

**DESENVOLVIMENTO DE ROTINAS EM AMBIENTE PYTHON
PARA O PROCESSAMENTO DE IMAGENS E DADOS ÓPTICOS
COLETADOS EM AMBIENTES AQUÁTICOS PARA
MONITORAMENTO DE SISTEMAS AQUÁTICOS CONTINENTAIS
POR SENSORIAMENTO REMOTO**

**RELATÓRIO FINAL DE PROJETO DE INICIAÇÃO CIENTÍFICA
(PIBIC/INPE/CNPq)**

Felipe Menino Carlos (Fatec Jessen Vidal, Bolsista PIBIC/CNPq)
E-mail: felipe.carlos@fatec.sp.gov.br

Cláudio Clemente Faria Barbosa (INPE, Orientador)
E-mail: claudio.barbosa@inpe.br

Julho de 2019

“Eu posso não ter ido onde gostaria, mas acho que cheguei onde deveria”.

Douglas Adams

RESUMO

A utilização de técnicas de Sensoriamento Remoto tem sido amplamente utilizada em diversas áreas, por conta da riqueza de informações que podem ser extraídas dos dados gerados. A obtenção destas informações depende diretamente de metodologias de processamento dos dados. Atualmente o Laboratório de Instrumentação de Sistemas Aquáticos (LabISA) utiliza dados de Sensoriamento Remoto para o estudo e caracterização de águas interiores, sendo necessário a aplicação de diferentes metodologias de correções. Portanto, este trabalho é motivado pela oportunidade da criação de rotinas que facilitem o processamento dos dados utilizados no LabISA. Para isto, fez-se a criação de rotinas em ambiente Python e MATLAB para a correção de efeitos atmosféricos e de sunglint, aplicação de agrupamento, automatização de simulações HydroLight e visualização de dados. Como resultado as metodologias de processamento podem ser aplicadas de forma mais simples e direta.

Palavras-chave: Correções. Python. MATLAB. Sistemas aquáticos.

LISTA DE FIGURAS

	<u>Pág.</u>
Figura 1 - Multiespectral X Hiperspectral.....	5
Figura 2 - Influências na REM	6
Figura 3 - Interferência atmosférica	7
Figura 4 - Geração do <i>sunlint</i>	8
Figura 5 - Divisão do espaço amostral	10
Figura 6 - Distâncias	11
Figura 7 - Etapas do K-Médias	12
Figura 8 - Metodologia	14
Figura 9 - API de correção atmosférica	17
Figura 10 - Interface principal do AtmosPy	18
Figura 11 - Seleção de geolocalização.....	19
Figura 12 - Comparativo de uso de memória	20
Figura 13 - Janelas da metodologia Kutser (2009)	22
Figura 14 - Janelas da metodologia Kutser (2013)	23
Figura 15 - Telas das correções Goodman, Hedley e Lyzenga	23
Figura 16 - Visualização dos grupos	25
Figura 17 - Figuras geradas pela biblioteca de plot	27
Figura 18 - Formas de utilização (HL).....	28
Figura 19 - Camadas do HydroLight-Py	29
Figura 20 - Atalhos da tabela HydroLight-Py	30
Figura 21 - Fluxo de funcionamento	30

LISTA DE TABELAS

	<u>Pág.</u>
Tabela 1 - Requisitos (AtmosPy)	17
Tabela 2 - Requisitos (rGlint)	21
Tabela 3 - Requisitos (K-Médias)	25

LISTA DE SIGLAS E ABREVIATURAS

GEE	<i>Google Earth Engine</i>
SR	<i>Sensoriamento remoto</i>
LabISA	<i>Laboratório de Instrumentação de Sistemas Aquáticos</i>
REM	<i>Radiação eletromagnética</i>
MSI	<i>Multispectral Instrument</i>
OLI	<i>Operational Land Imager</i>
OLCI	<i>Ocean and Land Colour Instrument</i>
IA	<i>Inteligência Artificial</i>
AM	<i>Aprendizado de máquina</i>
SAM	<i>Spectral Angle Mapper</i>
HL	<i>HydroLight</i>
CPU	<i>Central Process Unit</i>
GPU	<i>Graphics Processing Unit</i>

SUMÁRIO

	<u>Pág.</u>
1 INTRODUÇÃO.....	1
1.1 Objetivo geral	1
1.2 Objetivos específicos	2
2 Fundamentação teórica	3
2.1 Sensoriamento remoto orbital em ambientes aquáticos.....	3
2.2 Sensores imageadores	4
2.3 Imagens multiespectrais e hiperespectrais.....	4
2.4 Correções de imagens orbitais	5
2.4.1 Correção atmosférica.....	6
2.4.2 Métodos alternativos.....	7
2.4.3 Métodos físicos.....	8
2.4.4 Correção de Sunlint.....	8
2.5 Aprendizado de máquina	9
2.5.1 Árvore de decisão	10
2.5.2 K-Médias.....	11
3 Materiais e métodos utilizados.....	13
3.1 Materiais	13
3.2 Métodos	14
4 Desenvolvimento e resultados.....	16
4.1 Pré-processamento dos dados	16
4.1.1 Correção atmosférica.....	16
4.1.2 Remoção de sunlint	20
4.2 Classificação e análise dos dados	24
4.2.1 K-Médias em dados espectrais.....	24
4.2.2 Árvore de decisão e visualização de dados.....	26
4.3 Simulações com HydroLight	27

5	Conclusão.....	32
5.1	Contribuições e conclusões.....	32
5.2	Publicações.....	32
5.3	Trabalhos futuros.....	33

1 INTRODUÇÃO

A utilização de produtos de sensoriamento remoto é feita pelas mais diversas áreas, para as mais variadas aplicações, podendo ir desde o controle e proteção ambiental (MASCARENHAS et al, 2008), até o estudo e verificação de meios aquáticos (CAIRO, 2015).

Os produtos de sensoriamento remoto podem ser coletado de diferentes fontes (NOVO, 2010), porém para todos os casos, podem ser necessários formas de correção e análise dos dados, uma vez que, durante a coleta ruídos podem ser captados e inseridos nos dados e também, análises rápidas e eficientes facilitam a obtenção de informações que podem ser utilizadas na tomada de decisão, principalmente para o estudo e controle de diferentes ambientes.

Para a correção e análise destes dados existem diversas metodologias que podem ser aplicadas e testadas, porém as aplicações podem ser trabalhosas e suscetíveis a erros, quando formas distintas são utilizadas para a aplicação das metodologias.

Desta forma foi levantado pelos pesquisadores e alunos de pós-graduação do LabISA que, para o aumento da eficiência da aplicação dos métodos de correção e análise, é necessário o desenvolvimento de ferramentas que auxiliem a execução destas etapas, encapsulando o máximo possível os diversos passos necessários nos métodos de correção.

1.1 Objetivo geral

Desenvolver ferramentas que facilitem a correção e análise de dados envolvidos nas pesquisas de análise e caracterização de sistemas aquáticos, desenvolvidas pelo LabISA.

1.2 Objetivos específicos

Como forma de atingir o objetivo geral, foram definidos os seguintes objetivos específicos

- Desenvolver rotinas para o pré-processamento de imagens e dados ópticos;
- Desenvolver rotinas para a classificação e visualização de dados;
- Facilitar a aplicação de metodologias de visualização de modelos híbridos.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo revisa os principais conceitos utilizados no desenvolvimento das ferramentas deste Trabalho.

2.1 Sensoriamento remoto orbital em ambientes aquáticos

Segundo Novo (2010), sensoriamento remoto (SR) é a junção do trabalho de sensores imageadores multiespectrais e equipamentos para processamento e transmissão de dados, que são colocados a bordo de aeronaves, espaçonaves (orbital) e diversas outras plataformas, com o objetivo de estudar a superfície do planeta terra.

A utilização do SR envolve a detecção, aquisição e análise de radiação eletromagnética (REM), emitidas ou refletidas por alvos terrestres (MAIO et al, 2008) e registradas por sensores remotos. As características da REM emitidas ou refletidas estão diretamente relacionadas as propriedades físico-químicas e mesmo biológicas dos objetos que estão sendo imageados pelos sensores remotos (MAIO et al, 2008), o que possibilita o estudo e caracterização de objetos e ambientes através do SR, como citado anteriormente.

Em ambientes aquáticos, o SR é aplicado com os mais variados objetivos, a citar, a determinação e mapeamento de propriedades de qualidade de água, como a clorofila e alguns sólidos (BARSOSA et al, 2019), e através destas informações de qualidade de água caracterizar biofísica e estado trófico dos diferentes sistemas aquáticos.

Por conta das características de sistemas aquáticos, sua relação com a REM o SR quando aplicado no estudo de ambientes aquáticos a nível orbital utilizam em larga escala sensores imageadores passivos e a radiação solar espectral (BARBOSA et al, 2019).

2.2 Sensores imageadores

Sensores imageadores, são definidos como instrumentos eletro-óptico capazes de obter imagens em diferentes intervalos do espectro eletromagnético da superfície terrestre (MENESES et al, 2012).

Como apresentado no capítulo anterior, estes sensores são amplamente utilizados em aplicações de sensoriamento remoto orbitais. MENESES et al (2012) define três motivos que justificam esta ampla aplicação, sendo eles: (i) Dados em formato digital, permitindo processamento mais fácil; (ii) Facilmente aplicáveis em plataformas espaciais, o que permitiu sua aplicação em escala global; (iii) Imagens disponíveis em diferentes níveis espectrais (multiespectral), o que tornou sua utilização útil para diferentes áreas.

Estes sensores são compostos por diversos componentes detectores de radiação eletromagnética (MENESES et al, 2012), que através de diferentes técnicas e métodos, convertem a REM para número digital (MENESES et al, 2012).

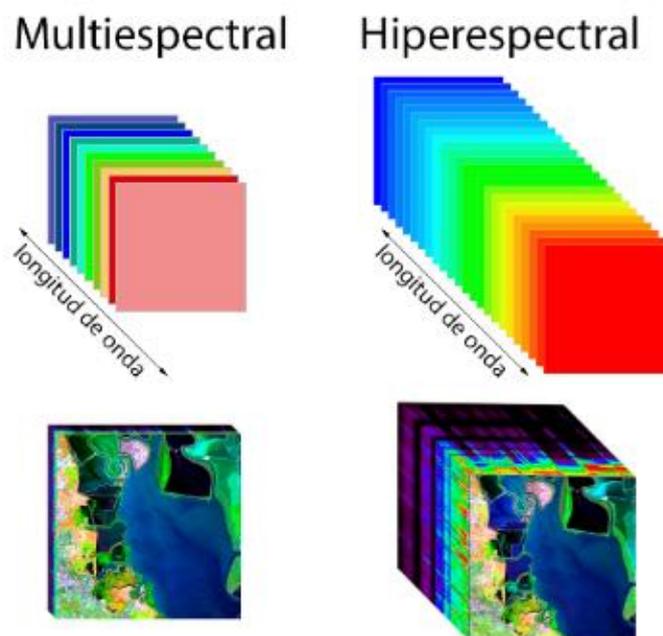
Atualmente existem diversos sensores sendo aplicados em plataformas de satélites, como por exemplo o *Multispectral Instrument* (MSI), que está a bordo do satélite Sentinel-2 e busca realizar a aquisição de imagens multiespectrais (ESA, 2018). Outro sensor também utilizado para a geração de produtos de SR é o *Operational Land Imager* (OLI), que com sua boa resolução radiométrica permite uma boa distinção dos alvos imageados (NASA, 2018). Por fim, existe também o *Ocean and Land Colour Instrument* (OLCI), que está a bordo do Sentinel-3 e oferece uma boa resolução espacial.

2.3 Imagens multiespectrais e hiperespectrais.

Os produtos do SR coletados através de sensores imageadores são as imagens, estas que por sua vez podem ser multiespectrais ou hiperespectrais (FURLANETTO, 2017). Estes tipos que podem variar de acordo com o tipo de sensor que está sendo utilizado.

A principal diferente entre estes dois tipos de produtos está na quantidade de espectros da REM registrado (FURLANETTO, 2017). No caso das imagens multiespectrais um pequeno conjunto de registros é feito, e para as imagens hiperespectrais tem-se centenas de espectros registrados. A Figura 1 apresenta um exemplo da quantidade de bandas captadas de uma mesma cena.

Figura 1 - Multiespectral X Hiperespectral



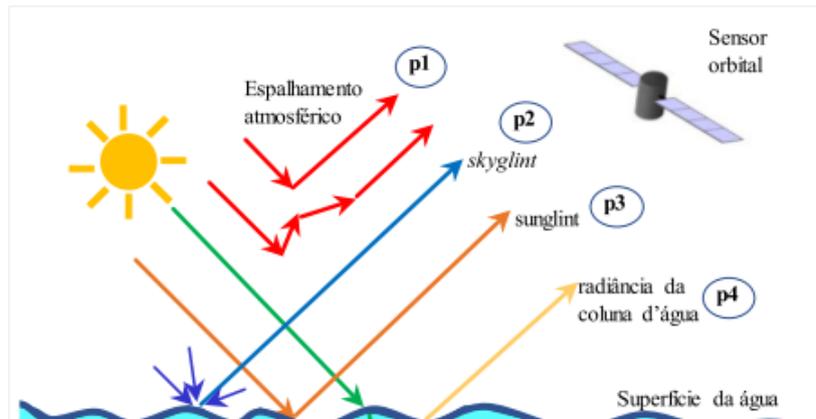
Fonte: Adaptado de Durán (2008)

Para os casos de imagens hiperespectrais tem-se uma quantidade muito maior de informações, o que pode significar um detalhamento muito maior do ambiente que está sendo imageado (FURLANETTO, 2017).

2.4 Correções de imagens orbitais

Durante o imageamento de uma certa região, o satélite orbital pode receber não somente a REM advinda dos objetos terrestres, mas também de uma série de elementos que influenciam o sinal recebido. A Figura 2 apresenta alguns tipos de sinais que o sensor orbital pode receber durante o imageamento.

Figura 2 - Influências na REM



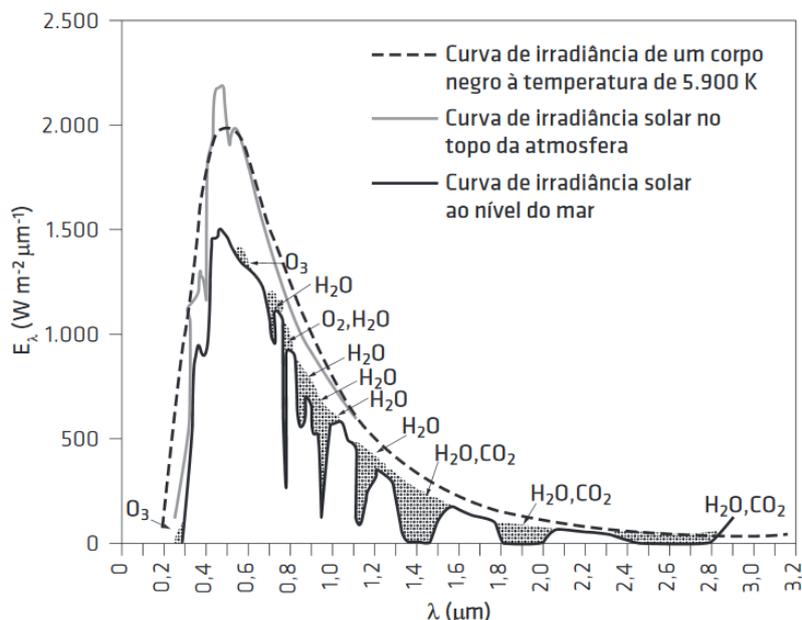
Fonte: Adaptado de Barbosa et al (2019)

O recebimento destes diferentes sinais pode influenciar nos dados que estão sendo gerados, uma vez que informações não advindas dos objetos imageados podem passar a fazer parte dos dados. Para remover estas influências dos dados, existem correções que podem ser aplicadas nas imagens.

2.4.1 Correção atmosférica

A correlação dos dados sensor-superfície quando influenciados pela atmosfera pode apresentar limitações, uma vez que esta é um meio extremamente dinâmico (LATORRE et al, 2002) que ao interagir com a REM, provoca alterações nos fluxos radiantes provenientes dos alvos (LATORRE et al, 2002). Esta interação (Figura 3), causa uma interferência espectral sobre o fluxo incidente na superfície terrestre (PONZONI et al, 2007).

Figura 3 - Interferência atmosférica



Fonte: Ponzoni et al (2007).

Assim, esta interação da REM com a atmosfera, pode causar incertezas nas análises, uma vez que, os dados presentes na imagem representam a soma dos efeitos atmosféricos e da realidade dos alvos capturados. Com isso torna-se necessário a remoção da interferência atmosférica nos valores presentes na imagem.

A remoção desta interferência pode ser feita através de alguns métodos, estes listados nas subseções a seguir.

2.4.2 Métodos alternativos

Métodos alternativos, são métodos que para a realização da correção utilizam como base somente a imagem, buscando identificar os valores dos pixels escuros, para quantificar a interferência dos efeitos atmosféricos nos demais valores da imagem (SOUZA, 2008).

2.4.3 Métodos físicos

Denomina-se, métodos físicos, os métodos que são fundamentados na teoria da transferência radiativa. Estes levam em consideração as propriedades físicas da atmosfera para a realização da correção (SOUZA, 2008).

2.4.3.1 6SV

O modelo 6SV é um código de transferência radiativa para a correção atmosférica de dados de diferentes satélites para uma grande variedade de condições climatológicas (VERMOTE et al., 1997).

Resultados apresentados por Martins (2017) indicam que, para certos grupos de águas interiores o 6SV apresentou melhores resultados de correção atmosférica que os demais métodos utilizados na correção.

2.4.4 Correção de Sunlint

De acordo com Streher (2013) o *sunlint*, pode ser definido como sendo a luz do sol que é refletida diretamente pela superfície da água, sem interagir com a coluna d'água, e atinge o campo de visada do sensor. A Figura 4 mostra a maneira de geração do *sunlint*.

Figura 4 - Geração do *sunlint*



Fonte: Adaptado de ESA (2012)

Este fenômeno está ligado diretamente relacionada ao vento (BARBOSA et al, 2019), uma vez que, quanto maior a velocidade do vento mais turbulenta está a água, o que aumenta as chances da ocorrência do *sunlint* (BARBOSA et al, 2019). O *sunlint* quando coletado pelos sensores imageadores, diminui a qualidade do dado resultante, uma vez que, a coleta do *sunlint* adiciona elementos de aleatoriedade aos dados (BARBOSA et al, 2019).

2.5 Aprendizado de máquina

Inteligência artificial (IA) é um campo da ciência da computação (ZUBEN, 2013), que busca a capacidade de raciocinar, perceber e agir para sistemas computacionais (WINSTON, 1992).

Para possibilitar as capacidades listadas a sistemas computacionais, diversas técnicas podem ser aplicadas, dentre elas se destaca o Aprendizado de Máquina (AM), que busca gerar um sistema inteligente através da experiência deste frente ao problema a ser resolvido (MICHALSKI; CARBONELL; MITCHELL, 2013), fazendo com que o sistema encontre uma forma de resolver o problema.

De acordo com Murphy (2012), os algoritmos de AM podem ser divididos nos grupos de aprendizado supervisionado e não supervisionado.

Para o grupo de aprendizado supervisionado, todo o resultado que o algoritmo deve produzir, frente a um conjunto de dados já conhecido. Assim o objetivo do algoritmo supervisionado é entender a relação dos dados e gerar um modelo que explique as relações entre as entradas e saídas que lhe são passadas (MURPHY, 2012).

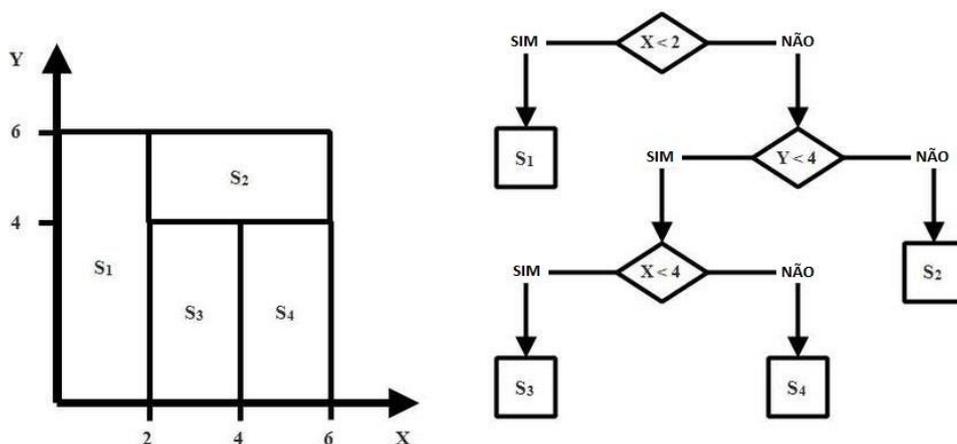
Já no grupo de aprendizado não supervisionado, os resultados a serem produzidos pelos algoritmos não são previamente conhecidos, fazendo com que os algoritmos sejam aplicados para gerar grupos (SILVA, 2017), e assim possibilitar o entendimento dos dados (MURPHY, 2012).

2.5.1 Árvore de decisão

Árvore de decisão é um modelo estatístico de algoritmo de AM supervisionado (SILVA, 2005). Estes algoritmos utilizam a técnica de dividir-e-conquistar, onde um problema é subdividido em problemas menores e para estes problemas menores, recursivamente aplica-se a mesma técnica (GAMA, 2004).

A representação de uma árvore de decisão é simples, o que permite uma fácil interpretação de seus resultados (SHIBA et al, 2005; GAMA, 2004; SILVA, 2005). A Figura 5 apresenta a forma como uma árvore de decisão separa um espaço amostral.

Figura 5 - Divisão do espaço amostral



Fonte: Adaptado de Grzonka (2016)

É possível perceber que, todo o espaço amostral apresentado no gráfico da Figura 5 foi dividido apenas com regras de “se-então” da árvore de decisão.

Para a criação das regras de uma árvore de decisão são utilizados índices que ajudam o algoritmo a entender que o atributo escolhido melhora sua resposta final ao problema ou não. Índices que podem ser aplicados são GINI e a entropia (GAMA, 2004)

2.5.2 K-Médias

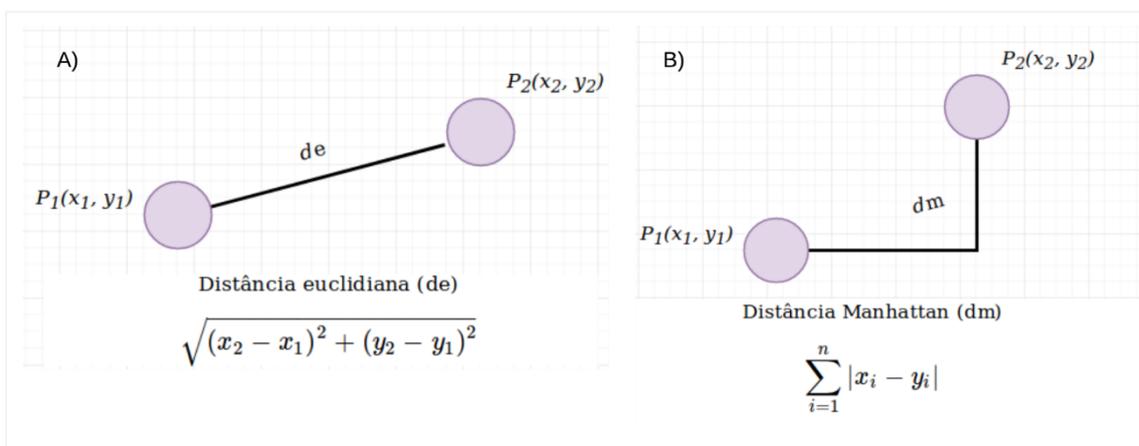
Para certos casos não haverá classes pré-definidas para os elementos de um conjunto de dados, nestes casos os algoritmos de AM não supervisionado são aplicados, como citado anteriormente. Estes algoritmos normalmente utilizam técnicas de análise de *cluster* para realizar a definição dos tipos de dados que são similares dentro do conjunto (SANTOS, 2012).

Onde, análise de *cluster* representa um conjunto de técnicas que busca separar um conjunto de dados em grupos (NUNES, 2015), de forma que, a homogeneidade dentro de cada grupo criada seja alta (NUNES, 2015).

Dentro os diversos algoritmos para a criação dos grupos, está o K-Médias (SANTOS, 2012), que através de uma análise de semelhança dos atributos do conjunto de dados, cria os grupos, segundo a regra de homogeneidade definida anteriormente. Para realizar a análise de semelhança o K-Médias utiliza uma métrica de distância (BEM et al, 2015) para verificar o quão próximo os elementos são e então agrupá-los.

Diferentes tipos de distâncias podem ser aplicados (BEM et al, 2015), variando de acordo com o problema a ser resolvido. Exemplos de distâncias aplicadas neste algoritmo são a Euclidiana e a Manhattan, que são representadas graficamente na Figura 6.

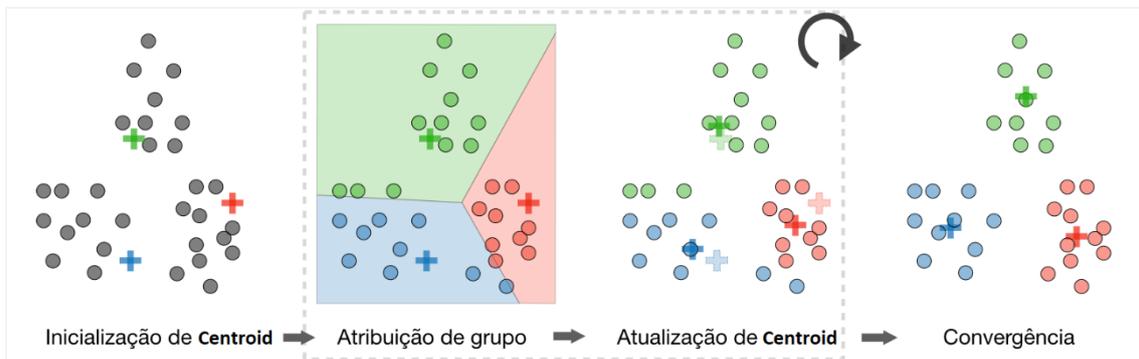
Figura 6 - Distâncias



Fonte: Produção do autor

Com estes elementos básicos do algoritmo, é possível entender os passos que são realizados para seu funcionamento, visto na Figura 7.

Figura 7 - Etapas do K-Médias



Fonte: Adaptado de Amidi (2018)

Nos passos apresentados na Figura 7, inicialmente, o algoritmo cria pontos de valores aleatórios (*Centroid*) e utiliza estes para realizar a atribuição de grupos. Com a atribuição de grupos feita, o ponto que foi inicializado aleatoriamente tem seu valor atualizado para a média de todos os valores de seu grupo, e este é um processo iterativo realizado até que não haja mais mudança no *centroid* de cada grupo. Ao final é dito que o algoritmo convergiu em seu resultado (AMIDI, 2018; HARTIGAN; WONG, 1979; SANTOS, 2012).

2.5.2.1 Spectral Angle Mapper

Como citado anteriormente, as formas de distâncias podem variar de acordo com o contexto que está sendo trabalhado e o tipo de dado. Para dados radiométricos, onde há a necessidade da classificação espectral pode-se utilizar o algoritmo de *Spectral Angle Mapper* (SAM) (BARBOSA et al, 2019). Este que calcula o grau de similaridade entre os espectros através do ângulo entre cada espectro (BARBOSA et al, 2019), desconsiderando a amplitude dos espectros (BARBOSA et al, 2019).

3 MATERIAIS E MÉTODOS UTILIZADOS

Abaixo é apresentado os materiais e métodos utilizados no desenvolvimento do presente Trabalho.

3.1 Materiais

Os materiais utilizados no desenvolvimento deste trabalho podem ser divididos em dois grupos: (i) Dados, (ii) Tecnologias

Nos dados tem-se imagens multiespectrais dos sensores OLI/Landsat-8, MSI/Sentinel-2 e imagens hiperespectrais.

Nas tecnologias para o desenvolvimento da ferramenta de correção atmosférica e classificação de dados com a árvore de decisão foi utilizado a linguagem de programação Python com o auxílio das bibliotecas GDAL, para o trabalho com dados geoespaciais (GDAL, 2018), scikit-learn para a aplicação dos classificadores, PyQt para o desenvolvimento de interfaces gráficas e o Py6S, para facilitar a utilização do 6SV na correção atmosférica. No caso do desenvolvimento das rotinas para a remoção de *sunglint* e classificação de dados com K-médias foi utilizado a linguagem de programação MATLAB.

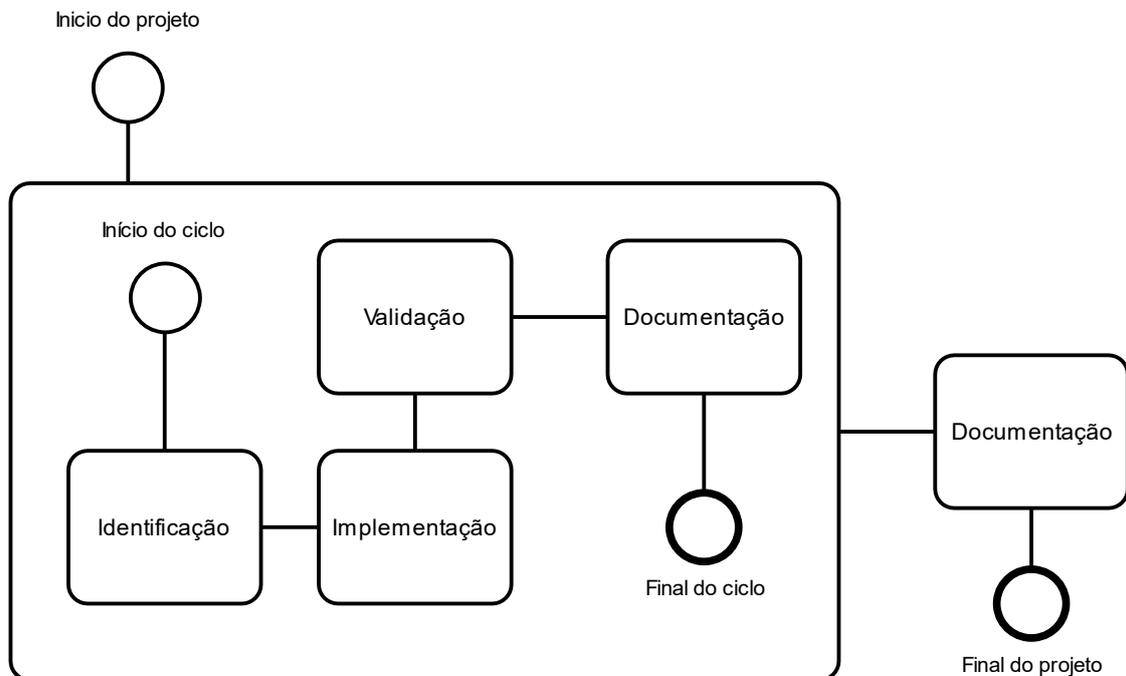
Além disto, para facilitar o processo de aquisição de parâmetros atmosféricos foi feita a utilização do *Google Earth Engine* (GEE), uma plataforma de análise global de grandes conjuntos de dados geoespaciais (Gorelick et al, 2017). Por fim, foi criado uma interface via tabelas para a ferramenta HydroLight (HL) de simulações de transferência radiativa em ambientes aquáticos.

3.2 Métodos

O método utilizado para o desenvolvimento deste trabalho foi dividido nas seguintes etapas: (i) Identificação, (ii) Implementação, (iii) Validação e (iv) Documentação.

A Figura abaixo demonstra a ligação entre cada uma destas etapas, e em seguida a descrição de cada uma delas.

Figura 8 - Metodologia



Fonte: Produção do autor

Ao iniciar o projeto, como na Figura 8, iniciou-se também um ciclo de desenvolvimento contínuo de diversas ferramentas, cada uma delas criadas de acordo com as necessidades do LabISA.

Dentro do ciclo de desenvolvimento, para cada uma das rotinas, inicialmente fez-se uma identificação, esta que junto aos alunos e pesquisadores do laboratório foram idealizadas. Após isto cada uma das ferramentas foi implementada e

validada, sendo que, cada uma destas etapas foi realizada com o acompanhamento dos colaboradores do laboratório. Ao final de todo o processo e validação das ferramentas desenvolvidas, elas foram documentadas para possibilitar a facilidade de uso e possivelmente de manutenção.

Com o final deste ciclo de desenvolvimento, fez-se a criação do documento do relatório final de todo o projeto desenvolvido.

4 DESENVOLVIMENTO E RESULTADOS

Este tópico irá apresentar as rotinas desenvolvidas neste Trabalho, suas etapas de implementação e os resultados obtidos.

4.1 Pré-processamento dos dados

Esta subseção apresenta as ferramentas implementadas para o auxílio das etapas de pré-processamento de dados do LabISA.

4.1.1 Correção atmosférica

Para águas interiores a correção atmosférica é extremamente importante uma vez que, a redução de influências no sinal captado do satélite melhora a identificação e resposta dos corpos d'água na imagem (MARTINS, 2017).

Existem diversos métodos que podem ser utilizados para a correção atmosférica, porém, o foco do LabISA de atividades realizadas com águas interiores trouxe a necessidade da aplicação de uma correção através do método físico 6SV. Mas a aplicação deste método em um conjunto de imagens pode ser muito trabalhosa uma vez que, para a aplicação do método, certos parâmetros atmosféricos precisam ser passados para que uma caracterização atmosférica seja feita (VERMOTE et al, 1997) e esta permita a correção atmosférica.

Com esta identificação realizada, dentro do ciclo de desenvolvimento deste projeto, apresentado na contextualização metodológica, fez-se o início do desenvolvimento de uma ferramenta capaz de facilitar o processo de aplicação do 6SV. Para isto a ferramenta deveria buscar os parâmetros necessários automaticamente e permitir que estes fossem utilizados na correção.

Para iniciar o desenvolvimento, um conjunto de requisitos do sistema foi levantado, apenas para garantir que, as necessidades do LabISA fossem atendidas com a ferramenta que estava sendo desenvolvida. Estes requisitos são evidenciados na Tabela 1.

Tabela 1 - Requisitos (AtmosPy)

REQUISITOS FUNCIONAIS	REQUISITOS NÃO FUNCIONAIS
Busca automática dos constituintes atmosféricos	Desenvolvimento em Python
Facilidade de utilização	Velocidade nas correções
Permitir que os resultados da simulação 6S sejam salvos	Facilidade de utilização
Informar ao usuário sobre as etapas que estão sendo realizadas	Correção atmosférica feita com a metodologia de Martins et al (2017).
Permitir inserção manual dos constituintes atmosféricos	

Fonte: Produção do autor

Com os requisitos levantados, foi realizada a implementação da ferramenta, que foi nomeada de AtmosPy. A implementação desta ferramenta foi dividida em duas partes: (i) Criação da *Application Programming Interface* (API), (ii) Desenvolvimento da interface gráfica.

Isto foi feito para que, inicialmente fosse criado uma API da ferramenta, para que esta pudesse ser aplicada em qualquer outro projeto que se utiliza Python e para permitir sua fácil ampliação, não criando dependências ou acoplamento com uma interface gráfica. O exemplo de código para a correção atmosférica com uma imagem é apresentado na Figura 9.

Figura 9 - API de correção atmosférica

```
01. # Geração dos coeficientes
02. atmos = AtmosPy()
03.
04. ## Geração automática de parâmetros
05. parameter = AutoParameter(sensor, image_path, lat, long, aero_profile, decimal_hour)
06.
07. # Geração dos coeficientes utilizando o 6S
08. atmos.generate_sixs(parameter)
09.
10. # Gera coeficientes atmosféricos
11. at_coefficient = load_coeffs_table(search_coeffs_table(image_path))
12. metadata = filter_metadata(read_metadata(image_path, sensor), sensor)
13.
14. # Aplica a Correção Atmosférica
15. atmos.correct_atmosphere(at_coefficient, metadata)
```

Fonte: Produção do autor

O código da Figura 9 é extremamente simples e permite uma fácil aplicação em qualquer outra ferramenta que esteja sendo desenvolvida.

Após a geração da API, uma interface gráfica que consome a API foi desenvolvida através do PyQt. A interface gráfica é dividida em duas partes (Figura 10), a primeira para a geração de coeficientes através do 6SV, e a segunda para a aplicação dos coeficientes gerados nas imagens, este que por sua vez pode ser aplicado utilizando Central Process Unit (CPU) ou mesmo processamento paralelo com Graphics Processing Unit (GPU).

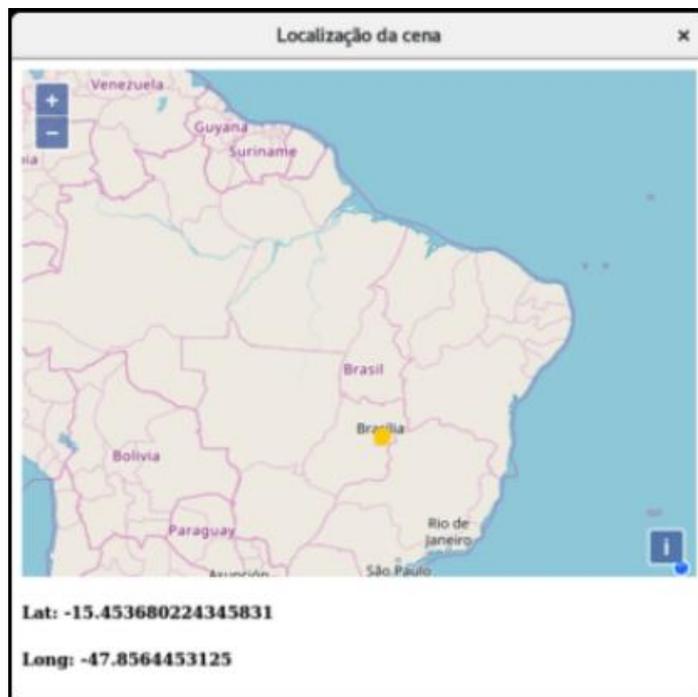
Figura 10 - Interface principal do AtmosPy



Fonte: Carlos et al (2019)

Como a geração dos coeficientes está diretamente relacionada a geolocalização da imagem a ser corrigida, a interface gráfica criada fornece também um mapa (Figura 11) que pode ser consultado para permitir que, um ponto seja escolhido e utilizado como referência para a geração dos parâmetros atmosféricos no 6SV.

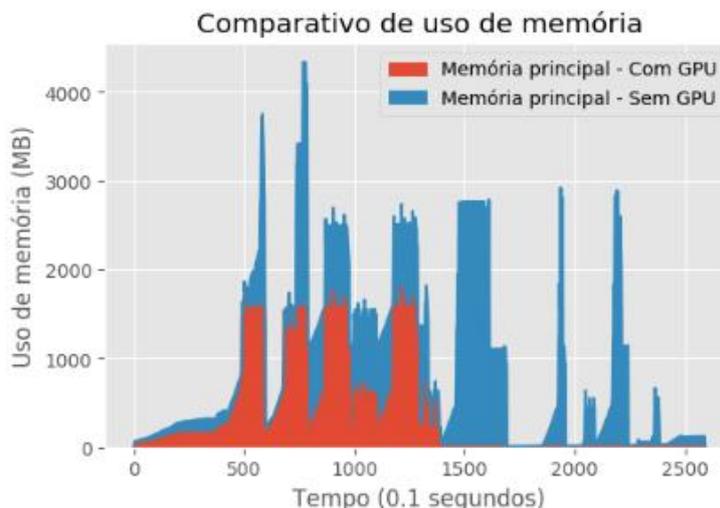
Figura 11 - Seleção de geolocalização



Fonte: Carlos et al (2019)

Anteriormente foi citado a possibilidade da aplicação de correções atmosféricas em GPU, isto foi realizado para garantir uma rápida aplicação em muitas imagens, principalmente a de sensores como o Sentinel-2/MSI, que possuem imagens grandes e podem demandar muito tempo computacional para seu processamento. Além da velocidade a aplicação das GPUs permitiu a diminuição da necessidade de muita memória disponível para a aplicação das correções. A Figura 12 apresenta um comparativo de tempo e utilização de memória, tanto com a CPU quanto para GPU.

Figura 12 - Comparativo de uso de memória



Fonte: Carlos et al (2019)

Desta forma, a ferramenta de correção atmosférica foi toda implementada para realizar correções das imagens dos sensores Sentinel-2/MSI, Sentinel-3/OLCI, LandSat-8/OLI, e ao final validada pelos usuários do LabISA. Atualmente o uso da ferramenta tem sido feito não apenas pelos pesquisadores do LabISA, mas também por alunos de pós-graduação da Universidade Federal de Santa Maria e Universidade Federal do Piauí.

4.1.2 Remoção de sunglint

A influência de *sunglint* em dados coletados por sensores imageadores pode comprometer a análise e trazer incertezas aos processos de pesquisa. Esta influência para imagens hiperspectrais pode ser um problema ainda maior que uma maior quantidade de dados será comprometida.

Assim faz-se necessário a utilização de metodologias que façam a remoção desta influência e permitam uma análise com maior grau de certeza e exatidão. Mas, a aplicação dos diferentes métodos para a remoção de *sunglint* em imagens hiperspectrais pode representar processos trabalhosos e demorados, uma vez que, dependendo da metodologia a ser aplicada diversos passos e

cálculos devem ser aplicados em cada uma das bandas. Por conta disto, o segundo ciclo de desenvolvimento deste Trabalho realizou a implementação de uma ferramenta que aplica diferentes metodologias para a correção de efeitos *sunlint* em imagens hiperspectrais, esta nomeada de rGlint.

Para iniciar o desenvolvimento, inicialmente fez-se a identificação dos principais requisitos funcionais e não funcionais da ferramenta a ser desenvolvida, estes são sumarizados na Tabela 2.

Tabela 2 - Requisitos (rGlint)

REQUISITOS FUNCIONAIS	REQUISITOS NÃO FUNCIONAIS
Permitir a seleção de bandas que serão trabalhadas	Desempenho nas correções
Visualização da imagem carregada	Facilidade de uso
Realizar correção Kutser (2009)	Condicionar o programa a trabalhar com quantidades variadas de banda
Realizar correção Kutser (2013)	Salvar resultados para serem utilizados no ENVI
Realizar correção Goodman (2008)	Utilizar MATLAB
Realizar correção Lyzenga (2006)	
Realizar correção Hedley (2005)	
Plot dos resultados	
Salvar os resultados	

Fonte: Produção do autor

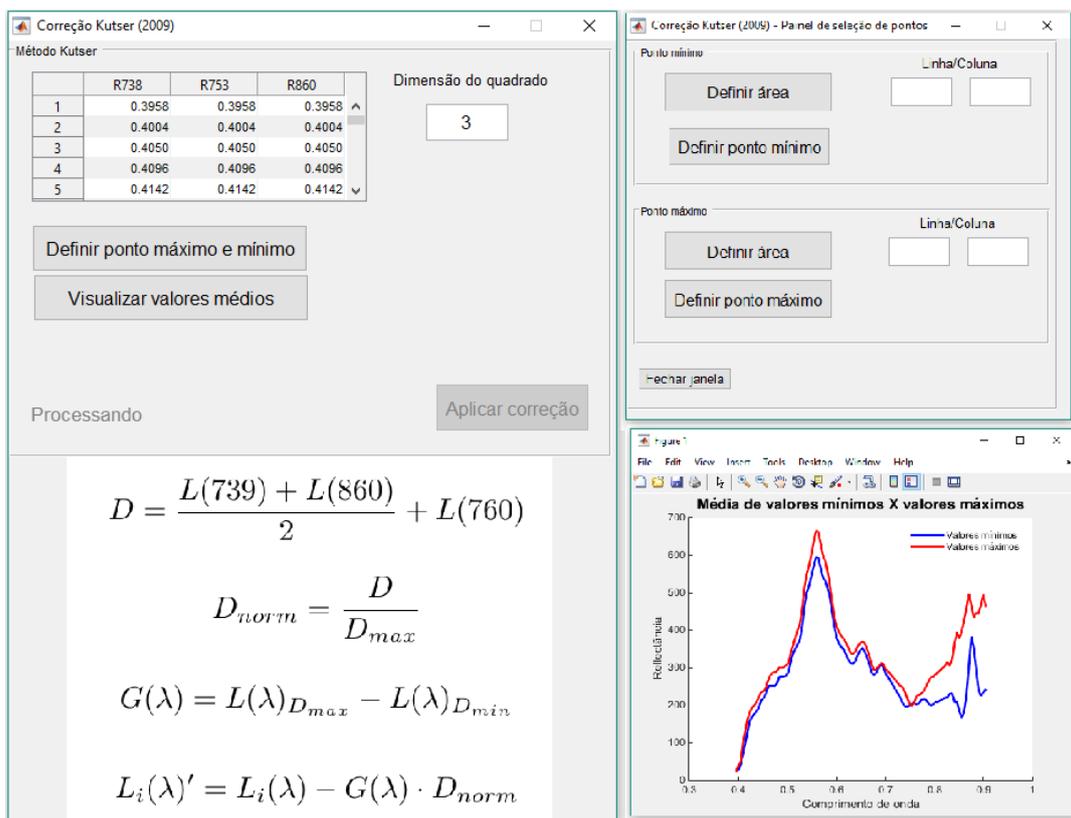
Após o levantamento de requisitos a implementação da ferramenta foi realizada. Toda a implementação foi feita utilizando MATLAB uma vez que a criação da rotina foi feita para que ela pudesse ser integrada com ferramentas já existentes no LabISA. Da mesma forma que o AtmosPy, o desenvolvimento foi dividido em dois, sendo o primeiro para a criação da API e o segundo para a criação da interface gráfica.

Por conta da vasta API padrão do MATLAB o desenvolvimento foi feito de forma rápida e sucinta.

Cada uma das metodologias implementadas na ferramenta tem seus passos bem definidos pela literatura, tornado assim necessário que, para cada uma dessas houvesse janelas organizadas que seguem o fluxo de aplicação. Além disso, para auxiliar o utilizador, cada uma das janelas possui descritas as equações utilizadas.

Por exemplo, para a janela da correção de Kutser (2009), tem-se a janelas da Figura 13.

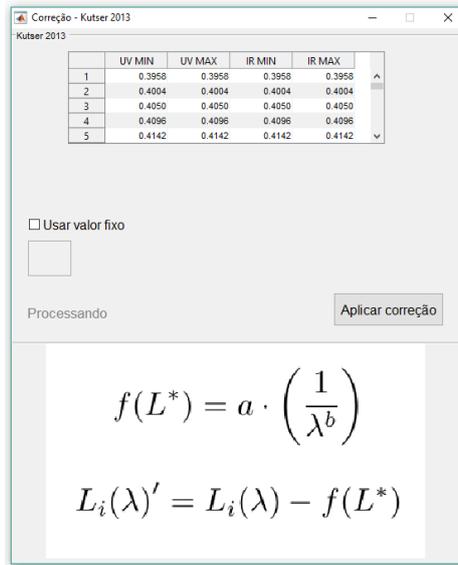
Figura 13 - Janelas da metodologia Kutser (2009)



Fonte: Produção do autor

Já para a metodologia de Kutser (2013), que apresenta outra abordagem metodológica, é utilizado a janela da Figura 14.

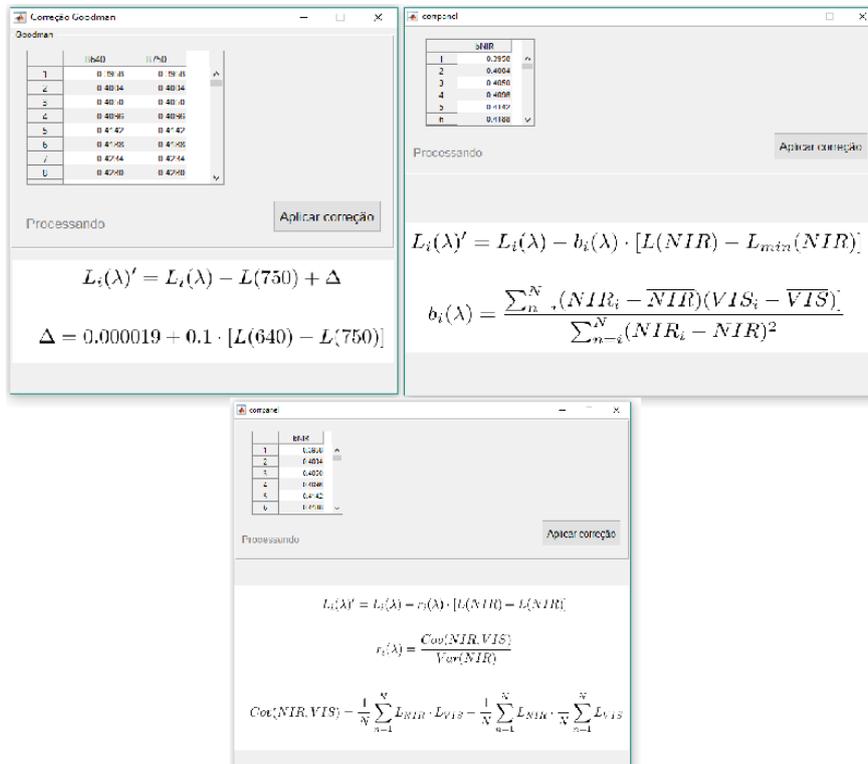
Figura 14 - Janelas da metodologia Kutser (2013)



Fonte: Produção do autor

Para as demais metodologias implementadas há as janelas apresentadas na Figura 15.

Figura 15 - Telas das correções Goodman, Hedley e Lyzenga



Fonte: Produção do autor

Com isto o desenvolvimento concluído a rotina foi validada pelos alunos e pesquisadores do LabISA. A utilização da rotina atualmente está sendo feita também por um doutorando da Universidade Federal da Bahia.

4.2 Classificação e análise dos dados

Esta subseção apresenta as ferramentas criadas para auxiliar as atividades de classificação e análise de dados no LabISA.

4.2.1 K-Médias em dados espectrais

Dados coletados *in situ* podem apresentar estruturas que necessitam de análise para a identificação dos tratamentos que podem ser aplicados e das informações que podem ser coletadas. Uma das formas de realizar este tipo de análise é através da classificação de dados, esta atividade que pode ser realizada através de diferentes métodos e algoritmos.

No LabISA surgiu a necessidade de realizar esta classificação de dados em um conjunto de dados que ainda não possuía caracterização. Para isto foi feita a escolha do K-Médias, com o formato de distância baseado nos espectros que resultou na seleção do algoritmo SAM para o agrupamento e caracterização dos dados através do K-Médias.

A implementação manual deste algoritmo pode levar um tempo, e a depender da forma do desenvolvimento, tornar a rotina pouco reutilizável. Por isto, no terceiro ciclo de desenvolvimento deste Trabalho, foi realizado a implementação do algoritmo K-Médias com a métrica de distância baseada no SAM. Com esta identificação feita, os requisitos desta ferramenta são resumidos através da Tabela 3.

Tabela 3 - Requisitos (K-Médias)

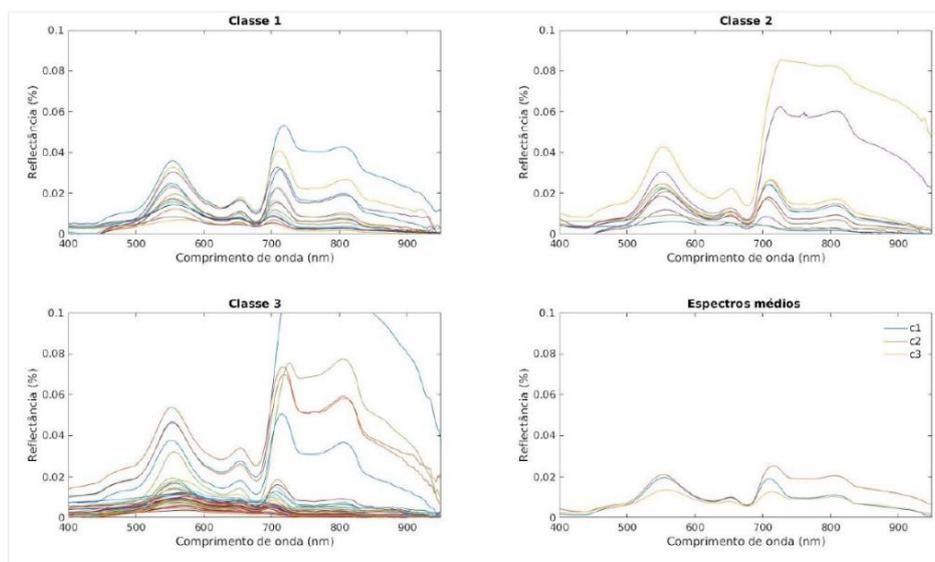
REQUISITOS FUNCIONAIS	REQUISITOS NÃO FUNCIONAIS
Permitir seleção dos dados utilizados	Desempenho na rotina
Visualização dos grupos criados	Utilizar MATLAB
Utilizar a métrica de distância SAM	Rotina para diferentes formatos de dados
Caracterização estatística de cada grupo criado	

Fonte: Produção do autor

Um dos requisitos apresentados na Tabela 3 é a necessidade de utilização do MATLAB, isto foi feito por uma necessidade de utilização desta rotina com outras já implementadas no laboratório. Por se tratar de uma rotina que deveria suportar diferentes tipos de dados, foi criada uma rotina que, necessita apenas receber uma matriz com os dados que eles já podem ser processados.

Desta forma a rotina foi implementada seguindo os requisitos levantados. Dentre os requisitos há também a visualização dos grupos criados. O gráfico criado para este requisito está na Figura 16.

Figura 16 - Visualização dos grupos



Fonte: Produção do autor

Desta forma a implementação da ferramenta foi realizada e validada junto a alunos de pós-graduação do LabISA. A implementação desta rotina tem auxiliado uma aluna de doutorado do INPE.

4.2.2 Árvore de decisão e visualização de dados

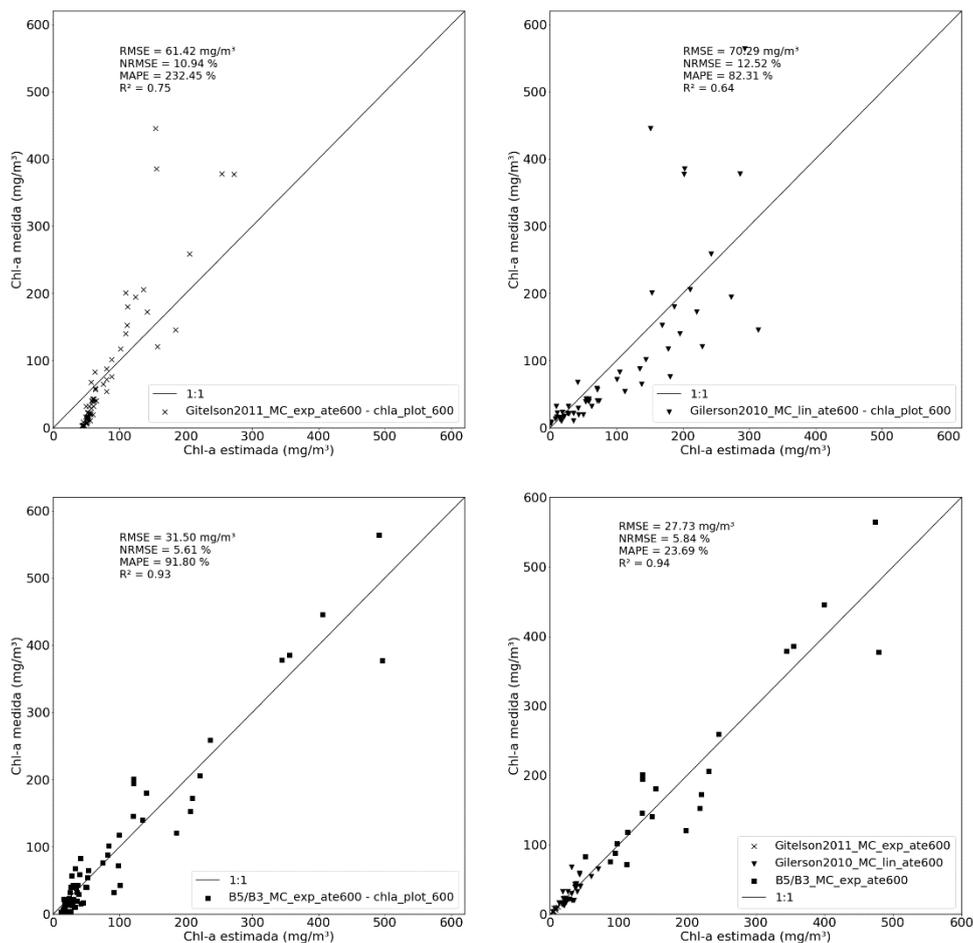
Em certos casos, quando estão sendo realizados o controle análises exploratórios em diversos conjuntos de dados, pode ser interessante aplicar métodos de classificação de elementos já conhecidos para que, conjuntos de características possam ser utilizados nas análises. Além disto, a análise visual dos dados é também extremamente importante.

Porém, nem sempre realizar classificações e visualizações dos dados para uma análise é uma tarefa rápida e automatizada. No caso do LabISA, há diferentes tipos de dados, de fontes limnológicas que são utilizados para a caracterização dos mais variados ambientes, que precisam de processos rápidos de classificação e de visualização. Desta forma, o terceiro ciclo de desenvolvimento deste Trabalho, focou na criação de uma rotina que através de uma utilização simples, permitisse a classificação e visualização dos dados do LabISA.

Para os casos de classificação, fez-se a aplicação das árvores de decisão que, são algoritmos rápidos de serem utilizados e seus resultados são facilmente interpretados pelos especialistas da área onde está sendo aplicada. Assim foi criada uma rotina que recebe dados limnológicos e através de um processo pré-definido para este tipo dado realiza a classificação.

Já no caso da visualização dos dados, foi criado uma biblioteca de *plot* que permite uma rápida visualização de modelos de estimativa de clorofila híbridos desenvolvidos por uma doutoranda do LabISA. Os exemplos das figuras geradas estão na Figura 17.

Figura 17 - Figuras geradas pela biblioteca de plot



Fonte: Produção do autor

Com isto as rotinas de classificação e visualização dos dados foi implementada e atualmente ajuda em uma tese de doutorado do LabISA.

4.3 Simulações com HydroLight (HL)

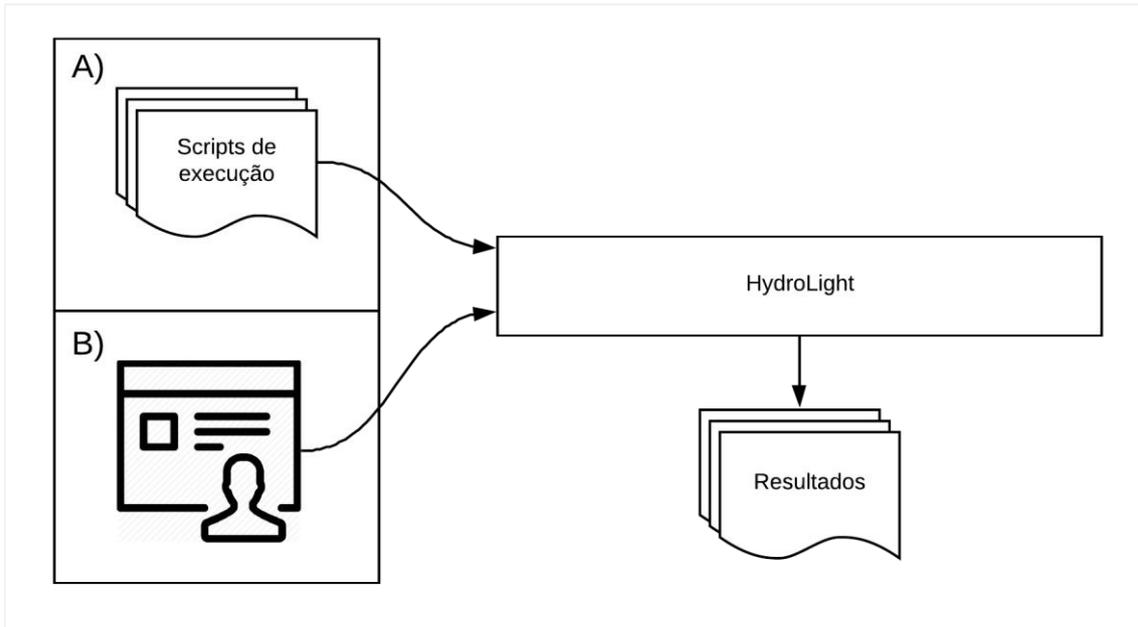
Realizar campanhas de campo para a coleta de dados nem sempre é possível, principalmente por conta dos fatores relacionados aos custos. Por isso para certos ambientes faz-se a estimativa dos parâmetros ópticos-limnológicos.

Dentre as diversas metodologias, há a empregada pelo HL, uma ferramenta que permite a simulação completo de sistemas aquáticos. O HL, mesmo permitindo uma simulação completa apresenta problemas na utilização e reprodutibilidade

dos resultados, isto por conta das formas de interação dos usuários com a ferramenta.

Por padrão os usuários do HL possuem somente duas formas de realizar as simulações, estas apresentadas na Figura 18.

Figura 18 - Formas de utilização (HL)



Fonte: Produção do autor

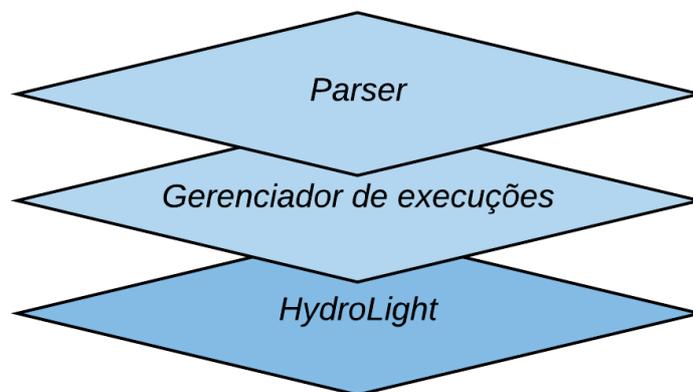
Na Figura 17 (A), o usuário pode inserir um *script* informando como a simulação deve ser feita. Este formato permite a reprodutibilidade dos resultados de forma simples, porém descrever este *script* é muito complicado já que ele segue uma semântica própria, o que demanda tempo de aprendizado e pode ocasionar problemas nos resultados após a execução. Já a Figura 17 (B) apresenta uma forma de executar as simulações feitas através de uma interface gráfica, esta diminui a reprodutibilidade dos resultados, além de não permitir a execução de diversas simulações ao mesmo tempo, como é feito pelo *script* de execuções.

Assim o quarto ciclo de desenvolvimento deste Trabalho realizou o desenvolvimento de uma biblioteca em Python, nomeado de HydroLight-Py, que

permite a escrita de tabelas de execução que podem ser facilmente descritas utilizando os principais editores de tabelas, como Excel e LibreOffice.

A biblioteca desenvolvida é dividida em três camadas (Figura 19): (i) HydroLight, sistema externo que fará as simulações; (ii) Gerenciador de execuções, que trabalha para controlar as execuções de simulações que estão sendo feitas; (iii) *Parser*, que é responsável por receber as entradas do usuário em formato de tabela e transcrever para *scripts* de execução do HL.

Figura 19 - Camadas do HydroLight-Py

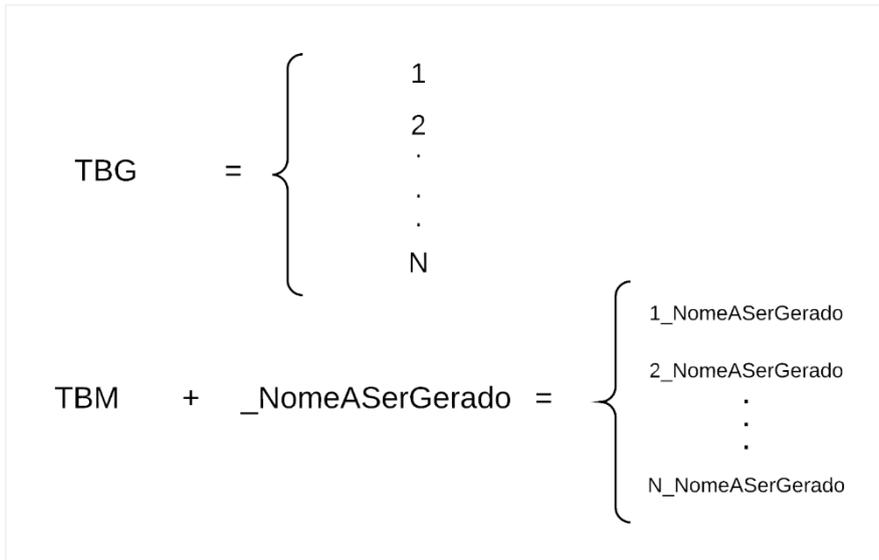


Fonte: Produção do autor

Como citado anteriormente, a entrada do usuário na biblioteca é feita através de uma tabela, para isto foi existe um *template* que deve ser utilizado para o funcionamento correto da biblioteca. Dentro da tabela, existem atalhos que podem ser aplicados para permitir que, durante o parse da tabela, dados sejam gerados sem a necessidade de o usuário repetir eles. Isto torna-se útil especialmente quando há a necessidade de execução de diversas simulações.

Os atalhos criados são, TBG e o TBM, sendo que, toda célula da tabela que for registrada com TBG, terá seus conteúdos preenchidos com números inteiros, de 1 a N, onde N é a quantidade de simulações, também definida pelo usuário. Para o TBM, são gerados números inteiros que são concatenados com algum texto também inserido pelo usuário. Os exemplos de funcionamento de cada um destes atalhos são apresentados na Figura 20.

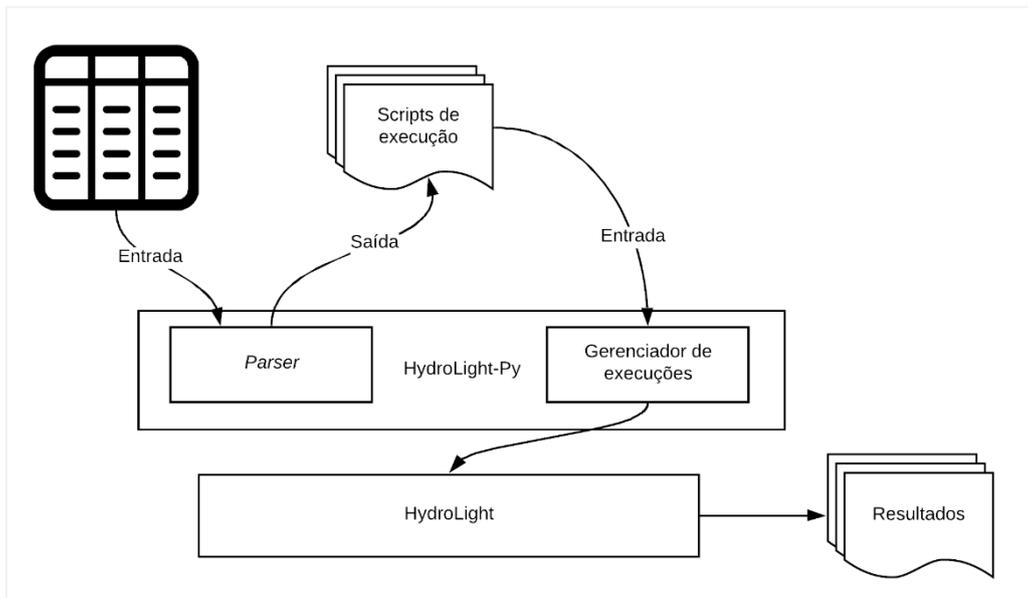
Figura 20 - Atalhos da tabela HydroLight-Py



Fonte: Produção do autor

Com a rotina desenvolvida o fluxo de simulações com o HL passou a ser feita de forma facilitada, e o resumo da utilização é apresentado na Figura 21.

Figura 21 - Fluxo de funcionamento



Fonte: Produção do autor

Com isto a rotina foi concluída e valida por alunos da pós-graduação do LabISA e atualmente tem sido utilizado em uma tese de doutorado no INPE.

5 CONCLUSÃO

Neste capítulo as contribuições e conclusões deste Trabalho são apresentadas, da mesma forma como as publicações e trabalhos futuros.

5.1 Contribuições e conclusões

As contribuições realizadas por este Trabalho são:

1. Rotinas que facilitam o pré-processamento de imagens Sentinel-2/MSI, Sentinel-3/OLCI e LandSat-8/OLI;
2. Rotinas para a classificação e visualização de dados ópticos e limnológicos de forma facilitada;
3. Uma biblioteca que permite a fácil execução de simulações do HydroLight.

Com isto, todo o desenvolvimento das bibliotecas e ferramentas criadas neste trabalho tiveram aplicações práticas para o auxílio de alunos e pesquisadores de diferentes instituições.

Além disto, durante o desenvolvimento, o autor aprendeu todo o *stack* de ferramentas necessárias para a criação de ferramentas completas de análise e processamento de dados utilizando as linguagens de programação Python e MATLAB.

Todo o ambiente de desenvolvimento criado e documentado, por auxiliar todos os que necessitem de uma base inicial para a análise de dados nos ambientes apresentados anteriormente.

5.2 Publicações

Com os resultados obtidos neste trabalho, as seguintes publicações puderam ser realizadas:

- Carlos, F. M., & Martins, Vitor De Souza; Barbosa, C. (2019). SISTEMA SEMIAUTOMÁTICO DE CORREÇÃO ATMOSFÉRICA

PARA MULTI-SENSORES ORBITAIS. *Simpósio Brasileiro de Sensoriamento Remoto*.

- Lobo, F. de L., Souza-Filho, P. W. M., Novo, E. M. L. de M., Carlos, F. M., & Barbosa, C. C. F. (2018). Mapping mining areas in the Brazilian amazon using MSI/Sentinel-2 imagery (2017). *Remote Sensing*, 10(8). <https://doi.org/10.3390/rs10081178>
- Cairo, C. T., Maciel, D. A., Júnior, R. F., Carlos, F. M., Barbosa, C. C. F., Lobo, F. D. L., & Novo, E. M. L. de M. (2019). *Estimativa da concentração da clorofila/SP, Resultados Preliminares. 1903–1906. Simpósio Brasileiro de Sensoriamento Remoto*.

5.3 Trabalhos futuros

Para dar continuidade ao desenvolvimento das ferramentas apresentadas neste trabalho, os seguintes tópicos podem ser considerados:

- Ambientes de execução do AtmosPy em *containers* Docker;
- Adicionar mais métodos de correção no AtmosPy;
- Adicionar mais métodos de correção no rGlint;
- Ampliar os sensores aceitos no AtmosPy;
- Disponibilizar todo o ambiente HydroLight em um *container* Docker;
- Possibilitar a visualização de dados diretamente de bancos de dados geográficos.

REFERÊNCIAS BIBLIOGRÁFICAS

- Amidi, S. (2018). Unsupervised Learning cheatsheet. Retrieved from <https://stanford.edu/~shervine/teaching/cs-229/cheatsheet-unsupervised-learning>
- Barbosa, C., Novo, E., & Martins, V. (2019). Introdução ao Sensoriamento Remoto de sistemas aquáticos.
- Bem, J. S. de, Giacomini, N. M. R., & Waismann, M. (2015). Utilização da técnica da análise de clusters ao emprego da indústria criativa entre 2000 e 2010: estudo da Região do Consinos, RS. *Interações (Campo Grande)*, 16(1), 27–41. <https://doi.org/10.1590/151870122015102>
- CAIRO, Carolline Tressmann. CARACTERIZAÇÃO TEMPORAL DAS PROPRIEDADES BIO-ÓTICAS DO RESERVATÓRIO DE IBITINGA/SP. 2015. 121 f. Dissertação (Mestrado) - Curso de Sensoriamento Remoto, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2015. Disponível em: <sid.inpe.br/mtc-m21b/2015/03.06.13.23-TDI>. Acesso em: 12 jul. 2018.
- Durán, D. M. (2008). Técnicas de oversampling aplicadas al análisis de imágenes hiperspectrales.
- ESA. MultiSpectral Instrument (MSI) Overview. Disponível em: <<https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/msi-instrument>>.
- Furlanetto, R. ., Nanni, M. ., Crusiol, L. G. T., Silva, G. F. C., & Oliveira Junior, A. de. (2017). Potencial de utilização de sensores multiespectral e hiperespectral no estudo de diferentes alvos agrícolas Material e Métodos.
- Grzonka, D., Suchacka, G., & Borowik, B. (2016). Application of Selected Supervised Classification Methods to Bank Marketing Campaign. *Information Systems in Management*, 5(1), 36–48.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*.
- GDAL.ORG. GDAL - Biblioteca de Abstrações de Dados Geo-Espaciais. Disponível em: <http://www.gdal.org/index_br.html>
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1), 100–108. Retrieved from <http://www.jstor.org/stable/2346830>
- LATORRE, Marcelo et al. CORREÇÃO ATMOSFÉRICA: CONCEITOS E FUNDAMENTOS. *Revista Espaço e Geografia*, São José dos Campos, v. 5, n. 1, p.153-178, 14 mar. 2002. Disponível em: <<http://www.lsie.unb.br/espacoegeografia/index.php/espacoegeografia/article/view/20>> . Acesso em: 06 jul. 2018.

MASCARENHAS, Luciane Martins de Araújo et al. SENSORIAMENTO REMOTO COMO INSTRUMENTO DE CONTROLE E PROTEÇÃO AMBIENTAL: ANÁLISE DA COBERTURA VEGETAL REMANESCENTE NA BACIA DO RIO ARAGUAIA. Sociedade e Natureza, Uberlândia, v. 21, n. 1, p.5-18, abr. 2009. Disponível em: <<http://www.scielo.br/pdf/sn/v21n1/v21n1a01.pdf>>.

MURPHY, Sam. Gee-atmcorr-S2. Disponível em: <<https://github.com/samsammurphy/gee-atmcorr-S2>>.

Maio, A. Di, Rudorff, B. F. T., Moraes, E. C. de, Pereira, G., Moreira, M. A., Sausen, T. M., & Florenzano, T. G. (2008). Sensoriamento remoto. 82.

MICHALSKI, Ryszard S; CARBONELL, Jaime G; MITCHELL, Tom M. Machine learning: An artificial intelligence approach. [S.l.]: Springer Science & Business Media, 2013.

Meneses, P. R., Almeida, T. de, Rosa, A. N. de C. S., Sano, E. E., Souza, E. B. de, Baptista, G. M. de M., & Brites, R. S. (2012). Introdução Ao Processamento De Imagens De Sensoriamento Remoto. Retrieved from http://s3.amazonaws.com/academia.edu.documents/33333156/MENESES__ALMEIDA_2012_INTRO_AO_PROCESSAMENTO_DE_IMAGEM.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1493063877&Signature=CSfQgHF6oDsiCY5cXA5MSVRi65U%3D&response-content-disposition=inline%3Bfile

Martins, Vitor Souza, DE CARVALHO, L.A.S, BARBOSA, CLAUDIO, NOVO, E. (2017). Avaliação da acurácia em produtos OLI/Landsat 8 em lagos amazônicos: Correção Atmosférica. SBSR, 18. Retrieved from <https://proceedings.galoa.com.br/sbsr/trabalhos/avaliacao-da-acuracia-em-produtos-olilandsat-8-em-lagos-amazonicos-correcao-atmosferica>

NASA. Operational Land Imager (OLI). Disponível em: <<https://landsat.gsfc.nasa.gov/operational-land-imager-oli/>>

Nunes, D. H. F. (2016). Um Estudo Sobre O Algoritmo K-Means. 60.

NOVO, Evlyn M. L. de Moraes. Sensoriamento Remoto: Princípios e Aplicações. 4. ed. São José dos Campos: Blucher, 2010. 388 p.

PONZONI, Flávio Jorge et al. Calibração Absoluta de Sensores Orbitais: Conceituação, Principais Procedimentos e Aplicação. São José dos Campos: Parêntese Editora, 2007.

SOUZA, Juarez Dantas de. MODELO FÍSICO-MATEMÁTICO DE CORREÇÃO ATMOSFÉRICA PARA IMAGENS TM - LANDSAT 5 E MODIS TERRA/AQUA. 2008. 201 f. Tese (Doutorado) - Curso de Meteorologia, Universidade Federal de Campina Grande, Campina Grande, 2008. Disponível em: <http://www.dca.ufcg.edu.br/posgrad_met/teses/JuarezDSouza_2008.pdf>.

Santos, R. D. C. dos. (2012). Minicursos ELAC 2012 (R. Rosa & N. Sant'anna, eds.).

SILVA, LUIZA MARIA OLIVEIRA DA. (2005). Uma Aplicação De Árvores De Decisão, Redes Neurais E Knn Para a Identificação De Modelos Arma Não-Sazonais E Sazonais. <https://doi.org/10.17771/PUCRio.acad.7587>

Shiba, M. H., Santos, R. L., Quintanilha, J. A., & Kim, H. Y. (2005). Classificação de imagens de sensoriamento remoto pela aprendizagem por árvore de decisão: uma avaliação de desempenho. Anais XII Simposio Brasileiro de Sensoriamento Remoto, 4319–4326.

STREHER, Annia Susin. OCORRENCIA E REMOCÃO DOS EFEITOS DE ~ SUNGLINT EM IMAGENS HIPERESPECTRAIS DE ALTA RESOLU , CAO ESPACIAL DO SENSOR SpecTIR. 2013. 110 f. Dissertação (Mestrado) - Curso de Sensoriamento Remoto, Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2013. Disponível em: <<http://urlib.net/8JMKD3MGP7W/3DR968E>>. Acesso em: 09 jul. 2018.

VERMOTE, E.f. et al. Second Simulation of the Satellite Signal in the Solar Spectrum, 6S: an overview. Ieee Transactions On Geoscience And Remote Sensing, [s.l.], v. 35, n. 3, p.675-686, maio 1997. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/36.581987>.

WINSTON, Patrick Henry. Artificial Intelligence. [S.l.]: Addison-Wesley, 1992. v. 3.

ZUBEN, Fernando J. Von. Introdução à Inteligência Artificial. p. 1–48, 2013.