

UNIVERSIDADE FEDERAL DE SERGIPE CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA DEPARTAMENTO DE COMPUTAÇÃO

Monitoramento de indicador de qualidade de ar por processamento de imagens de satélite

Trabalho de Conclusão de Curso

Victor Luiz da Fonseca Rocha



São Cristóvão – Sergipe 2022 UNIVERSIDADE FEDERAL DE SERGIPE CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA DEPARTAMENTO DE COMPUTAÇÃO

Victor Luiz da Fonseca Rocha

Monitoramento de indicador de qualidade de ar por processamento de imagens de satélite

Trabalho de Conclusão de Curso submetido ao Departamento de Computação da Universidade Federal de Sergipe como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

Orientador(a): Simone Soraia Silva Sardeiro Coorientador(a): Beatriz Trinchão Andrade de Carvalho

São Cristóvão – Sergipe

2022

Agradecimentos

Dedico a realização deste trabalho a meus pais, Antonio Luiz Santos da Rocha e Lourdes Rosendo da Fonseca Rocha, ao meu irmão, Fábio, a minha avó Maria e aos meus amigos que estiveram ao meu lado me apoiando em muitos momentos. Ademais, agradeço a minha orientadora Simone Soraia Silva Sardeiro, minha coorientadora Beatriz Trinchão Andrade De Carvalho, aos meus professores, em especial a Debora Maria Coelho Nascimento e Bruno Otavio Piedade Prado, por compartilhar seus conhecimentos e risadas em diversas ocasiões na UFS.

Resumo

A escassez no número de centros meteorológicos para amostragem no Brasil torna os dados sobre qualidade do ar escassos. A falta de acompanhamento frequente dos indicadores de qualidade do ar somente é notada quando chega a níveis nocivos à saúde. Existem diversos estudos sobre indicadores de qualidade do ar, focados principalmente no indicador de PM2,5. A contaminação por PM2,5, que é o material particulado de 2,5µm, pode causar problemas cardíacos, asma em crianças e diversos problemas respiratórios. Esses fatores de risco demonstram a importância do estudo desse indicador, visto que o tipo de partícula que ele representa pode afetar toda a população de uma área, prejudicando a qualidade de vida e saúde dos habitantes. O sensoriamento remoto é uma ferramenta que utiliza imagens de satélite para calcular as condições de diversos indicadores, como temperatura, qualidade do solo e qualidade do ar. Para termos uma boa noção da qualidade do ar, este estudo tem como proposta aprofundar o conhecimento sobre sensoriamento remoto e processamento de imagem de satélite, aplicados ao problema de estimar quantidade de PM2,5 no ar. Foi feita uma revisão sistemática das abordagens e métodos que estão sendo utilizados para esse problema. Após análise dessa revisão, o objetivo deste trabalho foi estudar o funcionamento dos sistemas existentes e conceitos, focando em métricas para avaliação e de estimativas. Por fim, um sistema para predizer PM2,5 foi desenvolvido, e utilizado na predição de PM2,5 em São Paulo, obtendo $R^2 = 0.112$ e RMSE = 17.15. Então, este trabalho levanta a possibilidade de aplicar os métodos estudados no Brasil, e faz uma análise da situação atual, chegando a conclusão que é necessário mais estações meteorológicas para amostragem no Brasil.

Palavras-chave: PM2,5. Landsat 8. Processamento de Imagens. Qualidade do Ar.

Abstract

The scarcity in the number of meteorological centers for sampling in Brazil makes data on air quality scarce. The lack of frequent monitoring of air quality indicators is only noticed when it reaches levels that are harmful to health. There are several studies on air quality indicators, mainly focused on the PM2.5 indicator. Contamination by PM2.5, which is 2.5µm particulate matter, can cause heart problems, asthma in children and many respiratory problems. These risk factors demonstrate the importance of studying this indicator, since the type of particle it represents can affect the entire population of an area, impairing the quality of life and health of the inhabitants. Remote sensing is a tool that uses satellite images to calculate the conditions of various indicators, such as temperature, soil quality and air quality. In order to have a good idea of air quality, this study proposes to deepen the knowledge about remote sensing and satellite image processing, applied to the problem of estimating the amount of PM2.5 in the air. A systematic review of the approaches and methods being used to address this problem was carried out. After analyzing this review, the objective of this work was to study the functioning of systems and concepts, focusing on measurements, for evaluation and estimation. Finally, a system to predict PM2.5 was developed, and used to predict PM2.5 in São Paulo, obtaining $R^2 = 0.112$ and RMSE = 17.15. Then, this work raises the possibility of applying the methods studied in Brazil, and a conclusion of the most current situation in Brazil, arriving at the need for the most recent meteorological studies for Brazil.

Keywords: PM2.5. Landsat 8. Image Processing. Air Quality.

Lista de ilustrações

Figura 1 –	Ilustração do sensoriamento remoto	13
Figura 2 –	Na esquerda, Landsat 5 com resolução espacial 30m, na direita, Spot com	
	resolução espacial 10m	14
Figura 3 –	Na esquerda, imagem colorida (bandas 4, 3 e 2), na direita imagem termal	
	(banda10)	15
Figura 4 –	Na esquerda, Imagem com resolução radiométrica de 2 bits, na direita, Imagem	
	com resolução radiométrica de 8 bits	17
Figura 5 –	Na esquerda, ilustração do satélite Aqua, na direita, ilustração do satélite Terra	18
Figura 6 –	Ilustração do Satélite Landsat 8	19
Figura 7 –	AOD no mundo	20
Figura 8 –	Imagem Landsat 8 sem tratamento	21
Figura 9 –	Imagem de NDVI	21
Figura 10 -	- Gráfico de uma regressão de PM2,5	29
Figura 11 -	- Estrutura de arquivos na pasta do programa	34
Figura 12 -	- Planilha Utilizada	35
Figura 13 -	- Trecho do código do arquivo main.py onde os metadados são obtidos	37
Figura 14 -	- Trecho do código do arquivo main.py onde o arquivo data.csv é lido	38
Figura 15 -	- Trecho do código do arquivo main.py referente à obter valor por localização	
	em WGS84	39
Figura 16 -	- Trecho do código do arquivo main.py referente à obter refletância no TOA .	39
Figura 17 -	- Trecho do código do arquivo main.py referente à obter refletância atmosférica	40
Figura 18 -	- Trecho do código do arquivo main.py referente ao cálculo e impressão do	
	modelo	40
Figura 19 -	- Trecho do código do arquivo main.py referente ao cálculo valores previstos .	41
Figura 20 -	- Trecho do código do arquivo main.py referente ao cálculo de RMSE e geração	
	do gráfico	41
Figura 21 -	- Gráfico comparando dados previstos (linha), observados(pontos) e o intervalo	
	de confiança (área transparente ao redor da linha)	42
Figura 22 -	- Console com resultados do programa	42

Lista de tabelas

Tabela 1 –	Características espectrais e espaciais dos dados obtidos por instrumentos do	
	MODIS	15
Tabela 2 –	Características espectrais e espaciais dos dados obtidos por instrumentos do	
	MODIS	16
Tabela 3 –	Características espectrais e espaciais dos dados obtidos por instrumentos do	
	Landsat 8	17
Tabela 4 –	Comparação entre diferentes abordagens e sua correlação	23
Tabela 5 –	Comparação dos métodos aplicados	25
Tabela 6 –	Comparação com resultados de Mishra, Agarwal e Shukla (2021)	43

Lista de abreviaturas e siglas

PM	Do inglês, particulate matter, ou seja material particulado
Landsat	Uma série de satélites artificiais estadunidenses
MODIS	Do inglês, Moderate-Resolution Imaging Spectroradiometer, ou seja, espec- trorradiômetro de imagens de resolução moderada.
ΙΟΤ	Do inglês, Internet of Things, ou seja, Internet das Coisas
IA	Inteligência Artificial
AOT	Do inglês, Aerosol Optical Thickness, ou seja, Espessura Óptica de Aerossol
TOA	Do inglês, Top of Atmosphere, ou seja, "no Topo da Atmosfera".
NDVI	Do inglês, Normalized Difference Vegetation Index, ou seja, Índice de Vegetação da Diferença Normalizada.
DN	Do inglês, Digital Numbers, ou seja, números digitais.
SR	Do inglês, Surface Reflectance, ou seja, Refletância da Superfície.
L1TP	Do inglês, Terrain Precision Correction, ou seja, Correção precisa de terreno.
L1GT	Do inglês, Systematic Terrain Correction, ou seja, Correção sistemática de terreno.
L1GS	Do inglês, Geometric Systematic Correction, ou seja, Correção sistermática geométrica.
ISS	Instituto Saúde e Sustentabilidade.
CIESIN	Do inglês, Center for International Earth Science Information Network, ou seja, Centro de Rede de Informação de Geociência.
OLI	Do inglês, Operational Land Imager, ou seja Imageador Operacional de Solo
USGS	Do inglês, United States Geological Survey, ou seja, Serviço Geológico dos Estados Unidos

Lista de símbolos

ho Letra grega rho minúscula, representa a refletânc	cia.
--	------

- λ Letra grega lambda minúscula, representa o comprimento de onda
- θ Letra grega theta minúscula, representa um ângulo.

Sumário

1	Intr	odução		11	
	1.1	Contex	xto	11	
	1.2	Motiva	ação	12	
	1.3	Objeti	vos	12	
		1.3.1	Objetivos específicos	12	
2	Refe	erencial	teórico	13	
	2.1	Sensor	riamento remoto	13	
		2.1.1	Resolução espacial	14	
		2.1.2	Resolução espectral	14	
		2.1.3	Resolução radiométrica ou quantização	17	
		2.1.4	Resolução temporal	18	
	2.2	Satélit	es	18	
		2.2.1	Aqua e Terra	18	
		2.2.2	Landsat 8	19	
		2.2.3	Números digitais	19	
		2.2.4	Alguns índices medidos por satélites	19	
			2.2.4.1 Refletância no TOA	20	
			2.2.4.2 AOT	20	
			2.2.4.3 NDVI	20	
3	Tral	b <mark>alhos</mark> r	relacionados	22	
	3.1	3.1 Estrutura de pesquisa		22	
	3.2	Anális	e dos trabalhos relacionados	22	
	3.3	Pré-pro	ocessamento dos dados	23	
	3.4	Defini	ção de métricas	24	
		3.4.1	Coeficiente de determinação	24	
		3.4.2	Raiz do erro quadrático médio	25	
		3.4.3	Variáveis	25	
4	Met	odologi	a	27	
	4.1	Ativid	ades	27	
	4.2	Materi	iais e métodos	28	
		4.2.1	Bibliotecas	28	
		4.2.2	Sistema de coordenadas	28	
	4.3 Estudo aprofundado				

		4.3.1	Regressão linear	29
		4.3.2	Efemérides	30
		4.3.3	Dados do Landsat 8	30
5	Des	envolvin	nento	32
	5.1	Requis	itos do sistema	32
		5.1.1	Requisitos funcionais	32
		5.1.2	Requisitos não funcionais	33
	5.2	Anális	e de dados	33
		5.2.1	Bases de dados	33
		5.2.2	Dados da estação	35
		5.2.3	Leitura do arquivo de metadados	36
		5.2.4	Leitura do arquivo .csv	38
		5.2.5	Obtendo valor em uma posição	38
		5.2.6	Calculo de refletância no TOA	38
		5.2.7	Calculo de refletância atmosférica	39
		5.2.8	Cálculo e impressão do modelo	40
		5.2.9	Cálculo dos valores previstos	40
		5.2.10	Cálculo de RMSE e gráfico	41
	5.3	Resulta	ados	43
	5.4	Aplica	ção no Brasil	43
6	Con	clusão .		44
	6.1	Traball	hos futuros	44

1 Introdução

1.1 Contexto

Um dos indicadores mais difíceis de se analisar atualmente no Brasil é o de qualidade do ar. Além de não termos muitos centros meteorológicos para amostragem, a percepção da qualidade do ar depende de fatores como idade, nível de educação, renda familiar e histórico de doenças respiratórias Chen et al. (2020).

Dentre os poluentes do ar, existe o **PM2,5**. PM, do inglês, *particulate matter*, ou seja material particulado. O valor 2,5 significa que seu tamanho é menor ou igual a 2,5 micrômetros. O PM2,5 fica suspenso no ar, e pode ser respirado por nós. Como é pequeno o suficiente para penetrar os brônquios e chegar na corrente sanguínea, ele está intimamente ligado a uma maior ocorrência de doenças respiratórias e cardiovasculares. Bu et al. (2021) afirmam que 4,58 milhões de mortes foram causadas por exposição a PM2,5 globalmente em 2017, e segundo Pozzer et al. (2020), a poluição dos materiais particulados contribuiu com aproximadamente 15% das mortes por COVID-19 mundialmente no ano do estudo. O PM2,5 em específico tem a mais alta correlação com a taxa de reprodução do vírus em relação a outros poluentes do ar, como visto em Milicevic et al. (2021).

De acordo com o ISS (2019), em 2018, das 27 unidades federativas do Brasil, apenas 6 estados e o Distrito Federal atendem ao regulamento vigente de monitoramento de qualidade do ar. Porém, o PM2,5 é monitorado apenas nos estados de São Paulo e Rio de Janeiro e somente em um quarto de suas estações. Com esta falta de estrutura, é muito difícil utilizar os dados disponíveis em uma gestão de recursos inteligente na saúde.

1.2 Motivação

Tendo em vista a necessidade de acompanhar com maior frequência dados de PM2,5 no Brasil, esse estudo propõe estudar a utilização de imagens obtidas por sensoriamento remoto junto a dados de estações meteorológicas em solo para estimar PM2,5 com o uso de processamento de imagem. De acordo com o CIESIN (2021), existe uma extensa literatura de físicos e biólogos que demonstra o desenvolvimento ou o potencial do uso de técnicas de análise de sensores de satélite para identificar propriedades ambientais, monitorar processos físicos e biológicos relevantes e outras mudanças ambientais de larga escala. Como visto em AQ-SPEC (2021), um sensor de qualidade para ser usado em campo, pode custar entre U\$180,00 e U\$1000,00 dólares estadunidenses. Devido ao custo em implantar sensores de PM2,5 em solo e a falta de cobertura existente, esta é uma alternativa válida, como demonstrado por Mishra, Agarwal e Shukla (2021). Desta forma, a motivação deste trabalho é pesquisar métodos eficientes para estipulação desses dados.

1.3 Objetivos

Este trabalho tem o objetivo de analisar métodos que utilizam processamento de imagem para estimar a concentração de PM2,5 no Brasil. Para isso, este trabalho visa pesquisar abordagens que utilizam processamento de imagem obtida por sensoriamento remoto, em especial, do satélite Landsat 8. Este satélite possui uma resolução de 30m, e suas imagens são oferecidas de forma gratuita USGS (2021a).

1.3.1 Objetivos específicos

- Analisar métodos para estimar PM2,5 com processamento de imagem;
- Definir métricas para avaliação das estimativas;
- Avaliar as estimativas;
- Analisar a possibilidade de aplicar os métodos escolhidos no Brasil.

2 Referencial teórico

Nesse capítulo, serão apresentados os conceitos básicos para um melhor entendimento deste trabalho.

2.1 Sensoriamento remoto

É a técnica de obtenção de informações sobre de um objeto, área ou fenômeno localizado na Terra, sem contato físico. As informações podem ser obtidas através de radiação eletromagnética, e são apresentadas na forma de imagens, sendo mais utilizadas, atualmente, aquelas captadas por sensores óticos orbitais localizados em satélites, como o Landsat 8 e o Terra e Aqua IBGE (2021a). Os estudos de sensoriamento remoto podem ter aplicação na climatologia, agricultura, monitoramento de desastres naturais, entre outras.





Fonte: IBGE (2016)

A resolução de imagem obtida varia de acordo com o satélite por qual a imagem é capturada. Conforme Meneses (2012), as resoluções de uma imagem de sensoriamento remoto são divididas em 4 tipos: resolução espacial, espectral, radiométrica e temporal.

2.1.1 Resolução espacial

É o tamanho individual do elemento mínimo de área imageada no terreno, ou seja, quanto menor, mais detalhes teremos na imagem. Como visto na Figura 2, uma imagem com resolução de 10m possibilita a captura de um nível de detalhes maior que uma de 30m.

Figura 2 – Na esquerda, Landsat 5 com resolução espacial 30m, na direita, Spot com resolução espacial 10m



Fonte: Meneses (2012, p. 26)

2.1.2 Resolução espectral

O termo se refere às bandas espectrais, e pode compreender características como: o número de bandas que um sensor pode capturar (no OLI, do inglês, *Operational Land Imager*, ou seja Imageador Operacional de Solo, que é um dos sensores do Landsat, são 9), a largura de onda das bandas (o intervalo de comprimentos de onda em µm que cada banda opera), e as posições onde as bandas estão situadas no espectro eletromagnético (por exemplo infravermelho). A largura de onda das bandas é estipulada como na Tabela 1, na Tabela 2 e na Tabela 3.

Na Tabela 3 as bandas espectrais dos sensores do Landsat 8 podem ser observadas. Existem espectros visíveis e invisíveis para os olhos humanos, e diferentes combinações deles resultam em imagens diferentes. Como exemplo, a combinação das bandas 4, 3 e 2 do Landsat 8 poderia resultar em uma imagem colorida e a banda 10 oferece a imagem termal, como visto na Figura 3.



Figura 3 – Na esquerda, imagem colorida (bandas 4, 3 e 2), na direita imagem termal (banda10)

Fonte: NASA (2021a)

Banda	Comprimento de onda(µm)	Resolução(m)
Banda 1	0.620 - 0.670	250
Banda 2	0.841 - 0.876	250
Banda 3	0.459 - 0.479	500
Banda 4	0.545 - 0.565	500
Banda 5	1.230 - 1.250	500
Banda 6	1.628 - 1.652	500
Banda 7	2.105 - 2.155	500

Tabela 1 – Características espectrais e espaciais dos dados obtidos por instrumentos do MODIS

Fonte: USGS (2021b)

Banda	Comprimento de onda(µm)	Resolução(m)
Banda 8	0.405 - 0.420	1000
Banda 9	0.438 - 0.448	1000
Banda 10	0.483 - 0.493	1000
Banda 11	0.526 - 0.536	1000
Banda 12	0.546 - 0.556	1000
Banda 13	0.662 - 0.672	1000
Banda 14	0.673 - 0.683	1000
Banda 15	0.743 - 0.753	1000
Banda 16	0.862 - 0.877	1000
Banda 17	0.890 - 0.920	1000
Banda 18	0.931 - 0.941	1000
Banda 19	0.915 - 0.965	1000
Banda 20	3.660 - 3.840	1000
Banda 21	3.929 - 3.989	1000
Banda 22	3.929 - 3.989	1000
Banda 23	4.020 - 4.080	1000
Banda 24	4.433 - 4.498	1000
Banda 25	4.482 - 4.549	1000
Banda 26	1.360 - 1.390	1000
Banda 27	6.535 - 6.895	1000
Banda 28	7.175 - 7.475	1000
Banda 29	8.400 - 8.700	1000
Banda 30	9.580 - 9.880	1000
Banda 31	10.780 - 11.280	1000
Banda 32	11.770 - 12.270	1000
Banda 33	13.185 - 13.485	1000
Banda 34	13.485 - 13.785	1000
Banda 35	13.785 - 14.085	1000
Banda 36	14.085 - 14.385	1000

Tabela 2 – Características espectrais e espaciais dos dados obtidos por instrumentos do MODIS

Fonte: USGS (2021b)

Banda	Comprimento de onda(µm)	Resolução(m)
Banda 1 - Aerosol	0.43 - 0.45	30
Banda 2 - Azul Visível	0.450 - 0.51	30
Banda 3 - Verde Visível	0.53 - 0.59	30
Banda 4 - Vermelho Visível	0.64 - 0.67	30
Banda 5 - Infravermelho Próximo	0.85 - 0.88	30
Banda 6 - Infravermelho Médio	1.57- 1.65	30
Banda 7 - Infravermelho Médio	2.11-2.29	30
Banda 8 - Pancromática	0.50 - 0.68	15
Banda 9 - Cirrus	1.36 - 1.38	30
Banda 10 - Infravermelho Termal	10.6 - 11.19	100
Banda 11 - Infravermelho Termal	11.5 - 12.51	100

Tabela 3 – Características espectrais e espaciais dos dados obtidos por instrumentos do Landsat 8

Fonte: USGS (2021a)

2.1.3 Resolução radiométrica ou quantização

Define o número de níveis de radiância que o detector pode discriminar. É normalmente quantificado em bits. No caso do Landsat 8, todas as bandas possuem resolução radiométrica de 12 bits, ou seja, 4096 níveis de radiância. A diferença entre duas imagens com resolução radiométrica diferente pode ser facilmente visualizada na figura Figura 4, onde a maior quantização torna a imagem mais detalhada. Como exemplos, em uma imagem com 2 bits, o pixel só pode assumir 4 valores, enquanto uma imagem com 8 pode assumir 256.

Figura 4 – Na esquerda, Imagem com resolução radiométrica de 2 bits, na direita, Imagem com resolução radiométrica de 8 bits



Fonte: Meneses (2012, p. 30)

2.1.4 Resolução temporal

É o intervalo de tempo em que o sensor revisita uma área. Como as imagens são obtidas por satélites orbitando a terra, a área que o sensor capta sendo limitada, é necessário orbitar algumas vezes a terra para poder obter imagens do mundo todo. Por isso temos um intervalo de tempo entre obtenção de imagens. No caso do Landsat, por exemplo, o sensor capta novamente um mesmo local depois de 16 dias. Porém, nem todas as imagens obtidas têm poucas nuvens, um fator que prejudica as predições, se tornando nesse quesito menos adequadas para uso no processo pesquisado. Por outro lado o MODIS, que tem a resolução de 24 hrs a 48 hrs, tem menor resolução espacial, não oferecendo tantos detalhes da área a ser estudada.

2.2 Satélites

O termo satélite se refere a satélite artificial quando, na astronomia, um objeto é posto em órbita ao redor de um corpo celeste, para observações científicas, telecomunicações ou pesquisa. Os sensores utilizados para obter as imagens estão em satélites que orbitam a terra. Existem satélites de diversos países e entidades privadas em órbita atualmente, como o Landsat e o MODIS da NASA (2021b) e o Spot da empresa francesa Airbus Defence and Space (2021).

2.2.1 Aqua e Terra

MODIS (do inglês, *Moderate Resolution Imaging Spectroradiometer*, ou seja, espectroradiômetro de imagens de resolução moderada) é o instrumento principal dos satélites Terra e Aqua, que podem ser vistos na Figura 5. O Terra orbita o planeta Terra de forma que cruza do norte para o sul da linha do equador no período da manhã, enquanto o Aqua cruza do sul para o norte da linha do equador no período da tarde. Assim, os sensores MODIS do Terra e do Aqua monitoram a superfície inteira da terra numa resolução temporal entre 1 e 2 dias. As bandas espectrais captadas pelo MODIS estão na Tabela 1

Figura 5 - Na esquerda, ilustração do satélite Aqua, na direita, ilustração do satélite Terra



Fonte: NASA (2021c)

2.2.2 Landsat 8

O programa Landsat da NASA tem como objetivo pôr satélites artificiais em órbita para aquisição de dados de recursos terrestres. O primeiro satélite Landsat começou a operar em 23 de julho de 1972, na época sob o nome de ERTS (do inglês, *Earth Resources Technology Satellite*, ou seja, Satélite Tecnológico de Recursos Terrestres). Os sensores utilizados a cada novo lançamento da série Landsat melhoram, aumentando suas resoluções espectrais, espaciais e radiométricas. No dia 11 de fevereiro de 2013 o Landsat 8 foi lançado, fornecendo imagens da Terra em 11 bandas espectrais, sendo 9 obtidas pelo sensor OLI e 2 obtidas pelo TIRS, que obtém imagens de infravermelho termal. As bandas espectrais captadas estão na Tabela 3 e o Landsat 8 pode ser observado na Figura 6.





Fonte: USGS (2021a)

2.2.3 Números digitais

Imagens do Landsat são calculadas em radiância absoluta usando dados obtidos em float 32-bit. Esses valores são convertidos para inteiros de 16 bits, possuindo valores que assumem entre 0 e 65.535, em números digitais, sem unidade. Esses valores são úteis para calcular radiância e refletância no TOA.

2.2.4 Alguns índices medidos por satélites

Alguns índices que podem ser calculados utilizando dados de satélites serão descritos nessa subseção.

2.2.4.1 Refletância no TOA

A refletância no TOA, do inglês *Top of Atmosphere*, ou seja, "no topo da atmosfera"é uma medida sem unidade que indica o quanto de radiação incidente numa determinada superfície foi refletida. Pode ser calculado como descrito na Equação 3.1 e Equação 3.2, utilizando os dados obtidos pelo satélite, que são recebidos em valores digitais. No caso do Landsat esses valores são inteiros de 16 bits. A refletância, que é calculada com o valor em números digitais, é um número real que pode assumir valores entre 0 e 1.

2.2.4.2 AOT

Do inglês, *Aerosol Optical Thickness*, ou seja, Espessura Óptica de Aerossol. É o grau com que os aerossóis previnem a transmissão de luz, absorvendo ou dispersando a mesma. Quando chamada de AOD, ou Profundidade Óptica de Aerossol, é definido como a integral dos coeficientes de extinção sobre uma coluna vertical de determinado tamanho na atmosfera. Um mapa de AOD no mundo pode ser visto na Figura 7

Figura 7 – AOD no mundo



Fonte: Observatory (2021)

2.2.4.3 NDVI

Do inglês, *Normalized Difference Vegetation Index*, ou seja, Índice de Vegetação por Diferença Normalizada. É um índice utilizado em análises ecológicas, pois ele pode ser um indicador da estrutura da vegetação (maiores valores de NDVI estão supostamente associados com vegetação mais complexa e valores menores com vegetação menos complexa), e também com análises de mudança no uso da terra ou incidência de queimadas. No Landsat 8, pode ser calculado com a Equação 2.1:

$$NDVI = (b5 - b4)/(b5 + b4)$$
(2.1)

onde:

b5 : Valor em números digitais da banda 5 do Landsat 8

b4 : Valor em números digitais da banda 4 do Landsat 8

Na Figura 8 podemos observar uma imagem Landsat 8, composta das bandas 6, 5 e 4, seguido de uma imagem NDVI na Figura 9.

Figura 8 – Imagem Landsat 8 sem tratamento



Fonte: USGS (2021c)





Fonte: USGS (2021c)

B Trabalhos relacionados

Este capítulo tem como intuito analisar trabalhos existentes sobre aplicações em que o processamento de imagens pode ser utilizado para predição de PM2,5.

3.1 Estrutura de pesquisa

Para o desenvolvimento deste trabalho foi feita uma busca de trabalhos relacionados nas bases de dados *IEEE*, *Science Direct*, *Research Gate* e Google Acadêmico. A string de busca utilizada foi (*"pm2.5"AND "remote sensing"*), e trabalhos de antes de 2015 foram excluídos, totalizando 93 artigos. Destes encontrados, 9 utilizavam soluções com IA, e 3 apresentavam soluções IOT para sensoriamento que não utilizavam dados de satélite para sensoriamento. A fim de delimitar os resultados aos critérios de inclusão, esses trabalhos foram desconsiderados. Também foram incluídos na pesquisa, os livros Meneses (2012) e USGS (2019)

3.2 Análise dos trabalhos relacionados

Um parâmetro utilizado por muitos trabalhos para estimar PM2,5 é a AOT. Esse parâmetro pode ser calculado a partir de imagens de sensores de satélites como o MODIS por Chu et al. (2003) e Landsat por Chen et al. (2014). Além da AOT, Pelletier, Santer e Vidot (2007) utiliza em sua pesquisa também condições meteorológicas, devido ao fato de que a concentração de PM2,5 pode ser influenciada por múltiplas variáveis, sendo umidade uma delas.

Zhao et al. (2018) discute a aplicação de uma técnica de fusão de imagens para adquirir uma imagem MODIS com resolução Landsat. O método foi proposto porque imagens na resolução espacial do Landsat (30 m) possuem maior acurácia na predição de PM2,5 numa escala de cidade, com relação a imagens do MODIS (1 Km), porém, com menor resolução temporal. Os dados do MODIS e do Landsat foram usados para criar um modelo de predição que depois foi aplicado em

imagens obtidas pelo MODIS, com uma técnica de fusão de imagem. Ele chega à conclusão que as imagens resultantes da fusão que tem resolução espacial de 30 m obtêm maior acurácia na predição de PM2,5 em escala de cidade. Uma comparação direta entre o uso de resolução 10 Km e 1 Km é feita por Li et al. (2005), que toma como conclusão que o coeficiente de correlação entre PM2,5 e AOT na imagem com resolução de 1Km é maior.

De acordo com Mishra, Agarwal e Shukla (2021), além da umidade, a temperatura também pode ser utilizada para o cálculo, obtendo maior correlação que a mesma abordagem multi-espectral sem utilizar fatores meteorológicos em consideração. A Tabela 4 compara algumas dessas abordagens e seus resultados.

Abordagem	Tipo de Parâmetro	Correlação
Mishra, Agarwal e Shukla (2021)	Único	0.86
Mishra, Agarwal e Shukla (2021)	Múltiplo	0.92
Chen et al. (2014)	Único	0.67
Chen et al. (2014)	Múltiplo	0.8279

Tabela 4 – Comparação entre diferentes abordagens e sua correlação

Fonte: Mishra, Agarwal e Shukla (2021) e Chen et al. (2014)

Apesar de utilizar variáveis diferentes, Mishra, Agarwal e Shukla (2021), Chen et al. (2014) e Zhao et al. (2018) utilizaram modelos de regressão multivariável para seus estudos, obtendo com sucesso cobertura de dados de PM2,5 em uma escala de cidade, utilizando poucos sensores em terra para o treinamento do modelo.

3.3 Pré-processamento dos dados

Antes do processo de obtenção da AOT, os trabalhos relacionados fazem um préprocessamento das imagens do Landsat, conforme citado no Guia do Landsat por USGS (2019). Esse processo tem como objetivo obter a refletância no TOA. Os processo começa na Equação 3.1

$$\rho_{\lambda'} = M_{\rho} * Q + A_{\rho} \tag{3.1}$$

onde:

 $\rho_{\lambda'}$: refletância Espectral Planetária no TOA de determinado comprimento de onda λ , sem correção para ângulos solares;

 M_{ρ} : fator multiplicativo de refletância, pode ser encontrado no arquivo de metadados de cada pacote de imagens como REFLECTANCEW_MULT_BAND_n, onde n representa o número da banda do Landsat 8 correspondente a λ , conforme Tabela 3;

Q: intensidade de um pixel na imagem de satélite, em números digitais;

 A_{ρ} : fator de soma de refletância, pode ser encontrado no arquivo de metadados de cada pacote de imagens como REFLECTANCE_ADD_BAND_n, onde n representa o número da banda do Landsat 8 correspondente a λ , conforme Tabela 3.

O valor de $\rho_{\lambda'}$ está sem correção de ângulo de elevação solar. Essa correção é necessária pois a luz pode vir do sol e refletir da Terra para o sensor em ângulos diferentes conforme as estações do ano mudam e pequenas variações na orbita do satélite ocorrem. Assim, é necessário o cálculo descrito na Equação 3.2

$$\rho_{\lambda} = \frac{\rho_{\lambda'}}{\cos \theta_{SZ}} = \frac{\rho_{\lambda'}}{\sin \theta_{SE}}$$
(3.2)

onde:

 ρ_{λ} : refletância Espectral Planetária no TOA;

 θ_{SE} : ângulo de elevação local do sol; o ângulo central de elevação do sol, em graus, está presente no arquivo de metadados de cada pacote de imagens;

 θ_{SZ} : ângulo de zênite solar local;

 $\theta_{SZ}: 90^\circ - \theta_{SE}.$

3.4 Definição de métricas

As métricas escolhidas para avaliar os trabalhos propostos foram: coeficiente de determinação (R^2), o RMSE e o número de imagens utilizadas para o treinamento da regressão. Essas métricas foram escolhidas pois como visto em Mishra, Agarwal e Shukla (2021) e Chen et al. (2014), elas afetam a qualidade da predição.

3.4.1 Coeficiente de determinação

O coeficiente de determinação, ou R^2 , é uma medida estatística que varia entre 0 e 1, expressando a quantidade da variância dos dados que é explicada pelo modelo, onde um valor alto representa um melhor ajuste. A definição geral do coeficiente de determinação é descrita na Equação 3.3:

$$R^{2} = \frac{\sum_{i=1}^{n} (\hat{y}_{i} - \bar{y}_{i})^{2}}{\sum_{i=1}^{n} (y_{i} - \bar{y}_{i})^{2}}$$
(3.3)

Onde:

 $\bar{y_i}$: média das observações y_i ;

 \hat{y}_i : valor estimado em y_i ;

y_i: valor observado em i;

n: número de observações.

3.4.2 Raiz do erro quadrático médio

O RMSE, do inglês *Root Mean Squared Error* ou raiz do erro quadrático médio é uma medida que representa a raiz da média dos erros entre os valores observados (medido em solo) e predições (calculado), onde um valor menor representa um melhor ajuste Equação 3.4:

$$RMSE = \sqrt{(1/n)\sum_{i=1}^{n} (y_i - x_i)^2}$$
(3.4)

Onde:

y_i: valor medido em uma observação i;

 x_i : valor previsto em uma observação i;

n: número de elementos observados.

3.4.3 Variáveis

Os trabalhos relacionados têm em comum o uso do PM2,5 calculado no solo e dados de diferentes bandas de satélite na predição de PM2,5. Além disso, as abordagens também utilizam outras variáveis meteorológicas, como pode ser visto na Tabela 5

Tabela 5 – Comparação dos métodos aplicados

Autor	Variáveis	
Mishra, Agarwal e Shukla (2021)	Refletância das bandas 1-4, Temperatura, Umidade	0.92
Mishra, Agarwal e Shukla (2021)	Refletância das bandas 1-4	0.86
Chen et al. (2014)	AOT	0.67
Chen et al. (2014)	AOT e LST	0.7222
Chen et al. (2014)	AOT e TVDI	0.7505
Chen et al. (2014)	AOT, LST e TVDI	0.8279

Fonte: Mishra, Agarwal e Shukla (2021) e Chen et al. (2014)

Ambos Mishra, Agarwal e Shukla (2021) e Chen et al. (2014) utilizam do processo de cálculo da refletância no TOA citado na Equação 3.1 e na Equação 3.2, porém o que segue difere nas abordagens seguidas pelos dois autores.

Mishra, Agarwal e Shukla (2021) descreveram um método mais simples de obtenção do valor de PM2,5. Após adquirir a refletância no TOA, os valores de refletância atmosférica foram calculados, como na Equação 3.5. O autor subtrai a refletância da superfície da terra da refletância no TOA para assim ter um valor aproximado de refletância da atmosfera. Os valores

então foram aplicados numa regressão linear, juntamente com a temperatura em Kelvin e umidade em porcentagem, observados na estação meteorológica, com o objetivo de achar correlação da refletância da atmosfera com valor de PM2,5.

$$R = R_{TOA} - R_{sup} \tag{3.5}$$

onde:

R: refletância atmosférica;

 R_{TOA} : refletância no TOA;

R_{sup}: refletância na Superfície.

Já Chen et al. (2014) utilizou a refletância da superfície nas bandas 2 e 4 para calcular AOT, utilizando uma LUT, do inglês *Look up Table*, ou seja, uma tabela que descreve a transformação dos valores. Devido a necessidade de dados em solo de AOT indisponíveis no Brasil esse método não será utilizado.

Além disso, Chen et al. (2014) utilizaram de apenas 2 imagens Landsat 8, uma para estimar os coeficientes do modelo e outra para a validação, enquanto Mishra, Agarwal e Shukla (2021) utilizaram 7 imagens, onde um subgrupo não explicitado foi utilizado para estimar os coeficientes, enquanto o resto foi utilizado para validação.

4 Metodologia

A partir de uma análise das abordagens citadas no Capítulo 3, nota-se a necessidade de utilizar mais de uma variável num modelo de predição para uma melhor estimativa da quantidade de PM2,5 no ar. Além disso, é possível perceber nesses estudos que as cidades estudadas têm um maior número de sensores em solo que a maior parte das cidades brasileiras, podendo tornar um desafio aplicar a mesma abordagem sem modificações. Percebe-se que um método de estimar PM2,5 no país é necessário para melhor orientar políticas públicas e privadas de controle a vírus que são transmitidos por material particulado, melhor alocação de recursos e evitar problemas ocasionados por exposição de longo prazo ao PM2,5.

4.1 Atividades

A realização das tarefas necessárias para alcançar os objetivos definidos na Seção 1.3 se deu de acordo com as seguintes atividades:

- Estudo aprofundado dos métodos disponíveis para estimar PM2,5;
- Definição de métricas para avaliação de estimativas;
- Busca por bibliotecas que contribuam para o projeto;
- Estudo e implementação do processamento de coordenadas;
- Aplicação de testes;
- Analisar a possibilidade de aplicar os métodos escolhidos no Brasil.

4.2 Materiais e métodos

O desenvolvimento e testes do modelo é realizado de maneira complementar à abordagem introduzida por Mishra, Agarwal e Shukla (2021). O desenvolvimento seguiu utilizando a linguagem de programação Python.

4.2.1 Bibliotecas

Foram escolhidas as seguintes bibliotecas: numpy, osgeo, math e statsmodels.

- Numpy foi escolhido para tratar das operações entre matriz;
- OSGeo foi escolhido para calcular as coordenadas baseado em latitude e longitude;
- Math foi escolhido para calcular operações como cosseno;
- Subprocess foi escolhido para interagir com a ferramenta externa gdallocationtool;
- Pathlib foi escolhido para lidar com locais de arquivos;
- Csv foi escolhido para interagir com arquivos csv;
- Seaborn foi escolhido para criar gráficos estatisticos;
- gdal foi escolhido para utilizar a ferramenta gdallocationinfo, que encontra o valor de um pixel pela coordenada.

4.2.2 Sistema de coordenadas

O sistema de coordenadas utilizado pelas imagens Landsat 8 é o WGS84 (IBGE (2021b)). O WGS84 é um sistema de referencia de coordenadas onde a terra é representada como um elipsoide e é dividida em latitude e longitude. É o sistema de referencia utilizado no GPS. Como a imagem recebida é uma matriz, uma ferramenta para calcular qual pixel representa um par de valores de latitude e longitude é necessária. A ferramenta gdallocationinfo com a opção -WGS84, fornecida no pacote gdal é um programa de terminal que pode realizar essa tarefa. Além de obter a coordenada do pixel, também recebemos o valor. É importante notar que essa ferramenta e as imagens de Landsat utilizam o padrão WGS84. Caso os dados de localização estejam em um padrão diferente, será necessária a conversão para WGS84.

4.3 Estudo aprofundado

O estudo foi desenvolvido utilizando os métodos disponíveis para estimar PM2,5. Para um melhor entendimento sobre as técnicas, métodos e dados utilizados nos nossos trabalhos relacionados, alguns conceitos complementares foram estudados e serão descritos nessa seção.

4.3.1 Regressão linear



Figura 10 – Gráfico de uma regressão de PM2,5

Fonte: Mishra, Agarwal e Shukla (2021)

Os modelos de regressão linear são ferramentas comuns na estatística cujo foco é a realização de inferências. A inferência consiste em, a partir de evidências encontradas para uma amostra, realizar generalizações de resultados para uma população. Ou, simplificando, verificar a correlação entre duas ou mais variáveis e testar o quanto se pode confiar nas estimativas encontradas Chein (2012, p. 7). Uma equação de regressão linear simples pode ser descrita como na Equação 4.1:

$$y = \alpha + \beta x \tag{4.1}$$

onde:

y: é a variável endógena (dependente); representa o que o modelo tentará prever;

 α : é uma constante, que representa a interceptação da reta com o eixo vertical;

 β : representa a inclinação (coeficiente angular) em relação à variável exógena;

x: variável exógena (independente).

Como pode ser visto na Figura 10, a regressão é uma reta que passa por pontos. Já uma equação de regressão linear de múltiplas variáveis pode ser descrita na Equação 4.2:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$
(4.2)

onde:

y: é a variável endógena (dependente); representa o que o modelo tentará prever;

 α : é uma constante, que representa a interceptação da reta com o eixo vertical;

 β_i : representa a inclinação (coeficiente angular) em relação à variável exógena x_i ;

x_i: variáveis exógenas (independentes).

Usaremos a forma matricial dessa equação, descrita na Equação 4.3:

$$Y = X\beta \tag{4.3}$$

onde:

Y: é uma matriz $n \times 1$, onde n é o número observações da variável endógena;

 β : é uma matriz p+1 × 1, onde p representa o número de coeficientes da regressão, sendo sua primeira coluna equivalente à constante α da Equação 4.2;

X: é uma matriz $n \times p+1$, onde n é o número de observações da variável exógena e p é o número de coeficientes da regressão.

A regressão linear é uma ferramenta utilizada por Mishra, Agarwal e Shukla (2021) pois estudos anteriores citados pelos mesmos traçaram uma correlação linear entre o nível de PM2,5 e AOT e entre AOT e os valores de refletância no TOA. Apesar disso, esse tipo de regressão pode estimar valores negativos ou valores muito acima do esperado, dependendo do quão bem os valores observados se adequarem à regressão.

4.3.2 Efemérides

Um efemérides é um conjunto de dados que fornece os locais dos corpos celestes (incluindo satélites artificiais) num intervalo regular. O efemérides da USGS (2021d) mostra a velocidade e posição de um corpo no momento que a imagem foi coletada, e mostra a velocidade e posição dos satélites em intervalos de um minuto.

Imagens recém adquiridas do Landsat 7 são processadas inicialmente com efemérides previstas, e então reprocessadas com as efemérides definitivas em até 48 hrs. Esse processo não existe no Landsat 8, já que os dados de correção do satélite são muito precisos.

4.3.3 Dados do Landsat 8

Segundo USGS (2021a) os dados do Landsat 8 estão separados para consulta em 2 níveis de correção, L1, e L2. Quanto maior o nível mais a imagem já foi processada. Cada nível tem sua aplicação, por exemplo, para calcular a refletância no TOA precisamos de dados do L1 e L2, enquanto para compor uma imagem colorida precisamos apenas do L2. O nível L1 fornece dados

que incluem efeitos atmosféricos, em números digitais. Já o nível L2 fornece dados corrigidos para remover os efeitos atmosféricos, resultando em imagens de refletância da superfície.

Os níveis dependem do uso de pontos de controle em solo, e de um modelo digital da elevação do terreno. O nível L1 se divide em três:

- L1TP, do inglês, *Terrain Precision Correction*, ou seja, correção precisa de terreno. É calibrado usando dados de pontos de controle no solo e de modelo digital de elevação. É o nível de qualidade L1 mais alto, viável para análise em nível de pixel. Pontos de controle no solo usados na correção são derivados do conjunto de dados GLS2000 (2021), do inglês, *Global Land Survey 2000*, ou seja, pesquisa terrestre global do ano 2000.
- L1GT, do inglês, Systematic Terrain Correction, ou seja, correção sistemática de terreno. É calibrado usando correções geométricas sistemáticas aplicadas usando os dados efemérides e de modelo digital de elevação. O L1GT é o maior nível de calibragem radiométrica e sistemática possível para cenas coletadas pelo Landsat 7 e 8 na Antártica.
- L1GS, do inglês, *Geometric Systematic Correction*, ou seja, correção sistemática geométrica. É calibrado apenas utilizando correções geométricas sistemáticas aplicadas utilizando dados efemérides.

Além dessa divisão, ainda é feita a divisão em níveis de qualidade, divididos em:

- Tier 1: são calibradas para a maior qualidade geométrica e radiométrica possível. Ideal para estudos que dependem de alta precisão.
- Tier 2: não preenchem os critérios do Tier 1 devido a grande cobertura de nuvens, controle de solo insuficiente, desvio de órbita, entre outros motivos.
- Real-Time Tier: disponibiliza dados em tempo real, imediatamente após a sua aquisição. Depois de seus dados serem reprocessados, a imagem é atribuida ao Tier 1 ou Tier 2. Dados em tempo real podem ser usados para reagir rapidamente a desastres em nível global.

Uma divisão de coleções foi criada para reprocessar imagens antigas, criando a Coleção 2 de imagens Landsat. A Coleção 2 tem diversas melhorias de processamento radiométrico e geométrico e soluciona alguns erros relacionados, dentre eles, imagens classificadas com nível de correção errado, datas diferentes e cenas repetidas. Com base nas opções dessa divisão, as imagens que mais se adequam para esse trabalho são as imagens L1TP de Tier 1, sendo consultados na Coleção 2, devido a maior precisão e as imagens terem menor cobertura de nuvens que as imagens de Tier 2.

5 Desenvolvimento

Um programa que visa calcular PM2,5 de forma que não requer conhecimento prévio de programação foi desenvolvido, considerando as características que mais se adequem às necessidades dos usuários.

5.1 Requisitos do sistema

Segundo Sommerville (2011), na engenharia de software, os requisitos de um sistema são as descrições do que o sistema deve fazer, os serviços que oferece e as restrições a seu funcionamento. Esses requisitos refletem as necessidades dos clientes para um sistema que serve a uma finalidade determinada. Os requisitos podem ser classificados como:

- Requisitos funcionais representam o que o sistema deve fazer;
- Requisitos não funcionais não estão relacionados diretamente com o que o sistema deve fazer, e sim com suas propriedades, como confiabilidade ou tempo de resposta.

Dessa forma utilizaremos a classificação especificada nesta seção para garantir que o sistema funcione de maneira confiável. Foi utilizada a técnica de estórias de usuários para obter os requisitos. Nessa técnica, a equipe e o cliente tentam identificar um conjunto de exemplos de uso, que abrange toda funcionalidade que será incluída no sistema final.

5.1.1 Requisitos funcionais

Para o levantamento dos requisitos funcionais, a colaboração de Simone Soraia Silva Sardeiro foi necessária, para alinhar os objetivos do programa com as necessidades dos usuários. Simone é uma pesquisadora da área de desenvolvimento e meio ambiente, sendo o objetivo desse trabalho auxiliar no estudo de qualidade do ar e promover a interdisciplinaridade. No Quadro 1, são apresentados os requisitos funcionais desse sistema.

Identificação	Descrição	
RF01	Como usuário, gostaria de receber o valor de R^2 da previsão	
DE02	Como usuário, gostaria de receber o valor de RMSE dos valores	
KI'02	medidos e previstos	
RE03	Como usuário, gostaria de receber um gráfico de valores medidos e	
N 105	previstos	
	Como desenvolvedor, devo desenvolver um sistema que utiliza dos	
RF04	dados de PM2,5 medidos em solo, a latitude e longitude de onde os	
	dados foram aferidos e imagens de satélite para calcular PM2,5	
	Como desenvolvedor, devo desenvolver um sistema que utiliza dos	
RF05	dados de PM2,5 medidos em solo e os dados previstos de PM2,5	
	pelo sistema do RF04, para calcular o valor de R^2 da previsão	
	Como desenvolvedor, devo desenvolver um sistema que utiliza dos	
RF05	dados de PM2,5 medidos em solo e os dados previstos de PM2,5	
	pelo sistema do RF04, para calcular o valor de <i>RMSE</i> da previsão	
	Como desenvolvedor, devo desenvolver um sistema que utiliza dos	
RF06	dados de PM2,5 medidos em solo e os dados previstos de PM2,5	
	pelo sistema do RF04, para criar um gráfico que compara os mesmos	

Quadro 1 - Requisitos funcionais

Fonte: Autor (2021).

5.1.2 Requisitos não funcionais

Como citado por Sommerville (2011), os requisitos não funcionais definem de que modo o sistema realizará os requisitos funcionais, de forma a definir limitações do projeto. No Quadro 2, são apresentados os requisitos não funcionais desse sistema.

5.2 Análise de dados

Esta seção tem como objetivo explorar a base de dados utilizada no sistema de predição de PM2,5 deste trabalho.

5.2.1 Bases de dados

Para que possamos predizer um valor de PM2,5 precisamos obter valores medidos em solo de PM2,5 para associarmos a uma regressão linear. Além disso também são necessárias imagens obtidas por sensoriamento remoto referentes a área estudada, que também serão usadas na regressão. Para este trabalho, dados de PM2,5 em solo foram obtidos do Qualar (2021), um sistema da CETESB (Companhia Ambiental do Estado de São Paulo) e imagens Landsat 8 foram

Identificação	Descrição					
DNE01	O software deve funcionar na linha de comando, sem argumentos,					
KINI'UI	para facilitar a entrada de dados do usuário					
DNE02	O software deve receber como entrada um arquivo data.csv ^{<i>a</i>} , que					
KINF02	contém dados referentes ao nome da imagem.					
DNE02	O arquivo .csv deve fornecer dados de latitude e longitude no sistema					
KINI'UJ	WGS84					
RNF04	O arquivo .csv deve fornecer dados de PM2,5 em µm/m ³					
RNF05	O arquivo .csv deve estar na pasta do programa					
	Os arquivos de imagem devem estar separados em uma pasta					
RNF06	correspondente ao identificador de produto Landsat, conforme					
	Figura 11					
DNE07	O gráfico e valor de RMSE devem se referir a imagem da ultima					
ININI'U7	coluna do .csv e os dados previstos					

O 1	0	D	• • .	~	C	•	•
()madro	2 -	Reo	unsitos	nao	tur	10101	າສາຮ
Quuuno	_	100	aisitos	muo	Iui	10101	iuis

^a <https://github.com/DCOMP-UFS/TCC-Monitoramento-de-indicador-de-qualidade-de-ar-por-processamento-de-imagem/ blob/main/data.csv>

Fonte: Autor (2021).

Figura 11 – Estrutura de arquivos na pasta do program	Figura 11 –	Estrutura	de ar	quivos	na	pasta	do	programa
---	-------------	-----------	-------	--------	----	-------	----	----------

Nome	Data de modificação	Тіро	Tamanho
LC08_L1TP_219076_20190817_20200827_02_T1	01/12/2021 18:40	Pasta de arquivos	
LC08_L1TP_219076_20190918_20200826_02_T1	13/12/2021 05:29	Pasta de arquivos	
LC08_L1TP_219076_20200328_20200822_02_T1	13/12/2021 05:29	Pasta de arquivos	
LC08_L1TP_219076_20200429_20200820_02_T1	13/12/2021 05:29	Pasta de arquivos	
LC08_L1TP_219076_20200531_20200820_02_T1	01/12/2021 18:39	Pasta de arquivos	
LC08_L1TP_219076_20210603_20210608_02_T1	30/11/2021 23:33	Pasta de arquivos	
LC08_L1TP_219076_20210822_20210827_02_T1	01/12/2021 18:39	Pasta de arquivos	
LC08_L2SP_219076_20190817_20200827_02_T1	01/12/2021 18:41	Pasta de arquivos	
LC08_L2SP_219076_20190918_20200826_02_T1	13/12/2021 05:29	Pasta de arquivos	
LC08_L2SP_219076_20200328_20200822_02_T1	13/12/2021 05:29	Pasta de arquivos	
LC08_L2SP_219076_20200429_20200820_02_T1	13/12/2021 05:30	Pasta de arquivos	
LC08_L2SP_219076_20200531_20200820_02_T1	01/12/2021 18:41	Pasta de arquivos	
LC08_L2SP_219076_20210603_20210608_02_T1	01/12/2021 01:52	Pasta de arquivos	
LC08_L2SP_219076_20210822_20210827_02_T1	01/12/2021 18:41	Pasta de arquivos	
🖈 data.csv	13/12/2021 08:53	Arquivo de Valore	1 KB
퉐 main.py	13/12/2021 18:28	Python File	8 KB

Fonte: Autor (2021)

obtidas do EarthExplorer (2021), um sistema da USGS, do inglês, *United States Geological Survey*, ou seja, Serviço Geológico dos Estados Unidos.

Na estrutura da pasta do programa (Figura 11), temos a planilha data.csv¹, o arquivo do programa main.py² e uma lista de pastas. O nome de cada pasta contém o identificador

¹ data.csv

² main.py

das imagens Landsat que serão utilizadas na predição, obtidas da base de dados do EarthExplorer (2021). Estes identificadores contêm informações referentes a quando a imagem foi obtida, processada, e seus níveis, como descrito na Subseção 4.3.3. Por exemplo, a imagem LC08_L1TP_219076_20190817_20200827_02_T1 foi obtida na data 17/08/2019 e processada em nível L1 e Tier 1 na data 27/08/2020. Além disso, o trecho "LC08"indica que é uma imagem Landsat 8.

5.2.2 Dados da estação

Dados da estação devem ser fornecidos em formato .csv para o sistema. Uma planilha pode ser criada por ferramentas gratuitas como Google Planilhas ou Libre Office Calc e salva no formato .csv. Um exemplo de planilha utilizada no cálculo pode ser observada em Figura 12, e deve seguir o padrão descrito a seguir. A primeira coluna se refere ao nome da estação de onde os dados de solo são obtidos, a segunda coluna se refere a latitude da estação, a terceira a longitude da estação e a partir da quarta temos na primeira linha o identificador de produto sem o identificador de nível, por exemplo, a imagem com identificador LC08_L1TP_219076_20190817_20200827_02_T1, resulta em 219076_20190817_20200827_02_T1, e nas linhas que se sucedem temos os valores observados de PM2,5 no dia e horário que a imagem foi obtida.

	A	В	C		D		E	
1	Estação	Latitude	Longitude	219076_201908	17_20200827_02_T1	219076_202005	31_20200820_02_T1	219076_202106
2	Congonhas	-23,61632	-46,663466		23		44	+
3	Grajaú - Parelheiros	-23,776266	-46,696961		27		51	
4	Guarulhos – Paço Municipal	-23,455534	-46,518533		24		38	i i
5	Itaim Paulista	-23,501547	-46,420737		28		65	i
6	Jundiaí	-23,192004	-46,897097		18		27	,
7	Marg. Tietê – Pte, Remédios	-23,518706	-46,74332		42		44	+
8	Mooca	-23,549734	-46,600417		23		35	i
9	Osasco	-23,526721	-46,792078		46		44	+
10	Parque D. Pedro II	-23,544846	-46,627676		24		41	
11	S. Bernardo – Centro	-23,698671	-46,546232		20		44	+
12	Santana	-23,505993	-46,62896		28		38	i
13	Santos - Ponta da Praia	-23,981295	-46,30051		19		30	1
	F		G		н		I	
2190	76_20210603_20210608_02_T	1 219076_2	0210822_202	210827_02_T1	219076_20190918_20	200826_02_T1	219076_20200429_20	200820_02_T1
		37		28		51		17
		31		30		38		12
		31		33		48		19
		49		42		44		17
		24		38		39		18
		44		43		51		21
		35		30		47		15
		42		61		59		20
		44		35		46		16
		36		28		49		16
		32		40		43		16
		22		23		10		11

Figura 12 – Planilha Utilizada

Fonte: Autor (2021)

5.2.3 Leitura do arquivo de metadados

O pré-processamento das imagens de nível L1 recebidas precisa ser realizado para obter a Refletância no TOA, como mostrado em Equação 3.1. Para esse pré-processamento, é necessário obtermos alguns dados incluídos no arquivo de metadados finalizado com "MTL.txt", que acompanha cada imagem. Esses dados são:

- SUN_ELEVATION representa o ângulo que a luz do sol incide sobre a área imageada com relação a posição do sensor do satélite;
- REFLECTANCE_MULT_BAND_n fator de multiplicação da refletância de uma banda n;
- REFLECTANCE_ADD_BAND_n fator de soma da refletância de uma banda n, n varia entre 1 e 4.

Na Figura 13 podemos observar um trecho do código do arquivo main.py que implementa essa funcionalidade.

Figura 13 – Trecho do código do arquivo main.py onde os metadados são obtidos.

```
def getToaCalArgs(path):
    dataDict = {}
    with open(path) as f:
        for line in f.readlines():
            if "SUN_ELEVATION" in line:
                dataDict.update({
                    "SUN ELEVATION": float(str(line).split("= ")[1])
                })
            elif "REFLECTANCE MULT BAND 1" in line:
                dataDict.update({
                    "REFLECTANCE_MULT_BAND_1": float(str(line).split("= ")[1])
                })
            elif "REFLECTANCE MULT BAND 2" in line:
                dataDict.update({
                    "REFLECTANCE_MULT_BAND_2": float(str(line).split("= ")[1])
                })
            elif "REFLECTANCE_MULT_BAND_3" in line:
                dataDict.update({
                    "REFLECTANCE_MULT_BAND_3": float(str(line).split("= ")[1])
                })
            elif "REFLECTANCE_MULT_BAND_4" in line:
                dataDict.update({
                    "REFLECTANCE_MULT_BAND_4": float(str(line).split("= ")[1])
                })
            elif "REFLECTANCE ADD BAND 1" in line:
                dataDict.update({
                    "REFLECTANCE_ADD_BAND_1": float(str(line).split("= ")[1])
                })
            elif "REFLECTANCE_ADD_BAND_2" in line:
                dataDict.update({
                    "REFLECTANCE_ADD_BAND_2": float(str(line).split("= ")[1])
                })
            elif "REFLECTANCE_ADD_BAND_3" in line:
                dataDict.update({
                    "REFLECTANCE ADD BAND 3": float(str(line).split("= ")[1])
                })
            elif "REFLECTANCE ADD BAND 4" in line:
                dataDict.update({
                    "REFLECTANCE_ADD_BAND_4": float(str(line).split("= ")[1])
                })
    return dataDict
```

Fonte: Autor (2021)

5.2.4 Leitura do arquivo .csv

Os dados de Latitude, Longitude e PM2,5 são carregados do arquivo .csv. O arquivo é uma tabela em UTF-8. Além disso, também é lido o identificador das imagens. Na Figura 14 podemos observar um trecho do código do arquivo main.py que implementa essa funcionalidade.

Figura 14 - Trecho do código do arquivo main.py onde o arquivo data.csv é lido

```
los de média móvel de PM2.5 das estações no mesmo dia e hora da obtenção das imagens, são
coords = []
pm = []
levelString = []
with open(imagePath + '\\data.csv', newline='', encoding='utf-8') as f:
   reader = csv.DictReader(f)
   pmAux = []
   for row in reader:
       coords.append([float(row['Longitude'].replace(',', '.')), float(row['Latitude'].replace(',', '.'))])
       row.pop('Estação')
       row.pop('Longitude')
       row.pop('Latitude')
       levelString = list(row.keys())
       pmAux.append(list(row.values()))
    for i in range(len(pmAux[0])):
       for pmValue in pmAux:
           pm.append(int(pmValue[i]))
```



5.2.5 Obtendo valor em uma posição

A imagem recebida pelo programa desenvolvido possui seu sistema de coordenadas em WGS84 (Subseção 4.2.2), e para obter o valor de um pixel localizado na imagem precisamos dos dados de latitude e longitude obtidos do Qualar (2021). Devido a variações de órbita do satélite, cada imagem pode começar e terminar numa latitude e longitude ligeiramente diferentes. Ou seja, se obtivermos a posição (x, y) de um pixel correspondente a um par de latitude e longitude da imagem, uma outra imagem do mesmo local em outro dia pode ter a posição (x, y) diferente. As informações quanto a posição da imagem podem ser obtidas utilizando a ferramenta gdallocationinfo, que recebe a imagem, latitude e longitude e retorna a posição (x,y) corrigida e o valor nessa posição. Na Figura 15 podemos observar um trecho do código do arquivo main.py que implementa essa funcionalidade, utilizando a biblioteca subprocess para executar um comando externo ao Python.

5.2.6 Calculo de refletância no TOA

Os imagens Landsat 8 de nível L1 são então utilizadas no calculo da Refletância no TOA em cada banda. Primeiramente a refletância sem correção em cada banda é calculada, conforme

Figura 15 – Trecho do código do arquivo main.py referente à obter valor por localização em WGS84



Fonte: Autor (2021)

Equação 3.1. Em seguida os valores obtidos são utilizados no calculo da Refletância no TOA corrigida em cada banda, conforme Equação 3.2. Podemos observar o processo no trecho de código do arquivo main.py presente na Figura 16.

Figura 16 – Trecho do código do arquivo main.py referente à obter refletância no TOA

#Obtendo caminho da imagem de nível 1
level1folder = imagePath + level1identifier + levelString[i] + "\\"
#Valores da imagem de nível 1, em números digitais
<pre>band1dn = getValuesFromCoords(level1folder + level1identifier + levelString[i] + "_B1.TIF", coords)</pre>
<pre>band2dn = getValuesFromCoords(level1folder + level1identifier + levelString[i] + "_B2.TIF", coords)</pre>
<pre>band3dn = getValuesFromCoords(level1folder + level1identifier + levelString[i] + "_B3.TIF", coords)</pre>
<pre>band4dn = getValuesFromCoords(level1folder + level1identifier + levelString[i] + "_B4.TIF", coords)</pre>
band1DNarray = np.array(band1dn)
band2DNarray = np.array(band2dn)
band3DNarray = np.array(band3dn)
band4DNarray = np.array(band4dn)
#Obtendo dados necessários para calculo de TOA Reflectance nos pontos
toaCalArgs = getToaCalArgs(level1folder + level1identifier + levelString[i] + "_MTL.txt");
#Calculo de TOA Reflectance nos pontos
<pre>sunCos = math.cos(toaCalArgs['SUN_ELEVATION']*math.pi/180)</pre>
band1toa = (band1DNarray*toaCalArgs['REFLECTANCE_MULT_BAND_1'] + toaCalArgs['REFLECTANCE_ADD_BAND_1'])/sunCos
band2toa = (band2DNarray*toaCalArgs['REFLECTANCE_MULT_BAND_2'] + toaCalArgs['REFLECTANCE_ADD_BAND_2'])/sunCos
band3toa = (band3DNarray*toaCalArgs['REFLECTANCE_MULT_BAND_3'] + toaCalArgs['REFLECTANCE_ADD_BAND_3'])/sunCos
band4toa = (band4DNarray*toaCalArgs['REFLECTANCE MULT BAND 4'] + toaCalArgs['REFLECTANCE ADD BAND 4'])/sunCos



5.2.7 Calculo de refletância atmosférica

Os valores das imagens de nível L2, que representam a refletância da superfície, são então utilizados no cálculo da Refletância Atmosférica em cada banda, conforme Equação 3.5. Esse processo é necessário pois como apresentado por Mishra, Agarwal e Shukla (2021), a refletância atmosférica representa a refletância na porção acima da superfície e abaixo do topo da atmosfera, representando a refletância no ar. Podemos observar o processo no trecho do código do arquivo main.py presente na Figura 17.

Figura 17 - Trecho do código do arquivo main.py referente à obter refletância atmosférica

```
#Obtendo caminho da imagem de nível 2
level2folder = imagePath + level2identifier + levelString[i] + "\\"
#Valores da imagem de nível 2, em refletância de superfície
band1sr = getValuesFromCoords(level2folder + level2identifier + levelString[i] + "_SR_B1.TIF", coords)
band2sr = getValuesFromCoords(level2folder + level2identifier + levelString[i] + "_SR_B2.TIF", coords)
band3sr = getValuesFromCoords(level2folder + level2identifier + levelString[i] + "_SR_B3.TIF", coords)
band4sr = getValuesFromCoords(level2folder + level2identifier + levelString[i] + "_SR_B3.TIF", coords)
band4sr = getValuesFromCoords(level2folder + level2identifier + levelString[i] + "_SR_B4.TIF", coords)
band2sRarray = np.array(band1sr)
band2sRarray = np.array(band1sr)
band2sRarray = np.array(band2sr)
band3sRarray = np.array(band3sr)
band4sRarray = np.array(band4sr)
#Calculo de Atmospheric Reflectance nos pontos
band1ar = (band1toa - band1SRarray)
band2ar = (band2toa - band2SRarray)
band4ar = (band4toa - band4SRarray)
band4ar = (band4toa - band4SRarray)
band4ar = (band4toa - band4SRarray)
```

Fonte: Autor (2021)

5.2.8 Cálculo e impressão do modelo

Os valores de Refletância Atmosférica são divididos por 1000 para o modelo obter menores coeficientes, não afetando o cálculo, pois na predição de um valor, conforme pode ser observado na Subseção 5.2.9, dividimos por 1000 novamente. Essa divisão é necessária para a biblioteca statsmodels não acusar erro devido ao valor da constante ser muito maior que o das variáveis exógenas. Podemos observar o processo no trecho do código do arquivo main.py presente na Figura 18.

Figura 18 - Trecho do código do arquivo main.py referente ao cálculo e impressão do modelo



Fonte: Autor (2021)

5.2.9 Cálculo dos valores previstos

Os valores dos coeficientes da regressão que foram calculados são então utilizados na Equação 4.3, onde o Y são os coeficientes calculados e o β é a ultima coluna de valores de PM2,5,

lida do .csv, conforme RNF07. Podemos observar o processo no trecho do código do arquivo main.py presente na Figura 19.

Figura 19 - Trecho do código do arquivo main.py referente ao cálculo valores previstos





5.2.10 Cálculo de RMSE e gráfico

Os valores previstos e medidos são então utilizados para calcular o RMSE entre eles, como descrito na Equação 3.4. Após o cálculo, os valores são impressos e um gráfico é exibido conforme definido na Seção 5.1. Podemos observar o processo no trecho do código do arquivo main.py presente na Figura 20. Na Figura 21 podemos observar um gráfico numa janela separada, que representa um gráfico comparando os valores de PM2,5 obtidos nas estações com a regressão calculada. O gráfico gerado pode ser salvo clicando no ícone do disquete. Já na Figura 22 estão presentes os resultados do programa no console, podemos observar os valores dos coeficientes da regressão linear em x_1, x_2, x_3, x_4 e const. Onde const é a constante α , e x_i é o valor do coeficiente i, ou β .

Figura 20 – Trecho do código do arquivo main.py referente ao cálculo de RMSE e geração do gráfico



Fonte: Autor (2021)



Figura 21 – Gráfico comparando dados previstos (linha), observados(pontos) e o intervalo de confiança (área transparente ao redor da linha)

Fonte: Autor (2021)

Figura 22 –	Console com	resultados	do	programa
-------------	-------------	------------	----	----------

C:\WINDOWS\syste	em32\o	md.exe					- C	×
C:\Users\Victor\Documents\TCC\codigo>python main.py OLS Regression Results								
Dep. Variable: Model: Mothod: Date: Time: No. Observations Df Residuals: Df Model: Covariance Type:	:	y OLS Least Squares Mon, 13 Dec 2021 18:37:35 72 67 4 nonrobust	R-sq Adj. F-st Prob Log- AIC: BIC:	uared: R-squared: atistic: (F-statisti Likelihood:	ic):	0.112 0.059 2.113 0.0888 -280.98 572.0 583.3		
	coet	std err	t	P> t	[0.025	0.975]		
const 43 x1 20 x2 -16 x3 -2 x4 0	. 3906 . 1433 . 2468 . 4536 . 5569	17.153 10.263 12.525 8.834 6.753	2.530 1.963 -1.297 -0.278 0.082	0.014 0.054 0.199 0.782 0.935	9.153 -0.341 -41.247 -20.087 -12.923	77.628 40.628 8.753 15.180 14.037		
Omnibus: Prob(Omnibus): Skew: Kurtosis:		0.063 0.969 0.014 2.717	Durb Jarq Prob Cond	in-Watson: ue-Bera (JB) (JB): . No.):	1.045 0.243 0.886 264.		
Notes: [1] Standard Errors assume that the covariance matrix of the errors is correctly specified. RMSE: 17.15161458502367μg/m ³								

Fonte: Autor (2021)

5.3 Resultados

Os valores de R^2 e RMSE foram calculados, com o programa, de acordo com todos os requisitos apresentados na Seção 5.1. Em contraste com Mishra, Agarwal e Shukla (2021), o resultado apresentado com o conjunto de dados utilizado obteve um R^2 mais próximo a 0 (quanto mais próximo a 1, melhor), representando que as variáveis utilizadas (bandas 1 a 4) não influenciam o valor de PM2,5 tanto quanto no estado da arte. Porém obteve menor RMSE (quanto mais próximo a 0, melhor), como pode ser visto em Tabela 6, isso demonstra que em valores absolutos, os valores medidos e previstos foram mais próximos que no estado da arte. Isso pode acontecer devido ao número limitado de imagens que podemos utilizar na predição, além de falhas nos sensores em terra, e uma diferença geográfica entre as cidades estudadas pela literatura, que em geral não são litorâneas, ao contrário da cidade estudada (São Paulo), o que pode surtir um efeito na predição.

Autor	Variáveis	R^2	RMSE($\mu g m^{-3}$)	Local
Mishra, Agarwal e Shukla (2021)	Refletância das bandas 1-4, Tempe- ratura, Umidade	0.92	20.97	Delhi
Mishra, Agarwal e Shukla (2021)	Refletância das bandas 1-4	0.86	42.85	Delhi
Autor	Refletância das bandas 1-4	0.112	17.15	São Paulo

Tabela 6 – Comparação com resultados de Mishra, Agarwal e Shukla (2021)

Fonte: Mishra, Agarwal e Shukla (2021) e Chen et al. (2014)

5.4 Aplicação no Brasil

O Brasil enfrenta alguns problemas para efetuar a predição de PM2,5, boa parte dos sensores operam somente no Rio de Janeiro e São Paulo, além de a maior parte do território não ser coberto por dados de PM2,5. Dito isso, a aplicação nesses estados específicos parece ser viável, desde que um número suficiente de amostras seja aplicado. Porém, pela proximidade com o litoral, a presença de nuvens nas imagens é frequente, dificultando o trabalho de coleta de dados e podendo influenciar negativamente na predição. De 195 imagens presentes na área estudada, apenas 10 tinham baixa cobertura de nuvens, e 4 dessas imagens podiam ser utilizadas devido a ausência de dados de estações locais para usar no calculo da regressão.

6 Conclusão

A busca por meios de medir poluição do ar tem uma grande importância na saúde e consequentemente na economia do país, principalmente depois do Covid-19 ter prejudicado toda a população. Esse estudo visa tornar simples o entendimento dos termos e do processo para o estudo de poluentes no ar com o uso de imagens de sensoriamento remoto, tornando possível mais pesquisadores estudarem os métodos de predição, obtenção de dados e a lógica do processo.

Durante este trabalho, um estudo dos métodos disponíveis para estimar PM2,5 foi realizado, além do estudo, um programa para aplicar um dos métodos foi criado. Para isso foi necessário estudar sensoriamento remoto, os sensores disponíveis e seus dados, além de como processar eles para obter informações sobre o assunto estudado.

Ademais, o Brasil tem uma pouca quantidade de sensores de PM2,5 e distribuição desigual, se faz necessário um maior investimento nesse tipo de instrumento, para podermos aplicar os métodos citados neste trabalho em outros estados e cidades do Brasil de uma forma mais precisa, já que, como visto na Seção 5.3 o modelo não teve um desempenho tão bom no Brasil.

6.1 Trabalhos futuros

Devido ao fato dos estudos sobre PM2,5 serem ainda escassos no país e a falta de estações de monitoramentos, uma continuidade desse estudo pode ser feita de forma a implementar outros algoritmos para calcular PM2,5, e até mesmo utilizar variáveis como temperatura e umidade. Além disso, o modelo pode ser treinado com imagens de outras cidades ao redor do mundo, onde existe um número maior de estações de monitoramento, e pode auxiliar nos estudos da área da geografia e saúde.

Referências

AIRBUS DEFENCE AND SPACE. 2021. Disponível em: https://www.intelligence-airbusds. com/imagery/constellation/spot/>. Acesso em: 01 jul 2021. Citado na página 18.

AQ-SPEC. 2021. Disponível em: http://www.aqmd.gov/aq-spec/evaluations/summary-pm. Acesso em: 01 jul 2021. Citado na página 12.

BU, X. et al. Global PM2.5-attributable health burden from 1990 to 2017: Estimates from the Global Burden of disease study 2017. *Environmental Research*, v. 197, p. 111123, 2021. ISSN 0013-9351. Disponível em: https://www.sciencedirect.com/science/article/pii/S0013935121004175>. Citado na página 11.

CHEIN, F. Introdução aos modelos de regressão linear. [S.l.]: Enap, 2012. Citado na página 29.

CHEN, Q. et al. Associations between individual perceptions of PM2.5 pollution and pulmonary function in Chinese middle-aged and elderly residents. *BMC Public Health*, v. 20, n. 1, p. 899, dez. 2020. ISSN 1471-2458. Disponível em: https: //bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-020-08713-6>. Citado na página 11.

CHEN, Y. et al. Estimating ground-level PM2.5 concentration using Landsat 8 in Chengdu, China. In: IM, E.; YANG, S.; ZHANG, P. (Ed.). *Remote Sensing of the Atmosphere, Clouds, and Precipitation V.* SPIE, 2014. v. 9259, p. 193–206. Disponível em: https://doi.org/10.1117/12.2068886>. Citado 6 vezes nas páginas 22, 23, 24, 25, 26 e 43.

CHU, D. A. et al. Global monitoring of air pollution over land from the Earth Observing System-Terra Moderate Resolution Imaging Spectroradiometer (MODIS). *Journal of Geophysical Research: Atmospheres*, v. 108, n. D21, 2003. Disponível em: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002JD003179>. Citado na página 22.

CIESIN. 2021. Disponível em: <<u>http://www.ciesin.org/TG/RS/satimprt.html</u>>. Acesso em: 01 jul 2021. Citado na página 12.

EARTHEXPLORER. 2021. Disponível em: <<u>https://earthexplorer.usgs.gov/></u>. Acesso em: 01 jul 2021. Citado 2 vezes nas páginas 34 e 35.

GLS2000. 2021. Disponível em: <a href="https://www.usgs.gov/core-science-systems/nli/landsat/global-land-survey-gls?qt-science_support_page_related_con=0#qt-science_support_page_related_con=0#qt-science_support_page_related_con=0. Acesso em: 01 jul 2021. Citado na página 31.

IBGE. Atlas geográfico escolar. [S.l.]: IBGE, 2016. Citado na página 13.

IBGE. 2021. Disponível em: https://atlasescolar.ibge.gov.br/conceitos-gerais/o-que-e-cartografia/sensoriamento-remoto.html. Acesso em: 01 jul 2021. Citado na página 13.

IBGE. 2021. Disponível em: http://geoftp.ibge.gov.br/informacoes_sobre_posicionamento_geodesico/sirgas/sisref_2.pdf. Acesso em: 01 jul 2021. Citado na página 28.

ISS. 2019. Disponível em: https://www.saudeesustentabilidade.org.br/wp-content/uploads/2019/06/An%C3%A1lise-do-Monitoramento-de-Qualidade-do-Ar-no-Brasil_ISS.pdf>. Acesso em: 01 jul 2021. Citado na página 11.

LI, C. et al. Retrieval, validation, and application of the 1-km aerosol optical depth from modis measurements over hong kong. *IEEE Transactions on Geoscience and Remote Sensing*, v. 43, n. 11, p. 2650–2658, 2005. Citado na página 23.

MENESES, T. d. A. P. R. *Introdução ao Processamento de Imagens de Sensoriamento Remoto*. [S.I.]: UnB, 2012. Citado 3 vezes nas páginas 14, 17 e 22.

MILICEVIC, O. et al. PM2.5 as a major predictor of COVID-19 basic reproduction number in the USA. *Environmental Research*, v. 201, p. 111526, 2021. ISSN 0013-9351. Disponível em: https://www.sciencedirect.com/science/article/pii/S0013935121008203. Citado na página 11.

MISHRA, R. K.; AGARWAL, A.; SHUKLA, A. Predicting Ground Level PM2.5 Concentration Over Delhi Using Landsat 8 Satellite Data. *International Journal of Remote Sensing*, Taylor & Francis, v. 42, n. 3, p. 827–838, 2021. Disponível em: https://doi.org/10.1080/2150704X.2020.1832279>. Citado 11 vezes nas páginas 6, 12, 23, 24, 25, 26, 28, 29, 30, 39 e 43.

NASA. 2021. Disponível em: <https://landsat.gsfc.nasa.gov/satellites/landsat-8/landsat-8-bands/ >. Acesso em: 01 jul 2021. Citado na página 15.

NASA. 2021. Disponível em: ">https://www.nasa.gov/>. Acesso em: 01 jul 2021. Citado na página 18.

NASA. 2021. Disponível em: https://science.nasa.gov/get-involved/toolkits/spacecraft-icons. Acesso em: 01 jul 2021. Citado na página 18.

OBSERVATORY, E. 2021. Disponível em: ">https://earthobservatory.nasa.gov/global-maps/MODAL2_M_AER_OD>. Acesso em: 01 jul 2021. Citado na página 20.

PELLETIER, B.; SANTER, R.; VIDOT, J. Retrieving of particulate matter from optical measurements: A semiparametric approach. *Journal of Geophysical Research: Atmospheres*, v. 112, n. D6, 2007. Disponível em: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006737>. Citado na página 22.

POZZER, A. et al. Regional and global contributions of air pollution to risk of death from COVID-19. *Cardiovascular Research*, v. 116, n. 14, p. 2247–2253, 2020. ISSN 0008-6363. Disponível em: https://doi.org/10.1093/cvr/cvaa288. Citado na página 11.

QUALAR. 2021. Disponível em: br/>. Acesso em: 01 jul 2021">https://qualar.cetesb.sp.gov.br/>. Acesso em: 01 jul 2021. Citado 2 vezes nas páginas 33 e 38.

SOMMERVILLE, I. *Engenharia de Software*. [S.l.]: Pearson Education do Brasil Ltda, 2011. Citado 2 vezes nas páginas 32 e 33.

USGS. Landsat 8 (L8) Data Users Handbook. [S.l.]: USGS, 2019. Citado 2 vezes nas páginas 22 e 23.

USGS. 2021. Disponível em: <https://www.usgs.gov/core-science-systems/nli/landsat/landsat-8? qt-science_support_page_related_con=0#qt-science_support_page_related_con>. Acesso em: 01 jul 2021. Citado 4 vezes nas páginas 12, 17, 19 e 30.

USGS. 2021. Disponível em: https://lpdaac.usgs.gov/data/get-started-data/collection-overview/ missions/modis-overview/>. Acesso em: 01 jul 2021. Citado 2 vezes nas páginas 15 e 16.

USGS. 2021. Disponível em: https://www.usgs.gov/media/images/landsat-surface-reflectance-and-normalized-difference-vegetation-index. Acesso em: 01 jul 2021. Citado na página 21.

USGS. 2021. Disponível em: https://www.usgs.gov/core-science-systems/nli/landsat/download-definitive-ephemeris. Acesso em: 01 jul 2021. Citado na página 30.

ZHAO, Y. et al. High spatiotemporal resolution PM2.5 concentration estimation with satellite and ground observations: A case study in New York City. In: *2018 IEEE International Conference on Environmental Engineering (EE)*. [S.1.: s.n.], 2018. p. 1–5. Citado 2 vezes nas páginas 22 e 23.