguns modelos, assim como seus números de bandas e faixas espectrais (MCDONALD, 2019).

Quadro 1: Câmeras multiespectrais utilizadas em VANT e suas características.

FABRICANTE	MODELO DE SENSOR	NÚMERO DE BANDAS	FAIXA ESPECTRAL (nm)
Bay Spec	OCI-M+	12	450 - 1000
Mapir	Survey 3	6	395 - 945
Mapir	Kernel	21	395 - 945
Micasense	RedEdge-MX	5	475 - 840
Micasense	Altum	5	475 - 840
Parrot	Parrot Sequoia +	4	550 - 790
Sentera	Multispectral Double 4K	5	386 - 860

Fonte: Adaptado de McDonald (2019).

As câmeras hiperespectrais são capazes de capturar centenas de bandas, cobrindo faixas espectrais muito mais amplas (350 – 13.400 nm) (MCDONALD, 2019). No entanto, câmeras com sensores hiperespectrais possuem alto valor comercial o que acaba sendo um fator limitante à sua aplicação. Por este motivo, neste estudo, será utilizada uma câmera multiespectral.

3.2.2 Uso na gestão ambiental de recursos hídricos

Segundo McDonald (2019), nos últimos 50 anos os VANTs tornaram-se uma tecnologia promissora para a gestão ambiental de recursos hídricos, pois a capacidade de coleta de imagens e vídeos permite ter acesso em tempo real a terrenos de difícil acesso, tornando esta tecnologia superior às tradicionais de coletas de dados. Os VANTs podem ser utilizados como um mecanismo de sensoriamento remoto no monitoramento de qualidade da água de corpos hídricos. Os dados coletados com câmeras multiespectrais e hiperespectrais acopladas em VANT mostraram-se bons indicadores de parâmetros de qualidade da água, como turbidez, SS e clorofila (MCDONALD, 2019).

Entre os estudos recentes envolvendo a utilização de VANT no monitoramento ambiental de corpos hídricos, podem ser destacados: Ehmann *et al.* (2018) avaliaram a aplicação de VANT equipado com câmeras multiespectrais para monitoramento de turbidez e vazão em um riacho no centro de Nova York; Woodget *et al.* (2017) utilizaram VANT para realizar uma classificação do tamanho do substrato fluvial no rio San Pedro (Chile); Su e Chou (2015) utilizaram sensores que capturam imagens em RGB e NIR em VANT para mapear o estado trófico de um pequeno reservatório em Taiwan; Veronez *et al.* (2018) utilizaram um sensor de baixo custo acoplado em VANT e Rede Neural Artificial (RNA) para monitorar a qualidade da água em um pequeno lago no Rio Grande do Sul; Guimarães *et al.* (2019) utilizaram AR e RNA para prever SS em um lago no Rio Grande do Sul, a partir de imagens de VANT; e Pölönen *et al.* (2014) utilizaram sensor hiperespectral para monitorar a qualidade da água de um pequeno lago no sul da Finlândia.

3.3 O Lago Igapó

O sistema de Lago Igapó está localizado no município de Londrina - PR (latitude 23° 20' 7,83" S e longitude 51° 9' 30,78" O) e é composto por 4 lagos (Lago Igapó I, Lago Igapó II, Lago Igapó III e Lago Igapó IV), no entanto, os objetos de estudo desta pesquisa são os lagos Igapó I e II.

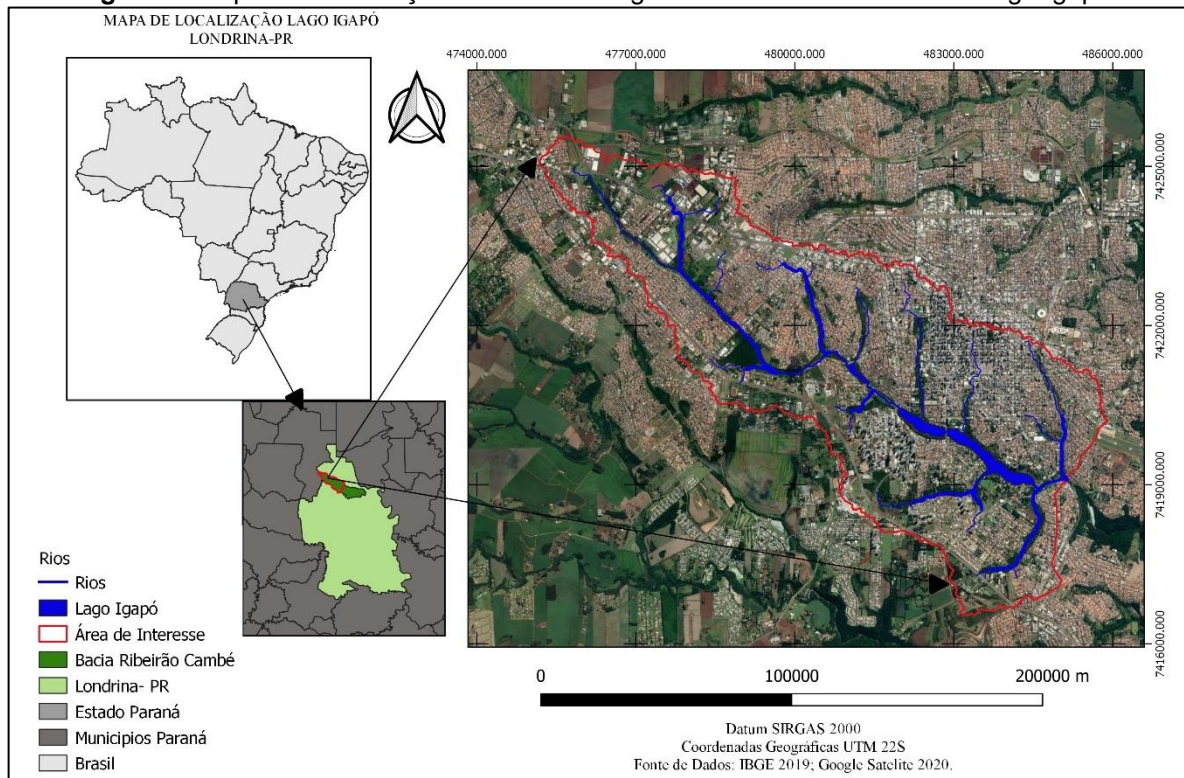
De acordo com o último censo, Londrina possui 506.701 habitantes, tendo como população estimada para 2020, um total de 575.377 habitantes (IBGE, 2020a). O clima predominante na região, segundo a classificação de Köppen-Geiger, é subtropical úmido (Cfa), caracterizado por temperatura média anual variando entre 18 e 22 °C, chuvas concentradas no verão e geadas pouco frequentes (CAVIGLIONE *et al.*, 2000). Os solos predominantes no município de Londrina são dos tipos Latossolo, Nitossolo e Neossolo (ITCG, 2006).

Segundo o Instituto Ambiental do Paraná (IAP) (2011), que a partir de 18 de dezembro de 2019 passou a ser denominado de Instituto de Água e Terra (IAT), e incorporou o Instituto das Águas do Paraná (AGUASPARANÁ) e o Instituto de Terras, Geografia e Geologia do Paraná (ITCG), a área urbana de Londrina é drenada por diversas microbacias hidrográficas, sendo elas: Jacutinga, Lindóia, Limoeiro, Cambé, Cafezal e Três Bocas. A área ocupada por estas bacias na área urbana totaliza 245,52 km².

Os lagos Igapó I e II, objetos deste estudo, estão localizados na Bacia Hidrográfica do Ribeirão Cambé. Esta bacia nasce a oeste de Londrina, no município de Cambé, e possui uma área aproximada de 77,20 km² tendo seu rio principal 25 km de extensão (ARAUJO, 2004) (Figuras 4 e 5). A bacia do Ribeirão Cambé está inserida majoritariamente em ambiente urbano, portanto, seus riachos comumente sofrem impactos negativos com lançamento de efluentes, redução da biodiversidade, desmatamento de mata ciliar, erosão, carreamento de solos e materiais particulados e assoreamento (KOSZMA, 2018).

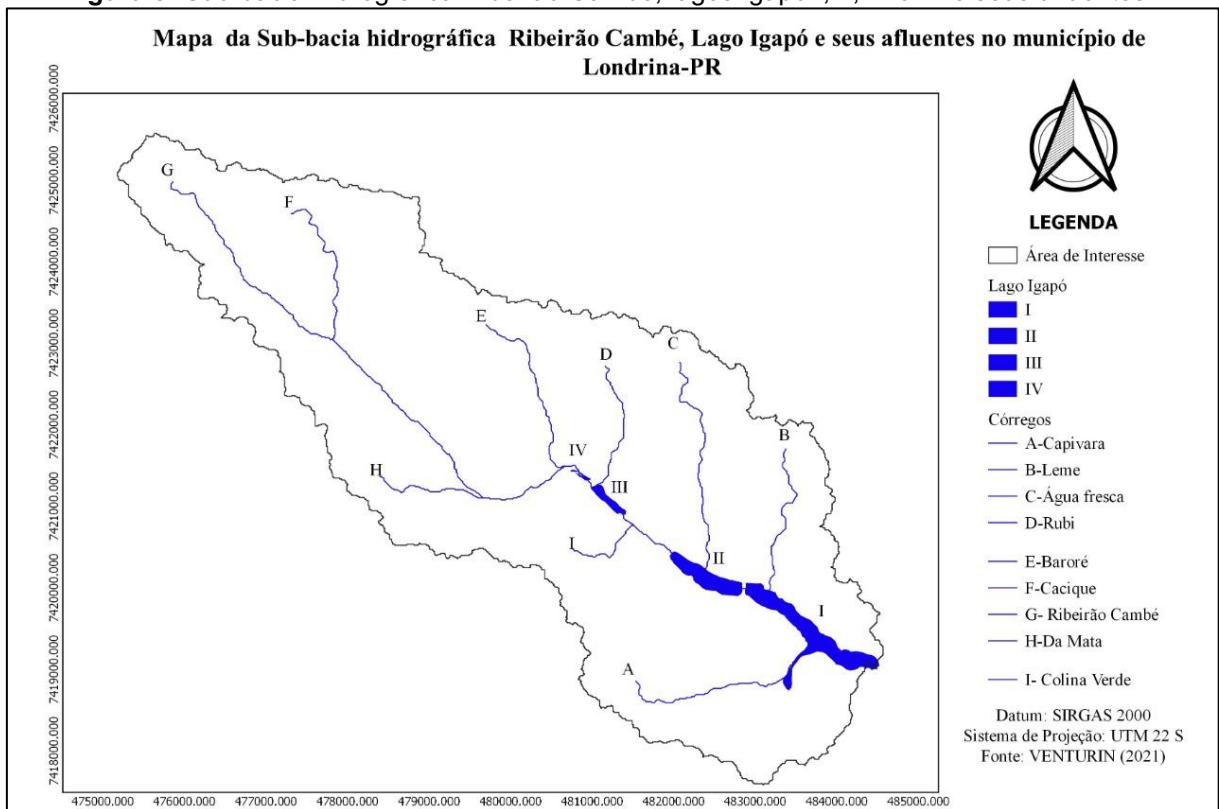
É importante destacar que, apesar de o local de estudo ser popularmente denominado de Lago Igapó, trata-se de um reservatório artificial, pois foi criado após o represamento de um trecho do Ribeirão Cambé (BORTOLO, 2011).

Figura 4: Mapa de Localização da Bacia Hidrográfica do Ribeirão Cambé e Lago Igapó.



Fonte: Autoria própria (2021).

Figura 5: Sub-bacia Hidrográfica Ribeirão Cambé, lagos Igapó I, II, III e IV e seus afluentes.



Fonte: Autoria própria (2021).

Segundo Lorenzo (2011), o Lago Igapó começou a ser idealizado na gestão do prefeito Antônio Fernandes Sobrinho em meados de 1957 e inaugurado em 1959. O Lago Igapó, que inicialmente era apenas um, foi subdividido em quatro lagos devido às intercepções de ruas e avenidas. Segundo medições realizadas pelo autor utilizando a base de dados do *Google Earth*® (2022), foi possível obter a área dos lagos, sendo, Lago I: 420.345 m², Lago II: 171.631 m², Lago III: 42.416 m² e Lago IV: 15.486 m².

Os Lagos Igapó I, II, III e IV originaram-se do represamento do Ribeirão Cambé, com a finalidade de solucionar os problemas de drenagem em parte do curso deste e ocupar uma área sem valor econômico para a época. Assim, esta área seria a maior área verde de Londrina, visando ainda o esporte e o lazer para a população londrinense (BORTOLO, 2011) (Figuras 6, 7 e 8).

Em meados de 1960 e 1970, começaram ocupações no entorno do Lago Igapó. A porção leste recebeu construções de residências maiores e mais luxuosas, enquanto que a porção oeste recebeu as primeiras casas populares da cidade. Com o decorrer dos anos, as ocupações no entorno do lago foram se expandindo, mantendo a divisão social entre lado leste e oeste (BORTOLO, 2011).

Figura 6: Barragem Lago Igapó (a) em fase de construção e (b) atualmente.



Fonte: (a) Iwanaga (1989); (b) Autoria Própria (2021).

Figura 7: Vista panorâmica do Lago Igapó I em 1960.



Fonte: Iwanaga (1989).

Figura 8: Vista panorâmica do Lago Igapó I em 2021.



Fonte: Aatoria Própria (2021).

Em 1990 houve a intensificação da urbanização na bacia do Ribeirão Cambé, devido a construção de um grande *shopping center* que agregou valorização urbana a esta região, que passou a ter construções verticalizadas. A urbanização acelerada gerou impermeabilização do solo e desmatamento da mata ciliar, que por sua vez contribuiu significativamente no ciclo hidrológico da bacia, aumentando o escoamento superficial e, conseqüentemente, provocando aumento de eventos de enchentes, alagamentos e inundações (VACARIO; MACHADO, 2017; VACARIO; MACHADO, 2019).

As ocupações urbanas não levaram em consideração os processos erosivos e as características biológicas da área, gerando degradação do ecossistema. Deste modo, tornou-se necessária a realização de obras de esvaziamento e limpeza, com a finalidade de recuperar a qualidade do Lago Igapó (FRESCA, 2010). Nos anos 1996 e 2001, os lagos Igapó I e II, respectivamente, passaram por obras de revitalização, onde foram esvaziados e desassoreados (VACARIO; MACHADO, 2019).

Segundo Vacario e Machado (2019), o problema de assoreamento nos lagos Igapó I e II sempre foram recorrentes, porém, intensificaram-se devido ao uso e ocupação da bacia, principalmente pela geração de resíduos da construção civil. Tais problemas provocaram enchentes, inundações e alagamentos causando prejuízos à população. Danos causados por inundações foram noticiados pela mídia local nos anos de 2011, 2012, 2014, 2016 e 2017.

Vacario e Machado (2019), com base no resultado de seus estudos, constataram que os lagos Igapó I e II encontram-se muito assoreados e em determinados lugares há formação de pequenas ilhas, que podem ser observadas nas épocas mais secas do ano. O assoreamento dos lagos Igapó I e II e a impermeabilização do solo ao entorno, associado com fatores ambientais e sociais, contribuem diretamente para a ocorrência de inundações, enchentes e alagamentos que se repetem com frequência.

3.4 Aprendizado de máquina

O AM é um subcampo da Inteligência Artificial (IA) que começou a evoluir na segunda metade do século XX. O AM mescla ideias da neurociência, estatística, matemática e física para fazer os computadores aprenderem. É capaz de desenvolver algoritmos de autoaprendizagem mais eficientes na captura de conhecimento de dados, melhorando gradualmente o desempenho de modelos preditivos, sendo capaz de tomar decisões baseadas em dados (RASCHKA, 2015; MARSLAND, 2015).

Existem 3 tipos básicos de AM, são eles: Aprendizado supervisionado; Aprendizado não supervisionado; e Aprendizado por reforço (RASCHKA, 2015). No aprendizado supervisionado o objetivo principal é ensinar um modelo a partir da dados de treinamento rotulados, que permitem fazer previsões sobre dados futuros, ou seja, o termo supervisionado refere-se a um conjunto amostral onde os dados de saída

desejados (rótulos) são previamente conhecidos (RASCHKA, 2015; MARSLAND, 2015). Esta categoria é subdividida em aprendizado supervisionado de classificação e de regressão (RASCHKA, 2015).

A categoria de classificação tem por objetivo prever rótulos de classes categóricas de novas instâncias, baseada em observações anteriores. O conjunto de rótulos de classe não necessita ser de natureza binária. Já a categoria de AR recebe diversas variáveis preditoras (explicativa) e ou mais variável resposta numérica (resultado) e assim o algoritmo é capaz de encontrar relação entre essas variáveis, sendo possível prever um resultado final (RASCHKA, 2015).

O aprendizado não supervisionado, diferente do supervisionado, não possui resposta previa para treinar o modelo, sendo assim, os dados não são rotulados de antemão. No entanto, usando algumas técnicas é possível explorar a estrutura dos dados e extrair informações significativas sem a orientação de uma variável de resultado conhecido, como por exemplo, agrupar observações com características semelhantes (RASCHKA, 2015; MARSLAND, 2015).

Por sua vez, o aprendizado por reforço tem como objetivo desenvolver um sistema (agente) que aprimore seu desempenho com base nas interações com o ambiente obtendo um sinal de recompensa. Por meio de interações com o ambiente, um agente pode aprender várias ações que maximizam a recompensa por meio de tentativa e erro ou planejamento deliberativo (RASCHKA, 2015).

Kupssinskü *et al.* (2020) utilizaram imagens do satélite Sentinel-2, imagens capturadas por VANT e diversas técnicas de AM com a finalidade de propor um modelo para predição de SST e Clorofila-a em um lago artificial. Sami *et al.* (2021) investigaram a confiabilidade de algoritmos de AM como ferramenta sustentável para a previsão de SST em um reservatório em Taiwan. Wang *et al.* (2021) aplicou modelos de AM para prever a turbidez de águas costeiras em resposta a fatores hidrodinâmicos e assim comparar os métodos e determinar qual fornece melhor resultado.

Alguns modelos de AM, mais especificamente de aprendizado supervisionado, serão utilizados neste estudo e, portanto, descrito nas seções subsequentes. São eles: Floresta Aleatória (FA); Regressão Linear (RL); Regressão de Ridge; Regressão de Lasso; e Regressão ElasticNet.

3.4.1 Floresta Aleatória

A FA ou *Random Forest* é uma combinação de preditores de árvores, em que cada uma depende de um vetor aleatório, amostrado independentemente e com mesma distribuição para todas as árvores da floresta. É a combinação de várias árvores de decisão, sendo então, um classificador que possui vários classificadores estruturados em árvores (BREIMAN, 2001; CUTLER; CUTLER; STEVENS, 2011). A técnica FA pode ser usada para uma variável categórica, denominada como classificação, ou como uma resposta contínua, denominada de regressão. Deste modo, as variáveis podem ser categóricas ou contínuas (BREIMAN, 2001).

O erro de generalização de uma floresta depende da força de cada árvore e da correlação entre elas. As estimativas internas monitoram o erro, a força e a correlação entre elas e são utilizadas para mostrar o aumento de recursos usados na divisão (BREIMAN, 2001).

De acordo com Cutler, Cutler e Stevens (2011), a FA é atraente, pois lida naturalmente tanto com a regressão quanto com a classificação, são relativamente rápidas de treinar, pode ser utilizada diretamente em grandes problemas e possui recursos adicionais que fornecem medidas estatísticas importantes.

3.4.2 Regressão Linear

Os modelos lineares são amplamente utilizados como parte do processo de aprendizado, tanto no planejamento quanto na análise de dados de uma pesquisa, fazendo previsões utilizando uma função linear, com base nas variáveis de entrada (RENCHEER; SCHAALJE, 2008; MÜLLER; GUIDO, 2016). A RL simples é utilizada para modelar a relação entre duas variáveis, por exemplo x e y , em que y é a variável dependente ou de resposta, e x é a variável independente ou preditora (Equação 1) (RENCHEER; SCHAALJE, 2008).

$$y = \beta_0 + \beta_1x + \epsilon \tag{1}$$

A variável aleatória ϵ é o termo de erro do modelo, que representa um termo estatístico de flutuações aleatórias ou efeito de fatores que estão fora de controle. Os

valores β_0 e β_1 são estimados através de inferências, conhecidos como intercepto e inclinação. Caso haja interesse em realizar uma previsão preditiva, é possível através da Equação 1, prever valores de y para um determinado valor de x (RENCHER; SCHAALJE, 2008).

Os modelos de Regressão Linear Múltipla (RLM) são utilizados para fins de predição, descrição ou explicação dos dados observados, estimativa de parâmetro, seleção de variáveis preditoras e controle de produção (RENCHER; SCHAALJE, 2008). Na RLM a resposta de y é influenciada por mais de uma variável preditora (RENCHER; SCHAALJE, 2008). O modelo de RLM pode ser expresso conforme a Equação 2,

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_kx_k + \epsilon, \quad (2)$$

em que β_0 , β_1 e β_k são chamados de coeficientes de regressão e variável aleatória ϵ , assim como na Equação (1), fornece variações aleatórias em y que não são explicadas pelas variáveis preditoras (RENCHER; SCHAALJE, 2008).

A RLM permite uma abordagem geométrica para três teorias importantes: a Regressão Ridge, e a Regressão de *Least Absolute Shrinkage and Selection Operator* (Lasso) e a Regressão *ElasticNet*, que apresentam resultados teóricos diferentes (PEREIRA, 2017).

3.4.3 Regressão de *Ridge*, Regressão de Lasso e Regressão *ElasticNet*

O método de Lasso foi proposto por Tibshirani (1996) tendo como principal característica ser um método automático de seleção de covariáveis. No entanto, este método possui alguns problemas que foram resolvidos com um novo método proposto por Zou e Hastie (2005) denominado de *ElasticNet*. Este é uma combinação entre o método de *Ridge* e Lasso.

A Regressão de *Ridge* é um modelo de RLM. Neste modelo, os coeficientes não são escolhidos apenas para prever os dados de treinamento, mas também para ajustar a uma restrição adicional. Esta restrição é chamada de regularização, que significa restringir explicitamente um modelo a fim de evitar *overfitting*. O tipo de regularização utilizada pela Regressão de *Ridge* é a regularização L2 (MÜLLER;

GUIDO, 2016). Além disso, este modelo busca coeficientes muito próximos de zero, fazendo com que cada recurso exerça o menor efeito possível sobre o resultado e ao mesmo tempo fazendo uma boa previsão (MÜLLER; GUIDO, 2016).

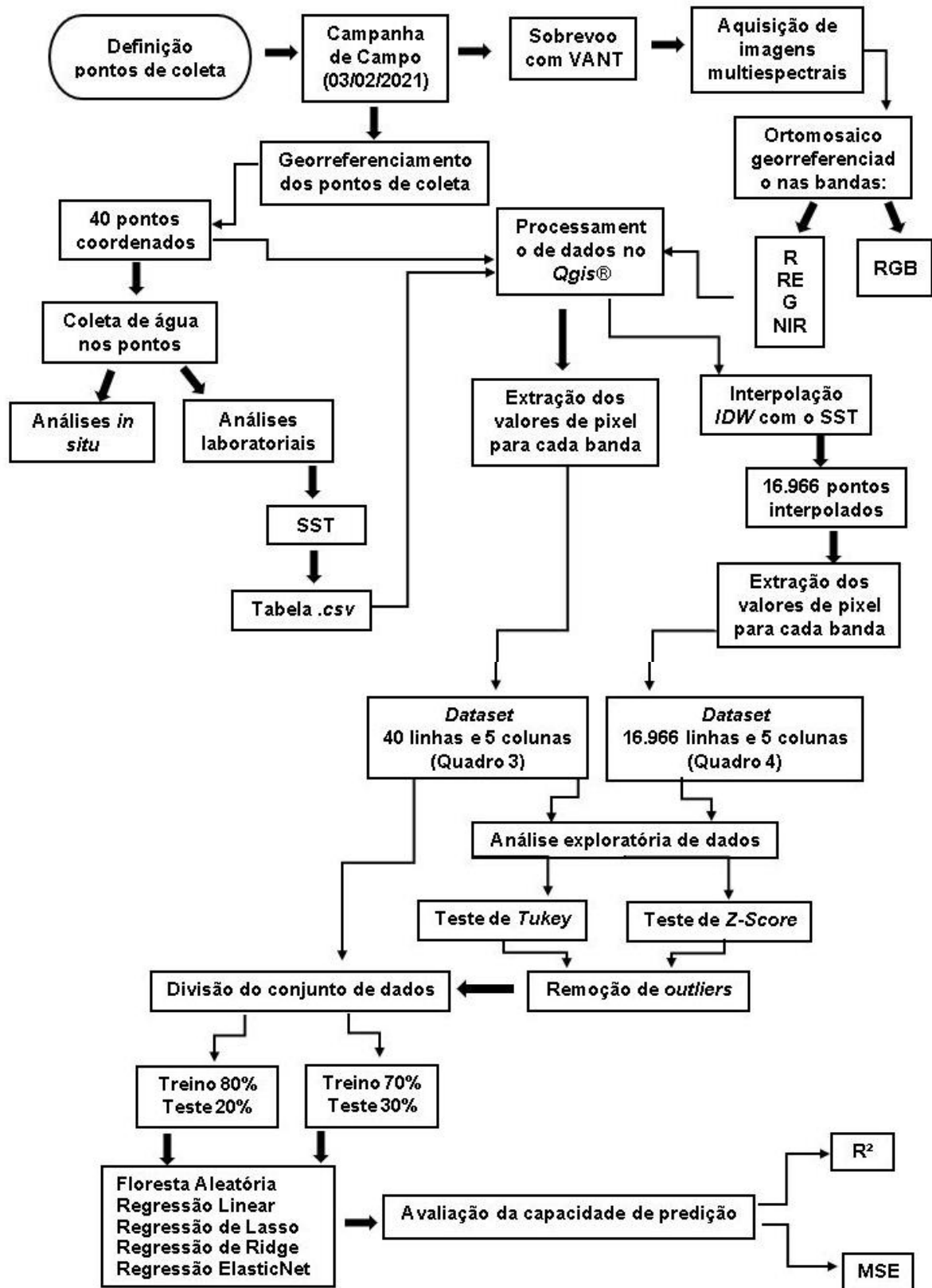
Outro modelo de regressão linear é a Regressão de *Lasso*. Da mesma forma que acontece na Regressão de *Ridge*, o uso do *Lasso* também restringe os coeficientes a serem próximos de zero, de uma forma ligeiramente diferente, denominada de regularização L1. Isso significa que alguns coeficientes são exatamente zero, de forma que certos recursos sejam totalmente ignorados pelo modelo, tornando o modelo mais fácil de interpretar e possibilita revelar os recursos mais importantes dentro do modelo em estudo (MÜLLER; GUIDO, 2016).

A Regressão de *Ridge* geralmente é a mais utilizada, entre esses dois modelos. No entanto, se o banco de dados tiver muitos recursos, e espera-se que apenas alguns deles sejam importantes, sendo capaz de obter um modelo mais fácil de ser interpretado, a Regressão de *Lasso* pode ser o melhor modelo. Porém, existe um modelo chamado *ElasticNet* que é uma combinação desses dois modelos (*Ridge* e *Lasso*), o que poderá funcionar melhor na prática (MÜLLER; GUIDO, 2016).

4 METODOLOGIA

A metodologia proposta neste trabalho envolve três etapas fundamentais: A primeira delas é a definição da área de mapeamento e os pontos de coletas de água para análises laboratoriais. A segunda etapa compreende o planejamento e execução de sobrevoos com VANT e a aquisição das imagens aéreas, que ocorrerá juntamente com a coleta de água. A terceira etapa compreende a aplicação das técnicas de AM na análise dos dados espectrais e sua relação com os dados de qualidade da água. Essas etapas são mostradas no fluxograma da Figura 9 e serão detalhadas na sequência.

Figura 9: Fluxograma do processo metodológico.



Fonte: Autoria própria (2021).

4.1 Área de estudo e campanha de campo

O local de estudo desta pesquisa foram os lagos Igapó I e II. Foi realizada uma campanha de campo para coleta de amostras de água, análises de parâmetros *in situ* e em laboratório, e ainda foram realizadas as capturas de imagens aéreas com VANT.

A campanha de campo ocorreu no dia 03 de fevereiro de 2021. Essa data foi escolhida para que a coleta sucedesse um período de chuva constante na região, de Londrina, como pode ser observado na Tabela 2. Os dados da Tabela 2 foram obtidos no aplicativo IAPAR Clima, criado pela área de agrometeorologia do Instituto Agrônomo do Paraná (IAPAR) e disponibilizado gratuitamente. Os dados disponíveis neste aplicativo são provenientes da estação automática do IAPAR localizada na Rodovia Celso Garcia Cid, km 375, Londrina - PR, Latitude 23°21'15"S e Longitude 51°09'42"O.

Tabela 2: Precipitação no período de 15 dias que antecederam a data da coleta.

Data	Precipitação (mm)	Tmed ¹ (°C)	Tmax ² (°C)	Tmin ³ (°C)	Ur ⁴ (%)
19/01/2021	32,6	21,7	26,7	19,9	96
20/01/2021	14,2	22,0	25,8	20,1	95
21/01/2021	1,4	22,7	27,4	18,9	89
22/01/2021	0,0	23,3	30,2	18,9	85
23/01/2021	32,4	23,4	30,0	19,0	84
24/01/2021	0,1	22,3	28,1	19,3	89
25/01/2021	0,9	21,9	27,3	19,2	90
26/01/2021	0,1	24,0	30,1	19,1	84
27/01/2021	0,0	23,8	29,6	20,4	85
28/01/2021	17,1	22,2	28,0	19,7	93
29/01/2021	0,9	22,9	28,7	19,7	89
30/01/2021	10,1	22,0	25,9	19,8	95
31/01/2021	25,8	21,3	26,0	18,9	96
01/02/2021	0,6	22,9	29,	18,8	90
02/02/2021	0,3	25,3	31,9	19,6	82
03/02/2021	10,9	24,1	31,9	20,6	86

Legenda: ¹Temperatura Média; ²Temperatura Máxima; ³Temperatura Mínima; ⁴Umidade Relativa.

Fonte: Adaptado do aplicativo IAPAR Clima (2021).

Os pontos de coleta foram definidos de maneira que representassem os diferentes padrões de ocupação no entorno dos lagos Igapó I e II, além de pontos distribuídos de maneira a garantir ampla cobertura da superfície dos lagos. Para o

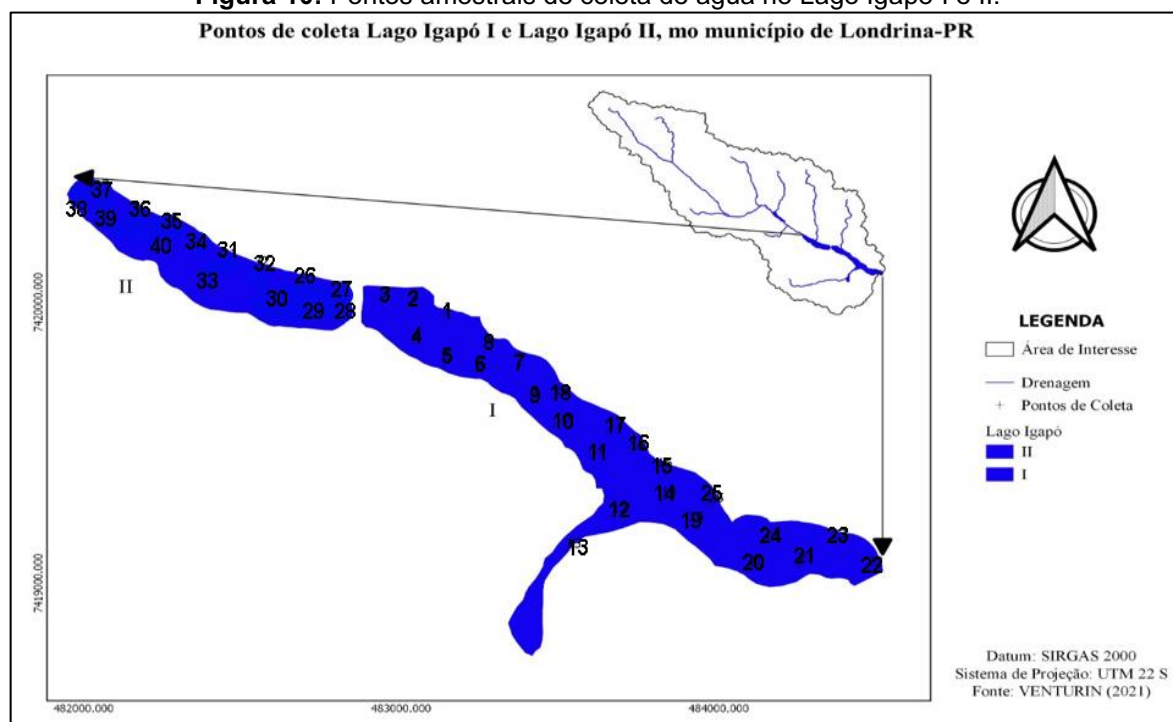
Lago Igapó I foram realizadas coletas em 25 pontos e para o Lago Igapó II em 15 pontos, totalizando 40 amostras coletadas. Todos os pontos foram georreferenciados com o auxílio de um GPS durante a amostragem (Figura 10) (Tabela 3).

Tabela 3: Localização dos pontos amostrais.

Ponto coletado	Coordenada Geográfica		Localização no lago
	Latitude	Longitude	
1	-23,329836	-51,164989	Lago Igapó I
2	-23,329456	-51,165989	Lago Igapó I
3	-23,329376	-51,166920	Lago Igapó I
4	-23,330763	-51,165958	Lago Igapó I
5	-23,331387	-51,164944	Lago Igapó I
6	-23,331650	-51,163998	Lago Igapó I
7	-23,331480	-51,162720	Lago Igapó I
8	-23,330872	-51,163673	Lago Igapó I
9	-23,332525	-51,162304	Lago Igapó I
10	-23,333429	-51,161575	Lago Igapó I
11	-23,334417	-51,160469	Lago Igapó I
12	-23,336315	-51,159801	Lago Igapó I
13	-23,337446	-51,161221	Lago Igapó I
14	-23,335793	-51,158287	Lago Igapó I
15	-23,334850	-51,158413	Lago Igapó I
16	-23,334299	-51,159264	Lago Igapó I
17	-23,333591	-51,159935	Lago Igapó I
18	-23,332525	-51,161594	Lago Igapó I
19	-23,336599	-51,157452	Lago Igapó I
20	-23,338026	-51,155582	Lago Igapó I
21	-23,337900	-51,154060	Lago Igapó I
22	-23,338030	-51,151924	Lago Igapó I
23	-23,337151	-51,153042	Lago Igapó I
24	-23,337162	-51,155125	Lago Igapó I
25	-23,335981	-51,156834	Lago Igapó I
26	-23,328890	-51,169426	Lago Igapó II
27	-23,329165	-51,168381	Lago Igapó II
28	-23,329849	-51,168339	Lago Igapó II
29	-23,329859	-51,169388	Lago Igapó II
30	-23,329418	-51,170635	Lago Igapó II
31	-23,328110	-51,171825	Lago Igapó II
32	-23,328360	-51,170826	Lago Igapó II
33	-23,328765	-51,172519	Lago Igapó II
34	-23,327703	-51,172725	Lago Igapó II
35	-23,326951	-51,173616	Lago Igapó II
36	-23,326635	-51,174621	Lago Igapó II
37	-23,325893	-51,175769	Lago Igapó II
38	-23,326103	-51,176258	Lago Igapó II
39	-23,326723	-51,175689	Lago Igapó II
40	-23,327419	-51,174068	Lago Igapó II

Fonte: Autoria Própria (2021).

Figura 10: Pontos amostrais de coleta de água no Lago Igapó I e II.



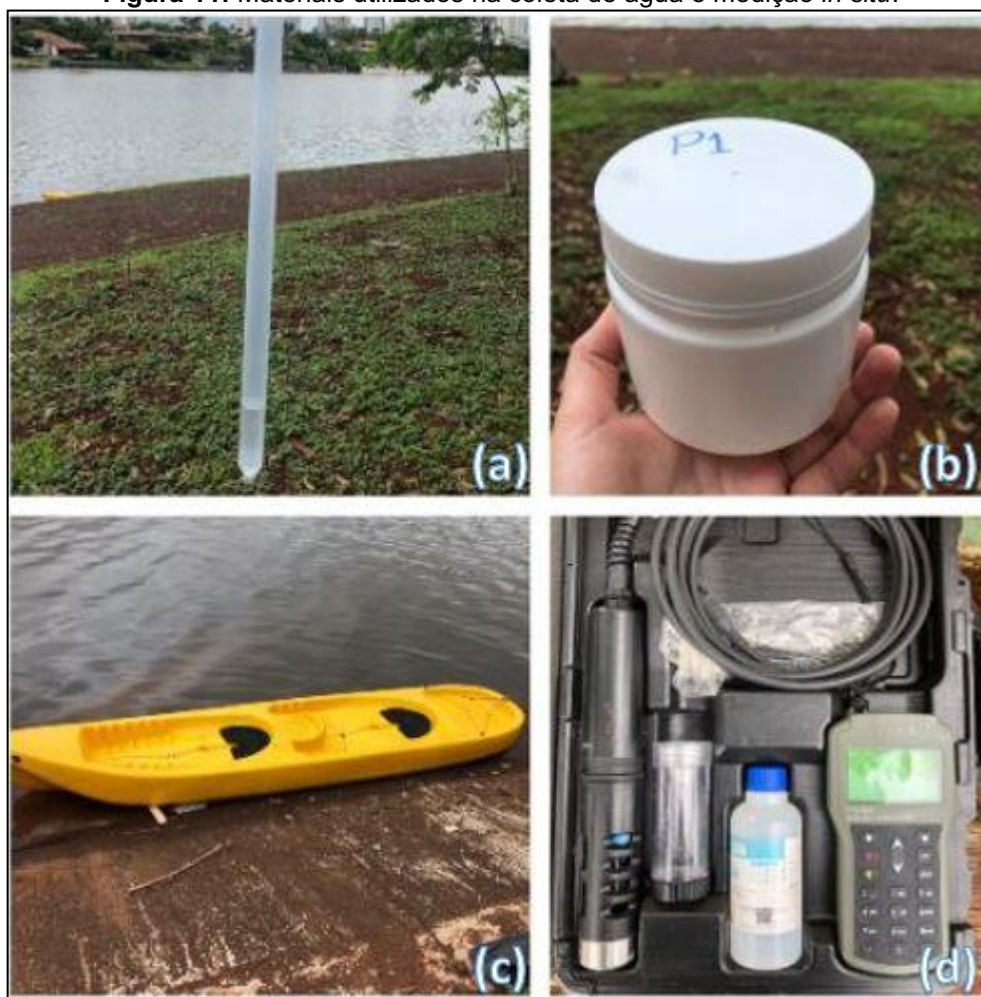
Fonte: Autoria Própria (2021).

4.2 Amostragem de água

Durante a campanha de campo, foram realizadas coletas de água para análises *in situ* e análises laboratoriais. Foram coletadas 500 mL de água em cada ponto amostral previamente definido. A amostra foi coletada com o auxílio de um tubo coletor de plástico transparente e em seguida transferida para um frasco plástico com tampa lacrável, devidamente identificada por um número sequencial e as coordenadas geográficas de cada ponto. As coletas em pontos afastados da borda foram realizadas com o auxílio de um caiaque.

As medições *in situ* foram realizadas com o auxílio de uma sonda multiparamétrica portátil modelo HI 9819X que foi inserida dentro do referido frasco por um minuto, para estabilização dos valores dos parâmetros (Figura 11). O medidor multiparâmetros encontra-se em conformidade com as normas IP67 e IP68 (HANNA INSTRUMENTS, 2015). Os parâmetros mensurados pela sonda multiparamétrica e analisados neste estudo foram: pH, CE, Temperatura, Porcentagem de OD (%OD), Concentração de OD, Salinidade; Pressão e SDT.

Figura 11: Materiais utilizados na coleta de água e medição *in situ*.



Legenda: (a): tubo coletor; (b): frasco plástico com tampa; (c): caiaque; (d): sonda multiparamétrica.
Fonte: Autoria própria (2021).

Após a medição *in situ*, os potes com as amostras foram lacrados e armazenados temporariamente em uma caixa térmica, a fim de manter as características das amostras para posterior análise laboratorial de turbidez e SST. Após a finalização das coletas, as amostras foram levadas ao Laboratório de Saneamento Ambiental da Universidade Tecnológica Federal do Paraná (UTFPR) no Campus Londrina, armazenadas em geladeira, para posterior análise de turbidez e concentração de SST. A turbidez foi mensurada por um turbidímetro digital *PoliControl AP 2000*® (Figura 12).

Figura 12: Turbidímetro digital *PoliControl AP 2000*®.

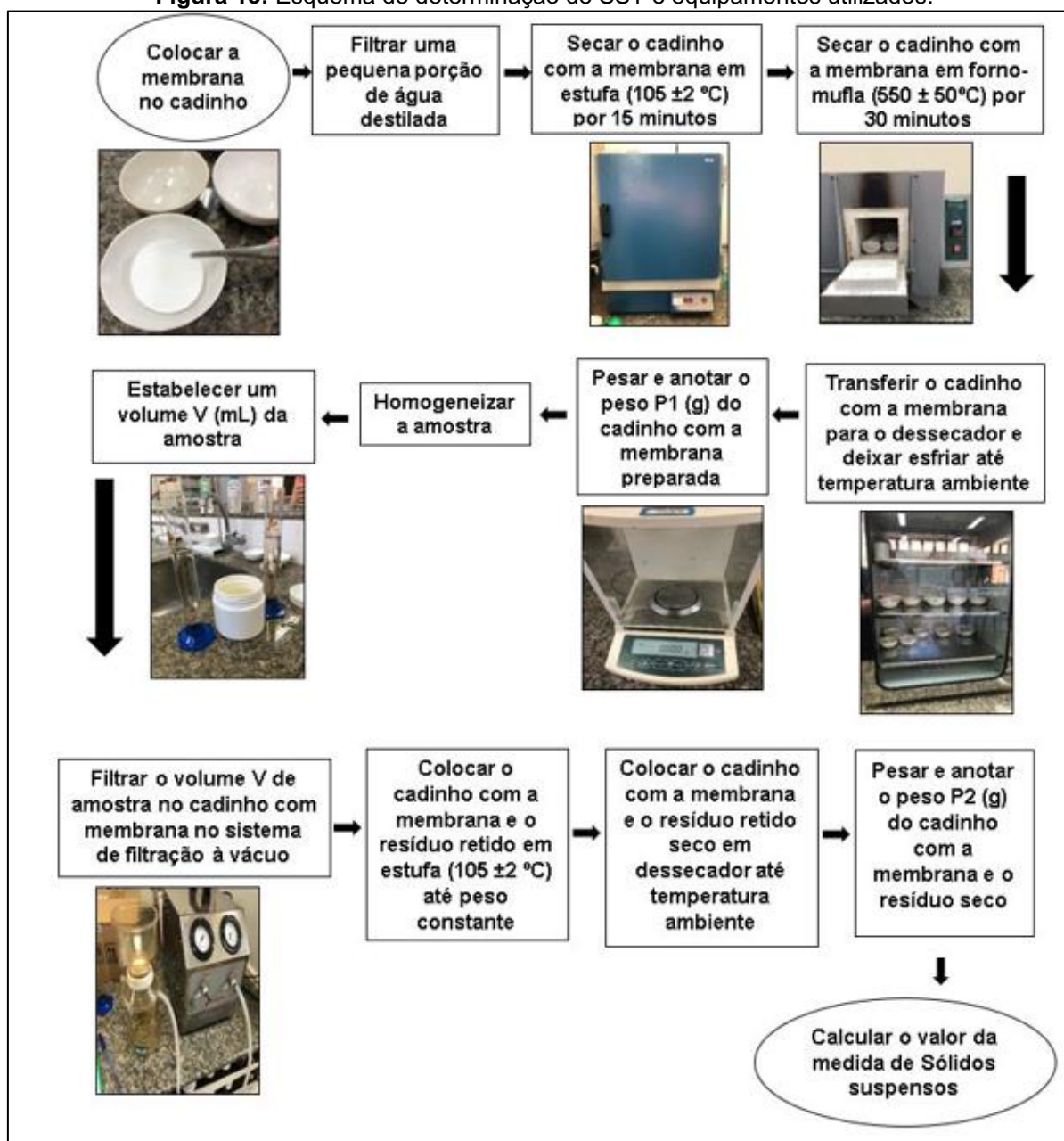


Fonte: Autoria própria (2021).

Já o SST foi determinado seguindo a metodologia proposta pela Companhia de Saneamento Básico do Estado de São Paulo (SABESP) (Figura 13) em sua Norma Técnica Interna SABESP NTS 013 (1999), que trata sobre os meios de determinação de sólidos. De acordo com a norma SABESP NTS 013, os sólidos em suspensão podem ser considerados como toda substância que após ser submetida a um processo de filtração e secagem permanece retida na membrana filtrante. Vale ressaltar que o SST foi determinado em duplicata para cada ponto de amostragem, de maneira que se utilizou o valor médio entre as duas amostras medidas calculado pela média aritmética.

Os equipamentos utilizados nesta etapa de mensuração do SST foram: cadinho de porcelana, forno mufla, dessecador à vácuo, balança analítica, proveta volumétrica de vidro, sistema de filtração a vácuo e estufa (Figura 13). Todos os equipamentos laboratoriais necessários para esta etapa estavam disponíveis no Laboratório de Saneamento Ambiental da UTFPR.

Figura 13: Esquema de determinação de SST e equipamentos utilizados.



Fonte: Adaptado de SABESP (1999).

O cálculo de determinação de SST, foi realizado seguindo a Equação (3):

$$\frac{P2-P1}{Vol.amostra} \times 1.000.000 \quad (3)$$

em que, $P1$ é a tara do cadinho mais a membrana (g), $P2$ é o cadinho com a membrana mais a amostra após filtragem e secagem (g), $Vol.amostra$ é o volume da amostra (mL), assim, o resultado foi expresso em mg/L. O volume de água utilizado para a determinação do SST foi de 100 mL.

Para a filtração da amostra de água, foi utilizada a membrana filtrante de fibra de vidro GF/C com 47mm de diâmetro e 1,2 μm de porosidade da marca *FilterPro* (Figura 14).

Figura 14: Membrana de fibra de vidro GF/C 47mm x 1,2 μm da marca *FilterPro*.



Fonte: Autoria própria (2021).

4.3 Aquisição das imagens aéreas com VANT

A aquisição das imagens aéreas foi realizada no mesmo dia da amostragem de água (03/02/2021), com o intuito de que as imagens aéreas pudessem fazer uma representação fidedigna das características dos lagos Igapó I e II no momento da coleta, para posterior treinamento dos modelos de AM.

Para a captura de imagens aéreas foi contratada a empresa G-drones®, que ficou responsável pelo imageamento aéreo, processamento das imagens e fornecimento de ortomosaico georreferenciado para a área de estudo.

Devido ao fato de o local de estudo estar nas proximidades do aeródromo de Londrina, fez-se necessário atender alguns requisitos burocráticos. Para iniciar a campanha de sobrevoo com VANT foi necessário solicitar à administração aeroportuária de Londrina, a elaboração de um ofício de anuência, para ser anexado à solicitação de autorização de voo junto ao Departamento de Controle do Espaço

Aéreo (DECEA) e ao fabricante do drone (*Dji*) (Anexo 1). Para tanto, por exigência dos responsáveis da administração aeroportuária e órgão de navegação aérea do aeroporto de Londrina, foi necessário que a empresa G-drones® elaborasse uma Avaliação de Risco Operacional (ARO) (Anexo 2).

Para a captura de imagens aéreas foi utilizado um VANT da marca *Dji Phantom 4 Pro*®. O equipamento possui massa de 1,375 kg, velocidade máxima de 72 km/h, sistema de posicionamento GPS/GLONASS, bateria de 5870 mAh e sua câmera original integrada possui sensor CMOS 1" com 20 pixels efetivos (DJI, 2021) (Figura 15).

Figura 15: VANT *Dji Phantom 4 Pro*® e câmera integrada.



Fonte: Autoria própria (2021).

As imagens foram capturadas por duas câmeras diferentes, a original do VANT que captura imagens nas bandas RGB e a câmera multiespectral *Parrot Sequoia*® que foi acoplada no VANT. A câmera multiespectral *Parrot Sequoia*® possui sensor RGB com 16 megapixel (MPx) de resolução (4.608 × 3.456 pixels), quatro sensores monocromáticos com 1,2 MPx de resolução (1.280 × 960 pixels) nas bandas G, R, *Red Edge* (RE) e NIR e GPS/GNSS integrado (MANUAL DO USUÁRIO PARROT SEQUOIA, 2017) (Quadro 2) (Figura 16).

Quadro 2: Características das bandas da câmera multiespectral *Parrot Sequoia*®.

Banda	Comprimento de onda (nm)	Largura de banda (nm)	Resolução (Mpx)
Verde (G)	550	40	1,2
Vermelho (R)	660	40	1,2
Red Edge (RE)	735	10	1,2
Infravermelho próximo (NIR)	790	40	1,2
RGB	-	-	16

Fonte: Adaptado de manual do Usuário *Parrot Sequoia*® (2017).

Figura 16: Câmera multiespectral *Parrot Sequoia*® que foi acoplada no VANT.

Fonte: Autoria própria (2021).

O VANT *Dji Phantom 4 Pro*® sobrevoou o local com altura de voo aproximada de 120 m do solo. As imagens RGB capturadas pela câmera original do VANT têm resolução espacial de 0,036 x 0,036 m. Já as imagens capturadas pela câmera *Parrot Sequoia*® nas bandas G, R, RE, NIR e RGB, têm resolução espacial de 0,14 x 0,14 m.

4.4 Análise de dados

Esta etapa metodológica foi similar à metodologia utilizada por Kupssinskü *et al.* (2020) que utilizou sensoriamento remoto e técnicas de AM para prever concentrações de SST em dois lagos, sendo um no Rio Grande do Sul e o outro em São Paulo.

4.4.1 Processamento das imagens aéreas

Primeiramente as imagens do Lago I e Lago II nas bandas G, R, RE e NIR foram carregadas no software Qgis 2.18.10, utilizando o sistema de Projeção SIRGAS 2000 Zona UTM 22S. Em seguida foi carregado o arquivo no formato CSV, correspondente às coordenadas de cada ponto de coleta (40 pontos). Os pontos coordenados foram carregados como *shapefile*, um formato de arquivos que contém dados espaciais, na Projeção DATUM SIRGAS 2000 Zona UTM 22S para que os pontos e as imagens possuíssem a mesma projeção cartográfica.

Feito isso, foi realizada a extração dos valores de pixels relativo a cada ponto coordenado para cada banda individualmente (G, R, RE e NIR) utilizando o complemento *Point Sampling Toll*. Assim, foi possível observar a tabela de atributos que continha informações do ponto coordenado, do valor do SST mensurado em laboratório e do valor de pixel relativo a cada uma das bandas. Os dados dessa camada foram exportados para o formato *MS Office Open XML spreadsheet*, para que pudessem ser visualizados em uma planilha do Excel. Com esse procedimento, obteve-se um conjunto de dados com as coordenadas, os valores do pixel por banda (R, RE, NIR e G) o SST, como apresentado no Quadro 3.

Quadro 3: Tabela de dados criada com os valores de pixel e SST (obtidos em laboratório) para cada ponto amostral.

Ponto coletado	Coordenada Geográfica		Valor do pixel por banda				SST (mg/L)
	Latitude	Longitude	R	RE	NIR	G	
1	-23,329836	-51,164989	0,03046	0,0111	0,01476	0,03445	0
2	-23,329456	-51,165989	0,03403	0,0283	0,02808	0,03852	5
3	-23,329376	-51,166920	0,03223	0,0509	0,03146	0,03542	10
4	-23,330763	-51,165958	0,02711	0,0266	0,02094	0,02840	0
5	-23,331387	-51,164944	0,02745	0,0201	0,02087	0,03021	5
6	-23,331650	-51,163998	0,02565	0,0423	0,01584	0,03532	25
7	-23,331480	-51,162720	0,02195	0,0085	0,00916	0,02890	20
8	-23,330872	-51,163673	0,02147	0,0077	0,00753	0,02873	15
9	-23,332525	-51,162304	0,02366	0,0238	0,01358	0,03463	5
10	-23,333429	-51,161575	0,02182	0,0723	0,01444	0,03346	10
11	-23,334417	-51,160469	0,02228	0,0508	0,01598	0,03035	5
12	-23,336315	-51,159801	0,01933	0,0152	0,01125	0,02854	15
13	-23,337446	-51,161221	0,02307	0,0085	0,01687	0,02657	10
14	-23,335793	-51,158287	0,01933	0,0074	0,00831	0,02802	5
15	-23,334850	-51,158413	0,01646	0,0075	0,00724	0,02359	5
16	-23,334299	-51,159264	0,01946	0,0082	0,00945	0,02926	10
17	-23,333591	-51,159935	0,01975	0,0093	0,00949	0,02900	10
18	-23,332525	-51,161594	0,02376	0,0125	0,01232	0,03248	15
19	-23,336599	-51,157452	0,01978	0,0103	0,01065	0,02776	20
20	-23,338026	-51,155582	0,02037	0,0168	0,01260	0,02662	10
21	-23,337900	-51,154060	0,01947	0,0097	0,01001	0,02837	10
22	-23,338030	-51,151924	0,01841	0,0087	0,01258	0,02693	50
23	-23,337151	-51,153042	0,01755	0,0078	0,00984	0,02525	5
24	-23,337162	-51,155125	0,01814	0,0077	0,00905	0,02720	5
25	-23,335981	-51,156834	0,01893	0,0082	0,00921	0,02744	10
26	-23,328890	-51,169426	0,03885	0,0246	0,02597	0,04041	15
27	-23,329165	-51,168381	0,04086	0,0304	0,03252	0,04752	10
28	-23,329849	-51,168339	0,04984	0,0319	0,03278	0,05279	10
29	-23,329859	-51,169388	0,04770	0,0290	0,03442	0,04889	10
30	-23,329418	-51,170635	0,04723	0,0274	0,03051	0,04662	15
31	-23,328110	-51,171825	0,05186	0,0278	0,03221	0,04914	10
32	-23,328360	-51,170826	0,03489	0,0235	0,02592	0,03809	15
33	-23,328765	-51,172519	0,06023	0,0352	0,03867	0,05583	5
34	-23,327703	-51,172725	0,05304	0,0348	0,03765	0,05531	25
35	-23,326951	-51,173616	0,04614	0,0331	0,03653	0,05201	20
36	-23,326635	-51,174621	0,05183	0,0331	0,03539	0,05210	15
37	-23,325893	-51,175769	0,05493	0,0361	0,03770	0,05275	30
38	-23,326103	-51,176258	0,05340	0,0367	0,03899	0,05233	20
39	-23,326723	-51,175689	0,05454	0,0370	0,04236	0,05951	25
40	-23,327419	-51,174068	0,04993	0,0332	0,03363	0,05325	10

Fonte: Autoria própria (2021).

Como o tamanho da amostra (40 pontos) é considerado relativamente pequeno quando se trabalha com AM, foi realizada a interpolação dos valores de SST desses pontos a fim de gerar um mapa de interpolação para os lagos Igapó I e II.

4.4.2 Interpolação IDW

O método de interpolação escolhido para este estudo foi o IDW, o mesmo método utilizado por Kupssinskü *et al.* (2020). Esse método consiste em realizar a estimativa de uma variável ao longo da área de estudo, ponderando os pesos dos valores mais próximos ao ponto onde está sendo calculado, sendo que o peso de cada valor é função do inverso de uma potência da distância, isso implica que quanto mais próximo do ponto a ser estimado, maior o peso a ser atribuído ao valor do ponto amostrado (SILVA; CAIADO; COSTA CAIADO, 2013).

Para realizar a interpolação, primeiramente foram carregadas no *software Qgis* as imagens do Lago Igapó I e Lago Igapó II nas bandas G, R, RE e NIR no sistema de Projeção SIRGAS 2000 Zona UTM 22S. Foi carregado também o arquivo no formato CSV com as coordenadas dos pontos coletadas e com os valores de SST obtidos, para os lagos I e II, totalizando 40 pontos. Os pontos coordenados foram importados como *shapefile* na Projeção DATUM SIRGAS 2000 Zona UTM 22S para que os pontos e as imagens possuíssem a mesma projeção cartográfica.

Em seguida foi utilizada a ferramenta disponível no *software Qgis* denominada de “Interpolação IDW”. Nessa ferramenta é necessário adicionar o tamanho do “*Raster de Saída*”, cujo valor considerado foi de 300 x 490 (linhas x colunas), e que gerou pixels no tamanho de 0,000051 x 0,000051 m (X, Y).

Assim, foi possível criar um *raster* de interpolação para toda a área dos lagos Igapó I e II. Feito isso, foi necessário converter o *raster* em vetor (*shape*). Para isso, foi utilizada a ferramenta *r.to.vector* que transformou toda a imagem em pontos vetoriais. No entanto, a fim de evitar pontos sombreados, foram desconsiderados todos os pontos que estavam a 10 metros da borda. Além disso, foi observado e retirado manualmente os demais pontos que não estavam dentro do *buffer* de 10 metros, mas que estavam sofrendo sombreamento de árvores ou de alguma edificação. Desse modo, foi possível obter um total de 16.966 pontos, distantes 5 metros entre si.

Posteriormente foram obtidas as coordenadas de todos os pontos através da ferramenta “Calculadora de Campo”, que gerou as coordenadas na projeção SIRGAS 2000 Zona UTM 22S. Desse modo, foi possível extrair o valor do pixel de cada ponto coordenado, utilizando a ferramenta “Amostrar valores de raster” para cada banda individualmente (G, R, RE e NIR). A camada gerada foi salva no formato *MS Office Open XML spreadsheet* para que pudesse ser visualizado como uma planilha do Excel. Assim, foi possível gerar uma nova tabela de dados com os 16.966 pontos, as respectivas coordenadas, os respectivos valores de SST e os respectivos valores de pixel para as bandas G, R, RE e NIR (Quadro 4).

Quadro 4: Tabela de dados criada com os valores de pixel e SST (obtidos em laboratório) para os pontos interpolados.

Ponto interpolado	Coordenada Geográfica		Valor do pixel por banda				SST (mg/L)
	Latitude	Longitude	R	RE	NIR	G	
1	-23,33	-51,18	0,049046	0,034147	0,037500	0,050669	24,57910
2	-23,33	-51,18	0,048916	0,033339	0,036635	0,052333	24,99377
3	-23,33	-51,18	0,049800	0,036450	0,040441	0,053965	25,40393
4	-23,33	-51,18	0,049530	0,037532	0,041269	0,053106	25,78572
5	-23,33	-51,18	0,048470	0,035049	0,038940	0,052966	26,11482
6	-23,33	-51,18	0,046729	0,033783	0,039187	0,051482	26,36987
7	-23,33	-51,18	0,050416	0,036015	0,040613	0,049872	24,21176
8	-23,33	-51,18	0,053002	0,037923	0,040112	0,051967	24,67963
9	-23,33	-51,18	0,051057	0,033656	0,038530	0,052358	25,17593
10	-23,33	-51,18	0,052817	0,037687	0,039114	0,052337	25,67409
[...]	[...]	[...]	[...]	[...]	[...]	[...]	[...]
16957	-23,34	-51,16	0,018937	0,000715	0,006434	0,014635	11,25603
16958	-23,34	-51,16	0,018375	0,001221	0,006634	0,013751	11,25603
16959	-23,34	-51,16	0,018274	0,001731	0,011294	0,015764	11,26610
16960	-23,34	-51,16	0,021270	0,001109	0,007958	0,017459	11,26563
16961	-23,34	-51,16	0,015672	0,003896	0,008038	0,011530	11,26528
16962	-23,34	-51,16	0,012988	0,005242	0,013307	0,005903	11,26506
16963	-23,34	-51,16	0,014276	0,000844	0,011275	0,012472	11,27474
16964	-23,34	-51,16	0,019346	0,001833	0,010483	0,015567	11,27433
16965	-23,34	-51,16	0,017725	0,002516	0,009521	0,014945	11,27404
16966	-23,34	-51,16	0,017797	0,004114	0,011401	0,014457	11,28271

Fonte: Autoria própria (2021).

4.4.3 Processamento dos dados

As análises dos dados foram realizadas no aplicativo de código aberto *Jupyter notebook*®, na linguagem *Python*®, utilizando diversas bibliotecas disponibilizadas gratuitamente.

Para dar início ao desenvolvimento do código, primeiramente foi carregada a tabela de dados com as variáveis independentes R, RE, NIR e G (características) e a variável alvo SST para cada ponto de coleta, na extensão CSV utilizando a biblioteca *pandas* (Quadro 5).

Quadro 5: Tabela de dados com as características e a variável.

Ponto	R	RE	NIR	G	SST (mg/L)
1	0,03046	0,01110	0,01476	0,03445	0
2	0,03403	0,02826	0,02808	0,03852	5
3	0,03223	0,05086	0,03146	0,03542	10
4	0,02711	0,02656	0,02094	0,02840	0
5	0,02745	0,02008	0,02087	0,03021	5
6	0,02565	0,04233	0,01584	0,03532	25
7	0,02195	0,00851	0,00916	0,02890	20
8	0,02147	0,00772	0,00753	0,02873	15
9	0,02366	0,02382	0,01358	0,03463	5
10	0,02182	0,07228	0,01444	0,03346	10
11	0,02228	0,05080	0,01598	0,03035	5
12	0,01933	0,01518	0,01125	0,02854	15
13	0,02307	0,00845	0,01687	0,02657	10
14	0,01933	0,00742	0,00831	0,02802	5
15	0,01646	0,00746	0,00724	0,02359	5
16	0,01946	0,00819	0,00945	0,02926	10
17	0,01975	0,00925	0,00949	0,02900	10
18	0,02376	0,01250	0,01232	0,03248	15
19	0,01978	0,01026	0,01065	0,02776	20
20	0,02037	0,01675	0,01260	0,02662	10
21	0,01947	0,00968	0,01001	0,02837	10
22	0,01841	0,00874	0,01258	0,02693	50
23	0,01755	0,00784	0,00984	0,02525	5
24	0,01814	0,00772	0,00905	0,02720	5
25	0,01893	0,00819	0,00921	0,02744	10
26	0,03885	0,02461	0,02597	0,04041	15
27	0,04086	0,03035	0,03252	0,04752	10
28	0,04984	0,03194	0,03278	0,05279	10
29	0,04770	0,02899	0,03442	0,04889	10
30	0,04723	0,02739	0,03051	0,04662	15
31	0,05186	0,02781	0,03221	0,04914	10
32	0,03489	0,02353	0,02592	0,03809	15
33	0,06023	0,03522	0,03867	0,05583	5
34	0,05304	0,03475	0,03765	0,05531	25
35	0,04614	0,03312	0,03653	0,05201	20
36	0,05183	0,03306	0,03539	0,05210	15
37	0,05493	0,03607	0,03770	0,05275	30
38	0,05340	0,03667	0,03899	0,05233	20
39	0,05454	0,03699	0,04236	0,05951	25
40	0,04993	0,03316	0,03363	0,05325	10

Fonte: Autoria própria (2021).

Com os dados inseridos, foi realizada a análise exploratória dos dados utilizando funções da biblioteca *pandas*, a fim de conhecer quais eram as características dos dados, bem como a correlação entre eles e principalmente descobrir qual banda possui maior correlação com o SST. Correlação é a relação entre duas variáveis, a variável independente (x) e a variável dependente (y) (LARSON; FARBER, 2010). Neste estudo, as variáveis independentes foram os valores dos pixels de cada banda R, RE, NIR e G, já a variável dependente (variável alvo) é o valor de SST. O coeficiente de correlação é a forma de mensurar a força da correlação entre as variáveis e possui amplitude de -1 até 1. Quando o coeficiente de correlação se encontra próximo de -1, diz-se que as variáveis x e y possuem correlação linear negativa forte, quando próximos de 1, as variáveis x e y possuem correlação linear positiva forte, e por fim, se o valor estiver próximo de 0 significa que não existe uma correlação linear entre as variáveis x e y (LARSON; FARBER, 2010).

Os *outliers* no banco de dados foram identificados utilizando os métodos de *Tukey* e *Z-score*. Porém, primeiramente foram realizadas as análises exploratórias com todos os dados e, em seguida, foram excluídos todos os pontos *outliers* do banco de dados, a fim de comparar os resultados obtidos.

Feito isso, os algoritmos de AM propostos foram aplicados utilizando a biblioteca *scikit-learn* que integra uma ampla quantidade de algoritmos de última geração para análises estatísticas através de códigos compilados (PEDREGOSA *et al.*, 2011). Portanto, nesta etapa foram utilizados os modelos Regressão Linear, Regressão de Lasso, Regressão de Ridge, Regressão ElasticNet e Floresta Aleatória, que estão integrados na biblioteca supracitada. Foram realizados testes para todos os modelos de AM com duas configurações diferentes, primeiramente considerando 80% dos pontos para treinamento e 20% dos pontos para teste do modelo, na segunda configuração foram considerados 70% dos pontos para treinamento e 30% para teste. Vale salientar que os pontos foram divididos de forma randômica para a etapa de treinamento e de teste.

A capacidade de predição dos modelos de AM testados foram avaliados pelos valores do Erro quadrático médio (EQM) e pelo coeficiente de determinação R-quadrado (R^2), sendo considerado o melhor modelo aquele que possuir maior valor de R^2 e menor valor de EQM (GUIMARÃES, 2019; KUPSSINSKÜ *et al.*, 2020). As estatísticas R^2 e EQM podem ser obtidas através das Equações 4 e 5:

$$R^2 = 1 - \frac{\sum(y_i - y'_i)^2}{\sum(y_i - \bar{y})^2} \quad (4)$$

$$EQM = \frac{1}{n} \sum_{i=0}^{n-1} (y_i - y'_i)^2 \quad (5)$$

em que, y_i representa o valor observado, y'_i o valor predito, \bar{y} a média das observações e n o número de amostras.

Segundo Sousa e Sousa (2010), R^2 é uma medida da proporção da variabilidade de uma variável que é explicada pela variabilidade de outra, e que é pouco provável que haja uma correlação perfeita ($R^2 = 1$). Optou-se por utilizar esses modelos de avaliação de predição, que foi utilizado com sucesso por Kupssinskü *et al.* (2020) e que serve como base metodológica para este estudo. O coeficiente de determinação R^2 também foi utilizado por Guimarães (2019), Bouzeria, Ghenim e Khanchoul (2017), Sousa e Sousa (2010) e o EQM por Cigizoglu (2002).

O objetivo específico desta etapa foi identificar o melhor modelo de AM capaz de prever as concentrações de SST nos lagos Igapó I e II a partir das informações obtidas pelo sensoriamento remoto e o SST mensurado em laboratório.

No entanto, percebeu-se que o banco de dados era pequeno, o que poderia não trazer bons resultados. Segundo Müller e Guido (2016), a complexidade de um modelo está ligada proporcionalmente a quantidade de dados de entrada. Portanto, conjuntos de dados maiores permitem a construção de modelos mais complexos. Por esse motivo, optou-se por fazer uma interpolação dos valores de SST mensurados em laboratório.

Deste modo, todo o processo supracitado foi repetido utilizando os dados da interpolação, com 16.966 linhas e 5 colunas, onde as linhas representam todos os pontos interpolados e as colunas representam os valores de pixels de cada banda (R, RE, G, NIR) e o TSS relativo a cada ponto interpolado.

5 RESULTADOS E DISCUSSÃO

5.1 Resultado da amostragem de água

Como resultado da coleta de amostras d'água *in situ* obteve-se os valores de pH, CE, Temperatura, Concentração de OD, %OD, Salinidade, Pressão Atmosférica, Turbidez, SDT e SST. Como o valor de SST foi medido em duplicata, foi considerado o valor médio de SST para cada ponto amostral (Quadro 6). O valor de CE foi expresso em Microsiemens por Centímetro ($\mu\text{S}/\text{cm}$), a Temperatura em Graus Celsius ($^{\circ}\text{C}$), a Concentração de OD em Miligramas por Litro de OD (mg/L de OD), a Salinidade em Unidade de Salinidade Prática (PSU), a Pressão Atmosférica em Libras por Polegadas Quadradas (PSI), a Turbidez em Unidade Nefelométrica de Turbidez (NTU) e por fim, o SST e o SDT foram expressos em Miligramas por Litro (mg/L).

Quadro 6: Resultados dos parâmetros analisados para cada ponto amostral.

Ponto de coleta	pH	Cond. elétrica (µS/cm)	Temp. (°C)	%OD	Conc. de OD (mg/L de OD)	Salinid. (PSU)	Press. Atm (Psi)	Turb. (NTU)	SDT (mg/L)	SST (mg/L)
1	6,28	110	28,14	88,7	6,46	0,05	13,971	3,21	55	0
2	6,86	109	28,13	84,8	6,26	0,05	13,971	2,05	54	5
3	6,91	108	28,20	85,0	6,29	0,05	13,971	2,46	54	10
4	7,02	108	27,72	82,1	6,12	0,05	13,978	3,51	54	0
5	7,02	107	27,96	83,7	6,20	0,05	13,979	2,06	54	5
6	7,01	104	28,13	84,5	6,25	0,05	13,983	2,44	52	25
7	7,07	101	28,15	85,4	6,32	0,05	13,993	2,91	50	20
8	7,02	104	28,19	86,1	6,38	0,05	14,000	3,85	52	15
9	7,19	101	28,5	92,4	6,81	0,05	14,004	2,90	51	5
10	7,07	100	28,53	93,3	6,88	0,05	14,013	1,08	50	10
11	7,10	97	28,60	92,9	6,83	0,04	14,025	1,26	49	5
12	7,13	98	28,64	93,8	6,92	0,04	14,075	1,39	49	15
13	7,06	98	28,69	92,9	6,85	0,04	14,092	1,80	49	10
14	7,16	98	28,74	95,1	7,01	0,04	14,116	1,83	49	5
15	7,18	99	28,90	97,3	7,17	0,05	14,147	1,69	50	5
16	7,15	102	29,13	100,3	7,33	0,05	14,168	2,36	51	10
17	7,24	102	29,27	99,0	7,28	0,05	14,184	2,51	51	10
18	7,23	102	29,38	100,5	7,38	0,05	14,200	1,71	51	15
19	7,23	96	29,28	99,4	7,27	0,04	14,111	1,69	48	20
20	7,33	97	29,48	93	6,77	0,04	14,108	1,50	48	10
21	7,34	99	29,18	95,5	7,02	0,04	14,135	1,93	49	10
22	7,24	97	28,92	95,3	7,06	0,04	14,186	2,04	49	50
23	7,29	96	29,08	97,2	7,21	0,04	14,210	2,51	48	5
24	7,41	97	29,36	93,8	6,91	0,04	14,234	2,40	48	5
25	7,37	99	29,44	97,8	7,21	0,05	14,243	2,24	50	10
26	8,37	118	28,98	111,6	8,13	0,05	13,952	4,76	59	15
27	8,17	106	27,88	110,8	8,21	0,05	13,934	5,54	53	10
28	8,05	104	27,37	110,5	8,25	0,05	13,917	6,49	52	10
29	7,80	105	27,82	113,2	8,33	0,05	13,904	5,34	53	10
30	7,87	106	27,79	109,0	8,07	0,05	13,893	4,55	53	15
31	7,71	105	27,40	99,6	7,47	0,05	13,878	4,52	53	10
32	8,05	105	27,08	106,3	7,95	0,05	13,867	5,26	52	15
33	7,29	103	26,81	97,9	7,36	0,05	13,861	5,27	52	5
34	6,83	108	26,52	90,9	6,85	0,05	13,862	6,97	54	25
35	8,20	103	27,02	107,6	8,05	0,05	13,865	5,87	52	20
36	7,71	107	27,50	100,3	7,51	0,05	13,866	5,36	53	15
37	7,84	106	27,08	104,3	7,78	0,05	13,860	6,27	53	30
38	7,55	100	27,03	103,2	7,71	0,05	13,861	8,68	50	20
39	8,41	102	27,10	106,1	7,95	0,05	13,858	6,24	51	25
40	8,4	103	27,26	111,2	8,28	0,05	13,856	4,85	51	10

Fonte: Autoria própria (2021).

Foram calculados os valores máximo, mínimo, médio e o desvio padrão dos parâmetros analisados das amostras de água (Tabela 4).

Tabela 4: Valores máximo, mínimo, médio e o desvio padrão dos parâmetros analisados.

Parâmetro	Unidade	Valor mínimo	Valor máximo	Valor médio	Desvio Padrão
SST	mg/L	0,00	50,00	12,88	9,26
SDT	mg/L	48,00	59,00	51,40	2,34
Temperatura	C	26,52	29,48	28,91	0,85
Turbidez	NTU	1,08	8,68	3,53	1,92
Condutividade elétrica	μS/cm	96,00	118,00	102,75	4,67
Concentração de OD	mg/L de OD	6,12	8,33	7,20	0,66
%OD	%	82,10	113,20	97,31	8,68
Salinidade	PSU	0,04	0,05	0,05	0,00
Pressão atmosférica	PSI	13,86	14,24	14,01	0,12
pH	-	6,28	8,41	7,40	0,49

Fonte: Autoria própria (2021).

Os dados obtidos serviram de base para uma avaliação sobre a qualidade da água dos lagos Igapó I e II, e principalmente, para a geração de modelos de predição utilizando o valor de SST, juntamente com as imagens capturadas pelo VANT.

Para a análise da qualidade da água, foi considerada a Resolução CONAMA 357/2005, que dispõe sobre a classificação dos corpos de água e diretrizes ambientais para o seu enquadramento, bem como estabelece as condições e padrões de lançamento de efluentes, e dá outras providências (BRASIL, 2022).

De acordo com a portaria de enquadramento dos cursos d'água do estado do Paraná do IAT, que os classifica seguindo a Resolução CONAMA 20/1986, a microbacia hidrográfica do Ribeirão Cambé, é classificada como Água Doce Classe 1 (IAT, 2020). Portanto, os parâmetros de qualidade da água analisados foram comparados com os padrões estabelecidos na Resolução CONAMA 357/2005 para Águas Doces de Classe 1.

Os corpos hídricos considerados Classe 1, podem ser destinados ao consumo humano após tratamento simplificado, à recreação de contato primário (natação, esqui aquático e mergulho), à irrigação de hortaliças e frutas que são consumidas cruas, à proteção de comunidades aquáticas, bem como à proteção de comunidades aquáticas presente em Terras Indígenas (BRASIL, 2022).

De acordo com a porcentagem de salinidade, disposta na Resolução CONAMA, os lagos Igapó I e II se enquadram na classificação de água doce (valor médio de a 0,5 PSU).

O valor médio de pH de 7,40 está contido no valor estabelecido pela Resolução CONAMA 357/2005, que é definido por pH entre 6 e 9. A turbidez média de 3,53 NTU está dentro do valor estabelecido de até 40 NTU. A concentração de OD médio de 7,20 mg/L de OD, atende o exigido na resolução que estabelece que em qualquer amostra, o valor não seja inferior a 6 mg/L de OD. O valor médio de SDT foi de 51,40 mg/L e está contido no valor exigido pela resolução CONAMA que estabelece o valor máximo de 500 mg/L (BRASIL, 2022).

Deste modo, foi possível constatar que os lagos Igapó I e II, objetos deste estudo, possuem parâmetros de qualidades de acordo com os padrões exigidos pela Resolução CONAMA 357/2005 para a Água Doce Classe 1.

5.2 Resultado da aplicação de técnicas de AM

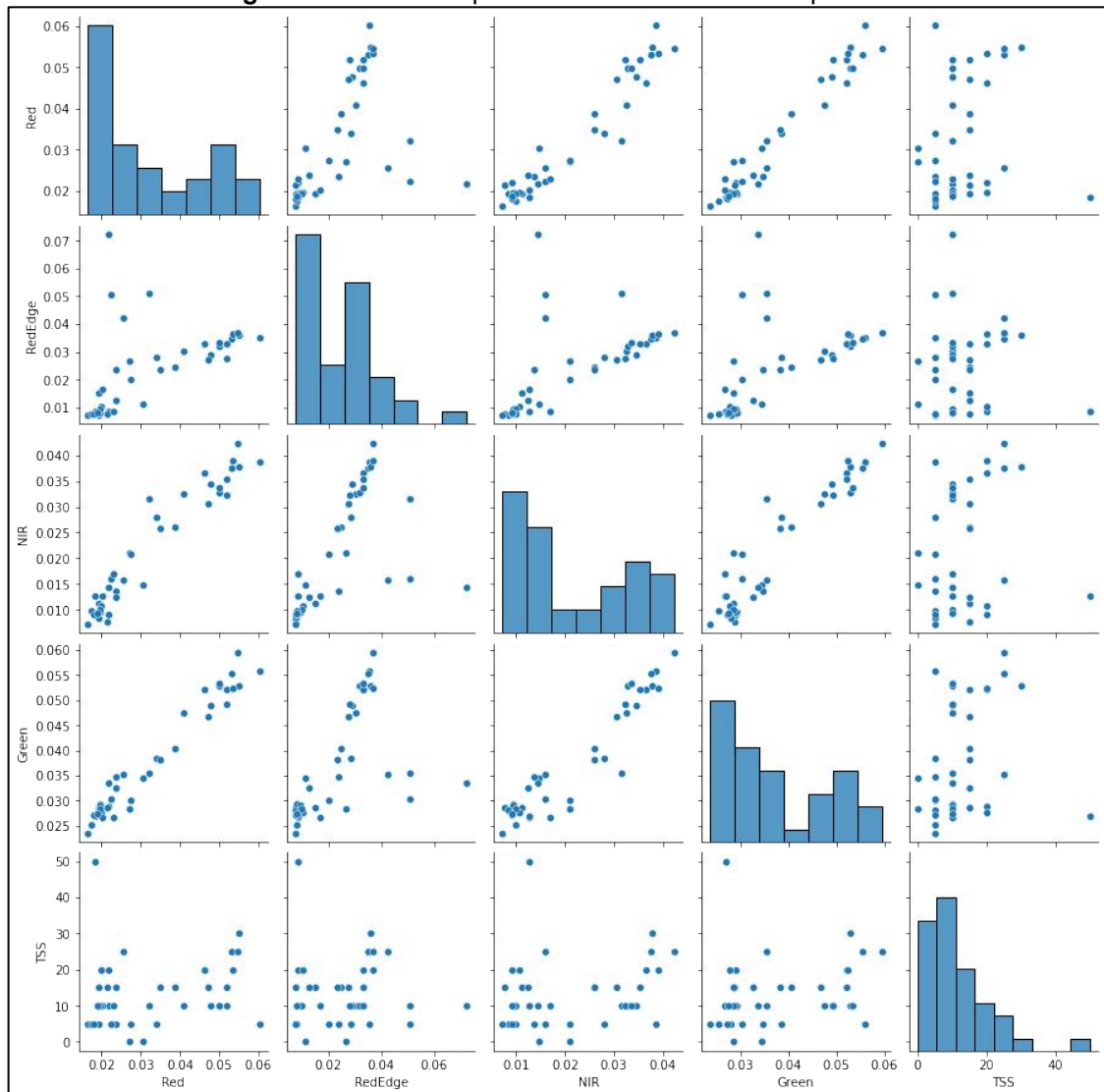
5.2.1 Utilizando os pontos medidos em campanha de campo

Os modelos de AM foram testados em diferentes configurações. Cada configuração utilizada em cada modelo de AM, assim como os resultados obtidos, serão detalhados a seguir. Toda essa etapa de análise exploratória de dados foi realizada utilizando a biblioteca *scikit-learn*. Dentro da análise exploratória de dados, é possível observar a correlação entre as variáveis em estudo, como será visto.

Primeiramente foram utilizados apenas os pontos amostrados (40 pontos), com os valores de pixels de todas as bandas e o resultado de SST (vide Quadro 5). Assim foi possível realizar a análise exploratória dos dados e observar a correlação entre as bandas e o SST (Figuras 17 e 18).

Figura 17: Correlação entre as bandas e o SST com 40 pontos.

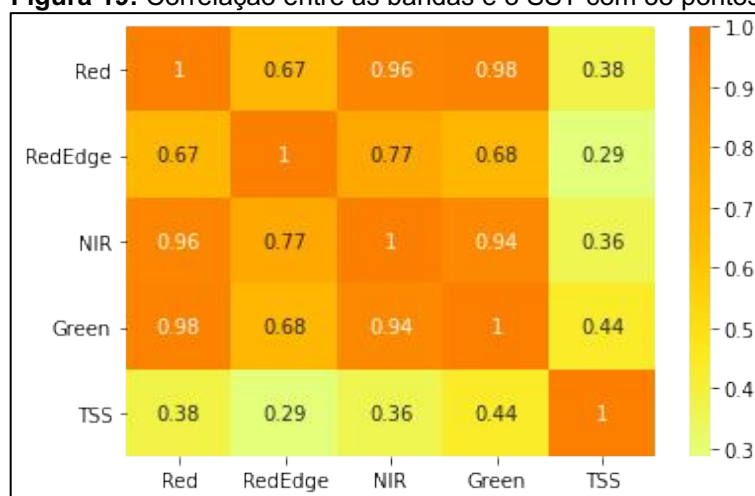
Fonte: Autoria própria (2021).

Figura 18: Análise exploratória dos dados com 40 pontos.

Fonte: Autoria própria (2021).

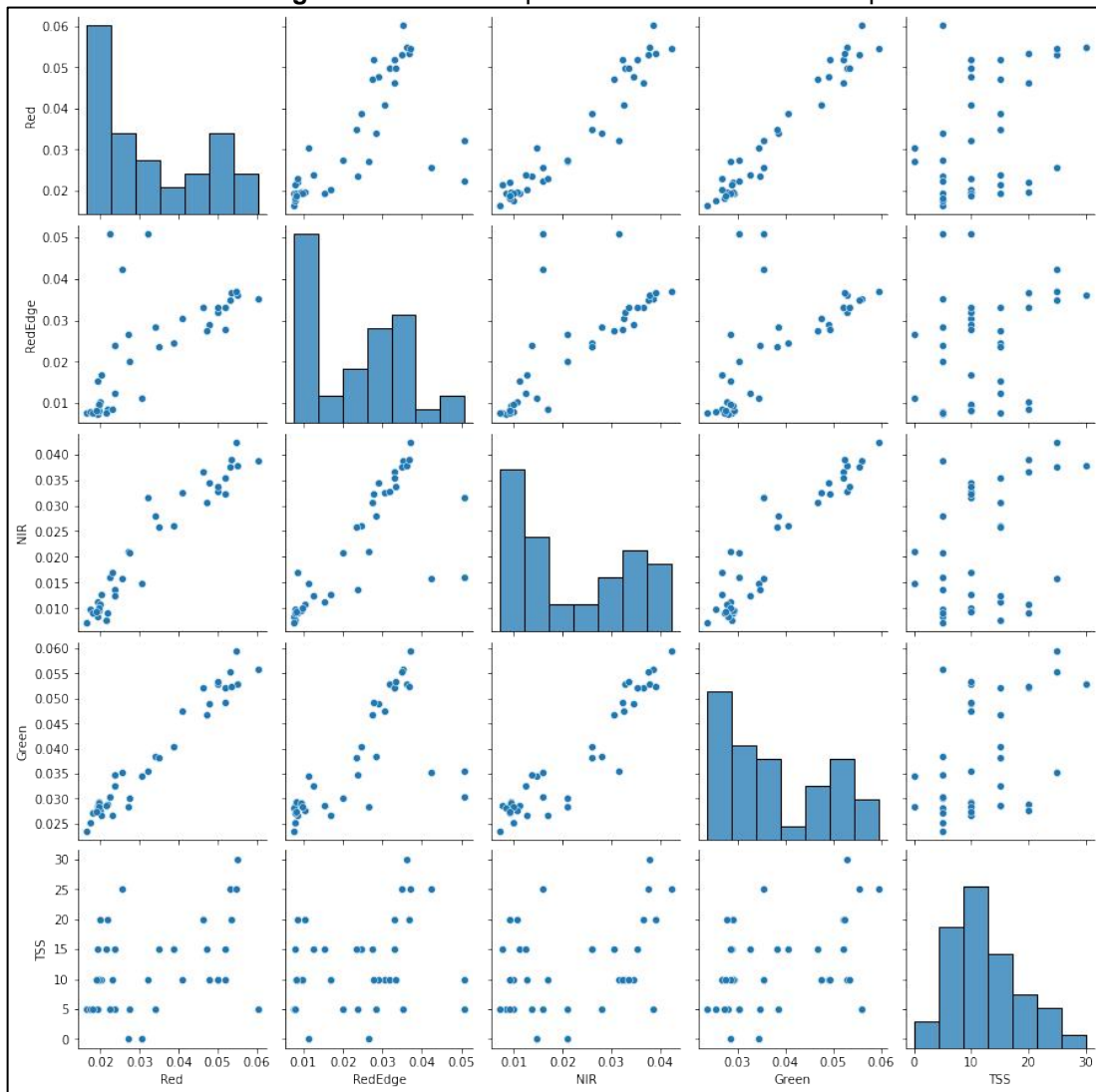
A análise exploratória foi realizada através do cálculo do coeficiente de correlação linear de Pearson (R_{xy}) entre os pares de bandas/SST. Os pontos correspondentes a esses pares de variáveis também foram colocados em um diagrama de dispersão, cuja observação já permite uma ideia preliminar da relação entre as variáveis. Observando a Figura 17 podemos constatar que as bandas G e NIR são as que possuem maior correlação com o SST, tendo uma correlação de 0,23 e 0,20, respectivamente. No entanto, este valor nos mostra que as correlações lineares entre as variáveis não são fortes, pois este valor encontra-se longe do valor ideal 1 (correlação linear positiva forte). Também é possível observar na Figura 18 uma forte correlação entre as bandas, pois os pontos nos gráficos da coluna 1 e linhas 2, 3 e 4 estão agrupados linearmente. No entanto, quando observamos o gráfico das bandas com o SST não há esse agrupamento linear visivelmente destacado, o que corrobora com os valores encontrados na Figura 17. Ainda, na Figura 18 é possível observar que existem pontos fora do padrão (*outliers*). Por exemplo, na linha do SST podemos observar pontos que estão muito distantes dos demais. O provável motivo dos resultados do coeficiente de correlação não demonstrarem fortes correlações entre as variáveis pode estar relacionado com os pontos *outliers*. Por este motivo, através dos testes de *Tukey* e *Z-Score* foi possível descobrir quais eram os pontos *outliers*, que foram removidos. Após a remoção dos *outliers* foram realizadas nova análise exploratória de dados e novo teste de correlação (Figuras 19 e 20).

Figura 19: Correlação entre as bandas e o SST com 38 pontos.



Fonte: Autoria própria (2021).

Figura 20: Análise exploratória dos dados com 38 pontos.



Fonte: Autoria própria (2021).

Observando a Figura 19 pode-se constatar que nesta etapa de análise, sem os pontos *outliers*, as bandas com maior correlação com o SST, passaram a ser as bandas G e R, com valor do coeficiente de correlação de 0,44 e 0,38, respectivamente, indicando que houve um aumento nos valores do coeficiente, devido à remoção dos *outliers*.

Para dar prosseguimento aos testes, foram aplicados os modelos de AM, em ambos os casos (com 40 pontos e com 38 pontos). Foram testados os modelos FA, RL, Regressão de Lasso, Regressão de Ridge e Regressão ElasticNet considerando duas configurações de conjunto de dados.

Por fim, e com o intuito de descobrir com qual modelo e qual configuração (80-20% ou 70-30%) seria capaz de trazer melhores resultados, foram mensurados os valores de R^2 e EQM, utilizando o valor de pixel de todas as bandas (G, R, RE e NIR),

o valor do SST e variando a quantidade de pontos, primeiramente com 40 pontos (Quadro 7) e depois, removendo os dois pontos *outliers* (Quadro 8).

Quadro 7: Resultados de R^2 e EQM para cada teste utilizando 40 pontos, o valor de pixel das bandas G, R, RE e NIR e o valor do SST.

Treino - Teste	80% - 20%		70% - 30%	
Métricas Estatísticas	R^2	EQM	R^2	EQM
Floresta Aleatória	-1,79443	76,1485	-3,96543	43,96479
Regressão Linear	-0,43006	38,96910	-3,92166	43,57728
Regressão de Lasso	-0,05752	57,10652	-0,00584	152,79728
Regressão de Ridge	-0,04689	56,53231	-0,00599	152,82039
Regressão ElasticNet	-0,05507	56,97403	-0,00586	152,80034

Fonte: Autoria própria (2021).

Observando o Quadro 7 e comparando os valores de R^2 para os diferentes modelos testados, pode-se dizer que, com 40 pontos, os métodos Regressão de Lasso, Regressão ElasticNet e Regressão de Ridge, possuem os melhores valores para R^2 quando utilizado 70% dos dados para treino e 30% dos dados para teste, possuindo os respectivos valores: -0,00584; -0,00586; e -0,00599. Considerando que o R^2 é uma medida da proximidade dos dados com o modelo ajustado, não se pode afirmar que foi um bom resultado, já que os valores encontrados foram negativos e próximos de zero. Do mesmo modo, analisando o EQM para todos os modelos testados, os valores encontrados foram muito altos, mostrando que os mesmos não fizeram boas predições.

Quadro 8: Resultados de R^2 e EQM para cada teste utilizando 38 pontos, o valor de pixel das bandas G, R, RE e NIR e o valor do SST.

Treino - Teste	80% - 20%		70% - 30%	
Métricas Estatísticas	R^2	EQM	R^2	EQM
Floresta Aleatória	0,27588	29,1455	-0,43809	91,87854
Regressão Linear	-0,13447	45,6625	-0,21588	77,68095
Regressão de Lasso	-0,02368	24,56832	-0,05619	43,27473
Regressão de Ridge	0,00004	23,99905	-0,06844	43,77640
Regressão ElasticNet	-0,07275	25,74617	-0,06779	43,75000

Fonte: Autoria própria (2021).

Analisando o Quadro 8, com os testes realizados com a remoção dos dois pontos *outliers*, mas mantendo todas as bandas (G, R, RE e NIR), observa-se uma melhora no valor dos parâmetros analisados. Neste caso, o melhor resultado, de acordo com o valor de R^2 e MSE, foi o teste de FA com 80% dos dados para treino e

20% para teste. Observa-se que o valor do R^2 deixou de ser negativo e subiu para 27,58%, enquanto o EQM diminuiu para 29,15. Porém, do ponto de vista de aplicação do modelo na predição de SST, não é possível afirmar que o resultado seja satisfatório, porque ainda temos um valor baixo para R^2 e um valor alto de EQM.

Repetindo os testes, agora considerando apenas as bandas G e R, e ainda desconsiderando os pontos *outliers* (vide Figura 19), observa-se mudança no desempenho dos modelos ajustados. De acordo com os resultados apresentados no Quadro 9, os valores de R^2 são positivos apenas para FA e RL. Neste caso, o melhor resultado foi obtido para RL, na configuração com 70% dos dados para treino e 30% para teste. Assim, o valor de R^2 (28,53%) foi ligeiramente maior do que o melhor resultado encontrado anteriormente, ao utilizar os valores de pixel de todas as bandas, enquanto o EQM (37,09) foi um pouco acima.

Quadro 9: Resultados de R^2 e EQM para cada teste utilizando 38 pontos, o valor do pixel das bandas G e R e o SST.

Treino - Teste	80% - 20%		70% - 30%	
Métricas Estatísticas	R^2	EQM	R^2	EQM
Floresta Aleatória	0,21807	44,76549	0,16484	43,35249
Regressão Linear	0,13638	49,44243	0,28535	37,09723
Regressão de Lasso	-0,69009	51,12532	-0,62484	55,00770
Regressão de Ridge	-0,49402	45,19438	-0,64104	55,55631
Regressão ElasticNet	-0,56206	47,25238	-0,64090	55,55131

Fonte: Autoria própria (2021).

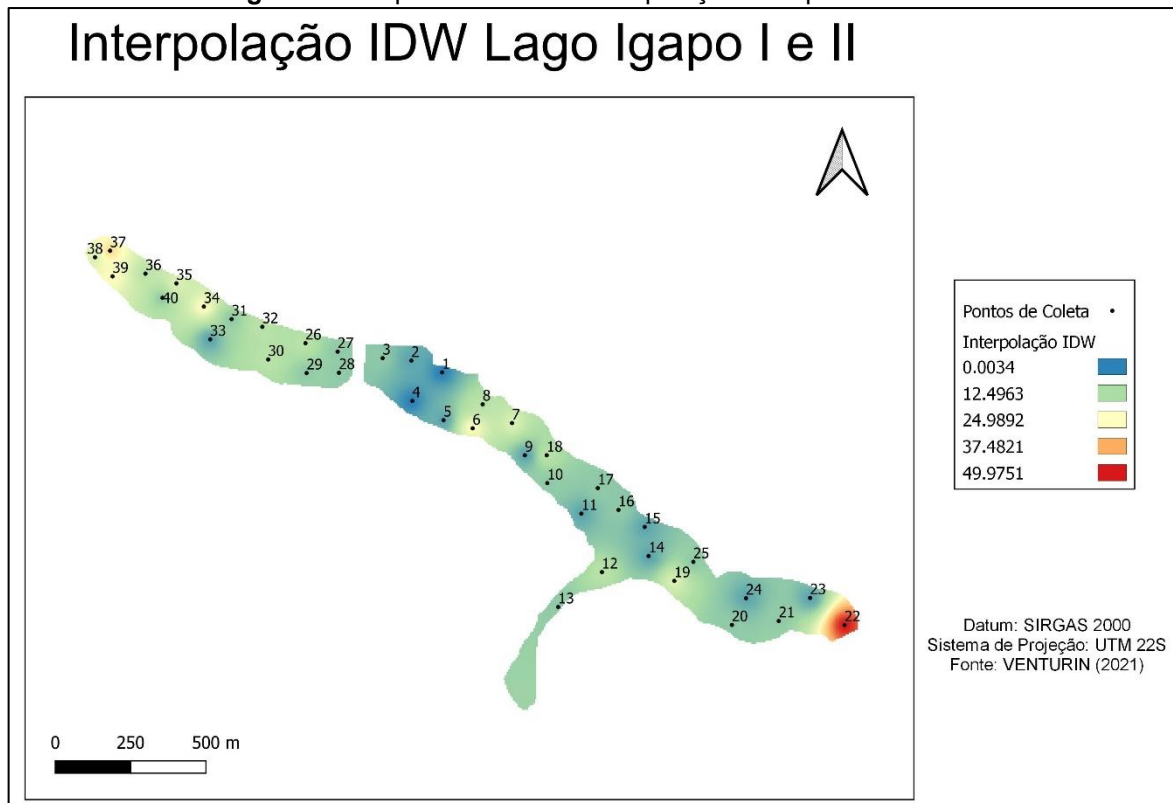
De um modo geral, em relação aos testes utilizando os pontos medidos *in situ*, pode-se observar que as bandas não possuem forte correlações com o SST, pois as correlações ficaram com valores abaixo de 20% para a maioria dos modelos testados. Entretanto, pode ser afirmado que o valor do coeficiente de correlação melhorou quando os *outliers* foram removidos, sugerindo uma maior atenção em relação aos pontos que efetivamente devem fazer parte da amostra. Pode-se observar ainda que os testes FA, RL, Regressão de Ridge, Regressão de Lasso e Regressão ElasticNet não apresentaram bons resultados, pois os valores de R^2 foram baixos e os valores de EQM foram altos. Uma das possibilidades é de que o baixo número de dados amostrais possa comprometer o ajuste. Por este motivo, optou-se por realizar a interpolação IDW para possibilitar uma maior quantidade de dados.

5.2.2 Utilizando os pontos interpolados e as bandas G, R, RE e NIR

O resultado da interpolação por IDW das medidas de SST para os lagos Igapó I e II está mostrado na Figura 21. Foram obtidos 16.966 pontos que compreendem toda a área de estudo. É possível notar no mapa (Figura 21) que as concentrações de SST, embora não apresentem homogeneidade espacial, também não apresentam padrões de predominância para qualquer dos lagos (Lagos Igapó I e II) ou partes específicas. A exceção é para um único ponto localizado próximo à barragem (Ponto 22), que apresentou valores muito acima da maioria das amostras.

Para dar início a etapa de análise exploratória dos dados e observar a correlação das bandas com o SST, foram utilizados todos os 19.966 pontos gerados na interpolação, que corresponde ao número de pixels de cada banda (G, R, RE e NIR). Os valores dos coeficientes de correlação entre os parâmetros analisados, bem como os correspondentes gráficos de dispersão, estão mostrados nas Figuras 22 e 23.

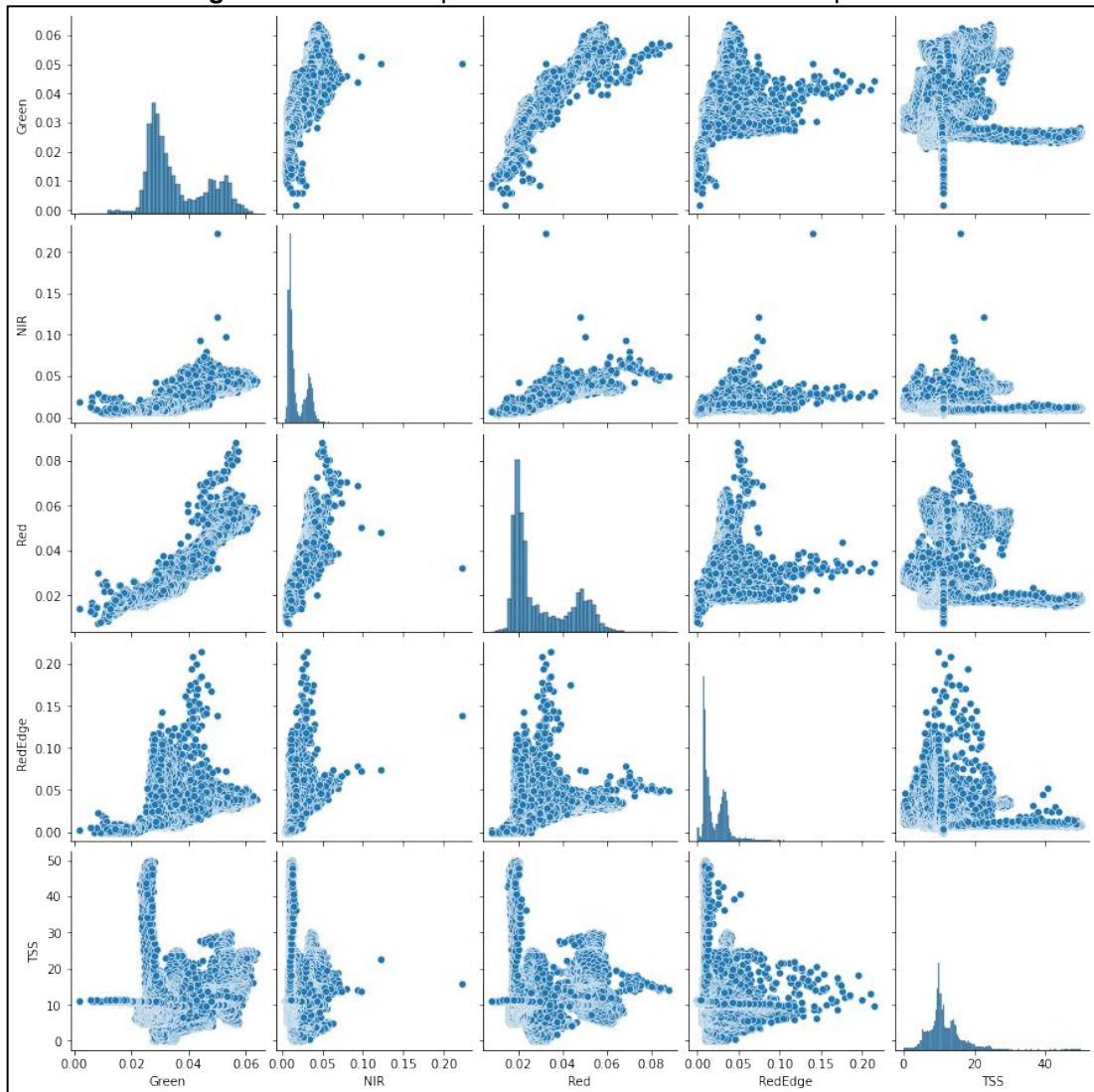
Figura 21: Mapa resultante da interpolação IDW para os SST.



Fonte: Autoria própria (2021).

Figura 22: Correlação entre as bandas e o SST com 16.966 pontos.

Fonte: Autoria própria (2021).

Figura 23: Análise exploratória dos dados com 16.966 pontos.

Fonte: Autoria própria (2021).

Analisando a Figura 22, pode-se observar que as bandas com maior correlação com o SST são as bandas G, R e NIR com valores de 0,21, 0,20 e 0,20, respectivamente. Os valores estão bem abaixo de 1,0 e, portanto, não mostram uma correlação linear positiva forte com o SST, já que a correlação é considerada forte quanto mais seu coeficiente se aproxima de 1 (LARSON; FARBER, 2010).

Na Figura 23 está representada a análise exploratória dos dados e pode-se observar que existem pontos fora do padrão (*outliers*). Por exemplo, na coluna do SST (coluna 5) e nas linhas do G, NIR, R e RE (linhas 1, 2, 3 e 4, respectivamente), é notória a existência de alguns pontos que estão isolados do agrupamento da maioria dos pontos. Um dos prováveis motivos para os resultados do coeficiente de correlação não demonstrarem fortes correlações entre as variáveis x e y , podem estar relacionados com a existência de *outliers*.

Para dar sequência às análises, foram aplicados os testes de *Tukey* e do Teste Z-Score. Como resultado dos testes foi possível descobrir quais eram os pontos *outliers*. Observa-se que os testes não concordaram entre si em relação ao número de pontos *outliers*. No teste de *Tukey* foram constatados 1.598 pontos *outliers*, restando, portanto, 15.368 pontos para análise. No teste de Z-Score, constatou-se 693 pontos *outliers*, restando 16.273 pontos para serem analisados.

Após a remoção dos *outliers* foram realizadas nova análise exploratória de dados e nova análise de correlação (Figuras 24, 25, 26 e 27).

Figura 24: Correlação entre as bandas e o SST com 15.368 pontos (*Tukey*).



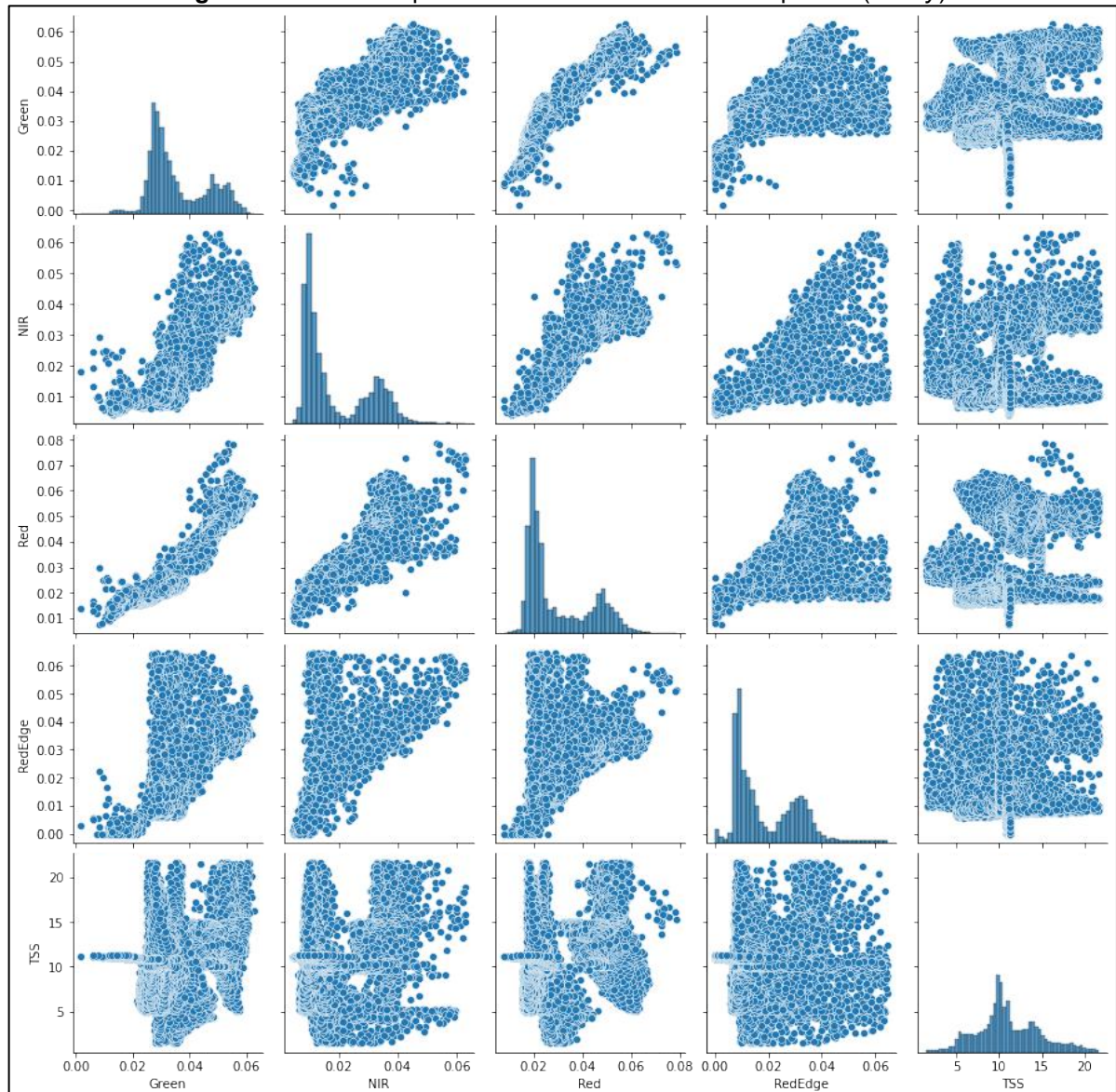
Fonte: Autoria própria (2021).

Figura 25: Correlação entre as bandas e o SST com 16.273 pontos (Z-Score).



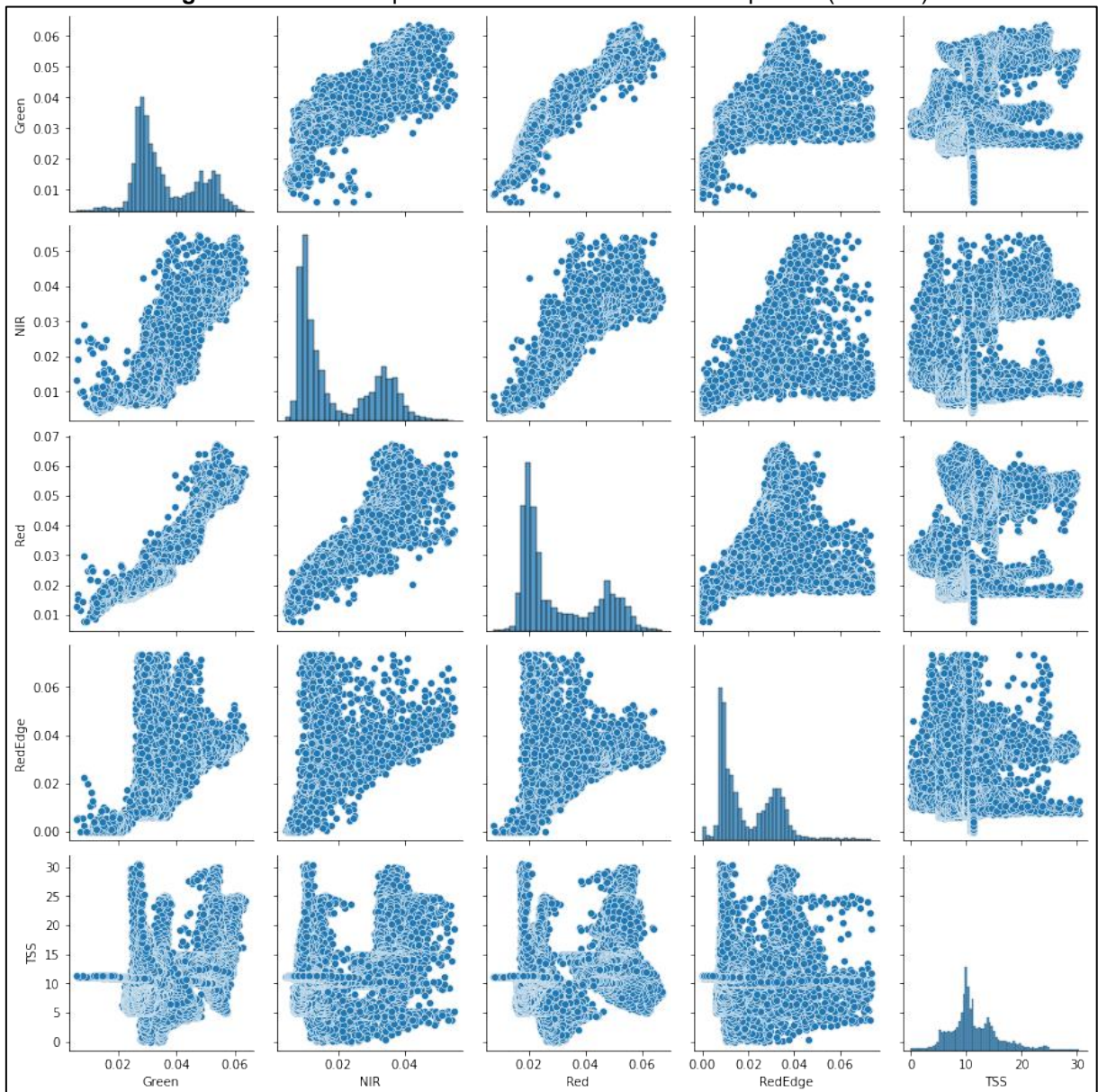
Fonte: Autoria própria (2021).

Figura 26: Análise exploratória dos dados com 15.368 pontos (Tukey).



Fonte: Autoria própria (2021).

Figura 27: Análise exploratória dos dados com 16.273 pontos (Z-Score).



Fonte: Autoria própria (2021).

Analisando a Figura 24, podemos observar que, no modelo em que foram excluídos os *outliers* resultantes do teste de *Tukey* (utilizando 15.368 pontos), as bandas que apresentam maior correlação com o SST são as bandas G e R com coeficiente de correlação de 0,33 e 0,32 respectivamente.

Da mesma forma, analisando a Figura 25, no modelo em que foram excluídos os *outliers* resultantes do teste de *Z-Score* (utilizando 16.273 pontos), as bandas que possuem maior correlação com o SST também são as bandas G e R, com coeficiente de correlação no valor de 0,39 e 0,37, respectivamente.

Os resultados apresentados sugerem que o melhor método para a remoção de *outliers*, de acordo com os dados disponíveis e com base no coeficiente de

correlação, é o método de *Z-Score*. Por outro lado, fazendo uma comparação entre os valores obtidos para o coeficiente de correlação, com 16.273 pontos (Figura 25) e com 38 pontos (vide Figura 19), pode-se observar que não houve melhora nos resultados, já que o coeficiente de correlação passou de 0,44 para 0,33 para a banda G e de 0,38 para 0,37 na banda R.

Deste modo, em ambos os casos não se pode afirmar que houve uma correlação linear positiva forte entre essas bandas e o SST (Larson e Farber, 2010). No entanto, comparando o teste utilizando todos os 16.966 pontos (vide Figura 19), com o teste utilizando os 16.273 pontos, quando os *outliers* identificados no teste de *Z-Score* foram removidos das análises (vide Figura 25), podemos dizer que o coeficiente de correlação teve uma mudança positiva em seu valor, passando de 20% para cerca de 40%.

Podemos notar, com base nos dados das Figuras 25 e 26, bem como nos gráficos de dispersão mostrados nas Figuras 26 e 27, que há uma forte correlação linear com as bandas entre si, diferentemente do que foi observado quando se compara as bandas (linhas 1, 2, 3 e 4) com o SST (coluna 5). Isso sugere que, embora as respostas espectrais de cada banda potencialmente representem características distintas da superfície, isso não se manifesta na totalidade da radiância medida. Muito provavelmente as diferenças estão em uma fração menor das radiâncias capturadas em cada banda, o que faz que, na totalidade, sejam semelhantes.

Para dar continuidade aos testes, em ambos os casos (com 15.368 e 16.273 pontos), foram realizados testes com os modelos de FA e RL, considerando primeiramente o conjunto de dados com 80% dos pontos para treinamento do modelo e 20% dos pontos para teste do modelo. Também foram realizados testes na forma 70% dos pontos para treinamento e 30% para teste. Vale salientar que os pontos foram divididos de forma randômica para a etapa de treinamento e de teste.

Assim, com o intuito de descobrir com qual modelo (FA ou RL) e quais configurações (80-20% ou 70-30%) seriam capazes de trazer melhores resultados na estimativa de SST, foram mensurados os valores de R^2 e EQM tanto para a análise com 16.966 pontos (Quadro 10), quanto com 15.368 pontos (Quadro 11) e 16.273 pontos (Quadro 12).

Quadro 10: Resultados de R^2 e EQM para cada teste utilizando 16.966 pontos, o valor de pixel das bandas G, R, RE e NIR e o valor do SST.

Treino - Teste	80% - 20%		70% - 30%	
Métricas Estatísticas	R^2	EQM	R^2	EQM
Floresta Aleatória	0,54961	18,53825	0,52647	17,73445
Regressão Linear	0,03500	39,72019	0,05979	35,21255
Regressão de Lasso	0,05367	35,44983	0,03629	39,34246
Regressão de Ridge	0,05367	35,44983	0,03632	39,34141
Regressão ElasticNet	0,05367	35,44990	0,03634	39,34038

Fonte: Autoria própria (2021).

Observando os resultados apresentados no Quadro 10, considerando todos os pontos interpolados (16.966 pontos), o valor do pixel de todas as bandas (G, R, RE e NIR) e o valor de SST, podemos dizer que o melhor modelo testado é o FA com configuração 80% dos dados para treino e 20% para teste com valores de R^2 e EQM de 54,96% e 18,53, respectivamente. Se observarmos os valores obtidos na configuração 70%-30%, podemos notar que não há grande diferença entre métricas estatísticas. Já em relação aos demais modelos de AM, há uma grande diferença nos valores de R^2 , que ficou entre 3% e 5% nos métodos de RL, Regressão de Lasso, Regressão de Ridge e Regressão ElasticNet. Portanto, observa-se uma melhora expressiva em relação aos testes utilizando 38 pontos (R^2 de 27,58% e EQM de 29,14 - vide Quadro 8).

Da mesma maneira, analisaremos os resultados encontrados para R^2 e EQM quando utilizado 15.368 pontos (Quadro 11).

Quadro 11: Resultados de R^2 e EQM para cada teste utilizando 15.368 pontos, o valor de pixel das bandas G, R, RE e NIR e o valor do SST.

Treino - Teste	80% - 20%		70% - 30%	
Métricas Estatísticas	R^2	EQM	R^2	EQM
Floresta Aleatória	0,62300	5,48537	0,61410	5,51627
Regressão Linear	0,11972	12,80808	0,12457	12,51370
Regressão de Lasso	0,13072	12,07290	0,11912	12,64810
Regressão de Ridge	0,13072	12,07289	0,11912	12,64809
Regressão ElasticNet	0,12453	11,89101	0,11912	12,64813

Fonte: Autoria própria (2021).

Observando os resultados dispostos no Quadro 11, que considera na análise apenas os pontos que restaram quando foram removidos os *outliers* do teste de *Tukey*

(15.368 pontos) o valor do pixel de todas as bandas (G, R, RE e NIR) e o valor de SST, podemos dizer que o melhor resultado foi apresentado pelo modelo FA. Com a configuração 80% dos dados para treino e 20% para teste, os resultados para R^2 e EQM foram de 62,30% e 5,48, respectivamente. Se observarmos os valores obtidos na configuração 70%-30%, podemos notar que não há grande diferença entre métricas estatísticas. Já em relação aos demais modelos de AM, há uma grande diferença nos valores de R^2 que ficou entre 11% e 13% nos métodos de RL, Regressão de Lasso, Regressão de Ridge e Regressão ElasticNet. Portanto, comparando os Quadros 10 e 11, podemos notar um aumento no valor de R^2 e diminuição do EQM quando os *outliers* identificados no teste de *Tukey* foram removidos da análise.

Na sequência, analisaremos os resultados encontrados para R^2 e EQM quando utilizado 16.273 pontos (Quadro 12).

Quadro 12: Resultados de R^2 e EQM para cada teste utilizando 16.273 pontos, o valor de pixel das bandas G, R, RE e NIR e o valor do SST.

Treino - Teste	80% - 20%		70% - 30%	
Métricas Estatísticas	R^2	EQM	R^2	EQM
Floresta Aleatória	0,59967	8,73393	0,58303	9,30702
Regressão Linear	0,15603	18,41266	0,16534	18,62983
Regressão de Lasso	0,16083	18,66213	0,16374	18,69292
Regressão de Ridge	0,16083	18,66213	0,16374	18,69291
Regressão ElasticNet	0,16083	18,66215	0,16374	18,69293

Fonte: Autoria própria (2021).

Analisando os resultados dispostos no Quadro 12, que considera na análise apenas os pontos que restaram quando foi removido os *outliers* do teste de *Z-Score* (16.273 pontos) o valor do pixel de todas as bandas (G, R, RE e NIR) e o valor de SST, podemos dizer que o melhor resultado foi apresentado pelo modelo FA na configuração 80% - 20%, pois resultou em um valor de R^2 de 59,96% e EQM de 8,73. Sendo assim, podemos afirmar que os valores de R^2 e EQM melhoram, em relação ao teste com todos os pontos, quando os *outliers* são removidos pelo método de *Z-Score*, já que o R^2 passou de 54,96% para 59,96%, enquanto o EQM passou de 18,53 para 8,73.

De modo geral, se compararmos os resultados obtidos nos Quadros 10, 11 e 12 podemos dizer que o melhor resultado foi obtido quando se considerou apenas os pontos que restaram ao remover os *outliers* determinado pelo teste de *Tukey*. Estes

resultados confirmam a hipótese de que o ajuste melhora com o aumento no número de dados amostrais (obtidos através da interpolação), bem como, com a remoção dos *outliers*.

5.2.3 Utilizando os pontos interpolados e as bandas G e R

Devido ao fato de todas as análises de correlação com os pontos interpolados apresentarem as bandas G e R como as que possuem maior correlação com o SST, foram realizados novos testes considerando apenas os valores de pixel dessas duas bandas e o valor de SST.

Assim, com o intuito de descobrir com qual modelo (FA ou RL) e quais configurações (80-20% ou 70-30%) seriam capazes de trazer melhores resultados, foram mensurados os valores de R^2 e EQM, variando a quantidade de pontos, utilizando todos os pontos interpolados (Quadro 13), os pontos restantes da remoção dos *outliers* resultante do teste de *Tukey* (15.368 pontos) (Quadro 14) e do teste de *Z-score* (16.273 pontos) (Quadro 15).

Quadro 13: Resultados de R^2 e EQM para cada teste utilizando 16.966 pontos, o valor do pixel das bandas G e R e o SST.

Treino - Teste	80% - 20%		70% - 30%	
	R^2	EQM	R^2	EQM
Floresta Aleatória	0,16290	30,14591	0,18648	31,59847
Regressão Linear	0,05600	33,99572	0,04039	37,27298
Regressão de Lasso	0,05153	34,63590	0,05429	34,31463
Regressão de Ridge	0,05153	34,63585	0,05429	34,31460
Regressão ElasticNet	0,05154	34,63579	0,05429	34,31461

Fonte: Autoria própria (2021).

Observando os resultados apresentados no Quadro 13, considerando todos os pontos interpolados (16.966 pontos), o valor do pixel das bandas G e R e o valor de SST, podemos dizer que o melhor modelo testado é o FA, na configuração 70% - 30%, pois resultou em um valor de R^2 de 18,64% e EQM de 31,59. Comparando este resultado com o resultado obtido utilizando todas as bandas nos testes (vide Quadros 10, 11 e 12), é possível notar uma grande diminuição no valor de R^2 e aumento no valor de EQM, o que permite concluir que ao utilizar apenas os valores de pixel das bandas G e R, os resultados são piores.

Da mesma maneira, analisaremos os resultados encontrados para R^2 e EQM quando utilizado 15.368 pontos (Quadro 14).

Quadro 14: Resultados de R^2 e EQM para cada teste utilizando 15.368 pontos, o valor do pixel das bandas G e R e o SST.

Treino - Teste	80% - 20%		70% - 30%	
Métricas Estatísticas	R^2	EQM	R^2	EQM
Floresta Aleatória	0,36270	9,06699	0,35670	9,01330
Regressão Linear	0,11581	12,57952	0,10899	12,48392
Regressão de Lasso	0,11335	12,27648	0,10045	12,67834
Regressão de Ridge	0,11340	12,27581	0,09996	12,68517
Regressão ElasticNet	0,11338	12,27604	0,09996	12,68518

Fonte: Autoria própria (2021).

Para os testes do Quadro 14 foram utilizados apenas os pontos que restaram ao retirar os *outliers* resultantes do teste de *Tukey*. Neste caso podemos dizer que os melhores resultados de R^2 e EQM foram obtidos utilizando o modelo FA na configuração 80% - 20%, pois resultou em um valor de R^2 de 36,27% e EQM de 9,06.

Apesar de os resultados deste teste (Quadro 14) serem melhores que o teste realizado com todos os pontos (Quadro 13), ainda é possível notar diminuições no valor de R^2 e aumento no valor de EQM quando comparados com os valores obtidos nos testes utilizando os valores de pixel de todas as bandas (vide Quadros 10, 11 e 12). Sendo assim, podemos dizer que ao utilizar apenas os valores de pixel das bandas G e R, obtivemos resultados piores do que quando utilizado o valor do pixel de todas as bandas (G, R, RE e NIR).

Por fim, observaremos os valores de R^2 e EQM encontrados nos testes utilizando apenas os pontos que restaram ao retirar os *outliers* resultantes do teste de *Z-Score* (Quadro 15).

Quadro 15: Resultados de R^2 e EQM para cada teste utilizando 16.273 pontos, o valor do pixel das bandas G e R e o SST.

Treino - Teste	80% - 20%		70% - 30%	
Métricas Estatísticas	R^2	EQM	R^2	EQM
Floresta Aleatória	0,38444	13,64148	0,59584	8,85546
Regressão Linear	0,15788	18,66229	0,15323	18,55357
Regressão de Lasso	0,13846	19,06298	0,14634	18,67319
Regressão de Ridge	0,13839	19,06456	0,14565	18,68816
Regressão ElasticNet	0,13795	19,07432	0,14554	18,69060

Fonte: Autoria própria (2021).

Podemos observar neste teste que os melhores resultados de R^2 e EQM foram obtidos com o modelo FA na configuração de 70% dos dados para treino e 30% dos dados para teste. O valor de R^2 neste caso foi de 59,58% e o EQM foi de 8,85, sendo estes valores os melhores encontrados utilizando apenas as bandas G e R. Podemos notar, que neste teste, onde utilizou-se os pontos que restaram ao remover os *outliers* determinado pelo teste de *Z-Score*, obtivemos o melhor resultado quando comparado com os testes com todos os 16.966 pontos e com 15.368 pontos.

Por fim, através dos resultados obtidos nos diferentes modelos e configurações, podemos dizer que os testes utilizando apenas os valores dos pixels das bandas com maior correlação com o SST, produzem resultados de R^2 e EQM, inferiores quando comparado com os testes utilizando os valores de pixel de todas as bandas.

5.3 Discussão dos resultados

Diante dos resultados apresentados anteriormente, podemos dizer que os melhores resultados encontrados de R^2 e EQM para cada um dos modelos de AM, foram encontrados quando se utilizou os pontos interpolados com a remoção dos *outliers* (Quadro 16).

Quadro 16: Melhores resultados encontrados de R^2 e EQM para cada modelo de AM, em ordem decrescente.

Métricas Estatísticas	R^2 (%)	EQM
Floresta Aleatória	62,30	5,48
Regressão Linear	16,53	18,62
Regressão de Lasso	16,37	18,69
Regressão de Ridge	16,37	18,69
Regressão ElasticNet	16,37	18,69

Fonte: Autoria própria (2021).

Dentre os melhores resultados, aquele que apresentou maior valor de R^2 e menor valor de EQM foi o modelo FA (62,30% e 5,48, respectivamente). Este resultado foi obtido ao se utilizar os pontos interpolados com a remoção de *outliers* através do teste de *Tukey*, ou seja, utilizando 15.368 pontos, o valor do pixel de todas

as bandas (G, R, RE e NIR) e os valores de SST, com configuração 80% dos dados para treino e 20% para teste.

Para os modelos de RL, Regressão de Lasso, Regressão de Ridge e Regressão ElasticNet, foram encontrados os melhores valores de R^2 e EQM ao utilizar os pontos interpolados com a remoção de *outliers* através do teste de Z-Score, ou seja, utilizando 16.273 pontos, o valor do pixel de todas as bandas (G, R, RE e NIR) e os valores de SST, com configuração 70% dos dados para treino e 30% para teste.

Na literatura foram encontrados poucos estudos que permitem alguma comparação. As especificidades do local de estudo (um lago artificial urbano), da metodologia (câmera a bordo de VANT, modelos de regressão) e parâmetro ambiental de interesse (SST) são fatores que explicam essa limitação.

Kupssinskü *et al.* (2020), em seu estudo para prever a concentração de SST e Clorofila-a, utilizou VANT e diferentes técnicas de AM dentre elas a RL, FA e a Regressão de Lasso. Para avaliar a capacidade de predição dos modelos de AM testados, foram mensurados os valores de R^2 e EQM, assim como feito neste estudo. Os melhores valores de R^2 e EQM encontrados por Kupssinskü *et al.* (2020), para os modelos FA, foram de 85,6% e 0,008, respectivamente; para RL foram de 22,38% e 0,046, respectivamente; e para Regressão de Lasso foram de 22,37% e 0,046, respectivamente. Portanto, os melhores valores de R^2 encontrados neste estudo diferem dos valores encontrados por Kupssinskü *et al.* (2020), já que, o melhor valor encontrado de R^2 foi de 62,30% e EQM de 5,48 ao utilizar o modelo FA (vide Quadro 16).

Dias *et al.* (2021) utilizaram sensores multiespectrais acoplados em VANT e diferentes modelos de AM, dentre eles o modelo FA, para estimar a concentração de SST em um reservatório e obteve R^2 igual a 87,50%. Este valor também é superior ao encontrado com FA neste estudo. Guimarães *et al.* (2019) utilizaram AR e RNAs para prever SST em um lago no Rio Grande do Sul, a partir de imagens de VANT. O melhor resultado de R^2 encontrado para a RL foi 16%. Este resultado é muito próximo o que foi obtido neste estudo, que foi de 16,53% (vide Quadro 16).

Considerando os três estudos encontrados na literatura, e que permitem comparação ao presente estudo, algumas diferenças podem ser destacadas e que podem explicar as diferenças observadas. No caso de Kupssinskü *et al.* (2020), observa-se que os autores analisaram um lago com área de 25.000 m², enquanto este estudo analisou o Lago Igapó I e II que possuem 420.345 m² e 171.631 m²,

respectivamente. Nota-se grande diferença entre as áreas dos objetos de estudo, o que pode ser uma justificativa para os resultados mais satisfatórios de R^2 e EQM encontrado por Kupssinskü et al. (2020). Já no estudo realizado por Dias *et al.* (2021), foram realizadas 4 campanhas de campo levando em consideração as diferentes estações do ano, o que de alguma forma produz maior diversidade de respostas espectrais. Guimarães *et al.* (2019) por sua vez, realizaram duas campanhas de campo e o objeto de estudo dos autores possui área de 25.000 m², porém, apesar dessas diferenças, os resultados de R^2 para RL se assemelham (16%).

6 CONCLUSÃO

Este trabalho se propôs à aplicação de técnicas de Aprendizado de Máquina (AM) e imagens multiespectrais de alta resolução espacial capturadas por VANT, para fins de estimativa da carga de SST nos lagos Igapó I e II, na cidade de Londrina.

Utilizando o *software Python*® e a biblioteca *scikit-learn*, foi possível realizar análises exploratórias de dados, observar a correlação entre as bandas e o valor de SST, bem como, gerar modelos de AM a fim de avaliar a capacidade de predição de cada modelo através do R^2 e do EQM.

Ao se tratar dos testes realizados com os 40 pontos medidos *in situ*, obtivemos que as bandas G e NIR são as que possuem maior correlação com o SST, tendo uma correlação de 0,23 e 0,20, respectivamente. No entanto, este procedimento mostrou uma correlação linear muito fraca entre as variáveis e o SST. Além disso, foi observado na análise exploratória dos dados, e através dos testes de *Tukey* e *Z-Score*, a existência de dois *outliers*, que foram removidos. Ainda utilizando os 40 pontos, os valores de pixel de todas as bandas (G, R, RE e NIR) e o valor de SST, foram aplicadas as técnicas de AM. Porém, os resultados não foram bons, com valores de R^2 negativos e valores de EQM altos. Isso permitiu concluir que os modelos de FA, RL, Regressão de Ridge, Regressão de Lasso e Regressão ElasticNet, não fizeram boas predições quando os 40 pontos foram utilizados.

Com a remoção dos *outliers*, novos testes foram realizados com os 38 pontos restantes. Neste caso, constatou-se que as bandas com maior correlação com o SST, passaram a ser as bandas G e R, com valor do coeficiente de correlação de 0,44 e 0,38, respectivamente. Nota-se um aumento no valor da correlação entre as bandas com o SST. No entanto, ainda não se observou uma forte correlação linear entre as variáveis. Ao aplicar as técnicas de AM utilizando os 38 pontos, os valores de pixel de todas as bandas (G, R, RE e NIR) e o valor de SST, o modelo com melhor resultado foi FA, que apresentou R^2 e EQM iguais a 27,58% e 29,14, respectivamente. Neste caso, foram utilizados 80% dos dados para treinamento do modelo e 20% dos dados para teste. No entanto, os valores de R^2 e EQM ainda não são capazes de mostrar que o modelo possui boa capacidade de predição.

Com os testes realizados apenas com 38 pontos e os valores de pixel apenas das bandas com maior correlação com o SST (bandas G e R), o melhor resultado de

R^2 e EQM foi obtido no modelo RL com 70% dos dados de treino e 30% dos dados de teste, que obteve o maior valor de R^2 (28,53%) e menor valor de EQM (37,09723).

Deste modo, em relação aos testes utilizando os pontos medidos *in situ* podemos afirmar que as bandas não possuem fortes correlações com o SST (os valores ficaram abaixo de 20%). O valor do coeficiente de correlação melhorou quando os *outliers* foram removidos, passando a girar entorno de 40%. Podemos afirmar ainda, que os testes FA, RL, Regressão de Ridge, Regressão de Lasso e Regressão ElasticNet não trouxeram bons resultados, pois os valores de R^2 foram baixos e os valores de EQM foram altos.

Considerando a limitação decorrente do baixo número de dados amostrais, foram realizados testes com os pontos interpolados (16.966 pontos) e as bandas com maior correlação com o SST (bandas G e R), que também não trouxeram bons resultados. Por outro lado, ao se aplicar as técnicas de AM utilizando os 16.966 pontos, os valores de pixel de todas as bandas (G, R, RE e NIR) e o valor de SST, os valores obtidos para R^2 e EQM foram de 54,96% e 18,53, respectivamente, para o modelo FA, utilizando 80% dos dados para treinamento do modelo e 20% dos dados para teste. Aqui já foi possível notar um aumento no valor de R^2 e diminuição do valor de EQM, mostrando que o modelo FA utilizando os pontos interpolados e todas as bandas, possui melhor capacidade de predição quando comparado com os resultados dos modelos utilizando apenas os pontos medidos *in situ*.

Ao aplicar as técnicas de AM utilizando apenas os pontos que restaram quando foram removidos os *outliers* do teste de *Tukey* (15.368 pontos), os valores de pixel de todas as bandas (G, R, RE e NIR) e o valor de SST, o modelo que gerou melhores resultados para R^2 (62,30%) e EQM (5,48537) foi o modelo FA ao utilizar 80% dos dados para treinamento do modelo e 20% dos dados para teste. Ao aplicar as técnicas de AM utilizando apenas os pontos que restaram quando foram removidos os *outliers* do teste de *Z-Score* (16.273 pontos), os valores de pixel de todas as bandas (G, R, RE e NIR) e o valor de SST, o modelo FA também gerou melhores resultados para R^2 (59,96%) e EQM (8,73393), com base em 80% dos dados para treinamento do modelo e 20% dos dados para teste.

Ao comparar os resultados obtidos ao considerar todos os pontos, os pontos que resultaram ao excluir os *outliers* pelo teste de *Tukey*, e ao excluir os *outliers* pelo teste de *Z-Score*, podemos dizer que o melhor resultado foi obtido quando foram

considerados apenas os pontos que restaram ao se remover os outliers determinados pelo teste de Tukey.

A utilização de câmeras multiespectrais acopladas em VANT para captura de imagens em diferentes bandas, em consonância à aplicação de técnicas de AM, mostrou-se um método promissor para a estimativa de SST. No entanto, o maior número de pontos, decorrente de maior número de campanhas de medidas e em diferentes condições de carga de sedimentos, podem favorecer a capacidade de predição dos modelos e produzir resultados ainda melhores.

REFERÊNCIAS

- ALMEIDA, A. J. F.; NETO, J. V. F. Análise do processo de urbanização e impermeabilização do solo da Bacia Endorreica do Tabuleiro do Martins, Maceió-AL. **Anáís do XI-ENANPEGE**, 5711-5722, 2015.
- AMANOLLAHY, J.; KABOODVANDPOUR, S.; MAJIDI, H. Evaluating the accuracy of ANN and LR models to estimate the water quality in Zarivar International Wetland, Iran. **Iran. nat. hazards**, 85, 1511–1527, 2017.
- ARAUJO, R. S. DE. **Micro bacia do Ribeirão Cambé-Londrina-PR: Levantamento ambiental utilizando técnicas de geoprocessamento e sensoriamento remoto**. 2004. 140 f. Bacharelado em Geografia – Universidade Estadual de Londrina, Londrina, 2004.
- BORTOLO, A. C. DE. A produção do espaço livre público do lago igapó no contexto urbano londrinense: análise de fatos e dados históricos como instrumentos para a construção do estudo. **Revista geografia em atos**, p. 97-109, 2011.
- BOUZERIA, H.; GHENIM, A. N.; KHANCHOUL, K. Using artificial neural network (ANN) for prediction of sediment loads, application to the Mellah catchment, northeast Algeria. **Journal Of Water And Land Development**, 33, p. 47-55, 2017.
- BREIMAN, L. Random forests. **Machine Learning**, 45, 1, p. 5-32, 2001.
- CAVIGLIONE, J. H.; KIIH, L. R. B.; CARAMORI, P. H.; OLIVEIRA, D. **Cartas climáticas do Paraná**. Londrina, IAPAR, 2000.
- CIGIZOGLU, H. K. Suspended sediment estimation and forecasting using artificial neural networks. **Turkish J. eng. env. sci**, 26,15-25, 2002.
- CUTLER, A.; CUTLER, D. R.; STEVENS, J. R. Random Forests. **Machine Learning – ML**, 45, 157-176, 2011.
- DECEA, (Departamento de Controle do Espaço Aéreo). Portaria DECEA nº 112/DGCEA, DE 22 DE MAIO. 1-57, 2020. Disponível em: <<https://publicacoes.decea.gov.br/?i=publicacao&id=5250>>. Acesso em: 22 out. 2020.
- DIAS, R. L. S.; SILVA, D. D. DA; FERNANDES-FILHO, E. I.; AMARAL, C. H. DO; SANTOS, E. P. DOS; MARQUES, J. F.; VELOSO, G. V. Machine learning models applied to TSS estimation in a reservoir using multispectral sensor onboard to RPA. **Ecological Informatic**, 11, 1574-9541, 2021.
- DJI. **Especificações Phantom 4 Pro V2.0**. Disponível em: <<https://www.dji.com/br/phantom-4-pro-v2/specs>>. Acesso em: 18 jan. 2021.
- DEZORDI, R.; KRAMER, G.; ROSA, C. N. DA; OSTO, J. V. DAL; FILHO, W.P. turbidez superficial a partir de imagens do Landsat-8 (sensor OLI) em compartimento aquático

do reservatório de Itaipu-PR, Brasil, **Anais do XIX simpósio brasileiro de sensoriamento remoto**, Santos-SP, 8, 2953-2955, 2019.

EHMANN, K.; KELLEHER, C.; CONDON, L. E. Monitoring turbidity from above: Deploying small unoccupied aerial vehicles to image in-stream turbidity. **Hydrological processes**, 2018.

FERREIRA, A. B.; PEREIRA FILHO, W. Avaliação da reflectância espectral de corpos d'água em Santa Maria-RS por meio de espectrorradiometria de campo. **Revista eletrônica do curso de geografia - UFG**, Jataí-GO, 13, 194-207, 2009.

FRANCO JUNIOR, J. C. **Modelagem BIM de infraestrutura urbana a partir de levantamentos aéreos com drone**. 2019. 216 f. Dissertação – Instituto de arquitetura e urbanismo da universidade de São Paulo, São Carlos – SP, 2019.

FRESCA, T. Mudanças recentes na expansão físico-territorial de Londrina. **Geografia**, 11, 2, 241–264, 2010.

GABRIEL FILHO, O. **Contribuições à análise de robustez de sistema de controle usando Redes Neurais**. 2004. 169 f. Tese – Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal do Rio Grande do Norte, 2004.

GUIMARÃES, T. T. **Utilização de imagens de satélite para predição de clorofila-a e sólidos suspensos em corpos d'água: estudo de caso da Represa do Lobo-SP**. 2019. 118 f. Dissertação – Programa de Pós-Graduação Área de Concentração em Ciências da Engenharia Ambiental da Escola de Engenharia de São Carlos da Universidade de São Paulo, 2019.

GUIMARÃES, T. T.; VERONEZ, M. R.; KOSTE, E. C.; SOUZA, E.M.; BRUM, D.; GONZAGA, L.; JR MAUAD, F.F. Evaluation of regression analysis and neural networks to predict total suspended solids in water bodies from unmanned aerial vehicle images. **Sustainability**, 11, 2580, 2019.

GUO, Q.; WU, X.; BING, Q.; PAN, Y.; WANG, Z.; FU, Y.; WANG, D.; LIU, J. Study on Retrieval of Chlorophyll-a Concentration Based on Landsat OLI Imagery in the Haihe River, China. **Sustainability**, 8, 758, 2016.

HANNA INSTRUMENTS. **HI 98194, HI 98195, HI 98196 Medidores Multi-parâmetros**. Portugal, 2015.

HSU, K.; GUPTA, H. V.; SOROOSHIAN, S. Artificial neural network modeling of the rainfall-runoff process, **Water resources research**, 31,10, 2517-2530, 1995.

IAP, (INSTITUTO AMBIENTAL DO PARANÁ). Monitoramento da qualidade das águas de microbacias urbanas, na região metropolitana de Londrina, norte do Estado do Paraná, no período de 2007 a 2009. Curitiba, 2011, **IAP**, 40p.

IAT, (INSTITUTO ÁGUA E TERRA). Portarias de enquadramento dos cursos d'água do estado do Paraná. Curitiba, 2020, **IAT**, 39p.

IBGE, (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA). **População no último censo. 2020a.** Disponível em: <<https://cidades.ibge.gov.br/brasil/pr/londrina/panorama>>. Acesso em: 21 set. 2020.

IBGE, (INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA). **População rural e urbana. 2020b.** Disponível em: <<https://educa.ibge.gov.br/jovens/conheca-o-brasil/populacao/18313-populacao-rural-e-urbana.html>>. Acesso em: 21 set. 2020.

INPE, (INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS). **Sensoriamento remoto: Sensores ópticos de alta resolução espacial. 2021.** Disponível em: <<http://www.dsr.inpe.br/DSR/areas-de-atuacao/sensores-plataformas/sensores-opticos-alta-resolucao-espacial>>. Acesso em: 02 jun. 2021.

ITCG (INSTITUTO DE TERRAS CARTOGRAFIA E GEOCIENCIAS). **Mapa de solos – Estado do Paraná.** Escala 1:2.000.000, 2006.

IWANAGA, C.H. **O Lago Igapó e o londrinense.** 1989. Trabalho de conclusão de curso - Departamento de História da Universidade Estadual de Londrina, Londrina – PR, 1989.

KISI, O.; DAILR, A. H.; CIMEN, M.; SHIRI, J. Suspended sediment modeling using genetic programming and soft computing techniques. **Journal of hydrology**, 450-451, 48-58, 2012.

KONG, J.; SHAN, Z.; CHEN, Y.; YANG, J.; HU, Y.; WANG, L. Assessment of remote-sensing retrieval models for suspended sediment concentration in the Gulf of Bohai. **Int. J. Remote sensing**, 40, 2324–2432, 2018.

KOSZMA, I. I. A. **Análise de risco ecológico da cabeceira de um ribeirão urbano.** 2018. 111 f. Dissertação Universidade Tecnológica Federal do Paraná, Londrina – PR, 2018.

KUPSSINSKÜ, L. C.; GUIMARÃES, T. T.; SOUZA, E. M.; ZANOTTA, D. C.; VERONEZ, M. R.; GONZAGA JR., L.; MAUAD, F. F. A method for Chlorophyll-a and suspended solids prediction through remote sensing and machine learning. **Sensors**, 20, 1-18, 2020.

LARSON, R.; FARBER, B. **Estatística Aplicada.** 4. ed. São Paulo: Pearson Education Brasil. Traduzido por: Luciene P. Viana; Fernanda C. Bonafini, 640, 2010.

LIU, L. W.; WANG, Y. M. Modelling reservoir turbidity using Landsat 8 satellite imagery by gene expression programming. **Water (Switzerland)**, 11, 1-16, 2019.

LORENZO, M. P. **Caracterização dos impactos ambientais negativos e medidas mitigatórias do processo de assoreamento do lago Igapó, Londrina – PR.** 2011. 68 f. TCC (Graduação) - Curso de Gestão Ambiental, Centro Universitário Filadélfia, Londrina, 2011.

MANUAL DO USUÁRIO PARROT SEQUOIA. **SEQUOIA**. V1.1. 2017. Disponível em: < <https://support.parrot.com/global/support/products/parrot-sequoia>>. Acesso em: 18 jan. 2021.

MARQUEZ, L. C. G.; BEJARANO, F. M. T.; ESPINOSA, A. C. T.; RODRIGUEZ, I. R. H.; GALLEGOS, H. B. R. Use of Landsat 8 images for depth and water quality assessment of El Guajaro reservoir, Colombia. **Journal of South American Earth Sciences**, 82, 231-238, 2018.

MARSLAND, S. **Machine Learning: An Algorithmic Perspective**. 2 ed. Chapman e Hall/Crc, 2015. 457 p.

MCDONALD, W. Drones in urban stormwater management: a review and future perspectives. **Urban water journal**, 1-14, 2019.

MENESES, P. R.; ALMEIDA, T. (org.). **Introdução ao processamento de imagens de sensoriamento remoto**. Brasília: UNB CNPQ, 2012. 266 p. Disponível em:<www.cnpq.br>. Acesso em: 24 set. 2020.

MORAES, E. C. de. Fundamentos de sensoriamento remoto: Capítulo 1. São José dos Campos, 2002. **Instituto Nacional de Pesquisas Espaciais (INPE)**. 23p.

MÜLLER, A. C.; GUIDO, S. **Introduction to Machine Learning with Python: a guide for data scientists**. United States Of America: O'reilly Media, 377 p., 2016.

MUSHTAQ, F.; LALA, M. G. N. Remote estimation of water quality parameters of Himalayan lake (Kashmir) using Landsat 8 OLI imagery. **Geocarto International**, 32, 274-285, 2017.

NOVAIS, J. DE J. **Mapeamento digital de solos da microbacia do Ribeirão Extrema, DF, a partir de imagens multitemporais ASTER e Biblioteca Espectral**. 2017. 211 f. Dissertação (Mestrado) – Programa de Pós Graduação em Agronomia, Universidade de Brasília, Brasília, 2017.

NOVO, E. M. L. DE M. Monitoramento de quantidade e qualidade da água e sensoriamento remoto. **XVII Simpósio Brasileiro de Recursos Hídricos**, 1-20, 2007.

NUNES, L. H. **Urbanização e desastres naturais, abrangência América do Sul**. São Paulo: Oficina de Textos, 112, 2015.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, 12, 2825-2890, 2011.

PEREIRA, L. DA S. **Geometria dos métodos de regressão LARS, LASSO e ElasticNet com uma aplicação em seleção genômica**. 2017. 167 f. Tese (Doutorado) – Programa de Pós Graduação em Estatística e Experimentação Agropecuária, Universidade Federal de Lavras, Lavras, 2017.

PIZANI, F. M. C.; MAILLARD, P.; FERREIRA, A. F. F.; AMORIM, C. C. DE. Estimation of water quality in a reservoir from Sentinel-2 MSI and Landsat-8 OLI sensors. **ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences**, 3, 401-408, 2020.

PÖLÖNEN, I.; PUUPPONEN, H.-H.; HONKAVAARA, E.; LINDFORS, A.; SAARI, H.; MARKELIN, L.; HAKALA, T.; NURMINEN, K. UAV-based hyperspectral monitoring of small freshwater area. **Remote sensing for agriculture**, Netherlands, 9239, 1-9, 2014.

RASCHKA, S. **Python Machine Learning**. Birmingham: Packt Publishing, 2015.

RENCHER, A. C.; SCHAALJE, B. Linear models in statistics. 2 ed. Canada: John Wiley & Sons, 672 p., 2008.

SABESP, (COMPANIA DE SANEAMENTO BÁSICO DO ESTADO DE SÃO PAULO). **Norma técnica interna SABESP: Sólidos – métodos de ensaio**. São Paulo, 1999.

SAMI, B. H. Z.; KHAI, W. J.; SAMI, B. F. Z.; FAI, C. M.; ESSAM, Y.; AHMED, A. N.; EL-SHAFIE, A. Investigating the reliability of machine learning algorithms as a sustainable tool for total suspended solid prediction. **Ain Shams Engineering Journal**, Egypt, 2021.

SILVA, C. A. **Consumo de pescado proveniente do Lago Igapó (Londrina/PR) por pescadores amadores – uma eventual exposição à contaminação por chumbo**. 2017. 57 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Engenharia Ambiental, Universidade Tecnológica Federal do Paraná, Londrina, 2017.

SILVA, J. N.; CAIADO, E. DA S.; COSTA CAIADO, M. A. Escolha de parâmetro para interpolação de precipitação diária utilizando o interpolador Inverse Distance Weight (IDW). **XX Simpósio Brasileiro de Recursos Hídricos**, Bento Gonçalves-RS, 1-8, 2013.

SOUSA, W. DOS S.; SOUSA, F. DE A. S. DE. Rede neural artificial aplicada à previsão de vazão da Bacia Hidrográfica do Rio Piancó. **Revista brasileira de engenharia agrícola e ambiental**, Campina Grande, 14, 2, 173-180, 2010.

SOUTO, C. K. B.; TELES, A. I. L.; ANDRADE, A. A.; XAVIER, B. V. M. P.; TORRES, E. E. S. A.; DIAS, E. C.; SILVA, G. P. C. DA; BARROS, K. DA C.; SOUZA, L. DA P. DE; PRATA, L. K. F.; GOMES, N. C. DA R.; FERNANDES, L. L. Fatores antrópicos de poluição hídrica na bacia do Tucunduba em Belém-PA. **Braz. J. of Develop.**, Curitiba, 5, 9, 13824-13834, 2019.

SU, T. C.; CHOU, H. T. Application of multispectral sensors carried on unmanned aerial vehicle (UAV) to trophic state mapping of small reservoirs: a case study of Tain-Pu reservoir in Kinmen, Taiwan. **Remote sensing**, 7, 10078-10097, 2015.

TIBSHIRANI, R. Regression Shrinkage and Selection via the Lasso. **Journal of the Royal Statistical Society**, 58, 1, 267-288, 1996.

TUCCI, C. E. M.; BERTONI, J. C. **Inundações urbanas**. Coleção Associação Brasileira de Recursos Hídricos. Porto Alegre: ABRH, 150, 2003.

VACARIO, E. P. L.; MACHADO, G. Inundações urbanas em Londrina PR: Um estudo da bacia do Ribeirão Cambé. **Geografia em Questão**, 12, 1, 120–141, 2019.

VACARIO, E. P. L.; MACHADO, G. Transbordamentos das águas superficiais em Londrina: o caso do Ribeirão Cambé. **Instituto de geociências – UNICAMP**, Campinas – SP, 16, 512-521, 2017.

VERONEZ, M. R.; KUPSSINSKÜ, L. S.; GUIMARÃES, T. T.; KOSTE, E. C.; SILVA, J. M. DA; SOUZA, L. V. DE; OLIVERIO, W. F. M.; JARDIM, R. S.; KOCH, I. É.; SOUZA, J. G. DE; JUNIOR, L. G.; MAUAD, F. F.; INOCENCIO, L. C.; BORDIN, F. Proposal of a Method to Determine the Correlation between Total Suspended Solids and Dissolved Organic Matter in Water Bodies from Spectral Imaging and Artificial Neural Networks. **Sensors**, 18, 159, 2018.

VILELA, M. A. M. A. **Metodologia para monitoramento da qualidade da água de reservatórios utilizando sensoriamento remoto**. 2010. 123 f. Dissertação – Universidade Federal de Uberlândia, Uberlândia, 2010.

WAGNER, F. E. WARD, J. Urbanization and Migration in Brazil. **American Journal of Economics and Sociology**, 39, 3, 249-259, 1980.

WANG, Y.; CHEN, J.; CAI, H.; YU, Q.; ZHOU, Z. Predicting water turbidity in a macrotidal coastal bay using machine learning approaches. **Estuarine, Coastal and Shelf Science**, 252, 2021.

WOODGET, A. S.; AUSTRUMS, R.; MADDOCK, I. P.; HABIT, E. Drones and digital photogrammetry: from classifications to continuum for monitoring river habitat and hydromorphology. **Wiley interdisciplinary reviews: Water**, 4, 1222, 2017.

YIGIT AVDAN, Z.; KAPLA, G.; GONCU, S.; AVDAN, U. Monitoring the water quality of small water bodies using high-resolution remote sensing data. **International journal of geo-information**, 8, 553, 2019.

ZOU, H; HASTIE, Y. Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society**, 67, 2, 301-320, 2005.

ANEXO 1 – Ofício autorização de voo



OFÍCIO Nº TALO-OFI-2020/00138

Londrina, 14 de dezembro de 2020.

Anderson Jorge
 Responsável por operações
 Gdrones Comércio e Serviços de Aeronaves Remotamente Pilotadas Eireli - EPP
 Rua Desembargador do Vale, 653 Pompéia - CEP 05010-040 - São Paulo-SP
 contato@g-drones.com.br
 tel: (11) 3673-1016

Assunto: Authorization to Modify DJI Flysafe

This letter is to inform that Empresa Brasileira de Infraestrutura Aeroportuária - INFRAERO, the concessionaire that manages and operates Londrina Airport, authorizes the Gdrones Comércio e Serviços de Aeronaves Remotamente Pilotadas Eireli - EPP to operate the drones described below to flight near Londrina Airport, as stated in the SARPAS authorizations. All operations will be coordinated with the local ATC (Air Traffic Control). Therefore, we allow that DJI Flysafe temporarily modify the software parameters of the drone stated below to be used inside de no fly zone near Londrina Airport, according to SARPAS flight plans.

The Operations are planned to occur in January 2021, in a single day, still to be determined, and won't go over 140m of height.

Airport information: IATA: LDB ICAO: SBLO City: Londrina, Paraná, Brazil

Administration: Empresa Brasileira de Infraestrutura Aeroportuária - INFRAERO

Drones and Operator data:

Name of Operator: Anderson Jorge

Name of Company: Gdrones CNPJ: 22.619.376/0001-11

Model: DJI Phantom 4 Advanced Serial Number: 0HACE6F0A20303

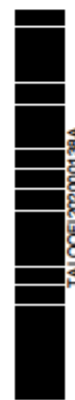
Operator Code (SARPAS): LLXP

Classif. documental 221.000

Empresa Brasileira de Infraestrutura Aeroportuária - Infraero
 Endereço : RUA TENENTE JOÃO MAURÍCIO DE MEDEIROS N 300 JD. AEROPORTO
 CEP:86039090 LONDRINA-PR-BRASIL
<http://www.infraero.gov.br>



Assinado com senha por SANDRA APARECIDA DE OLIVEIRA em 14/12/2020 15:03:43.
 Documento Nº: 1403661-8549 - consulta à autenticidade em
<https://sigadoc.infraero.gov.br/sigaex/public/app/autenticar?n=1403661-8549>



TALOOFI202000138A

SIGA

ANEXO 2 – Avaliação de Risco Operacional para Operação com RPA em Londrina-PR

Avaliação de Risco Operacional para Operação com RPA em Londrina-PR

1. Introdução:

Este documento de avaliação de risco operacional tem o objetivo de cumprir o estabelecido nos parágrafos E94.103(f)(2) e E94.103(g)(2) do **RBAC-E nº 94**, tendo sido elaborado conforme critérios contidos na Instrução Suplementar E-94-003, denominada *Procedimentos para elaboração e utilização de avaliação de risco operacional para operadores de aeronaves não tripuladas da ANAC – Agência Nacional de Aviação Civil*.

Esta avaliação de risco operacional foi realizada considerando-se especificamente as operações com ARP a serem realizadas no município de Londrina-PR em 13 de fevereiro de 2021, conforme condições descritas no corpo do presente documento, sendo válida unicamente para elas.

2. Legislação Aplicável:

- Código Brasileiro de Aeronáutica (Lei nº 7.565/1986)
- RBAC-E nº 94, da ANAC;
- IS nº E94-003 – Revisão A, da ANAC;
- Resolução nº 25/2008, da ANAC;
- ICA 100-40, reeditada em 03/01/2019, do DECEA;
- ICA 100-12, reeditada em 10/11/2016, do DECEA;
- ICA 100-37, reeditada em 10/11/2016, do DECEA;
- MCA 100-16, reeditada em 17/10/2016, do DECEA;
- AIC-N 23/18
- MCA 56-3
- Resolução Anatel nº 242, de 30 de novembro de 2000.

3. Cenário Operacional:

Voos deverão ser realizados com ARP multirrotor Phantom 4 Pro para a obtenção de fotografias aéreas em corpos d'água próximo ao Aeroporto de Londrina-PR.

Os voos serão realizados em VLOS, por meio de piloto remoto e com assistência de observadores em campo.

Os voos deverão ser realizados a no máximo 400 pés, aproximadamente, 120 m de altura em relação ao solo.

Os voos deverão ser realizados em modo automático, com a supervisão do piloto, o qual poderá e deverá a qualquer momento acionar a pilotagem manual de forma imediata, se necessário.

Deverá ser realizado o pouso da ARP, de forma conservadora, com no mínimo 20% de bateria disponível.

4. Localização da Operação:

Endereço: Os voos serão realizados na área urbana do município de Londrina-PR.

Segue mapa que apresenta as áreas a serem sobrevoadas e as sobreposições com as zonas de restrições a voos.

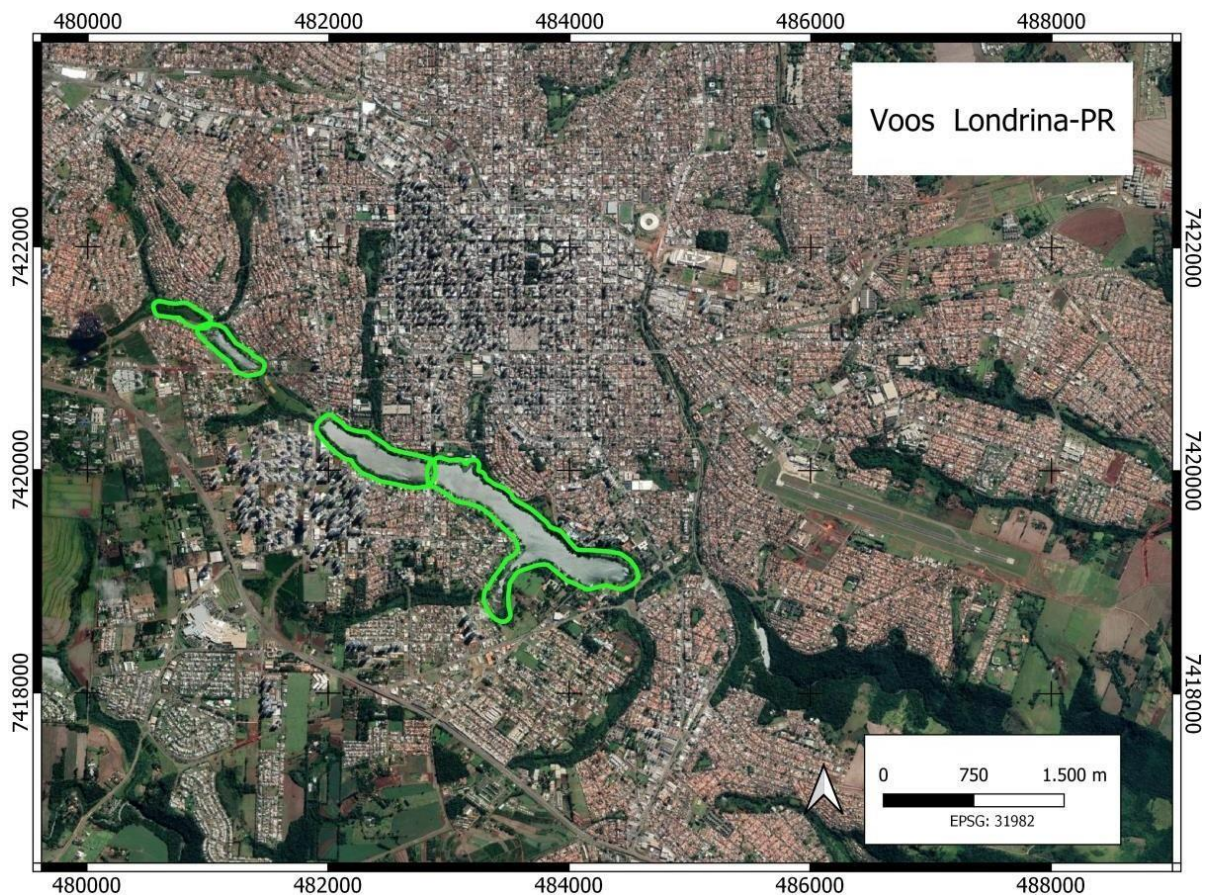


Figura 4-1: Áreas de voos

5. Responsável pela Operação, Operador em Comando e Observador:

Responsável: G Drones Comercio e Servicos de Aeronaves Remotamente Pilotadas Eireli

CNPJ / CPF: 22.619.376/0001-11

Operador em Comando: Anderson Jorge

CPF: 054.536609 - 70

ID SARPAS: LIXP

Contato: +55 11 960636807

Observador: Débora Moraes Duarte

CPF: 610.004.973 - 21

ID SARPAS: OKXZ

Contato: +55 11 997250909

6. Aeronave em Operação (RPA):

Fabricante e Modelo: DJI Phantom 4 Pro

Número de Série: OHACE6F0A20303

Cadastro ANAC: PP-201906653

Homologação ANATEL: 05260-16-07248

Certificado Seguro RETA: 2194/0003477/35

7. Condições para a realização da operação:

Os operadores possuem treinamento adequado provido pela G drones para operação de ARP multirrotor Phantom 4 Pro nas condições previstas.

Qualquer operação somente poderá ser realizada em horários e datas aprovadas por meio de autorizações expressas do DECEA, por meio do sistema SARPAS.

Operações em zonas de restrição ou “no fly zones” no entorno de aeródromos somente poderão ser realizadas com autorização e coordenação do órgão ATS ou Administrador Aeroportuário Local ou Órgão de Controle de Operações Aéreas Militares (OCOAM).

Os voos somente poderão ser realizados em horários e datas informados pelo órgão ATS ou Administrador Aeroportuário Local ou Órgão de Controle de Operações Aéreas Militares (OCOAM).

A operação deve adequar-se aos espaços aéreos condicionados, cabendo exclusivamente ao piloto verificar se a operação pretendida atende ao previsto e pela observância e cumprimento de todos os parâmetros previstos no AIC-N 23/18 e no MCA 56-3.

O piloto remoto da aeronave e observador estarão com seus telefones

celulares sempre ligados durante a operação para eventual comunicação com órgão ATS ou Administrador Aeroportuário Local ou Órgão de Controle de Operações Aéreas Militares (OCOAM).

Em caso de qualquer acidente com lesões em pessoas, o SAMU ou Corpo de Bombeiros deverão ser acionados. Os operadores efetuarão os voos munidos dos telefones de contatos necessários para comunicação com estes meios.

Em caso de perda de contato visual e de telemetria com o ARP, os operadores deverão comunicar imediatamente o órgão ATS.

8. Matriz de Risco:

Probabilidade da ocorrência, ou a frequência com que um evento, como consequência de um perigo existente, possa ocorrer. Está dividida em 05 níveis:

- **Nível 5 (frequente):** é provável que ocorra muitas vezes, ou historicamente tem ocorrido frequentemente;
- **Nível 4 (ocasional):** é provável que ocorra algumas vezes, ou historicamente tem ocorrido com pouca frequência;
- **Nível 3 (remoto):** é improvável, mas é possível que venha a ocorrer, ou ocorrer raramente;
- **Nível 2 (improvável):** é bastante improvável que ocorra e não se tem notícia de que tenha alguma vez ocorrido; e
- **Nível 1 (muito improvável):** é quase impossível que o evento ocorra.

Severidade da ocorrência, ou grau da consequência de um evento, como consequência de um perigo existente ou de uma situação insegura, tomando como referência a pior condição possível. Está dividida em 05 níveis:

- **Nível A (catastrófico):** morte de múltiplas pessoas;
- **Nível B (crítico):** morte de pessoa, lesões gravíssimas, capazes de deixar sequelas significativas e/ou incapacitantes, tais como cegueira, paralisia, amputações, etc.;
- **Nível C (significativo):** lesões sérias a pessoas, mas não incapacitantes nem com sequelas significativas;
- **Nível D (pequeno):** incidentes menores, danos a objetos, animais ou vegetação no solo, lesões leves;

- **Nível E (insignificante):** somente danos ao equipamento.

Tolerabilidade, ou grau de aceitabilidade em razão do resultado da avaliação de risco, determinada pela tabela seguinte e descritos à seguir:

			Severidade				
			Catastrófico	Critico	Significativo	Pequeno	Insignificante
			A	B	C	D	E
Probabilidade	Frequente	5	5A	5B	5C	5D	5E
	Ocasional	4	4A	4B	4C	4D	4E
	Remoto	3	3A	3B	3C	3D	3E
	Improvável	2	2A	2B	2C	2D	2E
	Muito improvável	1	1A	1B	1C	1D	1E

- **Risco extremo** (classificações 4A, 5A e 5B): a operação não deve ocorrer e, caso esteja ocorrendo, deve cessar imediatamente, enquanto persistir a condição ou até que medidas mitigadoras suficientes reduzam o risco para um nível aceitável pelo operador. Caso ainda se decida prosseguir com a operação, controles preventivos para mitigação do risco devem ser estabelecidos, devem estar em vigor e a aprovação da hierarquia mais alta da empresa (presidente) deve ser requerida.
- **Alto risco** (classificações 3A, 4B e 5C): a operação não deveria ocorrer e, caso esteja ocorrendo, deveria cessar imediatamente, enquanto persistir a condição ou até que medidas mitigadoras suficientes reduzam o risco para um nível aceitável pelo operador. Caso ainda se decida prosseguir com a operação, controles preventivos para mitigação do risco devem ser estabelecidos, devem estar em vigor e a aprovação da hierarquia de gestão da empresa (gerente ou diretor) deve ser requerida.
- **Risco moderado** (classificações 1A, 2A, 2B, 3B, 3C, 4C, 4D, 5D, 5E): a operação pode ocorrer com controles preventivos para mitigação do risco estabelecidos e que devem estar em vigor, conforme necessários. Operações neste nível de risco deveriam ser aprovadas por nível hierárquico imediatamente superior (chefia imediata).
- **Baixo risco** (classificações 1B, 1C, 2C, 2D, 3D, 3E, 4E): a operação pode ocorrer e controles preventivos para mitigação de risco e aprovação por nível hierárquico imediatamente superior (chefia imediata) são opcionais.
- **Risco muito baixo** (classificações 1D, 1E e 2E): a operação é

aceitável como concebida, e nenhum controle preventivo para mitigação de risco e aprovação é requerida para que ela ocorra.

9. Avaliação de Riscos:

Para o cenário de operação previsto, foram identificadas cinco situações de risco, as quais são descritas nos quadros a seguir:

Situação 1	Pane da ARP em voo
Probabilidade de Ocorrência	3 (remoto). A aeronave Phantom 4 Pro é um equipamento com baixo histórico de panes.
Severidade da ocorrência	C (significativo). Se houver pane e uma eventual queda, a aeronave pode atingir pessoas e patrimônio de terceiros, considerando-se a operação em área urbana.
Risco	3C
Tolerabilidade	Risco Moderado
Nível Hierárquico de Autorização da Operação	Operador em Comando
Medidas de Mitigação do Risco	Manutenção preventiva da ARP. Realização de check list de montagem e de voo antes da operação. (3) Calibração de sensores (sistema inercial, bússola e sensores anti colisão) quando houver erros de leitura acusados pelo sistema. (4) Monitoramento de mensagens dos sistemas de voo por meio de telemetria disponível no aplicativo de navegação da ARP durante a operação. (5) Retorno do drone e pouso se houver identificação de qualquer mensagem de erro no aplicativo de navegação.

Situação 2	Perda do enlace com aeronave (link)
Probabilidade de Ocorrência	3 (remoto). A aeronave Phantom 4 Pro possui enlace de comunicação dual band (2,4 e 5,8 Ghz) e em condições ideais, o alcance é de até 7 km. Entretanto a operação em área urbana aumenta o risco de perda de enlace.
Severidade da Ocorrência	E (Insignificante). Em caso de perda de link, a aeronave deverá retornar ao home point (ponto de decolagem) de forma automática.
Risco	3E
Tolerabilidade	Baixo Risco
Nível Hierárquico de Autorização da Operação	Operador em Comando
Medidas de Mitigação do Risco	(1) Análise do local de operação, evitando posicionar a aeronave em situação de sombra e com obstruções entre ela e o operador; Garantir que o <i>Home Point</i> foi registrado pela aeronave; (3) Configurar altura adequada de retorno ao <i>Home Point</i> , para evitar colisões com obstáculos; (4) Monitorar nível do sinal de comunicação e movimentar o rádio controle, buscando a melhoria do sinal.

Situação 3	Existência de Tráfego Aéreo no Local – fora das zonas de restrição e “no fly zones”.
Probabilidade de Ocorrência	1 (muito improvável). As operações ocorrem a baixa altitude, longe de rotas de aeronaves e em locais protegidos por edificações.
Severidade da Ocorrência	A (catastrófico). Se ocorrer uma colisão com aeronave tripulada pode ocorrer queda da aeronave e morte de pessoas.
Risco	1A
Tolerabilidade	Risco Moderado
Nível Hierárquico de Autorização da Operação	Responsável pela Operação
Medidas de Mitigação do Risco	(1) Operador e observador devem ficar atentos ao tráfego aéreo no local e pousar a aeronave se detectado tráfego na região da operação. (2) Operador e observador devem estar com comunicação disponível para órgãos ATS.

Situação 4	Existência de Tráfego Aéreo no Local – dentro das zonas de restrição e “no fly zones”.
Probabilidade de Ocorrência	3 (muito raro). Colisão de aeronaves é uma ocorrência muito rara, devido ao caráter tridimensional do espaço aéreo e porte da ARP. As operações ocorrem a baixa altitude também, mas em regiões de rotas de aeronaves.
Severidade da Ocorrência	A (catastrófico). Se ocorrer um encontro com aeronave tripulada pode ocorrer a queda da aeronave e morte de pessoas.
Risco	3A
Tolerabilidade	Alto Risco
Nível Hierárquico de Autorização da Operação	Responsável pela Operação
Medidas de Mitigação do Risco	(1) Operação deve ocorrer estritamente em horários e datas aprovadas pelo órgão ATS e DECEA e ser realizada de forma coordenada com estes. (2) Operador e observador devem ficar atentos ao tráfego aéreo no local e pousar a aeronave se detectado tráfego na região da operação. (3) Operador e observador devem estar com comunicação disponível para órgãos ATS.

Situação 5	Mudança climática abrupta com incidência de ventos fortes, nevoeiros, nuvens baixas e/ou chuvas.
Probabilidade de Ocorrência	3 (remoto). Mudanças meteorológicas abruptas não são comuns na região dos voos.
Severidade da Ocorrência	C (significativo). A aeronave possui resistência a chuvas leves, umidade e ventos até 12m/s, mas ela não deverá ser operada em quaisquer condições de chuva, nuvens baixas ou ventos fortes.
Risco	3B
Tolerabilidade	Risco Moderado
Nível Hierárquico de Autorização da Operação	Operador em comando
Medidas de Mitigação do Risco	(1) Verificar previsão do tempo e as condições climáticas e no momento da operação; (2) Observar indicadores de alteração abrupta do tempo, como mudança de direção e velocidade dos ventos; (3) Não realizar a operação nestas condições ou na possibilidade eminente delas; (4) Caso as condições meteorológicas e ambientais mudem durante a operação, esta deverá ser suspensa imediatamente. O drone deverá ser pousado em segurança.

10. Disposições Finais:

Declaro para os devidos fins que todos os operadores responsáveis conhecem e cumprem a legislação e regulamentação aplicáveis supracitadas.

Londrina, 18 de janeiro de 2020;