PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS Programa de Pós-graduação em Engenharia Elétrica

Maycon Gabriel Gomes Da Silva

PLATAFORMA WEB PARA A PREDIÇÃO DE TURBIDEZ E SÓLIDOS TOTAIS EM BACIAS HIDROGRÁFICAS USANDO OS MODELOS DE REGRESSÃO LINEAR E LASSO: UM ESTUDO DE CASO NO RESERVATÓRIO DE TRÊS MARIAS - MINAS GERAIS

Belo Horizonte 2022 Maycon Gabriel Gomes Da Silva

PLATAFORMA WEB PARA A PREDIÇÃO DE TURBIDEZ E SÓLIDOS TOTAIS EM BACIAS HIDROGRÁFICAS USANDO OS MODELOS DE REGRESSÃO LINEAR E LASSO: UM ESTUDO DE CASO NO RESERVATÓRIO DE TRÊS MARIAS - MINAS GERAIS

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Pontifícia Universidade Católica de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica.

Orientadora: Prof^a. Dr^a. Zélia Myriam Assis Peixoto

Linha de pesquisa: Processamento e Análise de Sinais

Belo Horizonte 2022

FICHA CATALOGRÁFICA Elaborada pela Biblioteca da Pontifícia Universidade Católica de Minas Gerais

S586p	 Silva, Maycon Gabriel Gomes da Plataforma web para a predição de turbidez e sólidos totais em bacias hidrográficas usando os modelos de regressão linear e LASSO: um estudo de caso no reservatório de Três Marias - Minas Gerais / Maycon Gabriel Gomes da Silva. Belo Horizonte, 2022. 68 f. : il.
	Orientadora: Zélia Myriam Assis Peixoto
	Dissertação (Mestrado) – Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Engenharia Elétrica
	1. Bacias hidrográficas - Três Marias, Reservatório de (MG). 2. Sensoriamento remoto - Três Marias, Reservatório de (MG). 3. Aprendizado do computador. 4. Eutrofização. 5. Turbidez da Água. 6. Reservatórios. 7. Vulnerabilidade a Desastres. I. Peixoto, Zélia Myriam Assis. II. Pontifícia Universidade Católica de Minas Gerais. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

CDU: 556.51 Ficha catalográfica elaborada por Elizângela Ribeiro de Azevedo - CRB 6/3393

Maycon Gabriel Gomes Da Silva

PLATAFORMA WEB PARA A PREDIÇÃO DE TURBIDEZ E SÓLIDOS TOTAIS EM BACIAS HIDROGRÁFICAS USANDO OS MODELOS DE REGRESSÃO LINEAR E LASSO: UM ESTUDO DE CASO NO RESERVATÓRIO DE TRÊS MARIAS - MINAS GERAIS

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Elétrica da Pontifícia Universidade Católica de Minas Gerais, como requisito parcial para obtenção do título de Mestre em Engenharia Elétrica.

Linha de pesquisa: Processamento e Análise de Sinais

Prof^a. Dr^a. Zélia Myriam Assis Peixoto (Orientadora) – PUC Minas

Prof. Dr. Luís Fabrício Wanderley Góes – University of Leicester – UK

Prof. Dr. Prof. Carlos Augusto Paiva da Silva Martins - PUC Minas

Prof. Dr. Maury Meirelles Gouvea Junior (Suplente) – PUC Minas

Belo Horizonte, 29 de Junho de 2022

Aos meus pais, Rosangela Ferreira da Silva e Geraldo Gomes da Silva.

AGRADECIMENTOS

Agradeço a Deus por ter me dado a oportunidade de estar estudando e me capacitando para estar sempre buscando fazer o melhor.

Agradeço à minha mãe Rosangela Ferreira e meu pai Geraldo Gomes, por me darem todo apoio, incentivo e suporte necessário.

Agradeço aos meus avós pelo apoio e orações que sempre fazem por mim.

Agradeço à minha família e amigos pela ajuda que sempre me deram quando eu precisei.

Agradeço à minha orientadora Profa. Zélia Myriam por ter guiado meus passos durante grande parte da minha caminhada acadêmica, dando-me todo o suporte necessário.

Agradeço ao Douglas Abreu, Mestre em Engenharia Elétrica pela PUC Minas, por ter me ajudado e orientado em vários momentos do meu curso de mestrado.

Agradeço aos meus professores e colegas do curso de Engenharia Eletrônica e de Telecomunicação da PUC Minas que me ajudaram nessa jornada acadêmica com trocas de conhecimento, apoio e incentivo.

Agradeço ao PPGEE (Programa de Pós-graduação em Engenharia Elétrica) e à PUC Minas pelo apoio e disponibilidade durante o difícil período de pandemia que infelizmente tivemos que enfrentar. Agradeço também pela bolsa que me foi concedida, sem a qual não teria sido possível realizar o mestrado.

Agradeço à CEMIG[®] por ter disponibilizado os dados históricos limnológicos do Reservatório de Três Marias - MG e também pela atenção e o respeito com que se colocaram à disposição para ajudar neste trabalho. "Se algo é importante o suficiente, você deve tentar. Mesmo se o resultado provável for o fracasso, as pessoas deveriam ir atrás do que as fazem felizes. Isso as fará ainda mais felizes do que qualquer outra coisa." - Elon Musk.

RESUMO

O método tradicional de avaliação da qualidade da água em reservatórios hídricos que consiste, basicamente, na coleta de amostras de água *in situ* e análises laboratoriais, é um método que demanda especialistas, altos recursos financeiros e longos intervalos de tempo de execução. Como uma alternativa, o sensoriamento remoto por meio de técnicas de processamento de imagens e *machine learning* vêm sendo aplicadas visando à redução de custos, possibilidade de execução em tempo real e o aumento da acessibilidade. Este trabalho trata do desenvolvimento de uma plataforma Web capaz de auxiliar no monitoramento e análise dos parâmetros limnológicos de bacias hidrográficas a partir de imagens de satélite, as quais são utilizadas para o treinamento de modelos de aprendizagem com base nas técnicas de regressão linear e regressão LASSO (Last absolute shrinkage and selection operator). A implementação da plataforma envolveu a associação de diversas tecnologias de desenvolvimento Web, dentre elas, API (Application Programming Interface) do GEE (Application Programming Interface), linguagens JavaScript e Python, e frameworks HTML5 (HyperText Markup Language) e CSS3 (Cascading Style Sheets) e Flask. Como um estudo de caso, o treinamento e testes dos modelos de aprendizagem foram aplicados à estimação dos parâmetros limnológicos turbidez e sólidos totais a partir de dados históricos medidos no Reservatório de Três Marias - MG, disponibilizados pela Companhia Energética de Minas Gerais S.A. (CEMIG[®]) devidamente correlacionados às imagens dos locais de medição, extraídas do Satélite Landsat 8 OLI, por intermédio da API do GEE. A validação dos modelos de aprendizagem foi realizada por meio de métricas estatísticas, o coeficiente de determinaçã R^2 , erro absoluto médio e o erro percentual absoluto médio, que indicaram uma boa aproximação em relação às medições disponíveis. Vale ressaltar que a metodologia proposta pode ser estendida para quaisquer reservatórios hídricos e outros parâmetros limnológicos desde que estejam disponíveis dados históricos das medições e informações espectrais das bandas do Satélite Landsat 8 OLI.

Palavras-chave: *Google Engine Earth, Machine learning*, Parâmetros limnológicos, Sensoriamento remoto, Reservatórios hídricos.

ABSTRACT

The traditional method of assessing water quality in water reservoirs, which basically consists of collecting water samples in situ and laboratory analysis, is a method that demands specialists, high financial resources and long runtime intervals. As an alternative, remote sensing through image processing and machine learning techniques has been applied to reduce costs, enable real-time execution and increase accessibility. This work deals with the development of a Web platform capable of assisting in the monitoring and analysis of limnological parameters of hydrographic basins from satellite images, which are used to train learning models based on linear regression and LASSO regression (Last absolute shrinkage and selection operator). The implementation of the platform involved the association of several Web development technologies, among them, API (Application Programming Interface) of the GEE (Application Programming Interface), JavaScript and Python languages, and HTML5 (HyperText Markup Language) and CSS3 (Cascading Style Sheets) frameworks) and Flask. As a case study, the training and testing of the learning models were applied to the estimation of the limnological parameters, turbidity and total solids, from historical data measured in the Três Marias Reservoir - MG, provided by Companhia Energética de Minas Gerais S.A. (CEMIG[®]) properly correlated to the images of the measurement sites, extracted from the Landsat 8 OLI satellite. The validation of the learning models was performed using statistical metrics, the coefficient of determination R^2 , mean absolute error and mean absolute percentage error, which indicated a good approximation in relation to the available measurements. It is worth mentioning that the proposed methodology can be extended to any water reservoir and other limnological parameters as long as historical data from measurements and spectral information from the Landsat 8 OLI satellite bands are available.

Keywords: Google engine earth, Machine learning, Limnological parameters, Remote sensing, Water reservoirs.

LISTA DE FIGURAS

FIGURA	1 -	Representação didática de uma bacia hidrográfica	17
FIGURA	2 -	Delimitação de uma bacia hidrográfica	18
FIGURA	3 -	Vista aérea do Reservatório de Três Marias	19
FIGURA	4 -	Comportamento espectral de alvos da superfície terrestre	22
FIGURA	5 -	Resposta espectral da água contendo sólidos em suspensão	25
FIGURA	6 –	Diagrama de blocos da metodologia proposta	42
FIGURA	7 –	Imagem com a presença de faixas de linhas ruidosas, causadas pela	
		descalibração dos detectores de bandas	43
FIGURA	8 -	Imagem da banda 5 do Satélite Landsat 8 OLI, Reservatório de Três	
		Marias - MG.	44
FIGURA	9 -	Pontos de coleta de amostras de água no Reservatório de Três Ma-	
		rias - MG	46
FIGURA	10 -	Página inicial da Plataforma Web.	49
FIGURA	11 –	Gráfico de resíduos relativos à predição dos sólidos totais usando a	
		regressão linear	53
FIGURA	12 -	Gráfico de resíduos relativos à predição da turbidez usando a re-	
		gressão linear.	53
FIGURA	13 –	Gráfico de resíduos relativos à predição dos sólidos totais usando a	
		regressão LASSO	54
FIGURA	14 -	Gráfico de resíduos relativos à predição da turbidez usando a re-	
		gressão LASSO	54
FIGURA	15 -	Valores preditos $\mathbf x$ valores reais sólidos totais usando regressão linear.	55
FIGURA	16 –	Valores preditos x valores reais sólidos totais usando a regressão	
		LASSO	55
FIGURA	17 -	Valores preditos x valores reais para a turbidez usando regressão	
		linear	56
FIGURA	18 -	Valores preditos x valores reais para a turbidez usando regressão	
		LASSO.	56
FIGURA	19 -	Correlação entre cor verdadeira e turbidez	57

FIGURA 20 – Correlação entre cor verdadeira e turbidez. \ldots \ldots \ldots \ldots	58
FIGURA 21 – Correlação entre banda 4 e turbidez	58
FIGURA 22 – Correlação entre banda 4 e turbidez	59
FIGURA 23 – Correlação entre s dt (sólidos dissolvidos totais) e sólidos totais. $\ .$.	59
FIGURA 24 – Correlação entre s dt (sólidos dissolvidos totais) e sólidos totais. $\ .$.	60
FIGURA 25 – Correlação entre banda 6 e sólidos totais	60
FIGURA 26 – Correlação entre banda 6 e sólidos totais.	61

LISTA DE TABELAS

TABELA	1 –	Características do Satélite Landsat 8 OLI	20
TABELA	2 -	Bandas espectrais do Satélite Landsat-8 OLI	20
TABELA	3 -	Coordenadas geográficas dos locais e datas das coletas no Reserva- tório de Três Marias - MG.	45
TABELA	4 -	Resultados dos coeficientes de determinação R^2 para os modelos de aprendizagem regressão linear e LASSO na etapa de treinamento	51
TABELA	5 -	Resultados do erro absoluto médio e erro percentual absoluto médio para os modelos de aprendizagem regressão linear e LASSO na	
		etapa de teste.	52

LISTA DE SIGLAS

- ANA Agência Nacional das Águas
- API Application Programming Interface
- CEMIG Companhia Energética de Minas Gerais
- COA Compostos Opticamente Ativos
- CSS Cascading Style Sheets
- GEE Google Engine Earth
- HTML HyperText Markup Language
- IQA Índice de Qualidade da Água
- LASSO Last Absolute Shrinkage and Selection Operator
- REM Radiação Eletromagnética
- RGB Red, Green, Blue
- R^2 Coeficiente de Determinação
- MAE Erro Absoluto Médio (Mean Absolute Error)
- MAPE Erro Percentual Absoluto Médio (Mean Absolute Percentage Error)
- MMQ Método dos Mínimos Quadrados (Least Squares Method)
- NDVI Índice de Vegetação por Diferença Normalizada
- NIR Infravermelho Próximo
- OLI Operational Land Imager
- OMS Organização Mundial da Saúde
- SDT Sólidos Dissolvidos Totais
- SWIR Ondas Curtas do Infravermelho e Infravermelho Termal
- TIRS Thermal Infrared Sensor

SUMÁRIO

1 INTRODUÇÃO	. 11
1.1 Problema Motivador	. 11
1.2 Objetivos	. 12
1.2.1 Objetivo Geral	. 12
1.2.2 Objetivos Específicos	. 12
1.3 Justificativa	. 12
1.4 Contribuições da Pesquisa	$. \bar{1}\bar{3}$
1.5 Estrutura do Trabalho	13
	. 10
ο δετεδενισιλι πεώδιος	15
2 REFERENCIAL IEURICO	. 1 3
	. 10
2.2 Aspectos gerais das características da agua em bacias hidrograficas	. 16
2.2.1 Composição de bacia hidrográfica - Conceitos fundamentais	. 17
2.2.2 Bacia Hidrográfica do Rio São Franscisco	. 18
2.3 O Satélite Landsat 8 OLI	. 19
2.3.1 Bandas espectrais	. 21
2.4 Sensoriamento Remoto	. 22
2.4.1 Sensoriamento remoto da água	$\bar{24}$
25 Goode Earth Engine	25
2.6 Mashing Loaming	- <u>2</u> 0 26
2.0 Muchine termining \ldots \ldots \ldots \ldots \ldots	. 20
2.0.1 Regressão integra	. 21
2.6.2 Regressão LASSO	. 27
2.7 Métricas Estatísticas	. 30
2.7.1 Erro absoluto médio (MAE)	. 30
2.7.2 Erro percentual absoluto médio (MAPE)	. 30
272 Coefficiente de determinação (B^2)	30
2, 1, 5 Complete ue ue ue ue ue ue ue ue	. 50
	. 51
3 TRABALHOS RELACIONADOS	. 33
3.1 Conclusão	. 39
4 METODOLOGIA E PROPOSTA DE DESENVOLVIMENTO	. 41
4.1 Dados limnológicos	41
4.2 Google Engine Earth	41
1.2 1 Bana da dadas	. 11
4.2. Machine Learning	. 44
4.3 Muchine Learning	. 41
4.3.1 Ireinamento, teste e valiaação aos modelos de aprenaizagem	. 48
4.4 Plataforma Web	. 48
4.5 Conclusao	. 50
5 RESULTADOS E DISCUSSÕES	51
51 Conclusão	58
	00
b Considerações Finais	. 63
6.1 Conclusoes	. 63
6.2 Propostas de Continuidade	. 63
BEFEBÊNCIAS	65
	. 00

1 INTRODUÇÃO

A água é uma substância química de vital importância para os seres vivos, fundamental às suas reações bioquímicas interiores e no seu entorno, em suas relações sociais e com a natureza. Nesse sentido, pode-se observar, historicamente, que as grandes civilizações se desenvolveram ao longo dos rios, utilizando a água como alimento, meio de transporte, comércio e desenvolvimento, dentre outros (MOTA; BRAICK, 1998). Nas últimas décadas, o problema da poluição hídrica tem se tornado cada vez mais significativo e uma das principais causas são os rejeitos industriais liberados diretamente no ambiente, o grande fluxo de esgoto doméstico das zonas urbanas e as erosões próximas aos leitos dos rios, decorrentes do manejo inadequado das matas ciliares e má utilização das regiões de proteção ambiental.

Os poluentes hídricos podem ser divididos em três principais classes, a saber físicos (mudança na temperatura, cor e turbidez), químicos (compostos inorgânicos, íons comuns e metais tóxicos e orgânicos) e biológicos (bactérias, vírus e parasitas). Segundo a Organização Mundial de Saúde (OMS), pelo menos dois milhões de pessoas morrem por ano em consequência da ingestão de águas contaminadas (OMS, 2011).

1.1 Problema Motivador

Apesar da água ser um recurso abundante na Terra, somente uma pequena fração encontra-se disponível para consumo humano, animal e manutenção da flora. Estima-se que 97,5% da água existente é salgada e, portanto, não é adequada ao consumo direto. Dos 2,5% de água doce, a maior parte (69%) é de difícil acesso desde que está concentrada em geleiras, 30% são águas subterrâneas (armazenadas em aquíferos) e 1% encontra-se nos rios ((ANA), b). Por esses motivos, a qualidade da água dos reservatórios e lagos vem assumindo, progressivamente, maior importância na atualidade.

Os parâmetros que são usados para avaliar o nível de poluição ou eutrofização dos lagos e reservatórios, denominada água bruta, ou seja, da água que é captada pelas empresas responsáveis pelo abastecimento das cidades e áreas agrícolas, têm que apresentar um nível mínimo de qualidade segundo os protocolos dos órgãos reguladores para que possa ser tratada e disponibilizada para a população ((ANA), a). O monitoramento é importante para a preservação da vida aquática nas bacias hidrográficas e, também, para os reservatórios que são usados para diversas funções além do consumo humano e animal, como pescaria, esportes náuticos e turismo, dentre outros.

Como a qualidade e degradação das águas de lagos e represas precisam ser monitoradas com frequência, os métodos tradicionais que envolvem a coleta *in situ* e análises em laboratórios são, normalmente, de alto custo e demorados. É, nesse contexto, que o sensoriamento remoto a partir de imagens de satélite correlacionadas a medições de amostras de água disponíveis pode constituir uma ferramenta auxiliar e acessível no processo de monitoramento da qualidade da água de ambientes lênticos.

1.2 Objetivos

1.2.1 Objetivo Geral

Este trabalho consiste do desenvolvimento de uma plataforma Web para a predição dos parâmetros limnólogicos em reservatórios hídricos com base em informações espectrais das bandas do satélite Landsat 8 OLI (*Operational Land Imager*) e dados históricos obtidos de medições previamente realizadas nos pontos de interesse.

Essa plataforma baseia-se em modelos de aprendizagem que utilizam as técnicas de regressão linear e LASSO (*Least absolute shrinkage and selection operator*), que, após as etapas de treinamento e testes, deverão permitir novas predições apenas utilizando informações espectrais das bandas do satélite Landsat 8 OLI ou informações limnológicas do reservatório, visando à redução de custos, análise e aumento da acessibilidade ao monitoramento da qualidade da água em bacias hidrográficas.

Nesse contexto, definiu-se como estudo de caso, o desenvolvimento de uma plataforma Web para a estimação da turbidez e sólidos totais a partir dos parâmetros limnólogicos do Reservatório de Três Marias - MG, disponibilizados pela CEMIG[®], devidamente correlacionados às informações espectrais das imagens do Landsat 8 OLI.

1.2.2 Objetivos Específicos

Dentre as etapas de desenvolvimento, são destaques da proposta deste trabalho:

- Elaborar um levantamento bibliográfico acerca dos trabalhos relevantes sobre o tema da pesquisa;
- Analisar e destacar os fundamentos teóricos relativos à constituição das bacias hidrográficas, características das bandas espectrais do satélite Landsat 8 OLI e uso do sensoriamento remoto aplicado com *machine learning* para fazer monitoramento, analise e predições de parâmetros limnológicos de reservatórios hídricos;
- Desenvolver uma plataforma Web, utilizando tecnologias relacionadas ao desenvolvimento Web e machine learning para disponibilizar aos usuários finais (empresas, pesquisadores, alunos, etc...) um recurso digital de fácil utilização para o monitoramento, análise e predições de parâmetros limnológicos de reservatórios hídricos;
- Publicar os resultados obtidos em congressos e/ou periódicos da área.

1.3 Justificativa

Ainda que o Brasil não disponha de uma legislação própria para a regulamentação da qualidade da água em represas e lagos, alguns estados adotaram o padrão internacional,

criado nos Estados Unidos (PNQA, 2021), que trata do Índice de Qualidade da Água (IQA). Mesmo nesses casos, observa-se que os estados brasileiros não adotaram o mesmo padrão e ainda não têm uma frequência fixa quanto à realização de coletas de dados em campo. De acordo com os levantamentos realizados acerca dos trabalhos relacionados ao tema, a serem apresentados no Capítulo 3, o sensoriamento remoto poderá contribuir como uma ferramenta mais acessível para auxiliar no monitoramento da qualidade de água das bacias hidrográficas, com base na predição dos parâmetros físicos e químicos opticamente ativos, ou seja, que atuem e modifiquem o comportamento espectral da água.

Nesse contexto, visando ao sensoriamento remoto com base em imagens das bandas do Satélite Landsat 8 OLI associadas às técnicas de regressão linear, regressão LASSO e tecnologias de desenvolvimento Web, além dos dados limnológicos do reservatório de Três Marias/MG, este trabalho propõe a criação de uma plataforma computacional que permitirá realizar predições de parâmetros limnológicos de reservatórios hídricos, apresentando um estudo de caso para o Reservatório de Três Marias - MG. Vale destacar que a metodologia aplicada pode ser extendida a outros reservatórios, desde que se tenha dados limnológicos históricos e informações espectrais disponíveis. Essa proposta visa ao aumento da acessibilidade ao controle da qualidade água por empresas usuárias dos recursos das bacias hidrográficas, pesquisadores e consumidores em geral.

1.4 Contribuições da Pesquisa

As principais contribuições já alcançadas durante a realização da pesquisa são destacadas a seguir:

- a) Criação de um banco de dados com dados limnólogicos e informações espectrais do satélite Landsat 8 OLI relacionadas ao reservatório de Três Marias/MG;
- b) Utilização das ferramentas da plataforma GEE (Google Earth Engine), uma plataforma geoespacial aberta que disponibiliza imagens de satélite que já incluem as etapas de pré-processamento;
- c) Criação da plataforma Web para a predição de parâmetros limnólogicos turbidez e sólidos totais no Rservatório de Três Marias - MG;

1.5 Estrutura do Trabalho

Após esta introdução, os capítulos 2 e 3 apresentam, respectivamente, uma revisão das principais técnicas utilizadas e trabalhos relacionados a esta pesquisa. O Capítulo 4 descreve a metodologia adotada e no Capítulo 5 são mostrados os resultados obtidos e discussões acerca da metodologia proposta. Por fim, o Capítulo 6 apresenta as considerações finais e propostas de continuidade da pesquisa.

2 REFERENCIAL TEÓRICO

Neste capítulo serão apresentados os principais fundamentos teóricos relativos aos objetos de estudo e técnicas utilizadas neste trabalho, dentre eles, a composição das águas em bacias hidrográficas, características do satélite Landsat 8 OLI, sensoriamento remoto para a predição de parâmetros limnológicos, as técnicas de regressão linear e LASSO (*Least absolute shrinkage and selection operator*), o *Google Earth Engine* e as métricas estatísticas aplicadas à validação.

2.1 Limnologia da água

Segundo (TUNDINISI J. G.; TUNDINISI, 2008), a limnologia é o estudo científico do conjunto das águas continentais em todo o planeta, incluindo lagos, represas, rios, lagoas, costeiras, áreas pantanosas, lagos salinos e, também, estuários e áreas pantanosas em regiões costeiras.

Processos e mecanismos químicos que ocorrem em águas interiores continentais dependem muito da geoquímica dos solos das bacias hidrográficas pois os sistemas aquáticos interagem essas bacias, os seus diversos subsistemas e componentes e, seu entorno. Conforme (BORMAN F.; LIKENS, 1979), a limnologia é o estudo integrado de bacias hidrográficas e sua relação com lagos, rios, represas e áreas alagadas. As características das bacias hidrográficas determinam, por exemplo, a origem do material que contribui para a formação e o funcionamento de lagos, rios e represas.

Entre os vários conceitos da limnologia, pode-se destacar (BALDI, 1949) e (TUNDI-NISI J. G.; TUNDINISI, 2008) que a definiram como a ciência que estuda as inter-relações de processos e métodos pelos quais a matéria e a energia são transformadas em um lago. Em todas essas definições, deve-se levar em conta dois aspectos fundamentais, o descritivo e o funcional bem como a síntese entre esses aspectos. Em uma abordagem ampla e sintética, a limnologia é a ciência das águas interiores estudadas como ecossistemas, os quais constituem uma unidade natural que consiste de componentes vivos (bióticos) e não vivos (abióticos), pertencentes a um sistema de fluxo de energia e ciclos de materiais (TUNDINISI J. G.; TUNDINISI, 2008).

Na análise estrutural, dois aspectos básicos encontram-se incluídos: a descrição dos componentes abióticos e suas propriedades (fatores físicos e químicos, concentrações, intensidades) e, a avaliação das comunidades bióticas (composição de espécies, abundância, biomassa, ciclos de vida).

A análise das inter-relações funcionais em um ecossistema inclui a investigação dos elementos responsáveis pelos ciclos de materiais, os processos dinâmicos nos sistemas abióticos, as relações dos organismos com os fatores ambientais e as relações dos organismos entre si.

2.2 Aspectos gerais das características da água em bacias hidrográficas

A água possui uma estrutura molecular composta por dois átomos de hidrogênio e um de oxigênio (H2O), onde cada átomo de hidrogênio liga-se covalentemente ao átomo de oxigênio e as moléculas se mantêm juntas pela atração mútua de átomos carregados positiva e negativamente (CRISTINA B. S. M; CLECI TERESINHA, 2013).

A água possui uma densidade relativamente elevada, cerca de oitocentas vezes superior à densidade do ar, a qual varia em funcão da temperatura, concentração de substâncias dissolvidas e da pressão.

A cor e a turbidez da água são características físicas importantes das bacias hidrográficas na medida em que interferem diretamente na penetração de luz dentro dos corpos d'água e, consequentemente, são essenciais para a produção da fotossíntese, garantindo a existência da biota aquática. A coloração da água é determinada a partir da reflexão da luz incidente e é denominada cor real quando o meio aquático possui substâncias dissolvidas que lhe conferem uma determinada cor ou denominada cor aparente, quando reflexos de outros materiais, como algas e rochas, definem uma coloração específica, própria dos materiais predominantes no corpo d'água.

A água, em contato com diferentes minerais e o substrato rochoso, provoca a solubilização. O processo de intemperismo químico das rochas pela água é de dissolução, decomposição e hidratação. Esse processo é evidenciado devido à presença dos ácidos orgânicos e inorgânicos e, ainda, pelo aumento da temperatura. A água também possui gases (como oxigênio e o dióxido de carbono) e sais minerais dissolvidos, o que favorece a realização da fotossíntese e a existência de organismos vivos nos diferentes ambientes (CRISTINA B. S. M; CLECI TERESINHA, 2013).

O pH da água é uma característica importante a ser observada, já que muitas das reações químicas que ocorrem na água são afetadas pelo seu valor. As comunidades aquáticas são diretamente afetadas pelo pH da água. Os peixes, por exemplo, só conseguem sobreviver em águas com pH entre 5 e 9 e apresentam produtividade máxima com pH entre 6,5 e 8,5 (BRAGA, 2005) e (CRISTINA B. S. M; CLECI TERESINHA, 2013).

Existem mais algumas substâncias que são muito importantes entre as características da água, como as substâncias químicas e os metais tóxicos, o oxigênio, a alcalinidade, os sólidos totais dissolvidos (cloreto, sulfatos, entre outros), dureza e matéria orgânica.

Os diferentes ambientes aquáticos abrigam uma grande diversidade de espécies de organismos que vivem interligados entre si numa complexa trama ecológica. As características biológicas da água dependem exclusivamente das características físicas e químicas dos ambientes onde se encontra.Muitos dos organismos que compõem a comunidade aquática são denominados bioindicadores ou indicadores biológicos, pois a presença ou ausência desses organismos pode indicar boa ou má qualidade da água.

As alterações genéticas, fisiológicas, ecológicas, comportamentais, morfológicas ou bi-

oquímicas podem revelar a presença de poluentes ou stress no ecossistema e, ainda, o tempo de ocorrência dessa perturbação no meio. Dentre alguns exemplos, pode-se citar os bioindicadores como fungos, algas e liquens (CRISTINA B. S. M; CLECI TERESINHA, 2013).

2.2.1 Composição de bacia hidrográfica - Conceitos fundamentais

Segundo (BARRELLA, 2002) e (CRISTINA B. S. M; CLECI TERESINHA, 2013), uma bacia hidrográfica é um conjunto de terras de drenagem de um rio e seus afluentes, formados a partir das regiões mais altas do relevo por divisores de água, onde as águas das chuvas podem escoar superficialmente ou infiltrar no solo para a formação de nascentes e do lençol freático. As águas superficiais escoam para as partes mais baixas do terreno, formando riachos e rios, enquanto as cabeceiras são formadas por riachos que brotam em terrenos íngremes nas serras e montanhas. À medida que as águas dos riachos seguem seu curso, juntam-se a outros riachos, aumentando o volume e formando os primeiros rios. Esses pequenos rios continuam seus trajetos recebendo água de outros tributários, formando rios maiores até desembocar no oceano. Assim, uma bacia hidrográfica é um meio natural espacialmente definido, cujos elementos físicos, biológicos e socioeconômicos mantêm relações dinâmicas entre si (CRISTINA B. S. M; CLECI TERESINHA, 2013).

O conceito de uma bacia hidrográfica pode ser exemplificado utilizando-se um telhado em meia-água de uma pequena casa, no qual a cumeeira é o divisor de águas: o telhado funciona como um sistema de captação de água (a bacia propriamente dita) e a calha é o rio. A figura 1 mostra um telhado de casa como uma representação didática de uma bacia hidrográfica, onde toda a água da chuva escorre para um único rio que, por sua vez, deságua no mar ou em lagos (PIAVA, 2009).

Figura 1 – Representação didática de uma bacia hidrográfica.



Fonte: Adaptado de (PIAVA, 2009)

A Figura 2 mostra a composição de uma bacia hidrográfica e seus elementos básicos constituídos pelas nascentes, pelos divisores de águas, também chamados de divisores topográficos, e os atributos dos cursos d'água, principais e secundários, denominados afluentes e subafluentes. Todo rio, por menor que seja, está inserido em uma bacia hidrográfica ((ANA), a).

A delimitação das bacias hidrográficas é realizada através das cartas topográficas. O



Figura 2 – Delimitação de uma bacia hidrográfica.

Fonte: Adaptado de ((ANA), a)

recorte da bacia é realizado através dos seus pontos mais altos, os chamados divisores topográficos ou divisores de água, os quais captam as águas pluviais e as desviam para um dos cursos d'água da bacia (CRISTINA B. S. M; CLECI TERESINHA, 2013).

2.2.2 Bacia Hidrográfica do Rio São Franscisco

A região hidrográfica denominada Alto São Francisco, em Minas Gerais, compreende as nascentes dos rios São Francisco, Pará, Paraopeba, Indaiá, Borrachudo e seus afluentes até o Reservatório de Três Marias como também à Bacia do Rio Abaeté, localizada à jusante do referido reservatório. A área de drenagem dessa região hidrográfica é de 57.097 km² e representa 24% da área de drenagem da Bacia do São Francisco, em Minas Gerais.

O Rio São Francisco nasce na Serra da Canastra, no município de São Roque de Minas, em Minas Gerais, a 1.460m de altitude. Após percorrer 570 km, é barrado formando o Reservatório de Três Marias. Os principais afluentes do Rio São Francisco à montante do Reservatório de Três Marias, pela margem direita são o Rio São Miguel, Ribeirão Santana, Rio Pará, Rio Paraopeba e Ribeirão da Extrema Grande e, pela margem esquerda, Rio Samburá, Rio Ajudas, Rio Bambuí, Ribeirão Jorge Grande, Rio Marmelada, Rio Indaiá e Rio Borrachudo. À jusante da barragem, pela margem esquerda, encontra-se o rio Abaeté (2010 ATLAS DIGITAL DAS áGUAS DE MINAS,).

De acordo com as Unidades de Planejamento e Gestão de Recursos Hídricos (UPGRH) estabelecidas pela Deliberação Normativa do CERH/MG nº 06/2002 e suas alterações, essa região hidrográfica contém quatro UPGRHs, quais sejam: SF1- Nascentes do Rio São Francisco até a confluência com o Rio Pará; SF2 - Bacia Hidrográfica do Rio Pará; SF3 - Bacia Hidrográfica do Rio Paraopeba e SF4 - Entorno do Reservatório de Três Marias (2010 ATLAS DIGITAL DAS áGUAS DE MINAS,). Conforme estudos realizados no âmbito do programa HIDROTEC (HIDROTEC,), a área de drenagem da Bacia Hidrográfica do Rio São Francisco em território mineiro corresponde a 234.554 km². Ocupa o primeiro lugar em termos de produção de água (vazões médias e mínimas) e contribui com 44,0% da vazão mínima produzida no Estado. Em termos de produtividade hídrica ocupa, no *ranking* estadual, o sexto lugar (2010 ATLAS DIGITAL DAS áGUAS DE MINAS,). A Figura 3 mostra a vista aérea do Reservatório de Três Marias - MG com destaque para sua dimensão, presença de vegetação e nuvens.



Figura 3 – Vista aérea do Reservatório de Três Marias.

Fonte: Adaptado de (2010 ATLAS DIGITAL DAS áGUAS DE MINAS,)

2.3 O Satélite Landsat 8 OLI

No início do ano de 2013, na Base Aérea de Vandenberg, Califórnia, foi lançado o Landsat 8. Um novo satélite artificial que consistia em dois instrumentos científicos, o OLI (*Operational Land Imager*) e o TIRS (*Thermal Infrared Sensor*). A partir desses sensores, foi possível obter uma cobertura sazonal da massa terrestre global com uma resolução espacial de 30 m que envolve os espectros visível, NIR (Infravermelho próximo) e SWIR (Ondas curtas do infravermelho e infravermelho termal). O Landsat 8 possuia uma expectativa de vida útil de 5 anos, a qual poderia ser extrapolada (NASA 2019, a). As características do Landsat 8, segundo (INPE 2013,), são apresentadas na Tabela 1.

O sensor OLI é do tipo *push-broom* com um telescópio de quatro espelhos e quantização de 12 bits, onde os dados são recolhidos das bandas espectrais visíveis, do infravermelho próximo e de ondas curtas, além da banda pancromática. A partir daí, são colhidos dados para duas novas bandas, a costeira/aerossol e cirrus, que são as bandas 1 e 9. Já o sensor TIRS obtém os dados das bandas infravermelho adicionais, as bandas 10 e 11 (NASA 2019, d). A Tabela 2 apresenta as características espectrais do Landsat 8, incluindo as

Instrumento/Sensor OLI e TIRS					
Instituição responsável	NASA				
País	Estados Unidos				
Situação atual	Ativo				
Data de lançamento	Data de lançamento $02/11/2013$				
Altitude $705 \ km$					
Inclinação	98,2°				
Órbita	Circular, Polar, Heliossíncrona				
Faixa imageada	$185 \ km$				
Tempo de duração de órbita 90 minutos					
Horário de passagem $10 \ h/12 \ h$					
Período de revista 16 dias					
Resolução espectral	Pancromático: 15 $m \mid$ Multiespectral: 30 m				
Resolução espectral	Termal: 100 m reamostrada para 30 m				
Fonte: Adaptado de (INPE 2013,).					

Tabela 1 – Características do Satélite Landsat 8 OLI

	Tabela 2 –	Bandas	espectrais	do	Satélite	Landsat-8	OLI
--	------------	--------	------------	----	----------	-----------	-----

Distância (m)	Faixa espectral	Comprimento de onda (μm)	Bandas
30	Coastal/Aerosol	0,435 - 0,451	Banda 1
30	Blue	0,452 - 0,512	Banda 2
30	Green	0,533 - 0.590	Banda 3
30	Red	0,636 - 0,673	Banda 4
30	NIR (infravermelho próximo)	0,851 - 0,879	Banda 5
30	SWIR-1 (ondas curtas)	1,566 - 1,651	Banda 6
30	SWIR-2 (ondas curtas)	2,107 - 2,294	Banda 7
15	Pan (tons de cinza)	0,503 - 0,676	Banda 8
30	Cirrus (nuvens de alta altitude)	1,363 - 1,384	Banda 9
100	TIR-1 (infravermelho termal)	10,60 - 11,19	Banda 10
100	TIR-2 (infravermelho termal)	11,50 - 12,51	Banda 11

Fonte: Adaptado de (NASA 2019, c).

novas bandas dos sensores OLI e TIRS, relacionando-as com o comprimento de onda e a distância da resolução espectral.

O Satélite Landsat 8 adquire 725 imagens por dia, o que aumenta a probabilidade de capturar cenas da superfície terrestre sem nuvens. As imagens captadas possuem resoluções espaciais pancromáticas de 15 metros e multiespectrais de 30 metros ao longo de 185 km de largura, capaz de distinguir centros urbanos, fazendas, florestas, etc. Possui melhor relação sinal/ruído em comparação com os satélites Landsat anteriores, o que proporciona melhor desempenho e maior confiabilidade (NASA 2019, c).

Das 11 bandas presentes no Landsat 8 OLI e TIRS, apenas as que possuem comprimentos de ondas curtos percebem a luz visível, ou seja, as cores reais que o ser humano é capaz de visualizar em uma imagem de satélite, as quais uma pequena parte do espectro que pode ser captado através de seus sensores (NASA 2019, c).

2.3.1 Bandas espectrais

O Landsat 8 possui no total 11 bandas espectrais, sendo as bandas 1, 10 e 11 provenientes da inserção dos sensores OLI e TIRS. A seguir, são detalhadas as informações de cada banda, segundo (NASA 2019, b):

- Banda 1: Capta as cores azul e violeta. A cor azul é difícil de ser captada do espaço, pois é facilmente espelhada por pequenos pedaços de poeira, moléculas de água ou ar e, por isso, os objetos muito distantes se tornam azulados, por exemplo, o céu. É conhecida como banda costeira/aerossol porque é comumente utilizada para obter imagens de águas rasas e rastrear partículas finas, como poeira e fumaça.
- Bandas 2, 3 e 4: Visíveis em azul, verde e vermelho, respectivamente, ou seja, definem a imagem no espaço de cores RGB (*Red, Green, Blue*).
- Banda 5: Captura o NIR que é refletida pela vegetação verde. Ao ser comparada com outras bandas, torna possível obter o NDVI (Índice de Vegetação da Diferença Normalizada) capaz de medir a saúde das plantas com maior precisão. O NDVI é um indicador gráfico simples que pode ser usado para analisar medições de sensoriamento remoto, avaliando se o alvo observado contém ou não vegetação verde viva .
- Bandas 6 e 7: Captam o SWIR para distinguir a terra úmida da terra seca, rochas e solos que usualmente se tornam semelhantes em outras bandas.
- Banda 8: Corresponde à banda pancromática que funciona como um filme preto e branco que ao invés de coletar cores visíveis separadamente, combina-as em um canal. Permite mais visualização de luz se comparada às outras bandas, por isso é mais nítida e possui uma resolução de 15 metros.
- Banda 9: Utilizada apenas para a captação de nuvens, por isso o nome Cirrus, que se refere às nuvens altas e finas. O comprimento de onda mínimo capturado está na faixa de 1370 ± 10ηm. Essa faixa é quase totalmente absorvida pela atmosfera, logo, o solo é pouco visível, o que garante que a visibilidade de um objeto indica que ele está refletindo muito intensamente e/ou está acima da atmosfera.
- Bandas 10 e 11: Captam o infravermelho termal, ou seja, elas captam calor, no caso, o calor do próprio solo.

Para obter características de uma determinada região, pode-se combinar as diferentes bandas espectrais e, assim alcançar melhores resultados para o estudo de uma determinada imagem (NASA 2019, b).

2.4 Sensoriamento Remoto

No contexto deste trabalho, o sensoriamento remoto baseia-se na aquisição de dados em diversas faixas do espectro eletromagnético, obtidos a partir de sensores acoplados a satélites, sendo que a análise dessas informações pode ser realizada visualmente ou por meio de processamento digital de imagens. O sensoriamento remoto explora a relação entre as características do objeto alvo e a energia eletromagnética refletida, emitida ou espalhada por ele. Através da identificação do comportamento espectral de superfícies como solo, água e vegetação, é possível relacionar informações qualitativas e quantitativas que permitam a caracterização desses alvos (JENSEN, 2011) e (GUIMARÃES, 2019). Assim, cada tipo de material ou superfície apresenta características únicas de comportamento espectral, ou seja, cada alvo é caracterizado pela sua assinatura espectral, a qual está diretamente ligada às suas propriedades físicas, químicas, biológicas ou geométrica (NOVO, 2010) e (GUIMARÃES, 2019). A Figura 4 mostra os diferentes comportamentos espectrais de alvos da superfície terrestre.





Fonte: (FLORENZANO, 2002).

A Figura 4 destaca as diferenças quanto ao comportamento espectral dos alvos terrestres. Enquanto o comportamento espectral dos solos e vegetação indica a existência de reflectância da energia eletromagnética ao longo de todo espectro de frequências, nos casos da água limpa e da água turva, os gráficos indicam que há energia refletida apenas na região do vísivel até próximo à regiao do infravermelho, o que permite concluir que essa é a região espectral mais adequada ao sensoriamento remoto nos estudos de corpos hídricos.

Em geral, a aquisição de dados para o sensoriamento remoto é feita por sensores e sistemas acoplados a satélites, drones, aviões, etc., os quais são responsavéis pela conversão

da radiação eletromagnética (REM) em um registro na forma de imagem ou gráfico que permite a interpretação do comportamento espectral do material de estudo (NOVO, 2010) e (GUIMARÃES, 2019).

Existem quatro características principais relativas à resolução de um sensor remoto que definem sua escolha e aplicação: resolução temporal, espectral, espacial e radiométrica.

A resolução temporal de um sensor consiste no intervalo de tempo entre dois imageamentos sucessivos ou duas aquisições dos dados (NOVO, 2010) e (GUIMARÃES, 2019). Esse parâmetro é essencial para estudos de processos cíclicos e de monitoramento ambiental, pois a análise de dados temporais auxilia no entendimento de como o elemento ou fenômeno está se comportando no decorrer do tempo (JENSEN, 2011) e (GUIMARÃES, 2019).

A resolução radiométrica está relacionada à sensibilidade do sensor em detectar, registrar e reproduzir as variações nos níveis da radiação eletromagnética refletida, emitida ou retroespalhada pelo objeto alvo (JENSEN, 2011), (NOVO, 2010) e (GUIMARÃES, 2019). Segundo (CRóSTA, 1992), (MENESES, 2012) e (GUIMARÃES, 2019), a resolução radiométrica é definida pela quantização ou seja, a quantidade de níveis de cinza usada para expressar os dados coletados pelo sensor. Dessa forma, resoluções radiométicas mais elevadas aumentam a probabilidade de se obter informações mais precisas acerca do fenômeno analisado (JENSEN, 2011) e (GUIMARÃES, 2019).

A resolução espectral de uma imagem é definida pelo número de bandas espectrais de um sistema sensor e pelos comprimentos de onda cobertos por cada banda (CRóSTA, 1992) e (GUIMARÃES, 2019). Além do número e largura das bandas, a posição das bandas ao longo do espectro eletromagnético também é um importante aspecto a ser considerado quanto à resolução espectral para instrumentos de sensoriamento remoto. Logo, quanto maior o número de bandas situadas em diferentes em regiões do espectro eletromagnético e com faixas mais estreitas de comprimento de onda, maior será a resolução espectral do sistema sensor (MENESES, 2012) e (GUIMARÃES, 2019).

A resolução espacial é definida como sendo a capacidade do sensor visualizar os alvos terrestres. Segundo (JENSEN, 2011) e (GUIMARÃES, 2019), há uma relação entre o tamanho de um objeto ou área a ser identificada e a resolução espacial de um sistema de sensoriamento remoto. Conforme (CRóSTA, 1992) e (GUIMARÃES, 2019), quanto menor for o objeto, maior deverá ser a resolução espacial da imagem. A altura do sensor remoto utilizado em relação ao alvo define o que se convencionou chamar de nível de aquisição de dados. Esses níveis dependem do veículo ou sistema de suporte (satélite, aviões, drones, dentre outros) para a operação de um sistema sensor (NOVO, 2010) e (GUIMARÃES, 2019).

Com base nessas características, neste trabalho trabalho foi utilizado o sensor OLI que está acoplado ao satélite Landast 8, a escolha deste sensor se deve principalmente pela resolução radiométrica de 12 bits, e a resolução espectral com muitas bandas disponíveis para fazer a predição da qualidade da água de reservatórios hídricos, e também porque muitos trabalhos relacionados a este tema usou o Landsat 8 OLI e todos eles obtiveram bons resultados.

2.4.1 Sensoriamento remoto da água

Para trabalhar com sensoriamento remoto para a análise de ambientes aquáticos é necessário entender que o fluxo de energia radiante que atravessa a interface ar/água está sujeito a dois processos principais: absorção e espalhamento. Esses processos são quantificados através de coeficientes específicos que, por sua vez, são conhecidos como propriedades ópticas inerentes do corpo d'água. Essas propriedades dependem somente do meio em questão, sendo os valores dos coeficientes de absorção e espalhamento diretamente relacionados à presença, tipo e concentração de substâncias chamadas compostos opticamente ativos (COAs), os quais afetam o espectro de absorção e espalhamento da água pura, podendo ser formados por partículas em suspensão, organismos vivos e substâncias orgânicas dissolvidas.

De acordo com (JENSEN, 2011) e (GUIMARÃES, 2019), em estudos que visam à aplicabilidade de sensoriamento remoto para corpos d'água, é primeiramente útil entender como a água pura seletivamente absorve e/ou espalha a radiação incidente ou a luz solar descendente na coluna d'água. Ainda, é importante considerar como a luz incidente é afetada quando a coluna da água não é pura, mas contém materiais orgânicos e inorgânicos.

Segundo (JENSEN, 2011) e (GUIMARÃES, 2019), a característica mais notável do comportamento espectral da água pura é que a quantidade mínima de absorção e espalhamento da luz incidente na coluna d'água ocorre na região do comprimento de onda da cor azul (0,4 a 0,5 μ m), com o valor mínimo localizado em aproximadamente 0,46 – 0,48 μ m.

As substâncias presentes na água são determinantes para sua análise, pois determinam a composição da água no momento em que é feita a coleta de dados. Assim, em termos do monitoramento por sensoriamento remoto, os comportamentos espectrais dessas substâncias devem ser considerados (VILELA, 2010) e (GUIMARÃES, 2019).

Segundo (NOVO, 2010) e (GUIMARÃES, 2019), corpos hídricos que apresentam partículas inorgânicas em suspensão (sólidos dissolvidos e em suspensão na água) tendem a apresentar curvas de comportamento espectral superiores à da água pura. Dessa forma, quanto maior a concentração de sedimentos em suspensão na água, maior será sua reflectância, tendo em vista que esses componentes aumentam o coeficiente de espalhamento do volume de água.

A Figura 5 ilustra a reflectância espectral da água clara pura e da água contendo diversas concentrações de sólidos em suspensão.

Segundo (JENSEN, 2011) e (GUIMARÃES, 2019), a faixa de comprimento de onda do



Figura 5 – Resposta espectral da água contendo sólidos em suspensão.

Fonte: (JENSEN, 2011).

infravermelho próxima de 714 a 880η m pode ser utilizada para determinar a quantidade de sedimentos em suspensão nas águas em que eles sejam os constituintes predominantes, tendo em vista que, nesse intervalo, os comportamentos das curvas das diferentes concentrações de sedimentos suspensos apresentam comportamentos semelhantes.

2.5 Google Earth Engine

O Google Earth Engine (GEE) é um serviço que possibilita realizar o processamento geoespacial em escala, com a tecnologia do Google Cloud Platform (A GOOGLE EN-GINE EARTH,). Basicamente, o objetivo do GEE é fornecer uma plataforma interativa para o desenvolvimento de algoritmos geoespaciais em escala, viabilizar a ciência de alto impacto orientada por dados e viabilizar o tratamento de informações envolvam grandes conjuntos de dados geoespaciais.

O GEE permite que os usuários executem algoritmos em imagens georreferenciadas, as quais se encontram armazenados na infraestrutura do *Google Comece* (B GOOGLE ENGINE EARTH,). A API (*Application Programming Interface*) do GEE fornece uma biblioteca de funções que pode ser aplicada aos dados para exibição e análise. O catálogo de dados públicos disponibilizado contém uma grande quantidade de imagens públicas e conjuntos de dados vetoriais e ativos privados também podem ser criados nas pastas pessoais dos usuários. O GEE pode ser manipulado através das linguagens de programação JavaScript e Python.

Além de possibilitar o processamento geoespacial, o GEE permite criar aplicações que são interfaces dinâmicas dos usuários, compartilháveis para análises do GEE. Com os apps, usuários especialistas e não especialistas podem, além de contar com os recursos analíticos do GEE, usar os elementos da interface de forma simples e aproveitar o catálogo de dados. Esses aplicativos podem ser acessados a partir de uma URL específica, gerada no momento da publicação. Não é necessária uma conta no GEE para visualizar ou interagir com os aplicativos publicados, os quais se encontram disponíveis em uma *App Gallery* específica do usuário, por exemplo, USERNAME.users.earthengine.app (C GOOGLE ENGINE EARTH,).

2.6 Machine learning

O aprendizado de máquina introduziu um novo paradigma para os meios digitais programáveis. Enquanto na programação clássica, os especialistas inserem regras através de programas computacionais e dados, a serem processados de acordo com essas regras, para a obtenção de respostas, com o aprendizado de máquina, os dados e as respostas esperadas são, usualmente, inseridas para a formulação das regras. A partir daí, essas regras podem então ser aplicadas a novos dados para produzir novas respostas.

Um sistema de *machine learning* não é explicitamente programado mas requer uma fase de treinamento dos modelos, a qual é determinante para o seu desempenho. O aprendizado de máquina está intimamente relacionado à estatística matemática, porém possibilita lidar com grandes e complexos conjuntos de dados, por exemplo, milhões de imagens, cada uma consistindo em dezenas de milhares de pixels, para as quais a análise estatística clássica seria impraticável. Como resultado, o aprendizado de máquina e especialmente, o aprendizado profundo, exige comparativamente menor embasamento matemático se comparado às técnicas clássicas de processamento e tratamento de sinais (CHOLLET, 2018).

Os algoritmos de *machine learning* são ferramentas baseadas em conceitos estatísticos que permitem extrair informações relevantes de correlação e de comportamento de um conjunto de dados. Alguns desses algoritmos também possuem a característica de treinamento para obtenção de predições a partir em dados de entrada previamente obtidos e podem ser classificados quanto aos tipos de abordagem, técnica supervisionada e não-supervisionada.

Nos processos de predição de parâmetros, usualmente, o treinamento é feito usando operações estatísticas que têm como, objetivo definir funções matemáticas que melhor se ajustem aos dados prévios disponíveis, chamadas de técnicas de regressão.Pode-se fazer predições de qualquer natureza, como por exemplo, preços de produtos para lançamento no mercado, meteorologia, produção agrícola e, no caso deste trabalho, parâmetros limnólogicos de reservatórios hídricos. Neste trabalho são usadas a técnicas de regressão linear e regressão LASSO, ambas apresentadas nas Seções 2.6.1 e 2.6.2.

2.6.1 Regressão linear

Os modelos de regressão linear fazem parte de um conjunto de ferramentas cujo foco é a realização de inferências, na maior parte das vezes, causais. A inferência consiste em, a partir de evidências encontradas para um conjunto de amostras previamente conhecidas, realizar a generalização dos resultados. Ou seja, há um interesse em verificar a correlação entre duas ou mais variáveis e testar o quanto se pode confiar nas estimativas encontradas (CHEIN, 2019).

O modelo de regressão linear é dado pela equação (1):

$$Y_i = b_0 + \sum_{j=0}^p x_{i,j} b_j + e_i$$
(1)

em que Y_i representa o valor da variável dependente, ou seja, o parâmetro que está sendo predito, $x_{i,j}$ indicam as amostras coletadas das variáveis independentes de entrada sendo $i = 0, 1, 2, ... \in \mathbb{R}$ e $j = 0, 1, 2, ..., p \in \mathbb{R}$, b_j são os coeficientes associados às variáveis independentes a serem ajustados pelo modelo de aprendizagem, o termo b_0 referese à polarização do modelo e e_i é a componente que contém os erros associados a cada observação que representa a variabilidade não explicada pelo modelo, também denominada resíduo.

O modelo da equação (1) pressupõe que todos os erros são independentes e identicamente distribuídos, com média 0 e variância constante. Além disso, que a relação entre a variável dependente e cada uma das variáveis independentes é linear nos parâmetros. Usualmente, também assume-se que os erros têm distribuição normal (HASTIE T., 2020) e (JUNIOR, 2021).

Para encontrar as estimativas dos parâmetros do modelo linear mostrado na equação (1), frequentemente, é utilizado o método dos mínimos quadrados (MMQ). De acordo com (STOCK J. H.; WATSON, 2010) e (CHEIN, 2019), o estimador de mínimos quadrados ordinários escolhe os coeficientes de modo que a linha de regressão estimada fique o mais próxima possível dos dados observados. Para isso, deve-se encontrar a solução da função objetivo indicada em (2), a qual permitirá que a soma dos quadrados dos erros seja minimizada:

$$minimize(b_0, b_j) = \sum_{i=1}^n (y_i - b_0 - \sum_{j=1}^p b_j x_{i,j})^2$$
(2)

2.6.2 Regressão LASSO

Para alguns temas de pesquisa, existe uma grande quantidade de dados, o que exige que técnicas mais avançadas de *machine learning* sejam mais indicadas. No caso dos métodos de regressão, isto pode significar uma base de dados com alta dimensionalidade, o que ocorre quando o número p de covariáveis (variáveis preditoras) seja maior do que o tamanho amostral n. Nesse ponto, técnicas clássicas de estimação de parâmetros como o método dos mínimos quadrados tornam-se problemáticas, pois quando n < p a solução que minimiza a soma do quadrado dos erros não é única (HASTIE T., 2015) e (JUNIOR, 2021). Quando o número de covariáveis é muito grande, também é difícil o uso de técnicas clássicas de seleção de covariáveis como, por exemplo a *regressão stepwise* (JAMES G., 2013) e (JUNIOR, 2021), pois demandam tempo e custo computacionais muitas vezes inviáveis.

A regressão LASSO (*Last absolute shrinkage and selection operator*) é um método que foi criado para resolver o problema de alta dimensionalidade, pela adição de uma penalização na equação de mínimos quadrados. O LASSO minimiza a soma de quadrados adicionada a um peso, ou penalização não negativa, de forma a criar esparsidade dentro do modelo, isto é, fazendo muitos coeficientes convergirem para zero. Dessa forma, o modelo LASSO, simultaneamente, estima os parâmetros e seleciona covariáveis para o modelo (JUNIOR, 2021).

O estimador LASSO no contexto da regressão linear é dado pela resolução do problema apresentado na equação (3):

$$\beta_l = minimize(\beta) \left(\sum_{i=1}^n (y_i - x_i\beta)^2 + \lambda \sum_{j=1}^p |\beta_j|\right), \lambda \ge 0$$
(3)

em que λ é um valor que deve ser escolhido previamente. Note que, se $\lambda = 0$, então o estimador LASSO será igual ao estimador de mínimos quadrados como na equação 2, isto é, o estimador de mínimos quadrados de β pode ser visto como um caso particular do estimador LASSO (RODRIGUES, 2018).

O estimador LASSO é basicamente o problema de minimização usual dos erros quadráticos acrescido de um penalização L_1 em relação aos parâmetros de posição do modelo de regressão. Um detalhe importante acerca da penalidade é que o parâmetro β_0 não é restringido pela penalização. De forma geral, esse problema não possui solução analítica como o problema de mínimos quadrados. Portanto, é necessário utilizar algoritmos para obter a estimativa LASSO em problemas reais (RODRIGUES, 2018).

A escolha de um valor λ é fundamental para que o método funcione adequadamente. Vale lembrar que, se $\lambda = 0$, então o método de estimação é simplesmente o método de mínimos quadrados e se for escolhido um valor extremamente grande para λ , então provavelmente o modelo não será sensível às covariáveis, isto é, só possuirá intercepto (RODRIGUES, 2018).

Um modo de escolher o valor de λ é pela validação cruzada, que é um método bastante fácil de entender e de implementar. Basicamente dividi-se, aleatoriamente, a amostra em k partes iguais ou aproximadamente iguais. Escolhe-se a primeira parte para os dados de validação e as demais como dados de treinamento. O LASSO será ajustado para $\lambda = \lambda_0$ utilizando-se dados de treinamento e esse modelo será aplicado para prever os dados de validação, quando será calculado o erro de predição. Esse mesmo procedimento é repetido mais k - 1 vezes para as outras partes restantes. Após terminar as k iterações, tem-se k erros de predição que definirão o erro médio de predição. Esse procedimento será feito para vários valores λ_0 distintos. Finalmente, escolhe-se o valor λ_0 que minimize o erro de predição médio (RODRIGUES, 2018).

(TIBSHIRANI, 1996) e (RODRIGUES, 2018) recomendam que k = 5 ou k = 10, mas quaisquer valores inteiros entre cinco e dez podem ser usados. Não há uma regra para escolha de k, se o número de observações deve for suficientemente grande as amostras disponíveis podem ser divididas entre um maior número de k partes.

É importante salientar que o método de escolha do valor de λ não depende do conhecimento do número de parâmetros do modelo e também não depende de uma estimativa do parâmetro de escala do modelo. Embora na situação tradicional em que se usa o estimador de mínimos quadrados, o número de parâmetros do modelo ajustado esteja bem definido e seja fácil achar uma estimativa para o parâmetro de escala, no caso do modelo LASSO isso não é trivial, pois λ também é uma quantidade que tem impacto no ajuste do modelo. Logo, o método exposto acima é proposto por (TIBSHIRANI, 1996) para contornar esses problemas (RODRIGUES, 2018).

O erro de predição relacionado a λ_0 é dado pela equação 4:

$$EP_{\lambda_0} = \frac{1}{n} \sum_{i=1}^k n_i EQM_i \tag{4}$$

em que n_i é o número de observações da i-ésima parte dos dados. Abaixo tem-se a equação 5, que mostra o erro quadrático médio dos dados de validação.

$$EQM_i = \frac{1}{n_i} \sum_{j=1}^{n_i} (y_j - y_j)^2$$
(5)

ou seja, EQM_i é o erro quadrático médio quando se usa a i-ésima parte dos dados como validação e o erro de predição é simplesmente uma média ponderada dos EQM's. Essa é a quantidade que definirá a escolha do valor de λ , quanto menor o erro de predição, mais adequado é o valor de λ (RODRIGUES, 2018).

Em geral, não se utiliza o valor de λ que minimiza o erro de predição. Em vez disso, utiliza-se o maior valor de λ cujo erro de predição esteja a um desvio padrão do erro de predição associado ao valor de λ que minimiza o erro de predição. (TIBSHIRANI, 1996) e (RODRIGUES, 2018) argumentam que como não se conhece a real forma da curva do erro de predição em termos do valor de λ , deve-se escolher o valor de λ como definido anteriormente, favorecendo um valor de λ maior e, consequentemente, um modelo mais simples.

2.7 Métricas Estatísticas

As métricas estatísticas são modelos matématicos que tem como objetivo medir a generalização de um modelo de aprendizagem de técnicas de machine learning, neste trabalho usou as métricas: Erro absoluto médio, erro percentual absoluto médio e coeficiente de determinação, as duas primeiras métricas tem como objetivo medir o erro entre os valores reais dos parâmetros limnológicos e os valores preditos pelo modelo de aprendizagem, e a ultima métrica é usada para medir o quão relacionadas estão as amostras em relação ao modelo de aprendizagem, essas métricas foram escolhidas pelo fato de medir de maneira mais prática o quão generalizado é o modelo de aprendizado.

2.7.1 Erro absoluto médio (MAE)

O MAE - *erro absoluto médio* é uma métrica que faz a análise do erro entre duas observações, neste trabalho essa métrica faz a comparação entre valores reais dos parâmetros limnológicos com os valores preditos pelo modelo de aprendizagem treinado pelas técnicas de regressão usadas neste trabalho, veja a equação 6, como calcular o MAE:

$$MAE(y, yp) = \frac{1}{N_{samples}} \sum_{i=0}^{N_{samples}-1} |y_i - y_{pi}|$$
(6)

onde y_i é o valor real, y_{pi} é o valor predito pelo modelo de aprendizagem e $N_{samples}$ o número de amostras usadas para treinar o modelo de aprendizagem.

2.7.2 Erro percentual absoluto médio (MAPE)

O MAPE - *erro percentual absoluto médio* é comumente usado como função de perda para problemas de regressão e na avaliação de modelos, devido à sua interpretação muito intuitiva em termos de erro relativo, veja a equação 7, como calcular o MAPE:

$$MAPE(y, yp) = \frac{1}{N_{samples}} \sum_{i=0}^{N_{samples}-1} \frac{|y_i - y_{pi}|}{|y_i|}$$
(7)

onde y_i é o valor real, y_{pi} é o valor predito pelo modelo de aprendizagem e $N_{samples}$ o número de amostras usadas para treinar o modelo de aprendizagem.

2.7.3 Coeficiente de determinação (R^2)

O coeficiente de determinação R^2 é uma métrica representa a proporção da variância da variável dependente (y) que foi explicada pelas variáveis independentes no modelo. Ele fornece uma indicação da qualidade do ajuste e, portanto, uma medida de quão bem as amostras não vistas provavelmente serão previstas pelo modelo, por meio da proporção da variância explicada (SCIKILEARN,), conforme a equação 8:

$$R^{2}(y, yp) = 1 - \frac{\sum_{i=0}^{n} (y_{i} - y_{pi})^{2}}{\sum_{i=0}^{n} (y_{i} - y_{m})^{2}}$$
(8)

onde y_i é o valor real, y_{pi} é o valor predito pelo modelo de aprendizagem e y_m é a média aritmética dos dados valores reais.

2.8 Conclusão

Este capítulo apresentou as bases teóricas fundamentais para o desenvolvimento da plataforma voltada à predição de parâmetros limnólogicos, proposta pelo trabalho, com destaque para os aspectos seguintes:

- O sensoriamento remoto a partir das imagens do Satélite Landsat 8 OLI, utilizado para a extração de informações espectrais a serem usadas no treinamento e teste para a validação dos modelos de aprendizagem para a predição dos parâmetros limnológicos;
- A plataforma GEE a ser utilizado para acessar as imagens do Landsat 8 OLI, já com as etapas de pré-processamento executadas nas imagens;
- As técnicas regressão linear e regressão LASSO a serem aplicadas para o treinamento e ajuste dos modelos de aprendizagem. As métricas estatísticas que permitirão validar os modelos obtidos;

As tecnologias citadas neste capítulo (Javascript, Python, Biblioteca de Machine Learning, etc.) serão aplicadas no desenvolvimento da plataforma Web para a predição dos parâmetros limnológicos de bacias hidrográgicas, a qual pode ser acessada através do link <<u>http://plattform.herokuapp.com/></u>.

3 TRABALHOS RELACIONADOS

O sensoriamento remoto é uma tecnologia que vem sendo utilizada para diversas finalidades. Em (MACHADO; BAPTISTA, 2016), é mostrado o uso dessa tecnologia para fazer o monitoramento do Índice de Estado Trófico de Carson no Lago Paranoá que sofre a eutrofização, principalmente, devido ao recebimento de esgoto doméstico de parte da população de Brasília. O autor usou os parâmetros fósforo total, clorofila-a e transparência da água, coletados em *in situ* e analisados em laboratório e, também, imagens do satélite Landsat TM 5 para obter informações espectrais do local sob análise. O trabalho inclui o método de correlação linear e mapas temáticos com o intuito de obter equações para predizer novos parâmetros, determinar o índice de estado trófico e acompanhar o nível de eutrofização do lago. Esse estudo mostrou que o lago vem melhorando, em relação a esse indicador, ao longo do tempo.

A série de satélites Landsat tem demostrado ser uma boa opção para o sensoriamento remoto para a análise da qualidade da água em relação a outros satélites disponíveis como o Meris, Sentinel 2A, dentre outros. Em (DESENVOLVIMENTO..., 2018), imagens dos satélites Landsat 5 e 7 foram utilizadas para fazer a predição de clorofila-a e transparência da água, no Lago Pampean na Argentina. A Banda 4 foi escolhida para fazer a correlação dos dados espectrais com os dados *in situ* para a transparência da água e o índice de vegetação por diferenças normalizadas (NDVI) para a clorofila-a. No ano de 2013, foi lançado o último satélite da série Landsat, denominado Landsat OLI 8.

Em (SILVA et al., 2021), os autores apresentam o desenvolvimento de uma proposta para a avaliação da qualidade da água de bacias hidrográficas com base em técnicas de processamento digital aplicadas a imagens de satélite. Dentre as técnicas de processamento de imagens utilizadas destacam-se a limiarização pelo Método de Otsu, binarização, expansão linear por saturação, Filtro Laplaciano, extração de características por meio de matrizes de co-ocorrência e classificação pelo Discriminante de Bayes. Essas técnicas também foram implementadas em uma plataforma computacional em ambiente MATLAB $(\mathbf{\hat{R}})$, responsável pela interface entre o sistema e os usuários. O sistema proposto apresentou uma taxa de sucesso aproximada de 70% quanto à classificação dos índices de qualidade da água.

Em (Khryashchev et al., 2018), os autores fazeram a detecção de objetos terrestres a partir de imagens do satélite Landsat OLI 8 aplicando redes neurais convolucionais (CNN). Foi utilizada a arquitetura da Rede U-NET para implementar o reconhecimento e a base de imagens denominada *Urban Atlas*. Esse atlas contém 21 classes, das quais foram consideradas 3 classes: água, floresta e agricultura. O algoritmo foi testado e comparado com marcações do local feitas por especialistas. Os resultados alcançados por esse projeto mostrou-se satisfatório, podendo ser estendido para outras aplicações. O resultado da classificação é de 81,7% para objetos como recursos hídricos, 92,3% para objetos da classe floresta e 96,1% para objetos da classe agricultura. O algoritmo considerado pode ser aplicado para a análise semântica de imagens do satélite, tais como alocação dos territórios das cidades, controle de edificações civis e outros.

Em (CHITERO et al., 2020), é apresentada uma proposta de avaliação do solo onde os autores tem como objetivo fazer o monitoramento para avaliar sua recuperação. Essa análise é feita por meio de atributos do solo e valores de referência, e utiliza uma rede neural artificial Perception Multicamadas (MLP) para a análise e classificação. O autor criou ainda uma interface gráfica onde o usuário pode fazer a interação com a rede e acompanhar todo o processo. Essa metodologia mostrou-se muito promissora pois apresentou um erro médio em torno de 0,0868 %, ou seja, a rede praticamente conseguiu classificar o solo corretamente em todos os testes realizados.

Em (Kupssinskü et al., 2019) são apresentados resultados de pesquisas sobre a eutrofização em dois diferentes locais do Brasil, um lago da Universidade Unisinos no estado do Rio Grande Do Sul e o Reservatório de Broa, no estado de São Paulo. Foram feitas coletas de amostras de água em *in situ* nos dois locais e análises de laboratório em relação aos parâmetros sólidos em suspensão e clorofila-a. Os valores do pH foram extraídos por meio uma sonda multiparamétrica, sendo esse parâmetro usado como informação na entrada de uma RNA *feedforward* para se obter mais informação para a validação dos dados. As imagens foram adquiridas a partir das bandas 2, 4, 8, 11 do satélite Sentinel-2 para a avaliação dos parâmetros sólidos em suspensão e clorofila-a e as imagens submetidas à correção atmosférica por meio do software SNAP 6.0. As imagens foram adquiridas em datas próximas às coletas *in situ* para uma melhor correlação com os dados de campo. Todas as amostras foram georreferenciadas, pois desde que a correlação é feita usando os dados espectrais da imagem de satélite, torna-se necessário conhecer o local exato de coleta.

Segundo os autores, os resultados alcançados em (Kupssinskü et al., 2019) foram considerados satisfatórios, sendo o erro quadrático médio (MSE) para clorofila-a próximo a 0,02223 %, para sólidos em suspensão 0,02251 %, o coeficiente de determinação (R^2) para clorofila-a próximo a 0,72 % e para sólidos em suspensão 0,72 %. Esses resultados mostraram que as redes neurais articiais (RNA) podem ser uma boa opção para predição de parâmetros da qualidade de água em relação aos métodos tradicionais de regressão, em função das não linearidades e complexidade nos dados de qualidade de água. Nesse trabalho, foram usadas as bandas de infravermelho como entradas da RNA, o que disponibilizou uma quantidade maior de informações e a melhoria da precisão em relação aos demais trabalhos disponíveis na literatura.

Em (MARINHO et al., 2021) busca-se fazer o monitoramento de sólidos suspensos (SSC) no Rio Negro, na bacia amazônica, com dados de sensoriamento remoto *in situ* e imagens de satélite Sentinel-2. Segundo os autores, a escolha desse rio foi definida para comprovar análise do monitoramento remoto em rios de águas negras. Foi identificado

que a maior quantidade de sólidos suspensos vinha do Rio Branco, um afluente do Rio Negro, constatada com base em dados de sensoriamento remoto *in situ*. A banda 4, correspondente à cor vermelha da região do visível, do satélite Sentinel-2 foi a melhor banda para a correlação dos dados. Nas etapas de processamento foram realizados a correção geométrica e o georreferenciamento das imagens do Sentinel-2 e, posteriormente, foram aplicados os métodos de regressão de distância ortogonal e regressão de mínimos quadrados ordinários. O primeiro mostrou que a estimativa de sólidos suspensos (SSC) usando bandas simuladas de Sentinel-2 MSI resulta em melhores valores de inclinação quando comparados aos obtidos por regressão por mínimos quadrados ordinários. Os resultados obtidos mostraram, ainda, que usar o Sentinel-2 para monitoramento de sólidos em suspensão é uma boa opção, pois conseguiu-se alcançar um coeficiente de determinação (R^2) superior a 0,86 e um erro inferior a 30% nos modelos ajustados.

Em (LOBO et al., 2021), busca-se trazer para os usuários informações da clorofila-a (Chl-a) e do índice de estado trófico (TSI) usando imagens do sentinel-2 MSI e dados coletados in situ. Foi implementado um aplicativo no Google Engine Earth (GEE) com uma interface interativa onde os usuarios podem adicionar informações e fazer o acesso aos resultados em vários reservatórios no Brasil e na América Latina. A metodologia proposta utiliza imagens do satélite Sentinel-2 corrigidas para efeitos atmosféricos e de reflexo solar, de forma a gerar uma coleção de imagens do Índice de Clorofila-a Diferença Normalizada (NDCI) para toda a série temporal de uma determinada área. Os dados NDCI recuperados das imagens são, então, comparados com Chl-a medido in situ. O NDCI é usado para estimar a concentração de Chl-a, com base em um modelo de ajuste não linear, e o índice TSI é processado com base em um modelo de árvore de decisão. A árvore de decisão TSI classifica cada pixel em cinco níveis para o Índice de Estado Trófico (Oligo, Meso, Eutrófico, Super e Hipereutrófico). A primeira etapa de processamento é a correção atmosférica baseada no método Satellite Invariant Atmospheric Correction (SIAC). Após a correção atmosférica, o efeito do reflexo do sol foi corrigido subtraindo a banda 12 (2190 nm) das bandas 2–11 e a próxima etapa consiste em mascarar os pixels com nuvens, aplicando-se os produtos Sentinel-2: Probabilidade de nuvem para a máscara de água e o produto Máscara de água do Joint Research Centre (JRC), disponíveis no GEE. Após todas as etapas de processamento descritas anteriormente, foi feito o cálculo do NDCI e geradas as imagens indicando a concentração de Chl-a no reservatório. Os dados in situ foram coletados e disponibilizados pela Companhia Ambiental do Estado de São Paulo (CETESB). Os resultados obtidos pelo modelo final foi um coeficiente de determinação (R^2) de 0.86 e um erro médio de porcentagem absoluta (MAPE - Erro percentual absoluto médio) próximo a 90%. A baixa acurácia indicada pelo valor MAPE pode estar relacionada ao fato de o algoritmo proposto apresentar maior incerteza para valores baixos de Chl-a ($< 5 \mu g / L$).

Em (SAGAN et al., 2020), os autores apresentam uma análise sobre a capacidade do

sensoriamento remoto para predição de parâmetros de qualidade da água, buscando verificar o potencial e limitações do mesmo. Os objetivos propostos por esse trabalho visam analisar as tendências atuais e os avanços em sensoriamento remoto da qualidade da água, identificar e avaliar uma variedade de métodos de estimativa populares em fontes, conjuntos de dados e as limitações dos sistemas atuais, métodos de estimativa e sugerir melhorias futuras. Para essas finalidades, foram coletadas mais de 200 conjuntos de dados sobre a qualidade da água, incluindo ficocianina de algas verdes azuis (BGA-PC), clorofila-a (Chla), oxigênio dissolvido (DO), condutividade específica (SC), matéria orgânica dissolvida fluorescente (fDOM), turbidez e sedimentos de poluição de 2016 a 2018 distribuídos em oito lagos e rios no meio-oeste dos Estados Unidos. Os dados foram emparelhados com espectros síncronos, imagens hiperespectrais montadas em tripés e dados de satélite. Foi usado o Google Engine Earth (GEE) para extrair as imagens dos satélites Landsat 8 OLI e Sentinel-2 MSI, em imagens coletadas entre maio de 2013 e dezembro de 2018, datas o mais próximas possível dos dados coletados em campo in situ. Foram usados dados de sensoriamento remoto em várias escalas de sensores proximais e baseados em satélite para estimar os principais indicadores da qualidade da água, incluindo parâmetros opticamente ativos, bem como variáveis não opticamente ativa, e várias técnicas de monitoramento, tais como, métodos de proporção, métodos empíricos, rede neural profunda (DNN) e meios de estimativa bio-óptica. Os resultados foram comparados com dados de campo coletados em vários riachos e lagos no meio-oeste dos EUA. Os métodos de aprendizado de máquina forneceram a melhor precisão geral para quase todos os estudos de caso. Os resultados indicaram que o aprendizado de máquina com base empírica e os métodos de aprendizado profundo superaram as taxas de bandas de frequência, a forma espectral e as técnicas de inversão bio-óptica.

O objetivo da proposta apresentada em (JIA; ZHANG; DONG, 2019) foi fazer o monitoramento rápido e automático de floração de cianobactérias no Lago Taihu na China, usando sensoriamento remoto e computação em nuvem através do processamento das imagens do satélite no Google *Engine Earth*. Foram usados os produtos MOD09GQ e MOD09GA que são informações espectrais de reflectância de superfície MODIS no Google *Engine Earth*. Nas etapas de processamento foram feitos os cálculos dos índices de algas flutuantes (FAI), usando as bandas infravermelho próximo (NIR) do MOD09GQ e uma banda de infravermelho de ondas curtas (SWIR) do MOD09GA, realizada a correção atmosférica usando todas as bandas SWIR do MOD09GA e o *land mask* e o mascaramento de nuvem para filtrar os pixels de terra e nuvens, respectivamente. Foi desenvolvido um fluxo de trabalho operacional de monitoramento da proliferação de cianobactérias com base no GEE, tendo sido processadas cerca de 7.000 imagens diárias MODIS e recuperados padrões espaciais e temporais de florações de cianobactérias no Lago Taihu, na China, ao longo de 19 anos (2000 a 2018). O fluxo de trabalho criado fornece uma base para o monitoramento de sensoriamento remoto automático de florescências de cianobactérias. Concluiu-se que a relação entre as características de floração de cianobactérias e fatores ambientais determinantes indicam que das perspectivas mensal quanto interanual, os fatores meteorológicos estão positivamente correlacionados com as características de floração de cianobactérias, mas quanto às cargas de nutrientes, eles são apenas positivamente correlacionados com as características de floração de cianobactérias de uma perspectiva interanual. Dessa forma, foi indicado que a redução do fósforo total juntamente com a restauração do ecossistema de macrófitas, seriam as estratégias de manejo de longo prazo necessárias para o Lago Taihu.

O GEE vem demostrando ser uma opção promissora para a predição e análise de parâmetros da qualidade da água, pois ele entrega grande parte das imagens de satélites já pré-processadas, como mostrado nas referências (LOBO et al., 2021), (SAGAN et al., 2020) e (JIA; ZHANG; DONG, 2019).

Em (BUMA; LEE, 2020), os autores buscaram fazer uma comparação entre as informações espectrais dos satélites Landsat 8 e Sentinel-2 para a análise e predição de clorofila -a (Chl a) no lago Chade na Africa. Foram usados quatro algoritmos para fazer a comparação entre eles, as bandas 2BDA e 3BDA, altura da linha de fluorescência (FLH) e o índice de clorofila de diferença normalizada (NDCI). A avaliação do algoritmo indicou eficiência na aplicabilidade dos algoritmos 3BDA e NDCI com um coeficiente de determinação (R^2) médio de 0,8, os algoritmos 3BDA e NDCI Chl a executaram com precisão com imagens S2 e L8. A análise de correlação mostrou que a banda 3BDA pode ser usada para estimar a concentração de Chl. Esta análise confirma a necessidade de uma análise de validação adicional entre as imagens do sensor, a fim de melhorar o desempenho na estimativa de Chl. Os resultados obtidos demonstraram a aplicabilidade potencial das imagens S2 e L8 como fontes de dados para mapear as concentrações de Chl.

O objetivo em (ZHU; MAO, 2021) é melhorar a precisão da predição do índice de estado trófico (TSI) em águas urbanas a partir de dados de qualidade da água coletados em *in situ* e a correlação com imagens de satélite Sentinel-2 MSI. Uma rede neural de retropropagação (BP-NN) foi aplicada para desenvolver o modelo de estimativa TSI usando sensoriamento remoto e dados de qualidade da água. O modelo foi treinado e validado por meio do TSI quantificado por cinco indicadores tróficos hídricos obtidos para o período entre 2018 e 2019. Toda a pesquisa foi desenvolvida na cidade de Gongqingcheng na China. Foram usados os softwares SNAP e ENVI para fazer as etapas de processamento das imagens do Sentinel-2, incluindo a reamostragem, reprojeção e a remoção de pontos de nuvem-pixel, entre outras. O desempenho BP-NN foi satisfatório com coeficiente de determinação (R^2) de 0,922, erro médio quadrático (RMSE) de 3,256, erro percentual médio absoluto (MAPE) de 2,494% e taxa de precisão de classificação de 86,364%.

Em (ANSARI; AKHOONDZADEH, 2020) os autores buscaram realizar o mapeamento e predição da salinidade da água do Rio Karun, no Irã, quanto ao nível de sanilidade que tem aumentado devido a alguns fatores, como por exemplo, condições climáticas severas 38

e fisiografia regional, fontes industriais, esgoto doméstico e urbano, irrigação de terras agrícolas, incubação de peixes, esgoto hospitalar e nível de maré alta do Golfo Pérsico. Esse estudo teve como objetivo construir modelos de regressão para verificar a salinidade da água através da relação entre a reflectância do Landsat-8 OLI e medidas in situ. Foram coletadas 102 medições de junho de 2013 a julho de 2018, e usada a análise de sensibilidade de Sobol para determinar a melhor combinação de bandas do ponto de vista de desempenho. Os resultados da análise de sensibilidade de Sobol revelaram que a banda 1- Coastal / Aerosol (0,433–0,453 µm), banda 2-Azul (0,450–0,515 µm) e banda 3 - Verde (0,525–0,600 µm) são as melhores combinações e mostraram também que a banda 2 do Landsat-8 OLI tem a correlação mais próxima com a salinidade. Foram aplicados os metódos Ordinary Least Square (OLS), Support Vector Regression (SVR) e Multi-Layer Perceptron (MLP). O número de camadas e nós na rede neural MLP é otimizado usando o Genético (GA), para verificar a correlação entre a sanilidade e as imagens do Landsat 8 OLI. O metódo que melhor correlacionou a salinidade e as imagens de satélite foi o Multi-Layer Perceptron (MLP) em conjunto com o Algoritmo Genético (GA), alcançando um coeficiente de determinação (R^2) de 0,78.

A dissertação apresentada em (GUIMARÃES, 2019), busca confirmar os métodos usados nos artigos (MACHADO; BAPTISTA, 2016), (DESENVOLVIMENTO..., 2018), (Khryashchev et al., 2018), (CHITERO et al., 2020), (Kupssinskü et al., 2019), (MARI-NHO et al., 2021), (BUMA; LEE, 2020), (ZHU; MAO, 2021) e (ANSARI; AKHOOND-ZADEH, 2020) com base na análise dos elementos limnologicos da qualidade da água, clorofila-a e sólidos em suspensão, da represa de Lobo em São Paulo. A ideia básica é utilizar esses parâmetros para análise da eutrofização da água, fazer a análise espectral da água na represa e a correlação da resposta espectral com os dados limnólogicos. Para realizar a correlação e a análise dos resultados foram utilizadas técnicas de processamento e extração de características das imagens, regressão linear e RNA *feedforward*. Essa dissertação traz um complemento importante que é a possibilidade de predizer componentes de qualidade da água que não são opticamente ativos a partir de componentes opticamente ativos, o que tornou possível predizer a temperatura, condutividade, PH e turbidez através dos valores de clorofila-a e sólidos totais em suspensão (SST - *Total Suspended Solids*).

Os resultados alcançados em (GUIMARÃES, 2019) foram satisfatórios com base nas métricas coeficiente de determinação (R^2) e o erro médio quadrático (RMSE - root mean square error). Para a análise da correlação dos dados, usando regressão linear, conseguiuse alcançar um R^2 de 0,82 % e RMSE de 1,12 % para clorofila-a e R^2 de 0,70 % e RMSE de 1,02 % para SST. Quando se usou RNA feedforward, foi possível obter um R^2 de 0,97 % e RMSE de 0,34 % para clorofila-a e R^2 de 0,98 % e RMSE de 0,16 % para SST.

Em (GUIMARÃES, 2019), pôde-se constatar que a RNA *feedforward* correlaciona de forma mais precisa os dados de campo e a resposta espectral da represa, usando as imagens

do satélite Sentinel-2. Segundo as conclusões desse trabalho, isso se deu pela capacidade das redes neurais artificiais conseguirem estabelecer relações lineares e não-lineares na correlação dos dados e, também, pelo fato de terem sido usados os parâmetros umidade e temperatura na entrada da rede, o que fez com que a rede pudesse distinguir as estações do ano quando foram realizadas as coletas dos parâmetros.

3.1 Conclusão

Este projeto apresenta um diferencial em relação aos trabalhos relacionados ao tema pesquisados na medida em que mostra que é possível realizar o monitoramento da qualidade da água em reservatórios hídricos usando sensoriamento remoto a partir de imagens de satélites, técnicas de regressão e dados históricos dos parâmetros limnológicos. Os modelos de aprendizagem já encontrados mostraram-se generalizáveis porém, não foi encontrada uma alternativa que disponibilizasse essa tecnologia para usuários não especialistas. A metodologia proposta nesse sentido compreende: sensoriamento remoto via imagens de satélites, técnicas de regressão e dados históricos de parâmetros limnológicos dos reservatórios hídricos, dentro de uma plataforma Web <<u>http://plattform.herokuapp.com/></u>. Essa plataforma fornece uma interface para o usuário onde é possível realizar a predição dos parâmetros limnológicos de reservatórios hídricos - a principal contribuição desta pesquisa comparada aos demais trabalhos disponíveis na literatura técnico-científica atual.

Os modelos de aprendizagem treinados com base nas técnicas de regressão linear e regressão LASSO mostraram capacidade de generalização segundo as métricas estatísticas utilizadas, ou seja, o coeficiente de determinação, erro percentual absoluto médio e erro absoluto médio, conforme os resultados apresentados nas tabelas 4 e 5, na Seção 5, se comparados com os resultados indicados na Seção 3. Vale ressaltar que, apesar pequena quantidade de dados para o treinamento dos modelos de aprendizagem, foi possível obter bons resultados.

4 METODOLOGIA E PROPOSTA DE DESENVOLVIMENTO

Este capítulo descreve a metodologia proposta para o desenvolvimento deste trabalho visando à criação de uma plataforma Web para a predição dos parâmetros limnológicos em reservatórios hídricos com base nos modelos de regressão linear e regressão LASSO, conforme descrito no Capítulo 2. A proposta desenvolvida é validada em um estudo de caso no Reservatório de Três Marias - MG, onde foi feita a predição dos parâmetros turbidez e sólidos totais. A descrição da metodologia foi dividida em quatro etapas, sendo: Dados Limnólogicos, *Google Earth Engine, Machine learning* - Modelos de Regressão e Plataforma Web. A Figura 6 apresenta o diagrama de blocos da proposta, descrita nas seções a seguir.

4.1 Dados limnológicos

O bloco Dados Limnológicos é composto pelas etapas relativas à definição dos pontos de amostragem, coletas prévias de amostras de água e análise laboratorial e, posteriormente, os cálculos dos valores numéricos de cada parâmetro de interesse.

A definição de pontos de amostragem é a etapa onde são escolhidos os locais estratégicos para serem feitas as coletas de amostras de água. Nessa etapa, faz-se o georreferenciamento dos locais para a correlação entre os dados limnológicos e as informações espectrais das imagens do Landsat 8 OLI, indicadas no bloco *Google Earth Engine*. Com base nos dados das coletas prévias das amostras de água e suas respectivas análises, obtém-se os parâmetros limnológicos ou parâmetros de qualidade de água, como, por exemplo, turbidez, sólidos totais, sulfato, etc.Tomando-se as coordenadas geográficas dos locais de coleta, extraídas do georreferenciamento, busca-se as informações espectrais do Landsat 8 OLI.

4.2 Google Engine Earth

O bloco GEE tem como finalidade fazer a extração de informações espectrais das bandas do satélite Landsat 8 OLI através da API do GEE. A primeira etapa é iniciada com o acesso à coleção de imagens do Landsat 8 OLI, disponibilizadas após serem préprocessadas. As técnicas de pré-processamento são, essencialmente, funções operacionais para remover ou corrigir os erros e as distorções introduzidos nas imagens pelos sistemas sensores devidos a erros instrumentais (ruídos espúrios), às interferências da atmosfera (erros radiométricos) e à geometria de imageamento (distorções geométricas) (MENESES, 2012).

O GEE entrega imagens incluídas as correções rádiométrica, atmosférica e geométrica, com essas etapas concluídas as imagens já são entregues aptas a serem usadas. As imagens podem conter erros aleatórios em relação aos pixels ou linhas de pixels que, normalmente, se mostram como pixels ou linhas com valores saturados (claros) ou sem sinal (escuros).



Figura 6 – Diagrama de blocos da metodologia proposta

Fonte: Elaborado pelo autor.

São denominados ruídos e, tipicamente, são decorrentes de erros instrumentais e as correções desses ruídos é feita pela correção radiométrica (MENESES, 2012). A Figura 7 ilustra um exemplo da presença de ruídos em uma imagem de satélite.





Fonte: (MENESES, 2012)

A atmosfera afeta a radiância medida em qualquer ponto da imagem de duas maneiras, aparentemente, contraditórias. Primeiro, ela atua como um refletor, adicionando uma radiância extra ao sinal que é detectado pelo sensor. Segundo, ela atua com um absorvedor, atenuando a intensidade de energia que ilumina o alvo na superfície. Quando a atenuação é acima de um valor, a transparência da atmosfera torna-se opaca em determinados comprimentos de onda, impossibilitando a obtenção de imagens. Portanto, uma imagem deve-se verificar a degradação que pode ter sido causada pelo espalhamento atmosférico, e corrente da radiação adicionada ao pixel, sem que tenha sido proveniente do alvo (MENESES, 2012).

As distorções geométricas são os desajustes da grade de pixels da imagem. Na imagem digital isso deve ser entendido como a mudança de posição do pixel na grade. Como o pixel é visto por um valor de brilho ou de nível de cinza, se ele mudou de posição pode-se, então, dizer que aquele pixel na grade teve uma mudança relativa ao valor de brilho. Devido à alta altitude dos satélites, uma pequena mudança da plataforma pode causar grandes distorções de posição dos pixels na imagem. As fontes de distorções geométricas são mais diversas do que as radiométricas, as principais sendo devido à rotação da terra durante o tempo de imageamento, instabilidades dos eixos de posicionamento da plataforma, curvatura da Terra e campo de visada do sensor (MENESES, 2012).

Após essas etapas de pré-processamento, é necessário filtrar a região de interesse usando as coordenadas geográficas dos locais de coletas das amostras da água, no caso, o Reservatório de Três Marias/MG. Em seguida, é necessário identificar as datas de imageamento do satélite, as quais devem ser o mais próximo, se possível, coincidentes com as datas de coletas das amostras de água realizadas no reservatório. Selecionadas as imagens, busca-se as informações espectrais das bandas de interesse do Landsat 8 OLI diretamente correlacionadas às datas das coletas das amostras de água, o que permitirá a criação dos bancos de dados para o treinamento e validação dos modelos de regressão, processos a serem descritos na Seção 4.3.

A Figura 8 mostra uma imagem do Reservatório de Três Marias - MG extraída do GEE, uma imagem em escala de nível de cinza, onde a região mais escura representa a água do resevatório, indicando assim que a banda 5 do Landsat 8 OLI absorve muito energia emitida pelos raios solares, sendo uma boa opção para identificar corpos hídricos.

Figura 8 – Imagem da banda 5 do Satélite Landsat 8 OLI, Reservatório de Três Marias - MG.



Fonte: GEE.

4.2.1 Banco de dados

A base de dados utilizada tem um total de 97 amostras dos dados limnológicos de interesse, do qual 70% foi utilizado para a etapa de treinamento dos modelos de regressão e 30% para a etapa de teste. A validação do modelo na etapa de treinamento foi realizada utilizando-se o coeficiente de determinação e a avaliação dos testes foi feita por meio do erro absoluto médio e erro percentual absoluto médio.

Estação de Coleta	Latitude	Longitude	Data de Coleta					
			2016	2017	2018	2019	2020	2021
			3/5	17/2	2/3	28/2	19/2	2/2
TM15	100 50, 17 00 10	450 107 1 60 W	25/8	,	17/5	22/5	7/5	,
1 1110	10- 00 17.20 0	40^{-12} 12 1.02 W	8/11		23/8	21/8	26/8	25/8
					22/11	20/11	1/12	
			3/5	17/2	2/3	28/2	19/2	2/2
ТМ90	189 52' 54"S	459 07' 15"W	25/8		17/5	22/5	7/5	
1 1/120	10 00 04 0	40 07 10 W	9/11		23/8	21/8	26/8	25/8
					22/11	19/11	1/12	
			3/5	17/2	2/3	28/2	19/2	2/2
TM95	$18^{\underline{0}} 49' 57"S$	45^{0} 07' 59"W	25/8		17/5	22/5	7/5	
1 1/120			9/11		22/8	21/8	26/8	25/8
					22/11	20/11	1/12	
			4/5	16/2	1/3	27/2	20/2	3/2
TM30	18º 34' 20"S	45º 15' 39"W	24/8		16/5	22/5	6/5	
1 100			8/11		22/8	22/8	27/8	26/8
					21/11	21/11	3/12	
			4/5	16/2	28/2	27/2	20/2	3/2
TM25	189 20' 21 42"S	450 05, 20 40111	25/8		16/5	21/5	6/5	
1 100	10 29 21.42 5	45 25 50.40 W	9/11		22/8	22/8	27/8	26/8
					21/11	21/11	3/12	
			29/4	16/2	27/2	27/2	21/2	3/2
TM40	18º 13' 26"W	45 ^o 15' 45"W	23/8		15/5	21/5	6/5	
1 1/140	18- 15 20° W		8/11		21/8	22/8	26/8	26/8
					20/11	20/11	1/12	

Tabela 3 – Coordenadas geográficas dos locais e datas das coletas no Reservatório de Três Marias - MG.

As etapas relativas à definição dos pontos de amostragem, coletas e análise e, cálculo dos parâmetros limnológicos, indicadas na Figura 6 no bloco Dados limnológicos, foram realizadas pelos pesquisadores da CEMIG[®].

A Tabela 3 destaca as informações relacionadas às coletas de dados limnológicos no Reservatório de Três Marias - MG, disponibilizados pela CEMIG[®], e utilizados neste trabalho. Basicamente, tem-se a identificação da estação da coleta, latitude e longitude relativas aos locais das coletas e as respectivas datas de realização.

O banco de dados para o treinamento dos modelos de aprendizagem foi criado com dados limnológicos do Reservatório de Três Marias - MG disponibilizadas pela CEMIG[®] e informações espectrais das imagens das bandas do Satélite Landsat 8 OLI. Essas imagens foram selecionadas utiliando-se como critério as datas de coletas de amostras de campo da Tabela 3, tendo sido escolhidas as imagens disponíveis com datas mais próximas às indicadas na tabela citada. As informações espectrais foram acessadas por meio da plataforma do GEE e, posteriomente, correlacionados para submissão aos modelos de regressão linear e regressão LASSO.

A Figura 9 mostra o Reservatório de Três Marias - MG e também os postos de coleta de amostra de água *in situ* ao disponibilizados pela $\text{CEMIG}^{\textcircled{R}}$.

Figura 9 – Pontos de coleta de amostras de água no Reservatório de Três Marias - MG.



Fonte: CEMIG[®].

4.3 Machine Learning

Nesta etapa são aplicadas as técnicas de regressão linear e regressão LASSO apresentadas na Seção 2 para a obtenção dos modelos de predição da turbidez e sólidos totais presentes na bacia hidrográfica de interesse, no caso, o Reservatório de Três Maris - MG.

Os dados limnológicos e as informações espectrais disponíveis no banco de dados construído são submetidos à etapa de pré-processamento, que consiste das seguintes ações:

- Exclusão dos parâmetros limnológicos que não contêm informações suficientes: A exclusão deverá ocorrer quando há casos em que a quantidade de informação é insuficiente para o treinamento do modelo de aprendizagem, por exemplo, se os dados apresentam valores muito próximos, ou seja, eles variam pouco entre si. A presença de variações permite que o modelo de aprendizagem possa ser mais generalizável, o que leva a predições mais precisas.
- Análise da correlação entre os dados limnológicos e espectrais. Esta etapa é onde se verifica quais parâmetros limnológicos estão mais próximos e também se estão bem relacionados com as informações espectrais das imagens do satélite Landsat 8 OLI. Essa análise é feita usando-se uma escala de valores que varia de -1 a 1, sendo -1 a correlação mais fraca e 1 uma forte correlação.
- Normalização das informações espectrais das bandas do satélite Landsat 8 OLI, a qual consiste em homogeneizar os dados espectrais do Landsat 8 OLI, pois para treinar os modelos de aprendizagem os valores númericos dos dados de entrada precisam estar próximos. No caso deste trabalho, faz-se uma divisão dos dados espectrais das bandas do Landsat 8 OLI por 10. Como os valores espectrais são expressos na casa das centenas ou milhares e os valores dos dados limnológicos em dezenas, essa divisão foi necessária para reduzir o distanciamento das faixas, obtendo-se um treinamento dos modelos de aprendizagem mais apropriado e, mais generalízavel para novos dados.
- Separação dos dados de treinamento e teste de forma a verificar se o modelo de aprendizagem que foi treinado é generalizavel. Utilizou-se 70% das informações do banco de dados no treinamento do modelo de aprendizagem e os outros 30% foram destinados para a etapa de teste. Esses últimos dados são submetidos às entradas do modelo de aprendizagem treinado anteriomente para a realização de novas predições, a serem usadas na validação do modelo.
- Exclusão de dados quando ocorrerem valores faltantes ou nulos em uma coluna de algum parâmetro limnológico ou dado espectral no banco de dados. Para resolver esse problema existem várias formas, dentre elas, tomar a média dos demais valores

na coluna e adicioná-los em substituição aos valores excluídos, este foi o método usado.

4.3.1 Treinamento, teste e validação dos modelos de aprendizagem

Após a etapa de pré-processamento, os dados podem ser utilizados para o treinamento e teste dos modelos de aprendizagem, a partir das técnicas de regressão linear e regressão LASSO que estão disponíveis na biblioteca *Scikit Learn*. Como citado, foram inicialmente reservados para a etapa de treinamento 70% do total disponível na base dados pré-processada e os 30% restantes, usados na etapa de teste.

A biblioteca *Scikit Learn* foi aplicada para realizar as etapas de treinamento e validação dos modelos de aprendizagem usando regressão linear e regressão LASSO. Essa é uma biblioteca da linguagem Python desenvolvida especificamente para aplicações práticas de *machine learning*. Esta biblioteca dispõe de ferramentas simples e eficientes para análise preditiva de parâmetros e possui código aberto e acessível, tendo sido construída sobre os pacotes NumPy, SciPy e matplotilib (SCIENCE,).

Após a definição dos modelos na etapa de treinamento com base em informações prévias acerca dos parâmetros limnológicos e dados espectrais das imagens do satélite Landsat 8 OLI correlacionadas, a etapa de teste consiste do uso destes modelos matemáticos com dados desconhecidos para a validação dos modelos de aprendizagem, ou seja, verificar se esses modelos conseguem fazer novas predições de parâmetros limnológicos, que no caso deste trabalho são: turbidez e sólidos totais.

Para fazer a validação dos modelos obtidos utilizou-se métricas estatísticas, conforme já apresentado na seção 2.7. Na etapa de treinamento, o coeficiente de determinação R^2 e na etapa de teste, as métricas erro médio absoluto e erro percentual médio absoluto.

4.4 Plataforma Web

O bloco Plataforma Web é uma aplicação onde o modelo de aprendizagem treinado, conforme descrito na seção 4.3, é utilizado para a realização de novas predições, a partir de novos dados espectrais do satélite Landsat 8 OLI ou dados limnológicos informados pelo usuário. Essa aplicação foi desenvolvida utilizando-se as linguagens Python, JavaScript, CSS, HTML, API do GEE e bibliotecas de *machine learning*.

A plataforma foi desenvolvida com base em técnicas de machine learning e desenvolvimento Web. Machine learning é a etapa onde é feito o acesso ao banco de dados que são pré-processados e, posteriormente, usados para fazer o treinamento dos modelos de aprendizagem com base nas técnicas de regressão linear e regressão LASSO e também são usados na etapa de teste e validação destes modelos como foi explicado na seção 4.3.1. Essa etapa foi feita usando a linguagem Python e algumas bibliotecas de machine learning, sendo elas: pandas, matplotilib, numpy, sklearn, sklearn.model.selection, train.test.split, sklearn.impute, SimpleImputer, sklearn.linear.model, LinearRegression, LassoCV, sklearn.metrics, mean.absolute.percentage.error, mean.absolute.error e Jupyter Lab. Jupyter Lab é o mais recente ambiente de desenvolvimento interativo baseado na Web para notebooks, código e dados. Sua interface flexível permite que os usuários configurem e organizem fluxos de trabalho em ciência de dados, computação científica, jornalismo computacional e machine learning (2022 JUPYTER,).

A plataforma implementada, mostrada na Figura 10, está disponível em <<u>http://plattform.herokuapp.com/></u>, juntamente com um guia de utilização que visa torná-la acessível a usuários não especializados.



Figura 10 – Página inicial da Plataforma Web.

Fonte: Próprio autor.

Para a interface gráfica da plataforma foi utilizada a linguagem CSS (*Cascading Style Sheets* ou Folhas de Estilo em Cascata), uma linguagem de estilo usada para descrever a apresentação de um documento escrito em HTML ou em XML (incluindo várias linguagens em XML como SVG, MathML ou XHTML). O CSS descreve como os elementos são mostrados na tela, no papel, na fala ou em outras mídias (2022 CSS,). A inclusão de conteúdos nas páginas foi realizada através da linguagem HTML (Linguagem de Marcação de HiperTexto) Outras linguagenss, além da HTML, podem ser usadas para descrever a aparência/apresentação (CSS) ou a funcionalidade/comportamento (JavaScript) de uma página da Web (2022 HTML,). A linguagem JavaScript foi usada para verificar se o usuário inseriu corretamente os dados para a predição dos parâmetros limnológicos turbidez e sólidos totais na plataforma desenvolvida (2022 JAVASCRIPT,).

4.5 Conclusão

Este capítulo apresentou os principais aspectos da metodologia empregada no desenvolvimento de uma plataforma Web para a predição de parâmetros limnológicos turbidez e sólidos totais do Reservatório de Três Marias - MG, o estudo de caso escolhido nesta proposta.

Vale destacar que a plataforma desenvolvida pode ser utilizada em outros reservatórios hídricos bem como adaptada para a predição de outros parâmetros, desde que se tenha informação dísponível para a criação de um banco de dados adequado para as etapas de treinamento e teste dos modelos de regressão linear e regressão LASSO.

5 RESULTADOS E DISCUSSÕES

Com base nos dados de medição *in situ* relativos à Represa de Três Marias - MG, disponibilizados pela CEMIG [®] e imagens do satélite Landsat 8 OLI correlatas, neste capítulo são apresentados os resultados obtidos a partir dos modelos de regressão, nas fases de treinamento e testes, visando à predição dos parâmetros limnológicos turbidez e sólidos totais.

São mostrados os resultados do coeficiente de determinação R^2 , usado para a validação dos modelos de regressão na etapa de treinamento e das métricas estatísticas erro absoluto médio e erro percentual absoluto médio para validação dos modelos de aprendizagem na etapa de teste. Também são apresentados os gráficos de resíduos que permitem visualizar a aproximação entre os valores reais dos parâmetros turbidez e sólidos totais e os valores preditos, obtidos por meio da Plataforma Análise da Qualidade da água, disponível em <http://plattform.herokuapp.com/>.

A seguir, são destacados os resultados do coeficiente de determinação R^2 , conforme a Tabela 4, alcançados pelos modelos na etapa de treinamento e os resultados das métricas estatísticas, erro absoluto médio e erro percentual absoluto médio, indicados na Tabela 5, na etapa de teste.

De acordo com a Tabela 4, pôde-se observar que, após a fase de treinamento, uma melhor aproximação entre os valores preditos e os valores reais quando utilizado o modelo de regressão LASSO, para ambos os parâmetros limnológicos turbidez e sólidos totais, onde os coeficientes de determinação R^2 apresentaram reduções de valores em torno de 0,154 e 0,014, respectivamente.

Segundo os resultados da Tabela 5, o erro absoluto médio e o erro percentual absoluto médio apresentados pelo modelo de regressão linear indicam um pior desempenho em relação ao modelo de regressão LASSO. As reduções dos erros absolutos médios foram de, aproximadamente, 0,573 e 0,186 e, em relação aos erros percentuais absolutos médios, em torno de 0,784 e 0,003 para a predição da turbidez e sólidos totais, respectivamente.

As diferenças entre os resultados das métricas indicam um melhor desempenho do modelo de regressão LASSO e podem ser justificadas em função da existência de alguma dependência linear entre as variáveis de entrada utilizadas nas fases de treinamento e teste, penalizada pelo referido modelo. Essas condições variam para cada parâmetro limnológico,

Tabela 4 – Resultados dos coeficientes de determinação R^2 para os modelos de aprendizagem regressão linear e LASSO na etapa de treinamento

Tipo de Modelo	Quantidade de dados	Parâmetro limnológico	Coeficiente de determinação
Rogrossão Linoar	67	Turbidez	0,7213
negressao Linear	01	Sólidos Totais	0,832
Bogrossão LASSO	67	Turbidez	0,567
	01	Sólidos Totais	0,818

Fonte: Próprio autor.

Tabela 5 – Resultados do erro absoluto médio e erro percentual absoluto médio para os modelos de aprendizagem regressão linear e LASSO na etapa de teste.

Tina da Madala	Quantidade de	Parâmetro	Erro absoluto	Erro percentual
Tipo de Modelo	dados	limnológico	médio	absoluto médio (%)
Rogrossão Linoar	30	Turbidez	2,511	2,152
Regressão Linear	50	Sólidos Totais	4,822	0,090
Rogrossão LASSO	20	Turbidez	1,938	1,368
Regressão LASSO	30	Sólidos Totais	4,636	0,087

Fonte: Próprio autor.

ou seja, alguns dados de entrada podem estar mais correlacionados que outros, o que influencia diretamente na predição de novos parâmetros. Além disso, a base de dados disponível é composta por uma quantidade reduzida de amostras se comparada com os processos usuais utilizados em modelos de aprendizagem, o que também compromete o desempenho dos modelos.

Com base nos resultados obtidos por meio da Plataforma Análise da Qualidade da Água do Reservatório Três Marias – MG, disponível em <http://plattform.herokuapp. com/>, a seguir são apresentados os gráficos de resíduos que permitem visualizar a aproximação entre os valores reais e de predição dos parâmetros turbidez e sólidos totais. Ou seja, quanto mais próximos os resíduos estiverem do eixo Resíduos = 0, há a indicação de que os valores preditos pelo modelo estão mais próximos dos valores reais.

As Figuras 11 a 14 mostram os gráficos de resíduos relativos ao modelo de aprendizagem, para as etapas de treinamento e teste, respectivamente. Em ambos as fases, pode-se observar que as medições apresentam maior ocorrência na faixa de 0 a 20 mg/l para a turbidez e na faixa de 40 a 80 mg/l para os sólidos totais. Os eixos x representam os valores de predição dos parâmetros limnológicos turbidez e sólidos totais a partir dos modelos de aprendizagem treinados com base nas técnicas de regressão linear e regressão LASSO e os eixos y representam os resíduos que são a diferença entre as predições do modelo de aprendizagem e os valores reais dos dados usados nas etapas de treinamento dos modelos de aprendizagem e teste desses modelos. Quanto mais próximos os dados de treinamento e teste estiverem em relação ao eixo Resíduos = 0, melhor a capacidade de generalização do modelo de aprendizagem.

As Figuras 11 e 12 apresentam os resíduos da predição dos parâmetros sólidos totais e turbidez, respectivamente, resultantes da aplicação do modelo de regressão linear. Com base nas regiões de maior ocorrência das predições nos gráficos, pode-se observar que os espalhamentos se concentram dentro de uma faixa de ± 10 e ± 5 resíduos, em um total aproximado de ± 20 medições, mesmo se considerada a presença de alguns *outliers*.

As Figuras 13 e 14 mostram os resíduos relativos à predição do parâmetro turbidez, resultantes do modelo de regressão LASSO. Com base nas regiões de maior ocorrência das predições nos gráficos, pode-se observar que os espalhamentos também se concentram

Figura 11 – Gráfico de resíduos relativos à predição dos sólidos totais usando a regressão linear.



Fonte: Próprio autor.

Figura 12 – Gráfico de resíduos relativos à predição da turbidez usando a regressão linear.



Fonte: Próprio autor.

dentro de uma faixa de ± 10 e ± 5 resíduos em um total aproximado de ± 20 medições, considerada a presença de alguns *outliers*.

Os gráficos de resíduos indicam uma boa aproximação entre as predições e os valores reais, principalmente, em função se considerada a limitação do banco de dados disponível. Ou seja, confirmam que a estimativa dos parâmetros na etapa de teste seguiu a mesma tendência dos resultados da predição dos dados de treinamento, indicando a viabilidade dos modelos de aprendizagem.

Figura 13 – Gráfico de resíduos relativos à predição dos sólidos totais usando a regressão LASSO.



Fonte: Próprio autor.

Figura 14 – Gráfico de resíduos relativos à predição da turbidez usando a regressão LASSO.



Fonte: Próprio autor.

Com base nos dados utilizados nas etapas de treinamento e teste, as figuras 15 a 18 mostram os resultados das predições dos parâmetros limnológicos turbidez e sólidos totais a partir dos modelos de aprendizagem treinados com base nas técnicas de regressão linear e regressão LASSO, usando dados prévios na etapa de treinamento e dados desconhecidos na etapa de teste para fins da avaliação dos resultados. Os eixos x mostram os valores preditos dos parâmetros limnológicos turbidez e sólidos totais e os eixos y indicam os valores reais dos parâmetros. Pode-se observar a mesma tendência em relação ao comportamento do modelo de aprendizagem quanto aos dados conhecidos e em relação aos dados novos, o que indica capacidade de generalização dos modelos obtidos. Ou seja, todos os gráficos mostram que os resultados de predição dos modelos de aprendizagem apresentam boa aproximação e o mesmo comportamento relativo aos dados reais de medição dos parâmetros turbidez e sólidos totais.

Figura 15 – Valores preditos x valores reais sólidos totais usando regressão linear.



Fonte: Próprio autor.

Figura 16 – Valores preditos x valores reais sólidos totais usando a regressão LASSO.



Comparação entre valores preditos x valores reais de sólidos totais

O treinamento dos modelos de aprendizagem para as predições dos parâmetros linológicos turbidez e sólidos totais foi realizado através da Plataforma Análise da Qualidade da água (<<u>http://plattform.herokuapp.com/></u>).

Para o treinamento dos modelos de aprendizagem do parâmetro turbidez, foram usados

Figura 17 – Valores preditos x valores reais para a turbidez usando regressão linear.



Figura 18 – Valores preditos x valores reais para a turbidez usando regressão LASSO.



Comparação entre valores preditos x valores reais de turbidez

Fonte: Próprio autor.

os dados limnlógicos: cor verdadeira, sulfato e turbidez zona fótica e, também, os dados espectrais: banda 2, banda 3 e banda 4.

Para o parâmetro sólidos totais foram utilizados os dados limnlógicos: sulfato e sólidos dissolvidos totais (sdt) e, também, os dados espectrais: banda 1, banda 2, banda 4, banda 5, banda 6 e banda 7.

A escolha dos dados espectrais e limnológicos para treianmento dos modelos foram escolhidos por meio de análise de correlação usando a bilioteca *Pandas* da linguagem Python em conjunto com o pacote *Sckit learn*, um pacote de *machine learning* usado

para fazer treinamento de modelos de aprendizagem usando técnicas de regressão e redes neurais artificiais.

A banda 4 e cor verdadeira apresentaram uma melhor correlação com o parâmetro turbidez, alcançando 0,6124 para a banda 4 e 0,9049 para cor verdadeira. As banda 6 e sdt foram os que tiveram uma melhor correlação com o parâmetro sólidos totais, com valores de correlação aproximados de 0,4277 para a banda 6 e 0,8986 para sdt em uma escala que varia de -1 a 1, onde -1 indica forte correlação decrescente e 1, indica forte correlação crescente.

As Figuras 19 a 22 mostram os gráficos dos dados banda 4 e cor verdadeira usados para fazer o treinamento dos modelos de aprendizagem no eixo x e a predição do parâmetro turbidez usando o modelo de aprendizagem treinado no eixo y. Também mostram a correlação crescente entre os dados banda 4 e cor verdadeira e o parâmetro turbidez como pode-se visualizar pela função plotada nos gráficos.

As Figuras 19 e 20 mostram os gráficos da correlação entre o dado limnológico cor verdadeira e o parâmetro turbidez para os métodos de regressão linear e regressão LASSO e, em ambos, os métodos a função indica uma correlação crescente.



Figura 19 – Correlação entre cor verdadeira e turbidez.

As Figuras 21 e 22 apresentam os gráficos da correlação entre o dado espectral banda 4 e o parâmetro turbidez para os métodos de regressão linear e regressão LASSO e, em ambos os métodos, a função indica uma correlação crescente.

As Figuras 23 a 26 mostram os gráficos dos dados *banda* 6 e *sdt* usados para fazer o treinamento dos modelos de aprendizagem indicados no eixo x e a predição do parâmetro *sólidos totais* usando o modelo de aprendizagem treinado indicado no eixo y e também mostra uma correlação crescente entre os dados *banda* 6 e *sdt* e o parâmetro *sólidos totais* como pode-se visualizar pela função plotada nos gráficos.



Figura 20 – Correlação entre cor verdadeira e turbidez.

Figura 21 – Correlação entre banda 4 e turbidez.



Fonte: Próprio autor.

As Figuras 23 e 24 mostram os gráficos da correlação entre o dado limnológico sdt e o parâmetro sólidos totais para os métodos de regressão linear e regressão LASSO e, em ambos, os métodos a função indica uma correlação crescente.

As Figuras 25 e 26 mostram os gráficos da correlação entre o dado espectral banda 6 e o parâmetro sólidos totais para os métodos de regressão linear e regressão LASSO e, novamente, a função indica uma correlação crescente em ambos os métodos.

5.1 Conclusão

Esse capítulo mostrou os resultados alcançados pelos modelos de regressão linear e regressão LASSO nas etapas de treinamento e teste dos modelos de aprendizagem, se-



Figura 22 – Correlação entre banda 4 e turbidez.

Fonte: Próprio autor.

Figura 23 – Correlação entre sdt (sólidos dissolvidos totais) e sólidos totais.



Fonte: Próprio autor.

gundo as métricas estatísticas. Com base nas tabelas 4 e 5, os resultados indicaram que a técnica regressão LASSO conseguiu apresentar um modelo de aprendizagem com melhor desempenho, ou seja, capaz de realizar predições mais precisas dos parâmetros limnológicos turbidez e sólidos totais, tendo sido, portanto, escolhido para ser implementado na plataforma desenvolvida.

As Figuras 11 a 14 mostraram os gráficos de resíduos com o objetivo de verificar a aproximação dos valores reais e os valores preditos pelo modelo de aprendizagem. Os gráficos indicaram bons resultados, pois, a maioria dos dados estão próximos do eixo 0, que indica que os resultados preditos pelo modelo de aprendizagem estão próximos dos



Figura 24 – Correlação entre sdt (sólidos dissolvidos totais) e sólidos totais.



Figura 25 – Correlação entre banda 6 e sólidos totais.



Regressão linear para relação entre Banda 6 x Sólidos totais

Fonte: Próprio autor.

valores reais, para ambas os modelos de regressão estudados, apesar da restrição quanto à quantidade de dados para um melhor treinamento do modelo de aprendizagem.

As Figuras 15 a 18, mostraram uma comparação entre a predição dos dados de treinamento e teste usando os modelos de aprendizagem já treinados para avaliar o comportamento dos modelos diante de dados nunca vistos (teste) e dados já conhecidos (treinamento). Como pôde ser visto nas figuras 15 a 18, o comportamentos do modelos de regerssão linear e LASSO são próximos para ambos os parâmetros limnológicos, indicando que os modelos de aprendizagem podem ser generalizados.

As Figuras 19 a 26 apresentaram os gráficos que trazem os parâmetros turbidez e



Figura 26 – Correlação entre banda 6 e sólidos totais.

sólidos totais, indicando a correlação crescente positica dos parâmetros com os dados espectrais usados no treinamento dos modelos de aprendizagem.

Com base nas técnicas de regressão linear e regressão LASSO, o parâmetro turbidez mostrou uma correlação crescente mais forte se comparada ao parâmetro sólidos totais, em todas as predições e conforme as métricas estatísticas, há a indicação de que os modelos de aprendizagem são generalizáveis, a despeito das restirções quanto quantidade disponível de informações nas bases de dados

6 CONSIDERAÇÕES FINAIS

Neste capítulo são apresentadas as conclusões e principais contribuições deste trabalho, as dificuldades que foram encontradas durante o desenvolvimento da plataforma Análise da Qualidade da Água - Represa de Três Marias/MG, a contribuição da API do GEE no âmbito da contrução dessa plataforma, dentre outras, e as melhorias futuras propostas.

6.1 Conclusões

O objetivo principal deste trabalho consistiu do desenvolvimento de uma plataforma Web para predição de parâmetros limnológicos de reservatórios hídricos, com foco em um estudo de caso no Reservatório de Três Marias/MG e os parâmetros turbidez e sólidos totais. As justificativas para esse desenvolvimento foram mostradas na Seção 3, onde constatou-se que a grande maioria dos trabalhos relacionados tinham conseguido alcançar resultados satisfatórios na predição de parâmetros limnológicos de reservatórios hídricos, porém eles não apresentavam uma um meio digital para que outros pesquisadores, empresas e pessoas não especialistas pudessem fazer avaliações da qualidade da água em bacias hidrográficas. Com base na associação das técnicas de regressão linear e de programação pôde-se desenvolver uma plataforma Web, a qual se encontra disponível em <http://plattform.herokuapp.com/>.

A Seção 4 apresentou a metodologia utilizada, onde vale destacar a API do GEE para estudos envolvendo imagens do satélite Landsat 8 OLI, a qual já disponibiliza imagens pré-processadas. Os modelos de aprendizagem mostraram-se capazes de generalização para a predição de novos valores de turbidez e sólidos totais. De acordo com as tabelas 4 e 5, da Seção 5, a técnica regressão LASSO obteve melhores resultados segundo as métricas estatísticas utilizadas.

Enfim, conclui-se que a análise e monitoramento remoto da qualidade da água por meio dos parâmetros limnológicos constitui uma ferramenta que poderá reduzir o número de coletas de amostras de água *in situ*, consequentemente, reduzindo custos e o aumento da acessibilidade a esse tipo de informação. A maior dificuldade encontrada para o treinamento dos modelos de aprendizagem referiu-se à quantidade de informações previamente disponíveis para a construção das bases de dados, uma restrição que deve ser superada para os futuros desenvolvimentos e melhorias.

6.2 Propostas de Continuidade

Destacam-se como propostas futuras a serem desenvolvidas:

- Aumentar a quantidade de dados limnológicos históricos para a construção de bases de dados robustas;
- Testar novas técnicas de regressão e redes neurais artificiais;

- Adicionar API do GEE à plataforma desenvolvida para o acesso aos dados espectrais das imagens do Satélite Landsat 8 OLI;
- Expandir a plataforma para a predição de outros parâmetros limnológicos ou informação espectral do satélite Landsat 8 OLI;
- Testar a plataforma desenvolvida em outros reservatórios brasileiros ou outros países.

REFERÊNCIAS

2010 ATLAS DIGITAL DAS áGUAS DE MINAS, . <htp://www.atlasdasaguas.ufv.br/alto_sao_francisco/resumo_alto_sao_francisco.html>. Acessado em Abil 2022. 18, 19

2022 CSS,
.<https://developer.mozilla.org/pt-BR/docs/Web/CSS>. Acessado em Maio 2022. 49

2022 HTML, . <https://developer.mozilla.org/pt-BR/docs/Web/HTML>. Acessado em Maio 2022. 49

2022 JAVASCRIPT, . <
https://developer.mozilla.org/pt-BR/docs/Web/JavaScript>. Acessado em Maio 2022. 49

2022 JUPYTER, . < https://jupyter.org/>. Acessado em Maio 2022. 49

A GOOGLE ENGINE EARTH, a. <u>Google Earth Engine</u>. <https://developers.google. com/earth-engine/>. Acessado em 12-04-2021. 25

(ANA), A. N. D. Águas. <u>Agência Nacional de águas e Saneamento básico</u>. <<u>https://www.ana.gov.br/>. Acessado em 29-05-2019. 11, 17, 18</u>

(ANA), A. N. D. Águas. <u>Situação da Água no Mundo</u>. <<u>https://www.ana.gov.br/</u>panorama-das-aguas/agua-no-mundo>. Acessado em 29-05-2019. 11

ANSARI, M.; AKHOONDZADEH, M. Mapping water salinity using landsat-8 oli satellite images (case study: Karun basin located in iran). <u>Advances in</u> <u>Space Research</u>, v. 65, n. 5, p. 1490–1502, 2020. ISSN 0273-1177. Disponível em: https://www.sciencedirect.com/science/article/pii/S0273117719308762>. 37, 38

B GOOGLE ENGINE EARTH, b. Introdução Google Earth Engine. < https://developers.google.com/earth-engine/guides/getstarted>. Acessado em 12-04-2021. 25

BALDI, E. a situation actuelle de la recherche limnologique, après le congrès de zurich. Schweiz. Z. Hydrol., Zurique, v. 11, p. 637–649, 1949. 15

BARRELLA, W. e. a. <u>As relações entre as matas ciliares, os rios e os peixes. In:</u> <u>BAUDRY, F. B. Ecología del paisaje: conceptos, métodos y aplicaciones</u>. [S.l.]: Madrid: Ediciones Mundi-Prensa, 2002. 17

BORMAN F.; LIKENS, G. E. <u>Pattern and Processes in a Forested Ecosystem</u>. [S.l.]: Nova York: Springer, 1979. 15

BRAGA, B. e. a. <u>Introdução à engenharia ambiental</u>. [S.l.]: São Paulo: Prentice Hall, 2005. 16

BUMA, W. G.; LEE, S.-I. Evaluation of sentinel-2 and landsat 8 images for estimating chlorophyll-a concentrations in lake chad, africa. <u>Remote Sensing</u>, v. 12, n. 15, 2020. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/12/15/2437>. 37, 38

C GOOGLE ENGINE EARTH, c. <u>Aplicativo Google Earth Engine</u>. <https://developers.google.com/earth-engine/guides/apps>. Acessado em 12-04-2021. 26

CHEIN, F. <u>Introdução aos modelos de regressão linear</u>. [S.l.]: Enap; ISBN: 978-85-256-0115-5, 2019. https://repositorio.enap.gov.br/bitstream/1/47f>. 27

CHITERO, J. et al. Análise da recuperação física de solos degradados via redes neurais artificiais por meio de uma interface gráfica analysis of the physical recovery of degraded soils via artificial neural networks using a graphical interface análisis de la recuperación física de suelos degradados a través de redes neuronales. <u>Research, Society and Development</u>, v. 9, n. 7, p. e257973719, 2020. 34, 38

CHOLLET, F. <u>Deep Learnig with Python</u>. [S.l.]: Manning Publications C; United States of America; ISBN 9781617294433, 2018. 26

CRISTINA B. S. M; CLECI TERESINHA, N. J. A. A. C. <u>Análise e Gestão de</u> <u>Bacias Hidrográficas</u>. [S.l.]: Biblioteca Dante Alighieri UNIASSELVI – Indaial, 2013. <<u>https://www.uniasselvi.com.br/extranet/layout/request/trilha/materiais/livro/livro.</u> php?codigo=15075>. 16, 17, 18

CRóSTA, A. P. Processamento digital de imagens de sensoriamento remoto. [S.l.]: Campinas, SP: IG/UNICAMP, 1992. 170 p., 1992. 23

DESENVOLVIMENTO de um modelo empírico para estimativa de profundidade do disco de clorofila-a e Secchi para um lago raso dos Pampeanos (Argentina). Elsevier, v. 21, n. 2, p. 183–191, 2018. 33, 38

FLORENZANO, T. Imagens de satélite para estudos ambientais. [S.l.]: São Paulo: oficina de textos, 97 p., 2002. 22

GUIMARÃES, T. T. <u>Utilização de imagens de satélite para predição de clorofila-a e</u> sólidos suspensos em corpos d\'água: estudo de caso da Represa do Lobo/SP. Tese (Doutorado) — Universidade de São Paulo, 2019. 22, 23, 24, 38

HASTIE T., T. R. e. W. M. Statistical learning with sparsity: The lasso and generalizations, chapman hall/crc. In: . [S.l.: s.n.], 2015. 28

HASTIE T., T. R. T. R. e. a. Best subset, forward stepwise or lasso? analysis and recommendations based on extensive comparisons. In: . [S.l.: s.n.], 2020. v. 1, n. 27, p. 579–592. 27

HIDROTEC, G. d. M. <u>Hidrotec</u>. [S.l.: s.n.]. <http://www.hidrotec.ufv.br/home.html>. Acessado em Junho 2022. 19

INPE 2013, D. <u>Geração de Imagens do Instituto Nacional de P. E. Landsat.</u> <http://www.dgi.inpe.br/documentacao/satelites/landsat>. Acessado em Junho 2019. 19, 20

JAMES G., W. D. H. T. e. T. R. An introduction to statistical learning. In: . [S.l.: s.n.], 2013. v. 112, Springer, n. 1. 28

JENSEN, J. R. Sensoriamento Remoto do Ambiente: Uma Perspectiva em recursos <u>Terrestres</u>. [S.l.]: Tradução José Carlos Neves Epiphanio (coordenador) et. al. São José dos Campos, 2011. 22, 23, 24, 25 JIA, T.; ZHANG, X.; DONG, R. Long-term spatial and temporal monitoring of cyanobacteria blooms using modis on google earth engine: A case study in taihu lake. <u>Remote Sensing</u>, v. 11, n. 19, 2019. ISSN 2072-4292. Disponível em: . 36, 37

JUNIOR, G. P. de A. <u>Avaliação do lasso e métodos alternativos em modelos de regressão</u> logística. Dissertação (Mestrado) — Universidade de São Paulo, São Carlos, 2021. 27, 28

Khryashchev, V. V. et al. Optimization of convolutional neural network for object recognition on satellite images. In: <u>2018Systems of Signal Synchronization, Generating</u> and Processing in Telecommunications (SYNCHROINFO). [S.l.: s.n.], 2018. p. 1–5. 33, 38

Kupssinskü, L. S. et al. Prediction of chlorophyll-a and suspended solids through remote sensing and artificial neural networks. In: 2019 13th International Conference on Sensing Technology (ICST). [S.l.: s.n.], 2019. p. 1–6. 34, 38

LOBO, F. d. L. et al. Algaemap: Algae bloom monitoring application for inland waters in latin america. <u>Remote Sensing</u>, v. 13, n. 15, 2021. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/13/15/2874>. 35, 37

MACHADO, M. T. d. S.; BAPTISTA, G. M. d. M. Sensoriamento remoto como ferramenta de monitoramento da qualidade da água do lago paranoá (df). <u>Engenharia</u> Sanitaria e Ambiental, SciELO Brasil, v. 21, n. 2, p. 357–365, 2016. 33, 38

MARINHO, R. R. et al. Dinâmica espaço-temporal de sedimentos suspensos no rio negro, bacia amazônica, a partir de dados de sensoriamento remoto in situ e sentinel-2. International Journal of Geo-Information, v. 10, 2021. 34, 38

MENESES, P. <u>Princípios de sensoriamento remoto.</u> [S.l.]: Brasília: UNB/CNPq, 2012. p. 1-33., 2012. 23, 41, 43

MOTA, M. B.; BRAICK, P. R. <u>História das cavernas ao terceiro milênio: programa</u> completo de: pré-história e de história antigua, medieval, moderna, contemporânea, da América e do Brasil. [S.l.]: Editora Moderna, 1998. 11

NASA 2019, L. . <https://landsat.gsfc.nasa.gov/landsat-8/>. Acessado em Maio 2019. 19

NASA 2019, L. . B.

https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-bands/>. Acessado em Maio 2019.
 21

NASA 2019, L. . O.

https://landsat.gsfc.nasa.gov/landsat-8/landsat-8-overview/>. Acessado em Maio 2019.
 20

NASA 2019, T. D. <https://landsat.gsfc.nasa.gov/about/technical-information/>. Acessado em Maio 2019. 19

NOVO, E. M. L. M. <u>Sensoriamento Remoto: Princípios e Aplicações</u>. [S.l.]: 4ed. São Paulo; Blucher, 2010. 22, 23, 24

OMS. Pelo menos 2 milhoes de pessoas morrem por ano no mundo por causa de água contaminada. Brasília: Agência Brasil, p. 2015–04, 2011. 11

PIAVA, P. <u>Caderno do educador ambiental. 2. ed. rev. e ampl.</u> [S.l.]: Blumenau: Fundação Agência de Água do Vale do Itajaí; FURB, 2009. 17

PNQA. <u>ANA lança PNQA e Portal da Qualidade das Águas</u>. 2021. <https://www.ana.gov.br/noticias-antigas/ana-lanassa-pnqa-e-portal-da-qualidade-das-aguas.2019-03-15.3251065579>. 13

RODRIGUES, K. Lasso clássico e bayesiano. <u>Instituto de Matemática e Estatística</u> Universidade de São Paulo, 2018. 28, 29

SAGAN, V. et al. Monitoring inland water quality using remote sensing: potential and limitations of spectral indices, bio-optical simulations, machine learning, and cloud computing. Earth-Science Reviews, v. 205, p. 103187, 2020. ISSN 0012-8252. Disponível em: https://www.sciencedirect.com/science/article/pii/S0012825220302336>. 35, 37

SCIENCE,
. D. T. I. A. . D. < https://didatica.tech/a-biblioteca-scikit-learn-py
hton-para-machine-learning/>. Acessado em Maio 2022. 48

 $\label{eq:scikit-learn.org/stable/modules/model_evaluation.html#regression-metrics>. Acessado em fevereiro 2022. 31$

SILVA, M. G. G. et al. Analysis of water quality at hydrographic basin scale using satellite images, co-occurrence matrices and Bayes classifier. <u>Water Supply</u>, v. 21, n. 8, p. 4418–4428, 06 2021. ISSN 1606-9749. Disponível em: https://doi.org/10.2166/ws.2021.192. 33

STOCK J. H.; WATSON, M. W. Introduction to Econometrics. [S.l.]: 3. ed. Addison-Wesley Series in Economics, v. 1. Addison-Wesley, 2010. 27

TIBSHIRANI, R. 'regression shrinkage and selection via the lasso', journal of the royal statistical society: Series b (methodological). In: . [S.l.: s.n.], 1996. v. 1, p. 267–288. 29

TUNDINISI J. G.; TUNDINISI, T. M. Limnologia. [S.l.]: São Paulo: Oficina de Textos, 2008. https://www.academia.edu/43630915/Limnologia_Tundisi_LIVRO_pdf>. 15

VILELA, M. A. M. A. <u>Metodologia para monitoramento da qualidade da água</u> <u>de reservatórios utilizando sensoriamento remoto</u>. [S.l.]: Dissertação (Mestrado em Engenharia Civil) – Faculdade de Engenharia Civil, Universidade Federal de Uberlândia, 2010. 24

ZHU, S.; MAO, J. A machine learning approach for estimating the trophic state of urban waters based on remote sensing and environmental factors. <u>Remote Sensing</u>, v. 13, n. 13, 2021. ISSN 2072-4292. Disponível em: https://www.mdpi.com/2072-4292/13/13/2498>. 37, 38