



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2018/01.30.16.56-TDI

MINERAÇÃO DE INFORMAÇÕES PARA BUSCA AUTOMÁTICA EM CATÁLOGOS DE IMAGENS DE SENSORIAMENTO REMOTO

Mikhaela Aloísia Jéssie Santos Pletsch

Dissertação de Mestrado do
Curso de Pós-Graduação em
Sensoriamento Remoto, orientada
pelo Dr. Thales Sehn Körting,
aprovada em 09 de fevereiro de
2018.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34P/3QF6JPP>>

INPE
São José dos Campos
2018

PUBLICADO POR:

Instituto Nacional de Pesquisas Espaciais - INPE

Gabinete do Diretor (GBDIR)

Serviço de Informação e Documentação (SESID)

Caixa Postal 515 - CEP 12.245-970

São José dos Campos - SP - Brasil

Tel.:(012) 3208-6923/6921

E-mail: pubtc@inpe.br

**COMISSÃO DO CONSELHO DE EDITORAÇÃO E PRESERVAÇÃO
DA PRODUÇÃO INTELECTUAL DO INPE (DE/DIR-544):****Presidente:**

Maria do Carmo de Andrade Nono - Conselho de Pós-Graduação (CPG)

Membros:

Dr. Plínio Carlos Alvalá - Centro de Ciência do Sistema Terrestre (COCST)

Dr. André de Castro Milone - Coordenação-Geral de Ciências Espaciais e Atmosféricas (CGCEA)

Dra. Carina de Barros Melo - Coordenação de Laboratórios Associados (COCTE)

Dr. Evandro Marconi Rocco - Coordenação-Geral de Engenharia e Tecnologia Espacial (CGETE)

Dr. Hermann Johann Heinrich Kux - Coordenação-Geral de Observação da Terra (CGOBT)

Dr. Marley Cavalcante de Lima Moscati - Centro de Previsão de Tempo e Estudos Climáticos (CGCPT)

Silvia Castro Marcelino - Serviço de Informação e Documentação (SESID)

BIBLIOTECA DIGITAL:

Dr. Gerald Jean Francis Banon

Clayton Martins Pereira - Serviço de Informação e Documentação (SESID)

REVISÃO E NORMALIZAÇÃO DOCUMENTÁRIA:

Simone Angélica Del Duca Barbedo - Serviço de Informação e Documentação (SESID)

Yolanda Ribeiro da Silva Souza - Serviço de Informação e Documentação (SESID)

EDITORAÇÃO ELETRÔNICA:

Marcelo de Castro Pazos - Serviço de Informação e Documentação (SESID)

André Luis Dias Fernandes - Serviço de Informação e Documentação (SESID)



MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

sid.inpe.br/mtc-m21b/2018/01.30.16.56-TDI

MINERAÇÃO DE INFORMAÇÕES PARA BUSCA AUTOMÁTICA EM CATÁLOGOS DE IMAGENS DE SENSORIAMENTO REMOTO

Mikhaela Aloísia Jéssie Santos Pletsch

Dissertação de Mestrado do
Curso de Pós-Graduação em
Sensoriamento Remoto, orientada
pelo Dr. Thales Sehn Körting,
aprovada em 09 de fevereiro de
2018.

URL do documento original:

<<http://urlib.net/8JMKD3MGP3W34P/3QF6JPP>>

INPE
São José dos Campos
2018

Dados Internacionais de Catalogação na Publicação (CIP)

Pletsch, Mikhaela Aloísia Jéssie Santos.

P716m Mineração de informações para busca automática em catálogos de imagens de sensoriamento remoto / Mikhaela Aloísia Jéssie Santos Pletsch. – São José dos Campos : INPE, 2018.
xxiv + 80 p. ; (sid.inpe.br/mtc-m21b/2018/01.30.16.56-TDI)

Dissertação (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2018.

Orientador : Dr. Thales Sehn Körting.

1. Mineração de informações de imagens. 2. Catálogo de imagens de sensoriamento remoto. 3. Metadados. 4. Big data. 5. Fusão de dados. I.Título.

CDU 528.854



Esta obra foi licenciada sob uma Licença [Creative Commons Atribuição-NãoComercial 3.0 Não Adaptada](https://creativecommons.org/licenses/by-nc/3.0/).

This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](https://creativecommons.org/licenses/by-nc/3.0/).

Aluno (a): **Michaela Aloisia Jéssie Santos Pletsch**
Título: "MINERAÇÃO DE INFORMAÇÕES PARA BUSCA AUTOMÁTICA EM CATÁLOGOS DE IMAGENS DE SENSORIAMENTO REMOTO".

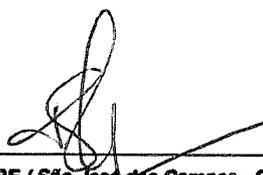
Aprovado (a) pela Banca Examinadora
em cumprimento ao requisito exigido para
obtenção do Título de **Mestre** em
Sensoriamento Remoto

Dra. Maria Isabel Sobral Escada



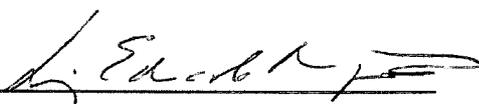
Presidente / INPE / SJCampos - SP
 Participação por Vídeo - Conferência

Dr. Thales Sehn Körting



Orientador(a) / INPE / São José dos Campos - SP
 Participação por Vídeo - Conferência

Dr. Luiz Eduardo Oliveira e Cruz de Aragão



Membro da Banca / INPE / São José dos Campos - SP
 Participação por Vídeo - Conferência

Dra. Camila Souza dos Anjos Lacerda



Convidado(a) / IFSULDEMINAS / Inconfidentes - MG
 Participação por Vídeo - Conferência

Este trabalho foi aprovado por:

maioria simples

unanimidade

São José dos Campos, 09 de fevereiro de 2018.

“Unless someone like you cares a whole awful lot, nothing is going to get better. It’s not”.

DR. SEUSS
em “*The Lorax*”, 2012

*A minha amada mãe **Roseli**, a meus pais científicos
Victor Velázquez e Thales Körting, e a meus
irmãos **Michel e Michelle***

AGRADECIMENTOS

Agradeço pelos desafios, derrotas, conquistas, momentos felizes e tristes. Pelos punções de orelha, conselhos e direcionamentos. Pelas conversas e histórias de vida. Pela chance de conhecer tantas pessoas nesse caminho, nem boas nem más, apenas únicas. Citar apenas um nome seria injusto com todo o universo de situações e pessoas que contribuíram para a minha formação como ser humano.

Agradeço também ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo auxílio financeiro ao longo de todo o mestrado, e ao Instituto Nacional de Pesquisas Espaciais (INPE), por todo o suporte e infraestrutura disponibilizados.

Por último, mas não menos importante, agradeço pela oportunidade de ter vivido o suficiente para completar mais esse ciclo e pela liberdade de escolher o meu próprio caminho.

Gratidão :)

RESUMO

A base de dados de observação da Terra está se aproximando da escala de *zettabyte* ($1\text{ZB} = 10^{12}\text{GB}$). Produzidos em alta velocidade, estes dados apresentam também grande variedade, devido a fatores como a diversidade dos sensores existentes. Frente a estas características, são enquadrados como *Big Data*, e apresentam tanto oportunidades, como a possibilidade de analisar e integrar um maior conjunto de dados, quanto desafios, principalmente no armazenamento e processamento para que possam ser disponibilizados aos usuários. A distribuição desta base é geralmente realizada por meio de catálogos, cujos parâmetros de buscas ainda são limitados aos metadados tradicionais de imagens de satélite, como data de aquisição, características do sensor e localização geográfica da área imageada. Assim, ainda existe uma demanda em oferecer aos usuários parâmetros com base no conteúdo das imagens em uma perspectiva de fusão de dados, permitindo buscas mais refinadas para dar suporte a estudos específicos e possíveis tomadas de decisão. Nesse contexto, o presente trabalho resultou em um protótipo de mineração de informações de imagens de Sensoriamento Remoto (*Remote Sensing Image Information Mining* - ReSIIM), capaz de realizar buscas inteligentes em grandes catálogos de imagens com base nos alvos de *nuvem*, *sombra de nuvem*, *área continental*, *água*, *floresta* e *solo exposto*. Por fim, este trabalho também apresenta testes sobre o uso do ReSIIM em 3 abordagens, espacial, temporal e espectral.

Palavras-chave: Mineração de informações de imagens. Catálogo de imagens de Sensoriamento Remoto. Metadados. *Big data*. Fusão de Dados.

INFORMATION MINING FOR AUTOMATIC SEARCH IN REMOTE SENSING IMAGE CATALOGS

ABSTRACT

The Earth Observation database is almost on the scale of Zettabyte (10^{12} GB). Produced at a rapid rate, those data also present great diversity, due to the range of sensor types. In such a manner, this kind of data is also classified as *Big Data*, and present opportunities, such as the possibility to analyse and integrate different data, as well as challenges, mainly regarding storing and processing steps in order to be available to users. The distribution of this database is normally through catalogues, which searching criteria is limited to traditional metadata, as acquisition date, sensor characteristics and geographical localization. Thus, there remained a demand for a tool which enables users to search for images based on phenomena in lieu of date or location in a data fusion perspective. In this manner, this work resulted in a Remote Sensing Image Information Mining (ReSIIM) prototype able to make smart searches in big databases based on well-known and basic targets found in Remote Sensing imagery: *cloud*, *cloud shadow*, *clear land* (continel area), *water*, *forest* and *bare soil*. Finally, this work also presents tests based on ReSIIM use in three approaches, spatial, temporal and spectral.

Keywords: Image Information Mining. Remote Sensing Image Catalogue. Metadata. Big Data. Data Fusion.

LISTA DE FIGURAS

| | <u>Pág.</u> |
|--|-------------|
| 2.1 Exemplos de descritores de cobertura do solo em uma imagem de Sensoriamento Remoto no nível 2. | 19 |
| 3.1 Estrutura da metodologia. | 31 |
| 3.2 Exemplo da aplicação do ReSIIM. | 33 |
| 3.3 Variação dos valores de NDVI em áreas identificadas como floresta pelo PRODES. | 35 |
| 3.4 Cena com baixo índice de acerto (ACo) para o limiar de <i>floresta</i> . A - Cena original (composição usando bandas do Landsat 8 - R5G4B2); B - Máscara de floresta segundo o PRODES; C - Área de <i>floresta</i> detectada (NDVI $\geq 0,8$); e D - Cena original, máscara do PRODES e <i>floresta</i> detectada. | 37 |
| 3.5 Área com presença de <i>solo exposto</i> (composição usando bandas do Landsat 8 - R4G3B2). | 38 |
| 3.6 Exemplo do processo de armazenamento no banco de metadados. | 39 |
| 3.7 Avaliação da acurácia do ReSIIM. | 41 |
| 3.9 Exemplo da busca combinada dos alvos <i>nuvem</i> e AC no ReSIIM. | 42 |
| 3.8 Exemplo da estrutura de armazenamento no banco de metadados em Python. | 42 |
| 4.1 Espacialização dos metadados gerados pelo ReSIIM no Brasil para os meses de Agosto e Setembro de 2017. | 45 |
| 4.2 Cena 231/067 (26/06/2004), Landsat 5. | 47 |
| 4.3 Espacialização dos resultados da busca combinada dos alvos <i>solo exposto</i> e <i>nuvem</i> | 49 |
| 4.4 Cenas retornadas a partir da intersecção da busca combinada de <i>solo exposto</i> acima de 6% e <i>nuvem</i> igual ou acima de 11%. | 51 |
| 4.5 Cenas retornadas a partir da intersecção da busca combinada de <i>solo exposto</i> igual ou acima de 29% e <i>nuvem</i> igual ou acima de 48%. | 52 |
| 4.6 Cenas retornadas a partir da intersecção da busca combinada igual ou acima de 29% para os alvos <i>solo exposto</i> e <i>nuvem</i> | 52 |
| 4.7 Cenas retornadas a partir da intersecção da busca combinada de percentual de [6;17] para o alvo <i>solo exposto</i> e de [11;18] para <i>nuvem</i> | 53 |
| 4.8 Cenas não retornadas a partir da intersecção da busca combinada de percentual de [6;17] para o alvo <i>solo exposto</i> e de [11;18] para <i>nuvem</i> | 54 |

| | | |
|------|---|----|
| 4.9 | Localização da cena do estudo de caso. | 55 |
| 4.10 | Variação da porcentagem dos alvos <i>nuvem</i> , <i>sombra de nuvem</i> , <i>área continental (AC)</i> , <i>água</i> , <i>floresta</i> e <i>solo exposto</i> ao longo do ano de 2017. . . . | 57 |

LISTA DE TABELAS

| | <u>Pág.</u> |
|---|-------------|
| 2.1 Plataformas de acesso a dados de Sensoriamento Remoto e os parâmetros de busca definidos pelo usuário. Nos parâmetros de busca disponíveis, Geográfico representa a área imageada, e X, a presença de determinado parâmetro de busca. | 11 |
| 2.2 Produtos disponíveis no catálogo do INPE. | 12 |
| 2.3 Aplicações em Sensoriamento Remoto de buscas por imagens com base em seu conteúdo. | 23 |
| 3.1 Análise da correspondência do alvo <i>água</i> . ACo: Acurácia da correspondência; EC: Erro de comissão. | 34 |
| 3.2 Análise da correspondência do alvo <i>floresta</i> . ACo: Acurácia da correspondência; EC: Erro de comissão; EOm: Erro de omissão. | 36 |
| 4.1 Resultados do ReSIIM para a cena 231/067 (26/06/2004), Landsat 5. . . | 48 |
| 4.2 Análise da relação entre os alvos <i>solo exposto</i> e <i>nuvem</i> | 48 |
| 4.3 Testes realizados na busca por imagens adequadas para a aplicação do MLME (DRef: Dado de Referência; DR: Dado de Ruído). As lacunas em azul representam valores menores ou iguais (\leq), e em laranja, maiores ou iguais (\geq). | 58 |

LISTA DE ABREVIATURAS E SIGLAS

| | | |
|---------|---|---|
| AC | – | Área Continental |
| ACo | – | Acurácia da Correlação |
| ASF | – | <i>Alaska Satellite Facility</i> |
| BAI | – | <i>Burned Area Index</i> |
| CAST | – | Academia Chinesa de Tecnologia Espacial |
| CBIR | – | <i>Content-Based Image Retrieval</i> |
| CBERS | – | <i>China-Brazil Earth Resources Satellite</i> |
| CLASS | – | <i>Comprehensive Large Array-Data Stewardship System</i> |
| CNN | – | <i>Convolutional Neural Network</i> |
| DAAC | – | <i>Distributed Active Archive Center</i> |
| DGI | – | Divisão de Geração de Imagens |
| DMI | – | <i>Deimos Imaging</i> |
| DMP | – | <i>Differential Morphological Profiles</i> |
| DR | – | Dado de Ruído |
| DRef | – | Dado de Referência |
| EC | – | Erro de Comissão |
| EE | – | <i>Earth Explorer</i> |
| EO | – | <i>Earth Observation</i> |
| EOm | – | Erro de Omissão |
| EOSDIS | – | <i>Earth Observing System Data and Information System</i> |
| ESA | – | <i>European Space Agency</i> |
| ETM+ | – | <i>Enhanced Thematic Mapper Plus</i> |
| EUA | – | Estados Unidos da América |
| GB | – | <i>Gigabyte</i> |
| GDAL | – | <i>Geospatial Data Abstraction Library</i> |
| GeoIRIS | – | <i>Geospatial Information Retrieval and Indexing System</i> |
| GIR | – | <i>Geographic Information Retrieval</i> |
| GloVis | – | <i>Global Visualization Viewer</i> |
| HSV | – | <i>Hue-Saturation-Value of Brightness</i> |
| IIM | – | <i>Image Information Mining</i> |
| INCRA | – | Instituto Nacional de Colonização e Reforma Agrária |
| INPE | – | Instituto Nacional de Pesquisas Espaciais |
| IR | – | <i>Information Retrieval</i> |
| ISRO | – | <i>Indian Space Research Organization</i> |
| KDD | – | <i>Knowledge-Discovery in Databases</i> |
| LAADS | – | <i>Level-1 and Atmosphere Archive & Distribution System</i> |
| LiDAR | – | <i>Light Detection and Ranging</i> |
| LLV | – | <i>LandsatLook Viewer</i> |
| LV | – | <i>Land Viewer</i> |
| MISS | – | Metadados em Imagens de Sensoriamento Remoto |

| | | |
|----------------|---|--|
| MLME | – | Modelo Linear de Mistura Espectral |
| MODIS | – | <i>Moderate Resolution Imaging Spectroradiometer</i> |
| NASA | – | <i>National Aeronautics and Space Administration</i> |
| NBR | – | <i>Normalized Burn Ratio</i> |
| NDWI | – | <i>Normalized Difference Water Index</i> |
| NDVI | – | <i>Normalized Difference Vegetation Index</i> |
| NIR | – | <i>Near infrared</i> |
| NOAA | – | <i>National Oceanic and Atmospheric Administration</i> |
| OBT | – | Observação da Terra |
| OLI | – | <i>Operational Land Imager</i> |
| PB | – | <i>Petabyte</i> |
| ReSIIM | – | <i>Remote Sensing Image Information Mining</i> |
| RF | – | <i>Relevance Feedback</i> |
| RGB | – | <i>Red-Green-Blue</i> |
| SAR | – | <i>Synthetic Aperture Radar</i> |
| Scientific Hub | – | <i>Sentinels Scientific Data Hub</i> |
| SIMR | – | <i>Satellite Image Matching and Retrieval System</i> |
| SNPP | – | <i>Suomi National Polar-orbiting Partnership</i> |
| SOM | – | <i>Self-Organized Map</i> |
| SS | – | <i>Satellite Search</i> |
| SSTL | – | <i>Surrey Satellite Technology Ltd</i> |
| TB | – | <i>Terabyte</i> |
| TM | – | <i>Thematic Mapper</i> |
| USGS | – | <i>United States Geological Survey</i> |
| VIIRS | – | <i>Visible Infrared Imaging Radiometer Suite</i> |
| ZB | – | <i>Zettabyte</i> |
| 3 V's | – | <i>Volume, Velocidade e Variedade</i> |

LISTA DE SÍMBOLOS

| | | |
|-----------|---|--|
| A | – | imagens retornadas pela busca no Banco de Metadados (BM) |
| e | – | erro |
| ha | – | hectare |
| m | – | metro |
| mm | – | milímetro |
| p | – | precisão |
| r | – | <i>recall</i> |
| R | – | conjunto de imagens relevantes para a busca |
| \bar{R} | – | conjunto de imagens irrelevantes para a busca |
| \cap | – | intersecção |

SUMÁRIO

| | <u>Pág.</u> |
|--|-------------|
| 1 INTRODUÇÃO | 1 |
| 1.1 Objetivo | 4 |
| 1.2 Organização da dissertação | 4 |
| 2 REFERENCIAL TEÓRICO | 7 |
| 2.1 Sensoriamento Remoto e geração de dados | 7 |
| 2.2 Catálogos de imagens de Sensoriamento Remoto | 8 |
| 2.3 Busca de imagens em grandes catálogos | 14 |
| 2.3.1 <i>Content-Based Image Retrieval</i> - CBIR | 14 |
| 2.3.2 <i>Image Information Mining</i> - IIM | 16 |
| 2.3.3 Estrutura de sistemas de busca de imagens por meio de metadados | 17 |
| 2.4 Aplicações em Sensoriamento Remoto e trabalhos relacionados | 22 |
| 2.4.1 Queimadas na Amazônia brasileira e a aplicação do Modelo Linear de Mistura Espectral | 26 |
| 3 METODOLOGIA | 31 |
| 3.1 ReSIIM: Mineração de Informações de Imagens de Sensoriamento Remoto | 32 |
| 3.2 Validação: Análise da correspondência dos metadados | 33 |
| 3.2.1 <i>Nuvem, sombra de nuvem e área continental</i> | 33 |
| 3.2.2 <i>Água</i> | 34 |
| 3.2.3 <i>Floresta</i> | 35 |
| 3.2.4 <i>Solo exposto</i> | 38 |
| 3.3 Banco de metadados | 38 |
| 3.4 Buscas por imagens: Avaliação da acurácia do ReSIIM | 39 |
| 3.5 A automatização em <i>Python</i> | 41 |
| 4 RESULTADOS E DISCUSSÕES | 43 |
| 4.1 Espacial: Aplicação do ReSIIM ao longo do Brasil | 43 |
| 4.1.1 Aplicação do ReSIIM em imagens com mudanças de cobertura da terra | 46 |
| 4.1.2 O conceito de <i>solo exposto</i> precisa de reajustes? | 48 |
| 4.2 Temporal: Aplicação do ReSIIM ao longo de um ano | 55 |
| 4.3 Espectral: ReSIIM como suporte de aplicação do Modelo Linear de Mistura Espectral | 58 |

| | |
|---|-----------|
| 5 CONCLUSÕES | 61 |
| 5.1 Indicação de trabalho futuros | 62 |
| REFERÊNCIAS BIBLIOGRÁFICAS | 63 |

1 INTRODUÇÃO

Os dados de Observação da Terra (do inglês: *Earth Observation* - EO) são compostos por uma grande quantidade de registros, cujo volume está se aproximando da escala de *zettabyte* ($1\text{ZB} = 10^{12}\text{GB}$) (QUARTULLI; OLAIZOLA, 2013). Produzidos em alta velocidade, os dados de EO também apresentam alta diversidade, decorrente da variedade de sensores remotos (DATCU et al., 2002; DATCU; SEIDEL, 2003a; DATCU; SEIDEL, 2003b; LI et al., 2016a), do modo de captura da interação espectral entre alvos e a radiação eletromagnética (KÖRTING et al., 2016), além do aspecto da atmosfera no momento da aquisição de imagens. Levando em consideração o grande Volume, Velocidade e Variedade (o que, segundo Laney (2001), é caracterizado como $3 V's$), os dados de EO podem ser enquadrados como *Big Data* (KÖRTING et al., 2016).

De modo geral, pode-se afirmar que o termo *Big Data* possui diferentes conotações de acordo com as funções existentes, como captar, processar e analisar dados (CHI et al., 2015). Como exemplo de um órgão que executa distintos serviços, cita-se o Instituto Nacional de Pesquisas Espaciais (INPE), responsável por atividades como receber e gravar dados brutos provenientes de sensores orbitais, transmitir e processar esses dados, bem como armazenar e distribuir imagens por meio de catálogos (FERREIRA; ARCANJO, 2007) para diferentes aplicações.

Ao longo de toda a cadeia de atuação frente a *Big Data*, é possível identificar tanto oportunidades quanto desafios. Primeiramente, o aumento da quantidade de fontes de imagens de Sensoriamento Remoto representa uma possibilidade de avanço no que se refere à automatização do processo de compreensão de imagens (YANG; NEWSAM, 2013), extraindo, assim, os dados de um conjunto de imagens ao invés de dados de imagens isoladas. A variedade de dados também possibilita que a comunidade acadêmica tenha acesso a uma grande quantidade de informações de distintas fontes, permitindo que objetos e padrões possam ser analisados, devido ao aumento das resoluções espacial (YANG; NEWSAM, 2013), temporal, radiométrica e espectral. A quantidade e variedade dos dados permitem ainda realizar processamentos espaço-temporais (LI et al., 2016a), onde a integração de dados viabiliza a análise e entendimento de sistemas complexos. Segundo Dong et al. (2009), o processo de combinar informações de diferentes sensores e fontes para inferências que não seriam possíveis com os dados de apenas um sensor ou fonte também é conhecido como fusão de dados. Embora a compreensão do potencial dessa grande quantidade de dados e da fusão de diferentes sensores ainda esteja sendo descoberta, na área de Sensoriamento Remoto já se tem utilizado dados de EO sob diferentes perspectivas

e aplicações (MURA et al., 2015). Os resultados provindos destes trabalhos são essenciais, uma vez que possibilitam dar suporte a tomadas de decisão importantes por parte de governos e organizações (CHI et al., 2015).

Nesse contexto, o acesso ao conteúdo de imagens é fundamental, já que amplia as opções de gestão de dados e a acessibilidade aos dados geoespaciais, viabilizando tomadas de decisões reais (LI et al., 2016a) ao retornar informações mais precisas aos usuários (LANEY, 2001). Devido a constante demanda por ferramentas mais robustas e refinadas para dar suporte aos 3 *V's*, surgem os desafios referentes a *Big Data* na área de Sensoriamento Remoto. Ademais, as resoluções, a quantidade de imagens, a heterogeneidade e a complexidade de cada conjunto de imagens continuam crescendo (SCHRÖDER; DIMAI, 1998; DATCU et al., 2002), tornando os processamentos, a visualização e a validação dos dados (LI et al., 2016a) cada vez mais complexos. Essas características tornam a análise de um conjunto de imagens de Sensoriamento Remoto, a cada consulta de usuários, um processo inviável, devido ao alto custo computacional e tempo despendido. Isso ocorre já que as buscas mais avançadas em uma vasta coleção de imagens podem gerar processamentos exaustivos, requerendo assim sistemas eficientes de busca de imagens de Sensoriamento Remoto (LI; NARAYANAN, 2006).

Embora existam diversos avanços frente a esses desafios, métodos mais inteligentes de buscas por imagens ainda são escassos (STEPINSKI et al., 2014). Grande parte dos sistemas de busca de imagens são baseados apenas em metadados simples, como a data de aquisição, o tipo de sensor e a localização geográfica (DATCU et al., 2002; DATCU et al., 2000; LI; NARAYANAN, 2006; STEPINSKI et al., 2014). Não há a possibilidade de refinar as buscas com base em temas específicos, ou alvos de Sensoriamento Remoto, como por exemplo a porcentagem de pixels com presença de *água* ou outro alvo. Essa é uma dificuldade encontrada em diferentes sistemas (CHI et al., 2015), inclusive no catálogo do INPE, o qual apresenta grande potencial de aprimoramento.

A técnica de encontrar rapidamente imagens dentro de grandes bases de dados, fornecendo informações para distintas aplicações, é conhecida como CBIR (do inglês: *Content-Based Image Retrieval*) (LI et al., 2016b). De acordo com Long et al. (2003) e Datcu e Seidel (2003b), nos sistemas CBIR, as imagens são localizadas com base em similaridades visuais com respeito a determinadas características. Entretanto, devido ao contínuo aumento da complexidade das imagens, esses tipos de sistemas não se mostraram eficientes e novas abordagens centradas em usuários foram propostas (DATCU; SEIDEL, 2003b).

KDD (do inglês: *Knowledge-Discovery in Databases*) é o processo de identificar padrões válidos, novos, potencialmente úteis e compreensíveis em dados (FAYYAD et al., 1996), sendo uma das etapas conhecida como mineração de dados (do inglês: *data mining*) (HAN et al., 2011). Devido à insatisfação com a abordagem CBIR, este termo foi fundido ao KDD e ao conceito de mineração de dados, engendrando assim o IIM (do inglês: *Image Information Mining*) (LI; NARAYANAN, 2006), também chamado de *mineração de informações de imagens*.

O campo da mineração de informações de imagens pertence a uma área de pesquisa em expansão (LI; NARAYANAN, 2006), uma vez que os sistemas antes desenvolvidos não eram robustos o suficiente, a ponto de compreender as diversas fontes de dados, bem como as peculiaridades de estudos específicos. Os sistemas IIM abriram novas perspectivas para a extração de informações de um grande conjunto de imagens (DATCU; SEIDEL, 2003b), como as de Sensoriamento Remoto. Apesar disso, os dados disponíveis ainda são sub-explorados, uma vez que uma grande parte nunca foi acessada (QUARTULLI; OLAIZOLA, 2013).

Considerando a demanda da comunidade científica de armazenar, processar e analisar dados, e a lacuna encontrada para nortear buscas inteligentes baseadas nas demandas dos usuários em grandes catálogos, ferramentas de refinamento de buscas de imagens de Sensoriamento Remoto são necessárias. Nesse contexto, o presente trabalho apresenta um protótipo de mineração de informações de imagens de Sensoriamento Remoto (*Remote Sensing Image Information Mining* - ReSIIM), a fim de refinar as buscas em grandes catálogos de imagens com base nos alvos de *nuvem*, *sombra de nuvem*, *área continental*, *água*, *floresta* e *solo exposto*. Assim, o ReSIIM não classifica determinada área, apenas analisa os alvos presentes nas cenas adquiridas por sensores remotos como um meio de dar suporte às buscas de imagens. Entre as diversas aplicações possíveis, este trabalho analisa o desempenho do ReSIIM em 3 principais abordagens. Em uma análise espacial, é possível integrar quali e quantitativamente as cenas ao longo de uma área, tendo como objetivo realizar inferências a respeito das diferenças dos metadados nas imagens ao longo desta região. Temporalmente, é possível analisar a persistência de determinado alvo ao longo do tempo, como nuvens ou floresta. Além disso, em uma abordagem espectral, é possível ainda analisar o desempenho do ReSIIM na busca de imagens adequadas para a aplicação do Modelo Linear de Mistura Espectral (MLME).

O MLME é usado para a detecção e mapeamento de cicatrizes de queimadas na região da Amazônia brasileira. Entretanto, essa aplicação requer uma etapa anterior

à identificação de cicatrizes: a seleção manual de imagens adequadas para a melhor performance do MLME. Esta fase é caracterizada como um processo trabalhoso e não automatizado, que leva em consideração critérios como a porcentagem de água e sombra, uma vez que estes elementos podem confundir os algoritmos de classificação, bem como nuvem, já que pode omitir certos pixels (ARAGÃO et al., 2016). Assim, ainda há uma carência na automatização desta etapa. A importância de estudos de cicatrizes de queimadas reside no fato de que o fogo torna os ambientes afetados mais vulneráveis à degradação florestal, além de reduzir a quantidade de biomassa da floresta (SATO, 2016).

1.1 Objetivo

O objetivo geral deste trabalho é desenvolver um algoritmo de busca inteligente aplicado ao contexto de catálogos de imagens de Sensoriamento Remoto.

Para contemplar o objetivo geral deste trabalho, os seguintes objetivos específicos foram delineados:

- Desenvolver um protótipo de mineração de informações de imagens de Sensoriamento Remoto (*Remote Sensing Image Information Mining - ReSIIM*) na linguagem de programação *Python*;
- Aplicar algoritmos de detecção e geração de metadados baseados nos alvos básicos de *nuvem, sombra de nuvem, área continental, água, floresta e solo exposto*;
- Validar a correspondência entre os metadados gerados e os dados de referência obtidos em trabalhos relacionados;
- Avaliar o desempenho do ReSIIM em três perspectivas: i) espacial - ao longo do Brasil no período de seca na Amazônia brasileira; ii) temporal - uma cena ao longo de 1 ano; iii) espectral - como suporte à aplicação do Modelo Linear de Mistura Espectral (MLME), principal abordagem em estudos de cicatrizes de queimadas na Amazônia brasileira.

1.2 Organização da dissertação

Na Seção 2, são apresentadas as fundamentações teóricas de: Sensoriamento Remoto e geração de dados, catálogos de imagens de Sensoriamento Remoto, sistemas de buscas por imagens relevantes em bancos de dados com base em seu conteúdo,

com destaque para as abordagens existentes, e exemplos de possíveis aplicações. Na Seção 3, é realizada uma descrição detalhada dos materiais e métodos utilizados para o desenvolvimento do protótipo e a metodologia da avaliação da correspondência entre os metadados e os alvos reais. Além disso, é apresentado ainda o processo utilizado para avaliar a acurácia do protótipo e a justificativa do uso da linguagem Python para a automatização.

Na Seção 4, são apresentados os resultados e discussões encontrados em três principais formas de busca: espacial - em cenas ao longo do Brasil no período de seca na Amazônia brasileira; temporal - na seleção de apenas uma cena e a análise temporal ao longo de um ano; e espectral - busca de cenas adequadas para a aplicação do MLME. Ademais, percepções encontradas ao longo do desenvolvimento e aplicação do protótipo são ressaltadas. Na Seção 5, são descritas as conclusões obtidas e sugestões para trabalhos futuros.

2 REFERENCIAL TEÓRICO

Na Subseção 2.1, é apresentada uma breve introdução ao Sensoriamento Remoto, bem como a geração massiva de dados. Alguns dos principais exemplos de catálogos de livre acesso, bem como os parâmetros de busca oferecidos aos usuários para refinar as buscas são apresentados na Subseção 2.2. A evolução dos sistemas de buscas por imagens em catálogos está apresentada na Subseção 2.3, enquanto que as aplicações em Sensoriamento Remoto e trabalhos relacionados ao ReSIIM são apresentados na Subseção 2.4.

2.1 Sensoriamento Remoto e geração de dados

O Sensoriamento Remoto pode ser definido como a arte e a ciência de, sem estar em contato físico direto com determinado objeto, obter informações a seu respeito (Instituto Brasileiro de Geografia e Estatística, 2001; JENSEN, 2009). Por meio da interação e análise da radiação eletromagnética e objetos da superfície terrestre, o Sensoriamento Remoto tem como objetivo estudar o ambiente terrestre (NOVO, 1992).

A energia eletromagnética mais relevante no Sensoriamento Remoto passivo é a energia solar, a qual ao incidir sobre um objeto pode ser parcialmente refletida, absorvida ou transmitida. A capacidade dos objetos de realizar estas interações são chamadas de reflectância, absortância e transmitância, respectivamente (Instituto Brasileiro de Geografia e Estatística, 2001). Muitos sistemas remotos registram apenas a energia refletida (JENSEN, 2009), transformando-a em um sinal que possa ser convertido em informação útil do objeto imageado. Esta informação pode fornecer indícios valiosos para reconhecer objetos terrestres (SABINS, 2007), já que está relacionada com as propriedades físico-químicas e biológicas do objeto, bem como com os fatores externos, como as condições atmosféricas no momento da aquisição do dado, descrevendo assim o *comportamento espectral* destes objetos (STEFFEN et al., 1996). Conforme Rosa (1990), o comportamento espectral representa a medida de reflectância de um alvo ao longo do espectro eletromagnético. Assim, é possível distinguir um objeto do outro, devido as diferenças nos níveis de reflectâncias entre os materiais em um determinado comprimento de onda (MENESES; ALMEIDA, 2012). Entretanto, alvos naturais com características comuns podem dificultar a identificação e individualização de alvos (BOWKER et al., 1985), como o caso de agricultura e diferentes tipos de florestas, os quais podem apresentar comportamentos semelhantes.

Os sistemas sensores de coleta de dados são acoplados em diferentes plataformas, como satélites, aviões (Instituto Brasileiro de Geografia e Estatística, 2001) ou drones.

Conforme o propósito da análise, os dados captados podem ainda ser integrados com outras informações para dar suporte a algum problema específico (CAMPBELL; WYNNE, 2011). Devido ao seu potencial, uma de suas principais limitações é a superestimação dos dados, já que apresentam apenas algumas e não todas as informações espaciais, espectrais e temporais dos objetos (JENSEN, 2009).

Um dos programas mais antigos de EO é a série Landsat (JENSEN, 2009), cujo primeiro satélite em órbita data da década de 70. Atualmente, o Landsat 8 é a mais nova plataforma, adquirindo cerca de 650 imagens da superfície da Terra diariamente. Possui dois sensores a bordo: OLI (*Operational Land Imager*) e TIRS (*Thermal Infrared Sensor*). Enquanto que o sensor OLI coleta 9 bandas espectrais, o sensor TIRS coleta 2 bandas na região do termal (USGS, 2016), sendo que cada banda espectral se refere a uma faixa no espectro eletromagnético (MOREIRA, 2005). Considerando que o Landsat 8 foi lançado em fevereiro de 2013, e cada coleta de dados possui cerca de 2 GB, até fevereiro de 2018, mais de 2.000 TB de dados do Landsat 8 foram gerados. Para gerir não apenas esses dados, mas também todo o volume de dados de Sensoriamento Remoto que estão sendo continuamente gerados, catálogos foram criados como uma forma de gerir e facilitar o acesso do conteúdo aos interessados.

2.2 Catálogos de imagens de Sensoriamento Remoto

Na Tabela 2.1, são apresentadas algumas das diversas plataformas de buscas em catálogos de imagens de Sensoriamento Remoto de livre acesso, juntamente com os parâmetros de busca disponíveis ao usuário. Em Quartulli e Olaizola (2013), é ainda possível analisar pormenorizadamente o estado da arte das ferramentas de busca de dados baseadas no conteúdo de arquivos de imagens de Sensoriamento Remoto. O nome na íntegra das organizações responsáveis, bem como os satélites/sensores podem ser visualizados na lista de Abreviaturas e Siglas presente no início deste trabalho.

O *Earth Explorer* (EE) permite que o usuário pesquise, baixe e ainda solicite dados presentes nos arquivos do USGS (*United States Geological Survey*). Como opção extra no parâmetro de busca, há a possibilidade de selecionar certos registros dos EUA, bem como selecionar a direção da órbita (ascendente ou descendente). No sistema EE, após selecionar algum dos resultados, os metadados da cena são apresentados, como nome, data de aquisição, linha/coluna, *bounding box*, satélite, sensor, tamanho da cena, número do talhão, agência responsável, dentre outros. Por fim, os dados podem ser selecionados em distintos formatos.

GloVis é uma rápida ferramenta de busca por imagens para selecionar dados de satélite e aéreos, cujos resultados apresentam os metadados das cenas. Já o *Landsat Look Viewer* é um protótipo desenvolvido para permitir rápida visualização e acesso aos arquivos de imagens da série Landsat. Ao final da busca é permitido visualizar em uma tabela as informações detalhadas das imagens selecionadas, como o nome da cena, o sensor, data e hora de aquisição, dia do ano, linha/coluna, elevação do sol, azimute e estação de recepção do dado.

O CLASS é uma biblioteca eletrônica de dados ambientais do NOAA (do inglês: *National Oceanic and Atmospheric Administration*). As buscas se restringem a coleções de metadados, como por exemplo o RADARSAT. A partir disto, outros parâmetros podem ser selecionados conforme a coleção desejada.

Desenvolvido pelo grupo EOS Data Analytics, a plataforma LV possibilita o acesso livre apenas a imagens do Landsat 8 e Sentinel-2. Como parâmetro de busca extra, cita-se a elevação do sol (0° - 90°). Nos resultados, é possível identificar o número de imagens correspondentes aos parâmetros da busca ao longo dos meses e realizar certas manipulações na nuvem, sem a necessidade de obter a imagem previamente ou alterar a imagem original. Além disso, o sistema permite o cadastro do usuário, envio de comentários, bem como o compartilhamento de mapas nas redes sociais.

Outra interface voltada para buscas por imagens do Landsat 8 e Sentinel-2 é organizado pelo Remote Pixel, o *Satellite Search*. As imagens são alimentadas pelo *Development Seed New Satellite API* e a visualização são do USGS *Earth Explorer* e do *Sentinel on AWS S3*, por meio da companhia Sinergise. Embora as buscas possam ser realizadas apenas por meio da localização, as cenas selecionadas apresentam a data de aquisição e a porcentagem de nuvem.

O ASF faz parte do Instituto de Geofísica da *University of Alaska Fairbanks*, tendo como missão tornar os dados de Sensoriamento Remoto acessíveis. Em sua plataforma, o principal atributo que pode nortear usuários é o nome do conjunto de dados de SAR e o período de aquisição dos dados.

O LAADS DAAC é um dos doze elementos da EOSDIS (do inglês: *Earth Observing System Data and Information System*) da NASA (do inglês: *National Aeronautics and Space Administration*), cujo foco é apoiar pesquisas e aplicações. Seus dados compreendem os sensores MODIS (do inglês: *Moderate Resolution Imaging Spectroradiometer*), SNPP (do inglês: *Suomi National Polar-orbiting Partnership*) e VIIRS (do inglês: *Visible Infrared Imaging Radiometer Suite*). Como opções extras, cita-se

os níveis de processamento das imagens, seleção de coleções de imagens e do intervalo de porcentagem em que não há informações, ou ainda coberturas realizadas apenas no período do dia, noite ou ambos. Além disso, ainda é possível salvar e baixar os parâmetros selecionados.

O Scientific Hub pertence ao Copernicus, sistema europeu de monitoramento da Terra, da ESA (do inglês: *European Space Agency*). Nele, é possível ter acesso completo e livre aos produtos do Sentinel (1, 2 e 3), e ainda há a presença de critérios de busca como o tipo de produto, número da órbita e o tipo da polarização.

O catálogo do INPE é gerido pela Divisão de Geração de Imagens (DGI), a qual faz parte da Coordenação Geral de Observação da Terra (OBT). O catálogo do INPE distribui imagens de diferentes satélites, sendo eles o S-NPP, UK-DMC-2, série LANDSAT, CBERS, MODIS e RESOURCESAT (Tabela 2.2). Nele, os principais parâmetros que podem ser definidos para a busca de imagens abrangem satélite, instrumento, data e cobertura máxima de nuvens (com variação de 10% e com base em 4 *quadrantes*). Além disso, as modalidades de busca podem ser definidas por município, órbita/ponto, região e interface gráfica, permitindo que o usuário navegue online conforme as coordenadas geográficas.

Apesar da diversidade, ainda há lacunas nos parâmetros de busca das imagens, como selecionar cenas considerando os alvos básicos presentes nelas e não apenas a localização geográfica e a data de aquisição.

Tabela 2.1 - Plataformas de acesso a dados de Sensoriamento Remoto e os parâmetros de busca definidos pelo usuário. Nos parâmetros de busca disponíveis, Geográfico representa a área imageada, e X, a presença de determinado parâmetro de busca.

| Nome | Sigla | Acesso | Parâmetros de Buscas Disponíveis | | | |
|--|----------------|---|----------------------------------|-------------------|------------|----------|
| | | | Geográfico | Data de Aquisição | Nuvens (%) | Sensores |
| Earth Explorer | EE | https://earthexplorer.usgs.gov/ | X | X | X | X |
| Global Visualization Viewer | GloVis | http://glovis.usgs.gov/ | X | X | X | X |
| LandsatLook Viewer | LLV | https://landsatlook.usgs.gov/ | X | X | X | X |
| Comprehensive Large Array-Data Stewardship System | CLASS | https://www.class.noaa.gov/saa/products/catSearch | X | X | - | - |
| Land Viewer | LV | https://lv.eosda.com/ | X | X | X | X |
| Satellite Search | SS | https://www.remotepixel.ca/projects/satellitesearch.html | X | - | - | - |
| Alaska Satellite Facility | ASF | https://www.asf.alaska.edu/get-data/get-started/how-to-get-sar-data/ | - | X | - | - |
| Level-1 and Atmosphere Archive & Distribution System - Distributed Active Archive Center | LAADS DAAC | https://ladsweb.nascom.nasa.gov/search/ | X | X | - | X |
| Sentinels Scientific Data Hub | Scientific Hub | https://scihub.copernicus.eu/dhus/ | X | X | X | X |
| Instituto Nacional de Pesquisas Nacionais | INPE | http://www.dgi.inpe.br/catalogo/ | X | X | X | X |

Tabela 2.2 - Produtos disponíveis no catálogo do INPE.

| Nome | Sigla | Sensores | Operadora / Instituição Responsável | Situação Atual | Lançamento | Resolução Espacial |
|---|----------------|-------------------------------|-------------------------------------|----------------------|------------|--|
| Aqua Project Science | AQUA | MODIS | NASA | Ativo | 04/05/2002 | 500 m |
| China-Brazil Earth Resources Satellite 2 | CBERS-2 | CCD; IRM; WFI | INPE; CAST | Inativo (10/01/2009) | 21/10/2003 | CCD: 20 m; IRMSS: 80 m (termal: 160 m); WFI: 260 m |
| China-Brazil Earth Resources Satellite 2B | CBERS-2B | CCD; HRC; WFI | INPE; CAST | Inativo (12/05/2010) | 19/09/2007 | CCD: 20 m; HRC: 2,7 m; WFI: 260 m |
| China-Brazil Earth Resources Satellite 4 | CBERS-4 | AWFI; IRS; MUX; PAN5M; PAN10M | INPE; CAST | Ativo | 07/12/2014 | PAN: 5 m (pancromática) e 10 m (multiespectral); MUX: 20 m; IRS: 40 m (pancromática) e 80 m (termal) |
| DEIMOS-1 | DEIMOS-1 | SLIM-6 | DMI | Ativo | 29/07/2009 | 22 m |
| Disaster Monitoring Constellation (DMC) | UK-DMC-2 | SLIM-6 | SSTL | Ativo | 29/07/2009 | 22 m |
| Global Land Survey | GLS-LANDSAT | TM; ETM+ | USGS; NASA | Inativo (2011) | 1972 | 30 m |
| Earth Resources Technology Satellite | Landsat (ERTS) | MSS | NASA | Inativo (01/1978) | 23/07/1972 | 30 m |

Continua ...

| Nome | Sigla | Sensores | Operadora / Instituição Responsável | Situação Atual | Lançamento | Resolução Espacial |
|---|------------------------|-----------------|--|-----------------------|-------------------|-----------------------------|
| Earth Resource Technology Satellite B | Landsat 2 (ERTS-B) | MSS | NASA | Inativo (05/02/1982) | 22/01/1975 | 30 m |
| Landsat 3 | Landsat 3 | MSS | NASA; NOAA | Inativo (03/1983) | 05/03/1978 | 30m |
| Landsat 5 | Landsat 5 | TM | NASA; USGS | Inativo (01/2013) | 01/03/1984 | 30 m |
| Landsat 7 | Landsat 7 | ETM+ | USGS | Inativo (05/2003) | 15/04/1999 | 30 m |
| Landsat 8 | Landsat 8 | OLI | NASA | Ativo | 11/02/2013 | 30 m |
| Indian Remote Sensing Satellite 1 | ResourceSat-1 (IRS-P6) | LISS3; AWIFS | ISRO | Inativo (30/09/2013) | 17/10/2003 | 23,5 m |
| Indian Remote Sensing Satellite 2 | ResourceSat-2 | LISS3; AWIFS | ISRO | Ativo | 20/04/2011 | LISS 3: 23,5 m; AWFI: 189 m |
| Suomi National Polar-Orbiting Partnership | S-NPP | VIIRS | NASA; NOAA | Ativo | 28/10/2011 | 750 m |
| Terra Project Science | TERRA | MODIS | NASA | Ativo | 18/12/1999 | 250 m |

2.3 Busca de imagens em grandes catálogos

A seguir, são apresentadas as principais abordagens de busca de imagens em grandes catálogos com base em seu conteúdo, e a estrutura de sistemas de busca de imagens por meio de metadados.

2.3.1 *Content-Based Image Retrieval* - CBIR

Embora o conceito CBIR já existisse desde a década de 70 (LONG et al., 2003) e 80 (MÜLLER et al., 2004), foi apenas a partir da década de 90 que passou a ser amplamente utilizado. Isso ocorreu devido à expansão das tecnologias de multimídia e internet, a qual gerou um rápido aumento de imagens disponíveis (EAKINS; GRAHAM, 1999; LONG et al., 2003; DATCU; SEIDEL, 2003b; ARAI, 2013), juntamente com uma demanda por ferramentas de gestão desse grande volume de dados.

Em um primeiro momento, as técnicas referentes à busca e gestão de imagens eram voltadas basicamente para a área de fotografia e eram baseadas em textos (do inglês: *Text-Based*, *Concept-Based* ou ainda *Description-Based Image Retrieval*). Nessa abordagem, as imagens são comentadas por meio de textos, que auxiliam nas buscas futuras (KHERFI et al., 2004). Assim, a busca por texto facilitava a navegação por entre as imagens e na busca por imagens (LONG et al., 2003). Entretanto, alguns pontos devem ser destacados. A descrição automática de uma grande quantidade de imagens não é possível (LONG et al., 2003), por ser um processo de anotação manual. Além de vagarosa, essa atividade é custosa e subjetiva, uma vez que depende da interpretação do operador, bem como da sua experiência para identificar certos elementos. Ademais, uma mesma palavra pode possuir uma quantidade significativa de sinônimos, dificultando o processo de busca de imagens. Assim, de acordo com o observador, a mesma imagem pode ser descrita e catalogada de maneira diferente. Por fim, certos elementos de uma imagem não são facilmente descritos por meio de palavras (KHERFI et al., 2004).

Levando em consideração a grande quantidade de imagens disponíveis, o método baseado em texto começou a enfrentar dificuldades. Assim, esforços foram concentrados visando identificar novas diretrizes para a gestão e busca de imagens (EAKINS; GRAHAM, 1999), como o CBIR (do inglês: *Content-Based Image Retrieval*).

Os sistemas CBIR têm como objetivo analisar e buscar imagens com base em seus conteúdos dentro de um catálogo principal (VEGANZONES et al., 2008; SCHUH et al., 2015). Mais especificamente, dentro de um banco de dados, esses sistemas buscam

encontrar imagens com uma determinada similaridade ou com uma certa distância definida no espaço do atributos (VEGANZONES et al., 2008). Segundo Long et al. (2003) e Datcu e Seidel (2003b), CBIR é uma técnica que utiliza conteúdos visuais na busca por imagens, sendo baseada em elementos como cor, textura, formato e distribuição espacial para representar e indexar imagens.

Embora diversos sistemas CBIR tenham sido propostos, uma quantidade significativa deles não obteve sucesso, uma vez que esses sistemas eram amplos e compreendiam diferentes tipos e fontes de imagens (SHYU et al., 2007). Desse modo, devido à potencialidade desse sistema, diversas tentativas foram realizadas com o objetivo de aplicá-lo à áreas específicas, como marketing, segurança, diagnósticos médicos, arquitetura, geologia e Sensoriamento Remoto (EAKINS; GRAHAM, 1999; VEGANZONES et al., 2008). Schuh et al. (2015) propuseram inclusive um sistema voltado para a busca de imagens do Sol. No que se refere ao Sensoriamento Remoto, a possibilidade de extrair dados de grandes bases de imagens é útil principalmente para aplicações que requerem procedimentos de reconhecimento de padrões (ARAI, 2013).

Apesar do termo primário ser CBIR, outras denominações foram geradas. Entre elas, destaca-se a sigla *GIR* (do inglês: *Geographic Information Retrieval*), sendo aquele que tem como objetivo fornecer acesso a fontes de informação georreferenciada (LARSON; FRONTIERA, 1996). Os dados geoespaciais descrevem elementos que tenham relação com o espaço geográfico, frequentemente com localização geográfica em um sistema de referência espacial (LI et al., 2016a).

Ainda que os sistemas de indexação sejam objetivos e automáticos, existe um intervalo semântico entre a máquina e o usuário. Visando diminuir essa lacuna entre sistemas computacionais e a subjetividade dos usuários, abordagens que focam no *feedback* (comentário) de usuários (do inglês: *Relevance Feedback* - RF) começaram a ser desenvolvidas. Para mais detalhes, Long et al. (2003) descreveram as técnicas fundamentais de CBIR voltadas para atributos visuais.

Apesar dos esforços e avanços, essas abordagens eram centradas em computadores (do inglês: *computer centered*) e não satisfaziam mais as demandas existentes. Assim, houve uma necessidade por desenvolver *designs* mais modernos tendo como centro o ser humano (do inglês: *human centered systems*). De acordo com Bijl-Brouwer e Dorst (2017), essas estratégias de inovação requerem abordagens holísticas, se adequando aos interesses do ser humano como parte do sistema, e contribuindo para impactos positivos. Assim, não se trata apenas dos usuários, mas clientes, provedores de serviços, cidadãos e tomadores de decisão nos diversos níveis organizacionais,

como setores privados e públicos.

O processo que identifica padrões válidos, novos, potencialmente úteis e compreensíveis em dados é conhecido como KDD (do inglês: *Knowledge-Discovery in Databases* (FAYYAD et al., 1996), sendo uma das 7 principais etapas conhecida como mineração de dados (do inglês: *data mining*), processo essencial onde métodos inteligentes são aplicados para extrair padrões (HAN et al., 2011). Entretanto, há autores que afirmam que o termo KDD caiu em desuso (PETRUSHIN; KHAN, 2007) e utilizam a mineração de dados com um significado similar, sendo a técnica de entender, navegar e explorar grandes e complexos bancos de dados (FAYYAD et al., 2001).

Independente do conceito, no final da década de 90, a abordagem de KDD e de mineração de dados foram difundidas ao CBIR, dando origem ao termo mineração de informações de imagens (do inglês: *Image Information Mining - IIM*) (DATCU; SEIDEL, 2003b).

2.3.2 *Image Information Mining - IIM*

Embora exista uma sobreposição entre os conceitos de CBIR e IIM (LI; NARAYANAN, 2006), este último tem como objetivo extrair as principais informações contidas em grandes bancos de dados (DATCU et al., 2000), levando em consideração distintas técnicas e atributos, enquanto que o CBIR apenas considera as similaridades visuais (LONG et al., 2003; DATCU; SEIDEL, 2003b). IIM pode assim ser considerado uma etapa de comunicação entre a fonte de informações (banco de dados) e os receptores (usuários), cuja acurácia depende dos modelos empregados (DATCU; SEIDEL, 2003b). Desse modo, o amplo conhecimento acumulado e compartilhado sobre dados e informações geoespaciais oferece uma oportunidade única de progresso frente ao IIM (SHYU et al., 2007). Sendo os dados de grande importância para a resolução de problemas complexos (GUPTILL, 1999), a aplicação de técnicas de IIM torna-se altamente relevante no refinamento de buscas de imagens em grandes catálogos de Sensoriamento Remoto.

Entretanto, a complexidade de IIM é maior se comparada aos sistemas CBIR e à mineração de dados (do inglês: *Data Mining*) (DATCU; SEIDEL, 2003a; DATCU; SEIDEL, 2003b), sendo um desafio aos profissionais envolvidos, devido aos seguintes fatores:

- grande volume de dados (TB para PB);
- variabilidade e heterogeneidade de imagens (como a diversidade de senso-

res, momento e condições da aquisição da imagem, e influências atmosféricas);

- conteúdo das imagens, cujo significado pode variar conforme o intérprete, tornando-se subjetivo;
- diversos interesses que os usuários podem apresentar, bem como a sua compreensão semântica e contextual.

Dessa maneira, para a sua aplicação e bom desempenho, dois desenvolvimentos devem caminhar juntos: a provisão de catálogos de dados e o uso de metadados, já que permitem aos usuários encontrar, compartilhar e usar dados geoespaciais (GUPTILL, 1999).

2.3.3 Estrutura de sistemas de busca de imagens por meio de metadados

De maneira geral, os sistemas de buscas de imagens podem ser divididos em três partes: interação com o usuário por meio de uma interface de busca, extração de atributos e a apresentação dos resultados (VEGANZONES et al., 2008). As buscas por imagens podem ainda apresentar distintos formatos, como buscas por categorias, conceitos, esboços e exemplos (LONG et al., 2003; VEGANZONES et al., 2008). Em uma mesma imagem, distintos objetos podem ser identificados (descritores), podendo ser considerados globais ou locais, referindo-se assim à imagem como um todo ou apenas à certas regiões/objetos, respectivamente (LONG et al., 2003). Uma classificação da complexidade nos níveis de busca foi proposto por Eakins e Graham (1999), sendo apresentada a seguir. O detalhamento, bem como os exemplos citados, foram adaptados para imagens de Sensoriamento Remoto, não representado necessariamente alguma tecnologia já desenvolvida.

- Nível 1 - baixa complexidade: leva em consideração as características mais primitivas das imagens, como cor, textura, formato e localização espacial dos elementos distribuídos em uma imagem. As informações podem ser extraídas a nível de pixel e não há semântica envolvida. Nesse nível, os sistemas não levam em consideração o intervalo semântico entre as expectativas dos usuários e a resposta do sistema (VEGANZONES et al., 2008). Além disso, os elementos visuais são extraídos e descritos por vetores multidimensionais, os quais são catalogados em um base de dados (LONG et al., 2003);

- Nível 2 - média complexidade: também chamados de busca por elementos lógicos, envolvem a presença ou a composição de tipos específicos de objetos com certo significado semântico. Podem ser obtidos por um conjunto de pixels e são utilizados com maior frequência para análises de cobertura do solo ¹ (*encontre imagens com 0% de nuvens e que possuam mais de 50% de floresta*);
- Nível 3 - alta complexidade: compreende a busca de atributos abstratos, possuindo, assim, alta complexidade semântica envolvida. Essas características são capazes de identificar os significados com um determinado conjunto de pixels que constituem um objeto. Além da cobertura, neste nível também é apresentado o uso do solo ² (*encontre imagens que apresentem plantações de feijão em época de safra*).

A Figura 2.1 apresenta parte de uma cena do sensor *Operational Land Imager* (OLI - Landsat 8), órbita/ponto 219/076, de fevereiro de 2015, sendo a composição das bandas R4G3B2 fusionada com a banda 8, pancromática. Nesse exemplo, é possível analisar alguns exemplos de descritores aplicados a uma imagem de Sensoriamento Remoto no nível 2. De uma maneira geral, a cobertura do solo na cena é representada como uma área urbana, mas, ao analisar os descritores locais, é possível identificar elementos como água e floresta.

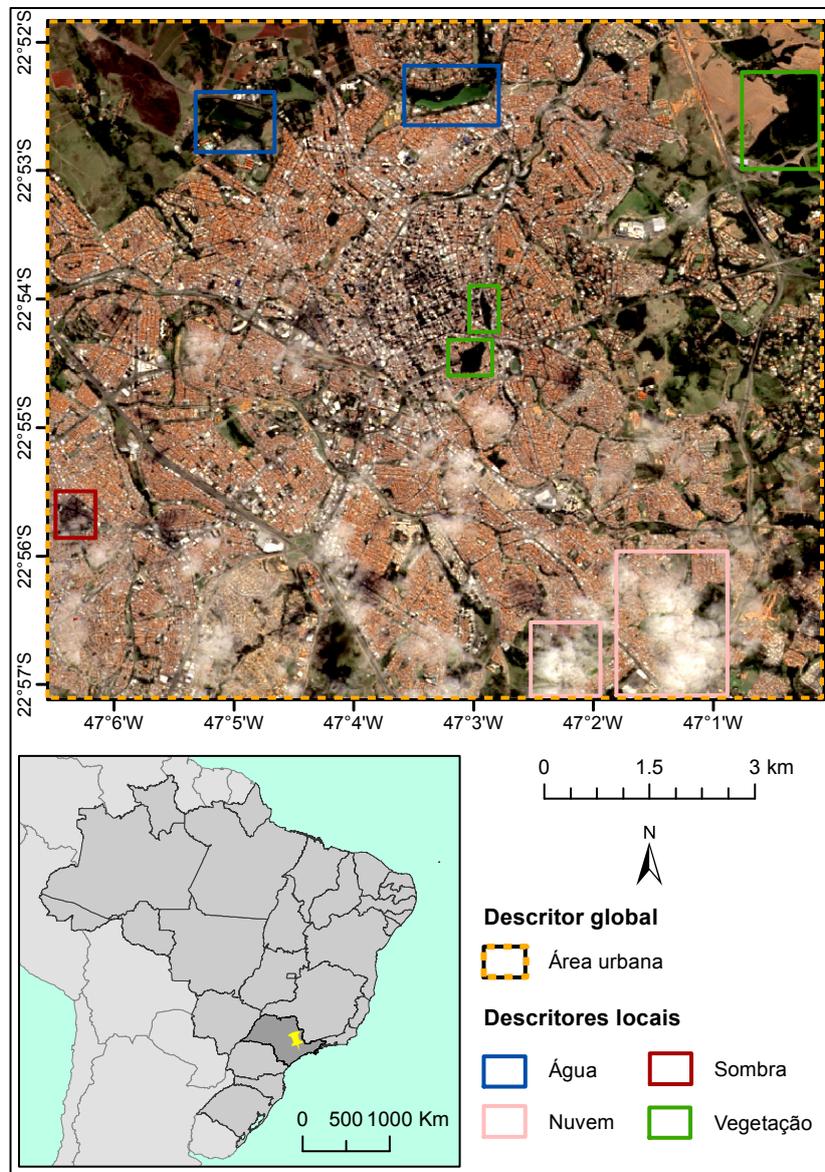
Um dado espacial está relacionado a qualquer informação com localização absoluta ou relativa. Nesse contexto, a coleção de informações de fenômenos geográficos formam bancos de dados espaciais (GUPTILL, 2015). Os descritores, amplamente chamados de metadados, apresentam informações sobre aspectos de um determinado elemento (GUPTILL, 1999). Assim, provêm atributos e características detalhadas de um determinado elemento (GUPTILL, 1999), sendo considerado *o dado sobre o dado*, parte essencial do arquivamento (EAKINS; GRAHAM, 1999).

Considerando que um dado espacial pode representar conceitos e inter-relações complexas do mundo real, os metadados deveriam transmitir de alguma forma essas abstrações (FRANK, 1994). Baseado em metadados de imagens de Sensoriamento Remoto, diferentes critérios de procura por imagens podem ser desenvolvidos em

¹Estado biofísico determinado pela direta observação da superfície terrestre, como corpos de água e floresta. Não é possível uma única área apresentar simultaneamente no tempo e espaço duas coberturas (MEYER; TURNER, 1994; FISHER et al., 2005; COMBER, 2008)

²Maneira como os elementos biofísicos são geridos, sendo uma interpretação sócio-econômica das atividades que são desempenhadas na superfície, como pesquisa ou agricultura. Uma mesma área pode apresentar diferentes usos (MEYER; TURNER, 1994; FISHER et al., 2005; COMBER, 2008)

Figura 2.1 - Exemplos de descritores de cobertura do solo em uma imagem de Sensoriamento Remoto no nível 2.



Destaque para o descritor global área urbana e descritores locais água, nuvem, sombra e vegetação localizados no Estado de São Paulo.

Fonte: Elaborada pela autora.

catálogos. Genericamente, um catálogo se refere a uma lista de descrição dos itens encontrados em uma coleção (FRANK, 1994). Um catálogo de dados é assim uma coleção de metadados arquivados, o qual é associado com ferramentas de buscas e gestão de dados (GUPTILL, 1999). Dessa maneira, um catálogo de imagens facilita

as operações de busca, compartilhamento e processamento de dados aos usuários.

Os esforços para a compreensão e geração de Metadados em Imagens de Sensoriamento Remoto (MISS) podem ser identificados através de algoritmos voltados para distintos objetivos. Para a detecção de nuvens e sombra de nuvens em imagens do sensor do Landsat *Thematic Mapper* (TM) e *Enhanced Thematic Mapper Plus* (ETM+), o algoritmo orientado a objeto *Fmask* foi desenvolvido por Zhu e Woodcock (2012). Ele foi idealizado, uma vez que não haviam métodos de identificação de nuvens em imagens de média resolução. Uma segunda versão do algoritmo foi desenvolvida por Zhu et al. (2015), o qual foi ajustado para as séries Landsat e Sentinel-2. Baseado em ambos os trabalhos ora descritos, Flood e Gillingham (2017) adaptaram o conjunto de comandos para a linguagem de programação *open source Python*, o qual foi utilizado na estrutura do ReSIIM. Como dado de entrada, todas as bandas da cena desejada são utilizadas, enquanto que o dados de saída é uma única cena raster temática com até 6 valores [0, 5], representando: *valor nulo*, *nuvem*, *sombra de nuvem*, *área continental (clear land)*, *neve*, e *água*. Área continental (AC) se refere a todos os pixels que não são classificados como os alvos mencionados e mesmo assim possui valor válido. Ou seja, AC são áreas que não são encobertas por nuvens ou sombra de nuvens e não apresentam água em sua superfície. Sendo assim, sob uma área continental, diferentes usos e coberturas de solo podem ocorrer, como floresta, solo exposto, área construída, agricultura, entre outros.

A informação sobre a presença de água nos MISS também pode ser extraída com base na metodologia desenvolvida por Namikawa et al. (2016), os quais extraíram automaticamente máscaras de corpos de água de 5m de resolução espacial de imagens do *RapidEye*³ ao longo do Brasil por 4 anos (2011, 2012, 2013 e 2014). Embora os corpos de água sejam normalmente identificados devido à sua baixa reflectância, no mundo real, diversos parâmetros interferem nessa condição, como os sólidos suspensos e a profundidade do corpo hídrico (NAMIKAWA et al., 2016). Nesse contexto, Namikawa et al. (2016) desenvolveram uma metodologia baseada na transformação de cor do RGB (*Red-Green-Blue*) para HSV (*Hue-Saturation-Value*) e uma radiância mínima de todas as bandas. Para isso, alguns fatores foram considerados, como as diferenças na iluminação e espalhamento ao longo de mais de 15.000 cenas do RapidEye. Como resultado, foi possível classificar 7 classes de água [1-7], sendo 1 mais confiável e 7 menos confiável. Além disso, também é possível filtrar os dados

³Mais detalhes sobre a constelação RapidEye disponível em: <https://www.planet.com/products/satellite-imagery/files/160625-RapidEye%20Image-Product-Specifications.pdf>.

com base na persistência do alvo água ao longo do tempo. De acordo com os autores, embora a metodologia descrita seja simples, é acurada para detectar corpos de água. Esta abordagem é considerada mais flexível quando comparada com um limiar único em grandes áreas, como o Índice de Diferença Normalizada da Água (*Normalized Difference Water Index* - NDWI). Apesar disso, Namikawa e Castejon (2017) identificaram alguns fatores que podem interferir no resultados. Entre eles, destacam-se os ruídos provindos de nuvens, sombra em áreas urbanas e o reflexo especular do sol.

Os algoritmos para estimar os parâmetros biofísicos e bioquímicos de vegetação podem ser classificados em três principais grupos, métodos estatísticos, físicos e híbridos. A primeira abordagem é baseada principalmente nas variedades de índices espectrais de vegetação, a segunda baseia-se em uma variedade de modelos de reflectância do dossel invertido, enquanto que os métodos híbridos abrangem ambos (LIANG, 2005). Os índices partem do pressuposto de que algumas combinações algébricas de bandas espectrais de Sensoriamento Remoto podem ressaltar alguma informação útil sobre a vegetação, como a estrutura, a quantidade de água e a densidade da folha. É considerado um bom índice, aqueles sensíveis aos fatores ora citados (LIANG, 2005). A grande variedade de índices espectrais de vegetação permite ainda que cada índice apresente um melhor desempenho de acordo com a necessidade do usuário (FORMAGGIO; SANCHES, 2017). Um exemplo de método estatístico é o *Normalized Difference Vegetation Index* (NDVI), apresentado detalhadamente a seguir.

Sendo um dos primeiros e mais difundidos índices de vegetação, o NDVI foi primeiramente proposto por Rouse et al. (1974). Variando entre $[-1, +1]$, esse índice tem como objetivo avaliar se o alvo é uma vegetação verde ou não. A definição dos limites no NDVI para o mapeamento de vegetação e solo exposto é ainda controverso, entretanto, os valores do NDVI entre 0,0 e 0,1 normalmente se referem a rochas e solo exposto. Valores acima de 0,1 indicam um aumento gradual do *verde* e a intensidade da vegetação (NOAA, 2017). Do outro lado, alguns autores usam o limite de 0,2 no valor do NDVI para a transição entre solo exposto e vegetação (JIN et al., 2014; LIANG et al., 2014b; LIANG et al., 2014a), sendo a zona de transição considerada entre 0,1 e 0,2 (LIANG et al., 2014a). O NDVI leva em consideração as bandas do satélite mais sensíveis a informação de vegetação, *near infrared* (NIR) e *red* (RED) (Equação 2.1). Assim, a variação do NDVI pode dar suporte a identificação de mudanças na cobertura da terra no que refere a oscilação entre áreas de solo exposto/rochas e densas vegetações.

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}} \quad (2.1)$$

2.4 Aplicações em Sensoriamento Remoto e trabalhos relacionados

Diversos estudos trabalham com refinamento de ferramentas de buscas por imagens em catálogos de imagens de Sensoriamento Remoto (SEIDEL; THERRE, 1994; SEIDEL et al., 1998). Entretanto, devido à complexidade e variedade das imagens de Sensoriamento Remoto, diferentes procedimentos podem ser adotados.

Algumas das abordagens utilizadas nesse tipo de imagem podem variar desde a utilização da informação espectral até mesmo características espaciais (VEGANZONES et al., 2008). O uso de dados espectrais para a classificação de imagens é uma das técnicas mais difundidas, embora apresente desafios, como mudanças das condições de iluminação e padrões espectrais semelhantes entre objetos distintos (SCHRÖDER; DIMAI, 1998). Levando em consideração a alta variabilidade de imagens de Sensoriamento Remoto, Napoletano (2016) analisou quais descritores seriam mais viáveis. O estudo abrangia elementos locais e globais em uma abordagem de CNN (do inglês: *Convolutional Neural Network*), baseado em aprendizagem de máquina. Os resultados apontaram que os atributos mais adequados variam conforme o tipo do conjunto de dados utilizados.

Assim, visando exemplificar trabalhos relacionados, independentemente do termo utilizado pelo autor (CBIR ou IIM), a seguir são apresentadas algumas aplicações e trabalhos relacionados em sistemas de busca de imagens (Tabela 2.3). Como as ferramentas extrapolam a discriminação meramente visual dos elementos, pode-se afirmar que são baseados em IIM.

Xie et al. (2008) desenvolveram uma abordagem em um esforço de identificar pinus australianos (*Casuarina spp*), espécie exótica proibida pelo Departamento de Proteção Ambiental da Flórida (EUA). Embora tenha este nome e realmente sejam nativas da Austrália, não são nem pinus e nem mesmo coníferas, são angiospermas e estão posicionadas na Categoria I de plantas mais invasivas de acordo com a organização *Florida Exotic Pest Plant Council* (Florida Exotic Pest Plant Council, FLEPPC, 2017). Nesse estudo, foram computados propriedades dos objetos para identificar qual combinação de atributos melhor representaria a *Casuarina spp*. Assim, foram computados os dados espectrais de NDVI, texturais (transformada de *wavelet* e variograma), e tridimensionais por meio de *LiDAR* (informação da altura da superfície do dossel). As imagens foram processadas por meio da segmentação multiresolução

Tabela 2.3 - Aplicações em Sensoriamento Remoto de buscas por imagens com base em seu conteúdo.

| Referência | Atributos utilizados | Foco da aplicação |
|---|---|---|
| (XIE et al., 2008) | <ul style="list-style-type: none"> • índice NDVI • dados texturais (transformada de <i>wavelet</i> e variograma) • informação da altura da superfície do dossel | Identificar pinus australianos (<i>Casuarina spp</i>) |
| (MOLINIER et al., 2005; MOLINIER et al., 2006; MOLINIER et al., 2007) | <ul style="list-style-type: none"> • cor • média dos valores de RGB • dados texturais • coordenadas geográficas • índice NDVI • histograma | Detectar mudanças provindas de atividades antrópicas |
| (CARRARA et al., 2004) | <ul style="list-style-type: none"> • dados espectrais | Indexar imagens de Sensoriamento Remoto de modo a subsidiar buscas em banco de dados de maneira mais rápida |
| (SHYU et al., 2007) | <ul style="list-style-type: none"> • histogramas de imagens pancromáticas • escala de cinza RGB • dados do infravermelho próximo • atributos agregados aos DMP • índice NDVI • formato de objetos | Sistema de busca de imagens com base em seus conteúdos |
| (SAMAL et al., 2009) | <ul style="list-style-type: none"> • comportamentos espectrais | Sistema de busca de imagens com base em seus conteúdos |
| (VEGANZONES et al., 2013; PLAZA, 2011) | <ul style="list-style-type: none"> • dados espectrais • dados espaciais | Sistema de busca de imagens hiperespectrais baseados em técnicas de modelo de mistura |

e, conforme critérios, passaram por uma filtragem do conteúdo. Por exemplo, a informação espectral foi principalmente utilizada para separar os objetos vegetais dos demais. O resultado demonstrou que para a detecção desta espécie, dados tridimensionais tiveram maior importância, devido às características físicas semelhantes a de pinheiros.

Nessa vertente, outra abordagem que merece destaque é chamada de PicSOM, a qual é baseada em buscas por imagens semelhantes. Assim, o usuário deve selecionar exemplos em um banco de imagens do próprio sistema ou em algum local externo. Esse processo é interativo e utiliza de comentários de usuários, ou seja, quanto mais o usuário interage, mais o sistema aprende as suas preferências. Essa possibilidade é útil principalmente pelo usuário não ter a necessidade de determinar diretamente os atributos ou limites, além do sistema se adaptar a diferentes interesses. O método de indexação utilizado no PicSOM é baseado no algoritmo mapa auto-organizável (do inglês: *Self-Organized Map* - SOM), o qual pode ser utilizado para visualizar dados multidimensionais em uma grade de duas dimensões. Assim, o SOM define uma rede de pontos que se ajustam ao espaço que é inserido. Esse sistema foi desenvolvido para ser capaz de se adaptar a diferentes tipos e tamanhos de imagens, embora em um primeiro momento o foco tenha sido imagens genéricas (LAAKSONEN et al., 1999; LAAKSONEN et al., 2000; KOSKELA et al., 2000). Devido a sua potencialidade, também foi aplicado o PicSOM para imagens de Sensoriamento Remoto. Essas abordagens tiveram como foco a detecção de mudanças em atividades antrópicas. Além dos atributos já utilizados, como cor, média dos valores de RGB e textura, também foram levados em consideração NDVI, histograma e as coordenadas geográficas (MOLINIER et al., 2005; MOLINIER et al., 2006; MOLINIER et al., 2007).

Em Carrara et al. (2004), é proposto um modelo de indexação de imagens de Sensoriamento Remoto. Nesse modelo, os pixels são representados pelos valores vetoriais por meio do comportamento espectral. Dessa forma, os descritores das imagens são as propriedades espectrais dos pixels. Para o teste do método, imagens Landsat da Itália foram utilizadas. Assim, o desempenho do modelo em dois sensores foram avaliados, *TM* e *ETM+*, a bordo do Landsat 5 e 7, respectivamente. O diferencial do sistema reside na modelagem explícita da incerteza intrínseca entre os padrões encontrados em imagens e a sua correspondência na cobertura do solo. As especificações de busca por imagens foram baseadas em procuras textuais e comportamentos espectrais. O resultado apresenta uma lista de imagens associadas ao grau de relevância. Dentre os resultados adquiridos na avaliação para quatro classes (vegetação, água, nuvem e solo exposto), aquele que melhor foi classificado foi a vegetação. Isso

ocorreu devido à alta reflectância na região do infravermelho próximo e à baixa reflectância na região correspondente ao vermelho no espectro eletromagnético. Solo exposto foi o segundo melhor identificado, seguido das classes água e nuvens.

Shyu et al. (2007) desenvolveram o GeoIRIS (do inglês: *Geospatial Information Retrieval and Indexing System*), o qual inclui extração automática de atributos, mineração de conteúdos visuais dentro de um banco de dados e indexação que propicie o rápido retorno do resultado na procura imagens. Os atributos foram extraídos orientados a células (256 x 256m) e a objetos. Os atributos espectrais e texturais foram computados em três conjuntos, histogramas de imagens pancromáticas, escala de cinza RGB e dados do infravermelho próximo. Enquanto que os atributos agregados ao DMP (do inglês: *Differential Morphological Profiles*) são provenientes dos conjuntos de vegetais e não-vegetais, determinados pelo índice NDVI. Os objetos são processados para extração de atributos espectrais e de formato. Assim como o exemplo de Carrara et al. (2004), o resultado do GeoRIS também apresenta um rank de imagens relevantes no banco de dados.

Um sistema voltado também para imagens de satélite foi proposto por Samal et al. (2009). Denominado de SIMR (do inglês: *Satellite Image Matching and Retrieval System*), esse sistema representa uma imagem por meio de um comportamento espectral. Assim, com base em uma medida de auto-correlação espacial, estruturas espaciais em diferentes escalas são determinadas. SIMR usa dos comportamentos espectrais integradas em um processo de busca de dados para associar a imagem analisada com as imagens já catalogadas no banco de dados. Além disso, SIMR utiliza diferentes medidas de distância para determinar as similaridades entre dois comportamentos, ordenando as imagens conforme a similaridade.

Embora existam diferentes abordagens, as proposições são recentes em aplicações de sensores hiperespectrais. Em Vezanzones et al. (2013), é apresentada uma proposta de metodologia de RF por meio da dissimilaridade entre os elementos. Assim, foram analisadas as distâncias dentro de uma matriz de distância espectral. Essa dissimilaridade foi baseada em duas abordagens. Primeiramente, foram considerados os atributos espaciais e espectrais extraídos por técnicas de um modelo de mistura espectral (SHIMABUKURO; SMITH, 1991). Esse tipo de modelo é importante para a exploração de dados, uma vez que a coleta de dados por sensores orbitais é composta por uma mistura dos diversos materiais existentes na superfície da Terra (PLAZA, 2011), além das possíveis interferências ao longo da trajetória da radiação eletromagnética (mais detalhes sobre este tipo de modelo são apresentados na

Subseção 2.4.1). Em segundo lugar, foram considerados os dicionários extraídos por algoritmo de compressores baseados em dicionário (do inglês: *dictionary-based compressors*), onde dado um sinal, esse algoritmo atribui ao padrão encontrado um determinado vocábulo, sendo que o conjunto dessas palavras únicas são chamados de dicionários. De maneira análoga, em Plaza (2011) também é possível encontrar uma aplicação baseada em modelo de mistura para dados hiperespectrais como uma ferramenta de busca por metadados em imagens.

2.4.1 Queimadas na Amazônia brasileira e a aplicação do Modelo Linear de Mistura Espectral

Os eventos de queimadas e incêndios florestais tornam os ambientes impactados mais expostos a subsequentes impactos climáticos e antropogênicos, reduzindo a quantidade de biomassa da floresta e conseqüentemente os estoques de carbono (SATO, 2016). Durante a combustão da matéria orgânica, libera-se o dióxido de carbono, principal gás de efeito estufa, representando cerca de 90% dos gases liberados (LASHOF, 1991). Desse modo, a queima de biomassa afeta não apenas a floresta no nível local, mas também o ciclo do carbono, influenciando as trocas radiativas terrestres e as mudanças no clima (ANDERSON et al., 2015; SHIMABUKURO et al., 2015; PADILLA et al., 2017).

No passado, a ocorrência de fogo na Amazônia é considerada rara (BUSH et al., 2007). Entretanto, juntamente com o aumento das taxas de desmatamento na região, esse fenômeno começou a se tornar mais frequente, já que possui estreita relação com o processo de desmatamento e manejo de áreas agrícolas e de pecuária (SHIMABUKURO et al., 2015; ARAGÃO et al., 2016).

Aragão et al. (2014) e Gatti et al. (2014) destacaram que com a crescente demanda internacional por reduções nas emissões de carbono, controlar o uso indiscriminado de fogo na região da Amazônia pode ser uma eficiente estratégia para reduzir essas emissões. Apesar disso, ainda há muitas incertezas associadas às detecções e mapeamentos de áreas queimadas por meio de Sensoriamento Remoto, como a problemática da super estimativa de áreas queimadas (ANDERSON et al., 2005), devido à confusão de pixels com resposta espectral semelhante, a variabilidade e desenvolvimento temporal das características espectrais das áreas queimadas (SHIMABUKURO et al., 2015), além da própria resolução espacial das imagens e produtos (ANDERSON et al., 2015; ANDERSON et al., 2017). Em Aragão et al. (2016), estudos para estimar áreas de floresta queimadas são brevemente descritos, variando desde os primeiros testes na década de 80 até algoritmos que separam áreas queimadas de outros

fenômenos como corte seletivo de árvores. Além disso, uma breve compilação dos sensores voltados para a identificação de áreas queimadas está disponível em Lima (2013).

De maneira geral, duas abordagens são usadas para identificar áreas de florestas queimadas, alterações nas propriedades biofísicas do material carbonizado e a liberação de calor durante a queima. No primeiro caso, o processo de identificação de áreas queimadas é baseado principalmente nas regiões do espectro eletromagnético do visível ($\approx 0.4 - 0.7\mu\text{m}$), *near infrared* ($\approx 0.7 - 1.5\mu\text{m}$) e *mid-infrared* ($\approx 1.5 - 4\mu\text{m}$). Dessa maneira, diferentes combinações de canais (índices) podem ser aplicados visando ressaltar complexos fenômenos (BANNARI et al., 1995; KEY; BENSON, 2006). Alguns dos principais índices espectrais para detectar áreas queimadas são o *Burned Area Index* (BAI) (MARTÍN, 1998), e o *Normalized Burn Ratio* (NBR) (GARCIA; CASELLES, 1991; KEY; BENSON, 1999; KEY; BENSON, 2006). Entretanto, algoritmos que geram estimativas de áreas queimadas operacionalmente para toda a superfície terrestre podem apresentar incertezas (HAWBAKER et al., 2017). Conforme identificado por Padilla et al. (2015), erros de comissão para áreas queimadas variam de 42% a 94%, enquanto os de omissão variam de 68% a 93%, dependendo do sensor usado.

No Brasil, existem diferentes projetos que monitoram os fenômenos de queimadas e incêndios florestais nos diferentes biomas, uma vez que cada ambiente exige uma abordagem própria. Um exemplo é o Programa de Monitoramento de Queimadas - INPE, o qual detecta áreas queimadas em dados de média resolução espacial na região do Cerrado (Instituto Nacional de Pesquisas Espaciais, 2017c). Atualmente, pesquisas de áreas queimadas na floresta Amazônica são comumente desempenhadas com base nos produtos do MODIS⁴, e na aplicação do Modelo Linear de Mistura Espectral (MLME) (ANDERSON et al., 2005; SHIMABUKURO et al., 2009; LIMA et al., 2012; CARDOZO et al., 2013; ANDERE et al., 2015; ANDERSON et al., 2015; JUNIOR et al., 2017).

A bordo das plataformas Terra e Aqua, o MODIS abriu novas perspectivas, uma vez que apresenta componentes específicos para a detecção do fogo e o monitoramento em escala global (ANDERSON et al., 2005; ROY et al., 2008; CARDOZO et al., 2013), como produtos de detecção de anomalias termais e fogos ativos (GIGLIO, 2015). Levando em consideração a sua importância, esforços constantes são realizados objetivando o melhor desempenho destes produtos (GIGLIO et al., 2016). Embora apresente alta

⁴<https://reverb.echo.nasa.gov/>

resolução temporal (1 a 2 dias), a resolução espacial é de 250m a 1km, não é indicada para estudos que exijam detalhamento. Para esses casos, Key e Benson (2006) indicaram imagens de média resolução espacial, as quais variam de 10 à 50m (de acordo com Ehlers Ronald Janowsky (2002)), como o sensor OLI (Landsat 8) com 30m (WULDER et al., 2012).

Em dados de resolução espacial mais fina, existe o *problema de mistura* (ANDERSON et al., 2005; SHIMABUKURO et al., 2009), já que um pixel representa a média da resposta espectral de todos os elementos localizados naquele pixel. Nessa perspectiva, o MLME foi desenvolvido com objetivo de descrever a heterogeneidade no nível de subpixel (SHIMABUKURO; SMITH, 1991), ou seja, o MLME considera que o valor do pixel de uma imagem é a combinação linear da resposta de cada elemento dentro da área do pixel imageado. No caso de áreas florestais, o modelo considera a presença de três principais componentes dentro de cada pixel, copa das árvores, solo e sombra (Instituto Nacional de Pesquisas Espaciais, 2006). Assim, a base para a geração do modelo é proveniente da resposta espectral de alvos puros, também chamados de *endmembers* (SHIMABUKURO; SMITH, 1991; Instituto Nacional de Pesquisas Espaciais, 2006), selecionados por um especialista (ANDERSON et al., 2005).

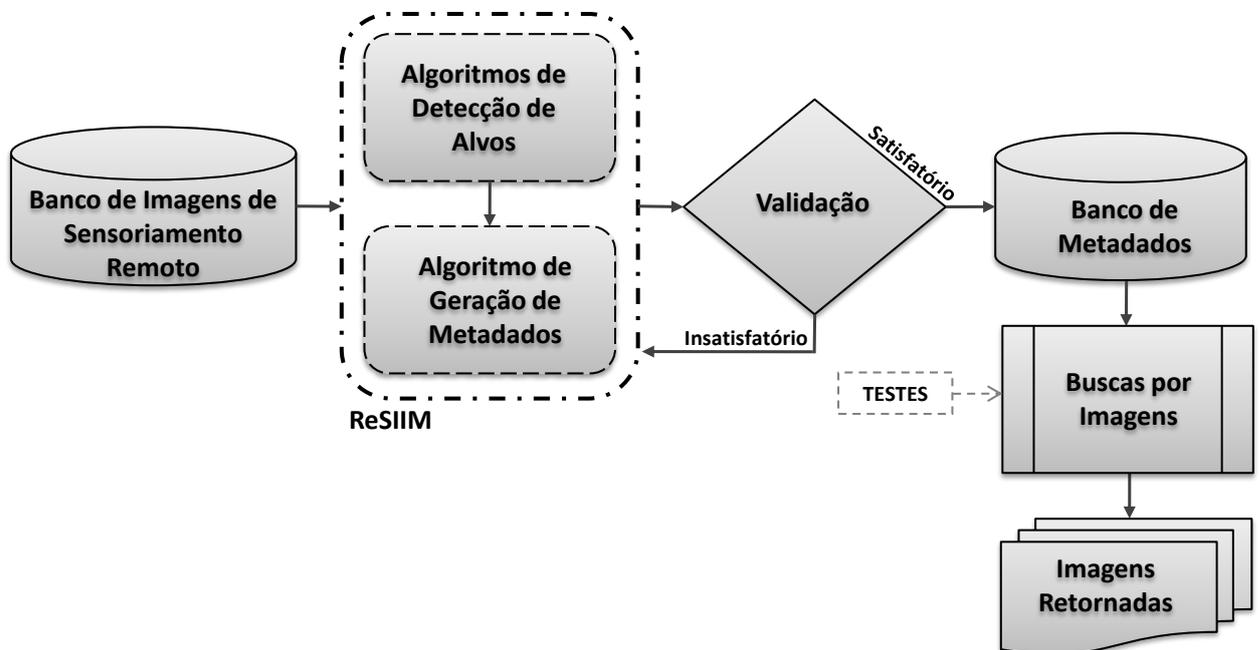
A aplicação do MLME apresenta resultados satisfatórios nos estudos de detecção e mapeamento de cicatrizes de áreas de queimadas (ANDERSON et al., 2005; LIMA et al., 2012; CARDOZO et al., 2013; ANDERE et al., 2015; JUNIOR et al., 2017). Entretanto, embora apenas a fração sombra derivada do MLME seja utilizada, uma vez que apresenta resposta espectral semelhante ao das queimadas, respostas espectrais semelhantes podem interferir nos resultados (ANDERE et al., 2015). Por exemplo, citam-se aqui água e sombra de nuvem, que exibem também baixa reflectância, conforme já reportado por Chuvieco e Congalton (1988). Além disso, a cobertura de nuvens e a fumaça provinda de queimadas podem omitir pixels com a resposta espectral afetada pelo fogo (ARAGÃO et al., 2016), enquanto que nuvem e sombra de nuvem influenciam negativamente alguns processos como correções atmosféricas e a classificação da cobertura da terra (ZHU; WOODCOCK, 2012). Assim, há etapas de filtragem necessárias tanto antes quanto depois da aplicação do MLME. Nos filtros posteriores são utilizadas máscaras de vegetação e de água, por meio de índices espectrais como por exemplo o NDVI e NDWI (*Normalized Difference Water Index*), respectivamente, e mapeamentos oficiais (ANDERE et al., 2015), além de máscaras geradas a partir do algoritmo *Fmask* (PLETSCH et al., 2018). Entretanto, a filtragem prévia das cenas apropriadas para a performance adequada do MLME ainda é manual em grandes catálogos de imagens. Isso ocorre, uma vez que essa seleção de

imagens baseia-se principalmente em sombra de nuvem, água e vegetação, indisponíveis atualmente como critérios nas ferramentas de busca por imagens em catálogos de imagens de Sensoriamento Remoto, representando assim uma lacuna de atuação.

3 METODOLOGIA

A estrutura da metodologia é apresentada na Figura 3.1. A partir da aplicação do protótipo de Mineração de Informações de Imagens de Sensoriamento Remoto (*Remote Sensing Image Information Mining - ReSIIM*) em um conjunto de imagens (Subseção 3.1), foi analisada a validação, ou seja, a correspondência dos metadados gerados com alvos reais por meio de dados de validação (Subseção 3.2). Caso os resultados fossem satisfatórios, com acertos acima de 90%, os metadados eram adicionados em um banco final. Nos casos em que os resultados fossem insatisfatórios, a etapa do ReSIIM com ajustes era reaplicada (Subseção 3.3). Ao final do processo, a fim de analisar a eficácia desta ferramenta na abordagem espectral, retorno de cenas adequadas para a aplicação do MLME, testes foram realizados na etapa de *Buscas por Imagens*. A abordagem utilizada para esta análise está descrita na Subseção 3.4 e os resultados estão apresentados na Subseção 4.3. A automatização das etapas ora citadas foi realizada com base em algoritmos escritos na linguagem de programação *Python* (Subseção 3.5).

Figura 3.1 - Estrutura da metodologia.



Fonte: Adaptada de Pletsch e Körting (2017).

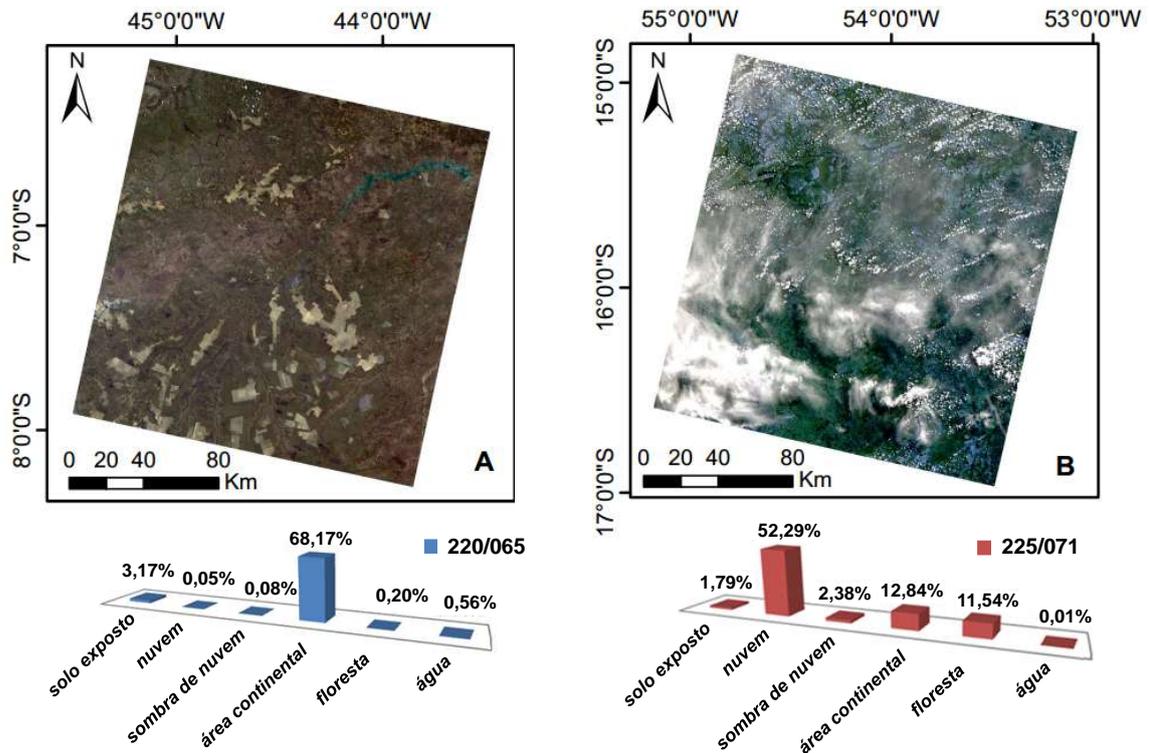
3.1 ReSIIM: Mineração de Informações de Imagens de Sensoriamento Remoto

O protótipo ReSIIM é constituído por dois principais conjuntos de algoritmos, um de detecção de alvos e outro de geração de metadados. A detecção de alvos consistiu em aplicar no conjunto de cenas os algoritmos *Fmask*, estruturado em Python por Flood e Gillingham (2017), e o índice NDVI. No segundo algoritmo, foram atribuídos aos dados gerados os metadados correspondentes. Assim, a partir do *Fmask*, foram identificados os alvos de *nuvem*, *sombra de nuvem*, *área continental* e *água*, conforme estrutura já definida no pacote utilizado, enquanto que o NDVI foi responsável por identificar os alvos de *floresta* e *solo exposto*. Os limiares utilizados no NDVI foram $\geq 0,8$ para *floresta* e o intervalo entre 0,0 e 0,2 para *solo exposto*, detalhados nas Subseções 3.2.3 e 3.2.4, respectivamente. O ReSIIM não classifica as áreas presentes nas cenas, ao invés disso, ele classifica os elementos presentes nas cenas. Sendo assim, a soma dos alvos identificados pelo ReSIIM não necessariamente será 100%. Além disso, os alvos de *floresta* e *solo exposto* deveriam ocorrer apenas nas áreas continentais das cenas, conforme já explicado na Subseção 2.3.3.

Devido à grande quantidade de dados analisados pelo ReSIIM, foi necessário desenvolver uma abordagem que não sobrecarregasse a memória computacional disponível. Dessa forma, o sistema abre o pacote de dados de determinada cena, aplica as etapas do ReSIIM e logo em seguida elimina todos os dados primários, permanecendo assim apenas os resultados provindos do ReSIIM.

Um exemplo de aplicação do ReSIIM em imagens de Sensoriamento Remoto pode ser analisado na Figura 3.2, onde A está localizada no bioma cerrado e B, na Amazônia brasileira. A Figura 3.2-A (A220/065 - 24/09/2013) apresenta baixo percentual de *nuvem* (0,05%), em contraste à Figura 3.2-B (225/071 - 10/04/2015), cujo percentual de *nuvem* chega a 52,30%. Entre ambas, a mais adequada para a aplicação do MLME em estudos de queimadas seria a Figura 3.2-A. Na imagem A, é notável a grande porcentagem do alvo *área continental*, embora porcentagens ainda superiores foram esperadas. Entretanto, as imagens brutas apresentam valores nulos no entorno da cena, e a porcentagem aqui apresentada é referente à cena na íntegra. A imagem B, embora repleta de *nuvem*, apresenta maior percentual de *floresta* do que a A, uma vez que este alvo foi definido de acordo com o conceito de *floresta* encontrado na Amazônia (Subseção 3.2.3).

Figura 3.2 - Exemplo da aplicação do ReSIIM.



A. Cena presente no Cerrado; B. Cena presente na Amazônia brasileira.

Fonte: Pletsch e Körting (2017).

3.2 Validação: Análise da correspondência dos metadados

Os metadados gerados foram avaliados e classificados como satisfatórios ou insatisfatórios. De forma a apresentar estas análises de maneira clara, conforme a disponibilidade de dados de validação, a análise da correspondência entre os metadados gerados e alvos reais foi subdividida em 4 etapas: i) *nuvem*, *sombra de nuvem* e *área continental* - AC (Subseção 3.2.1); ii) *água* (Subseção 3.2.2); iii) *floresta* (Subseção 3.2.3); e iv) *solo exposto* (Subseção 3.2.4). Quando necessário, foram utilizadas ao longo da Amazônia brasileira cenas aleatórias do sensor OLI (Landsat 8), cuja resolução espacial é de 30m.

3.2.1 *Nuvem, sombra de nuvem e área continental*

Para os alvos de *nuvem*, *sombra de nuvem* e *área continental*, (*clear land*), não foi possível encontrar na bibliografia dados de validação além daqueles também gerados com base no algoritmo *Fmask* (Subseção 2.3.3). Devido ao seu desempenho, o al-

goritmo *Fmask* é amplamente utilizado e foi inclusive adotado para ser aplicado na série Landsat (FOGA et al., 2017). Dessa forma, os dados foram diretamente alocados no banco de metadados final.

3.2.2 Água

A análise de correspondência para o alvo de *água* foi realizada com base nos dados desenvolvidos por Namikawa et al. (2016), detalhados na Subseção 2.3.3, com resolução espacial de 5m. A filtragem dos dados de referência foi realizada considerando a alta persistência deste alvo ao longo de 4 anos (NAMIKAWA et al., 2016). A análise foi baseada em duas abordagens: acurácia da correspondência (ACo), a qual representa o índice de acerto, e erro de comissão (EC), que são os dados erroneamente classificados como *água*. Para isso, 10 cenas aleatórias ao longo da Amazônia brasileira foram utilizadas. Os resultados da análise podem ser conferidos na Tabela 3.1.

ACo foi alta, variando entre 76% e 99%, e a sua média foi de 90%. Apenas na cenas 224/066 e 229/069 ACo foi abaixo de 80%, devido à diferença na resolução espacial dos dados. EC foi baixo, uma vez que em nenhum dos casos mais de 24% dos dados foram classificados erroneamente. A média do EC foi de 10%. Nesta análise não foram considerados erros de omissão, já que os dados de referência possuem resolução espacial de 5m e os utilizados para a aplicação do ReSIIM são de 30m, e essa diferença na resolução espacial poderia enviesar os erros de omissão.

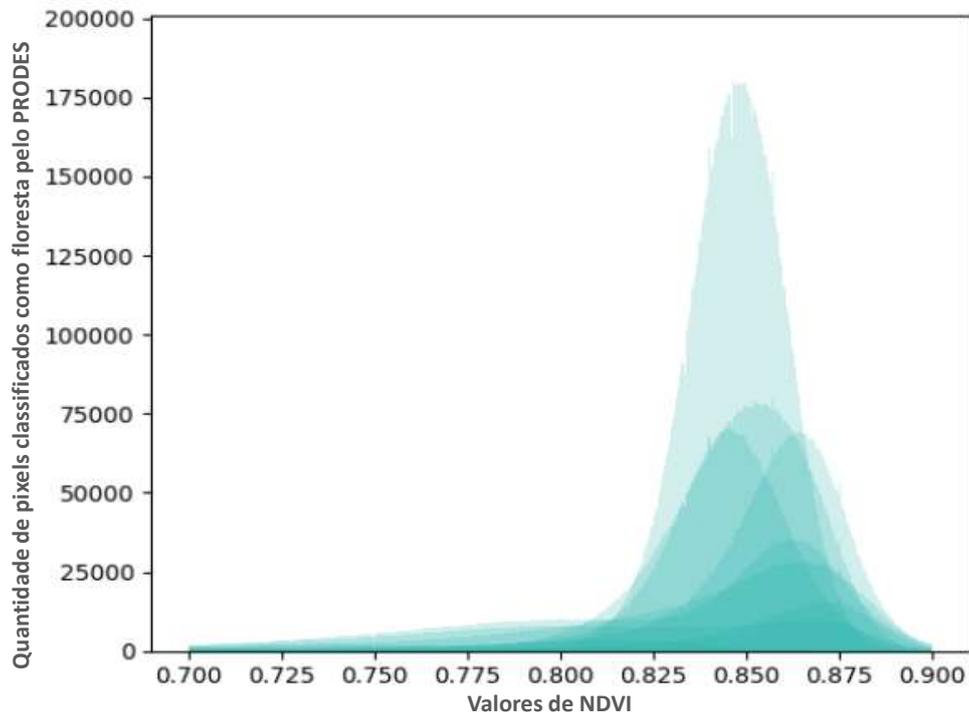
Tabela 3.1 - Análise da correspondência do alvo *água*. ACo: Acurácia da correspondência; EC: Erro de comissão.

| Órbita/ Ponto | Água Detectada (m^2) | ACo (%) | EC (%) |
|------------------|--------------------------|---------|--------|
| 002/066 | 83,18 | 84,48 | 15,52 |
| 224/065 | 16,51 | 89,30 | 10,7 |
| 224/066 | 9,30 | 78,47 | 21,53 |
| 228/067 | 177,82 | 99,86 | 0,14 |
| 228/068 | 16,29 | 91,85 | 8,15 |
| 229/061 | 631,86 | 98,27 | 1,73 |
| 229/069 | 22,45 | 76,13 | 23,87 |
| 232/063 | 2.042,42 | 98,63 | 1,37 |
| 232/064 | 336,50 | 98,54 | 1,46 |
| 232/067 | 18,30 | 84,15 | 15,85 |

3.2.3 Floresta

Levando em consideração os diferentes significados que a cobertura de terra *floresta* pode apresentar (CHAZDON et al., 2016), neste trabalho o limiar do NDVI (Subseção 2.3.3) foi definido conforme análise prévia na região da Amazônia brasileira. Para isso, foram utilizadas máscaras de floresta adquiridas pelo Projeto PRODES¹, o qual realiza o monitoramento por satélites do desmatamento por corte raso na Amazônia Legal (Instituto Nacional de Pesquisas Espaciais, 2017d). A máscara de floresta foi assim utilizada para verificar qual seria o limiar do NDVI adequado a ser aplicado no ReSIIM para os alvos de *floresta*, segundo valores da Amazônia brasileira. A distribuição dos valores de NDVI dos pixels é apresentada na Figura 3.3, onde quanto mais intensa a cor, maior é a recorrência de um determinado valor. Assim, de acordo com esta análise foi considerado o limiar de NDVI $\geq 0,8$, devido à sua alta correspondência com a densa vegetação, embora seja um limiar rígido.

Figura 3.3 - Variação dos valores de NDVI em áreas identificadas como floresta pelo PRODES.



Cada curva representa a dispersão dos valores do longo das áreas de floresta em uma cena. Quanto mais intensa é a cor, mais frequentemente esse valor de NDVI se repete.

Fonte: Elaborada pela autora.

¹<http://www.dpi.inpe.br/prodesdigital/>

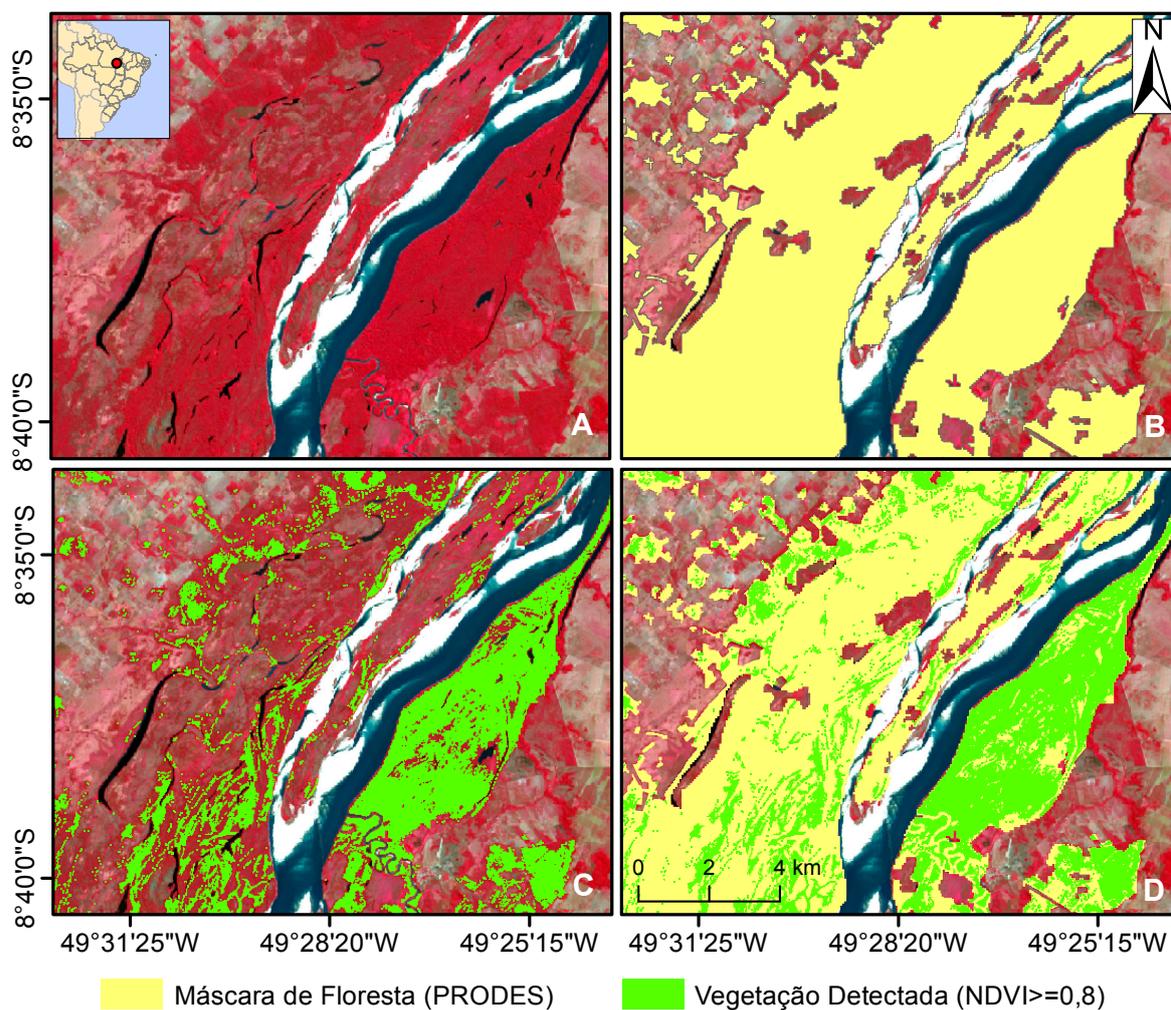
Para a análise da correspondência entre o metadado gerado e os alvos reais, foram aplicadas três principais abordagens: acurácia da correspondência (ACo), erro de comissão (EC) e erro de omissão (EOM). ACo representa o índice de acerto, e EC apresenta os dados erroneamente classificados como *floresta*. EOM se refere aos dados que deveriam ser classificados como floresta, mas não foram. Para esta análise, foram selecionadas 10 cenas aleatoriamente ao longo da Amazônia brasileira. As cenas, bem como os resultados são apresentados na Tabela 3.2.

Tabela 3.2 - Análise da correspondência do alvo *floresta*. ACo: Acurácia da correspondência; EC: Erro de comissão; EOM: Erro de omissão.

| Órbita/ Ponto | Floresta Detectada(m^2) | ACo(%) | EC(%) | EOM(m^2) |
|------------------|-----------------------------|--------|-------|--------------|
| 001/066 | 24.003,97 | 98,98 | 1,02 | 248,12 |
| 003/066 | 25.102,00 | 97,09 | 2,91 | 751,34 |
| 223/066 | 1.626,16 | 58,00 | 42 | 1.177,79 |
| 226/066 | 23.555,17 | 98,19 | 1,81 | 433,68 |
| 227/066 | 20.453,93 | 97,96 | 2,04 | 426,46 |
| 227/067 | 8.995,53 | 94,22 | 5,78 | 551,78 |
| 228/064 | 25.408,61 | 98,41 | 1,59 | 410,10 |
| 230/067 | 23.521,12 | 97,15 | 2,85 | 689,98 |
| 231/065 | 24.065,44 | 98,33 | 1,67 | 409,53 |
| 232/067 | 10.282,20 | 92,03 | 7,97 | 891,05 |

A média da ACo foi de 93% e a do EC foi de apenas 7%. Entretanto, é importante destacar que embora ACo foi superior a 92% em cada uma das cenas, apenas uma cena apresentou exceção, cuja ACo foi de 58%, EC, 42% e EOM superior a 1.000 m^2 . Esta cena está localizada entre os estados do Pará e Tocantins (órbita/ponto 223/066), e a discrepância foi causada devido à grande extensão de área de várzea, demonstrando que embora o limiar do NDVI tenha sido adequado para grande parte da Floresta Amazônica, não compreendeu esse tipo de região, devido à sua resposta espectral. A Figura 3.4 apresenta esse caso especial em quatro etapas: A - dado original; B - máscara de floresta segundo o PRODES; C - área de *floresta* de acordo com o ReSIIM ($NDVI \geq 0,8$); e D - composição das três camadas ora citadas. Ao longo destas áreas inundáveis, o valor do NDVI variou principalmente entre 0,4 e 0,7.

Figura 3.4 - Cena com baixo índice de acerto (ACo) para o limiar de *floresta*. A - Cena original (composição usando bandas do Landsat 8 - R5G4B2); B - Máscara de floresta segundo o PRODES; C - Área de *floresta* detectada (NDVI $\geq 0,8$); e D - Cena original, máscara do PRODES e *floresta* detectada.

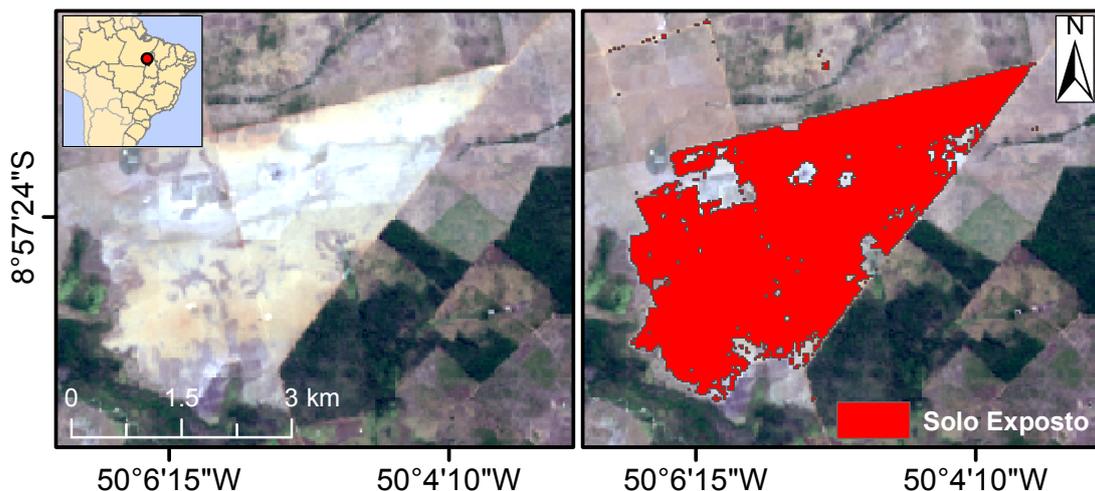


Fonte: Elaborada pela autora.

3.2.4 Solo exposto

Para o alvo de *solo exposto* não foi possível identificar dados de validação. Assim como destaca Liang et al. (2014b), processos de validação nesta área de pesquisa são restritos devido às fontes escassas de dados e tecnologias. Embora valores de NDVI baixos não representem necessariamente a falta de vegetação, uma vez que diferenças ao longo das estações e dos anos podem interferir nos valores de NDVI em áreas vegetadas (NOAA, 2017), foram levados em consideração os limiares de NDVI encontrados na literatura para este tipo de alvo (Subseção 2.3.3). Assim, para a aplicação do ReSIIM, o intervalo utilizado para o alvo de *solo exposto* foi $0,0 \leq \text{NDVI} \leq 0,2$. Um exemplo desta correspondência pode ser analisado na cena órbita/ponto 223/066 (Figura 3.5).

Figura 3.5 - Área com presença de *solo exposto* (composição usando bandas do Landsat 8 - R4G3B2).



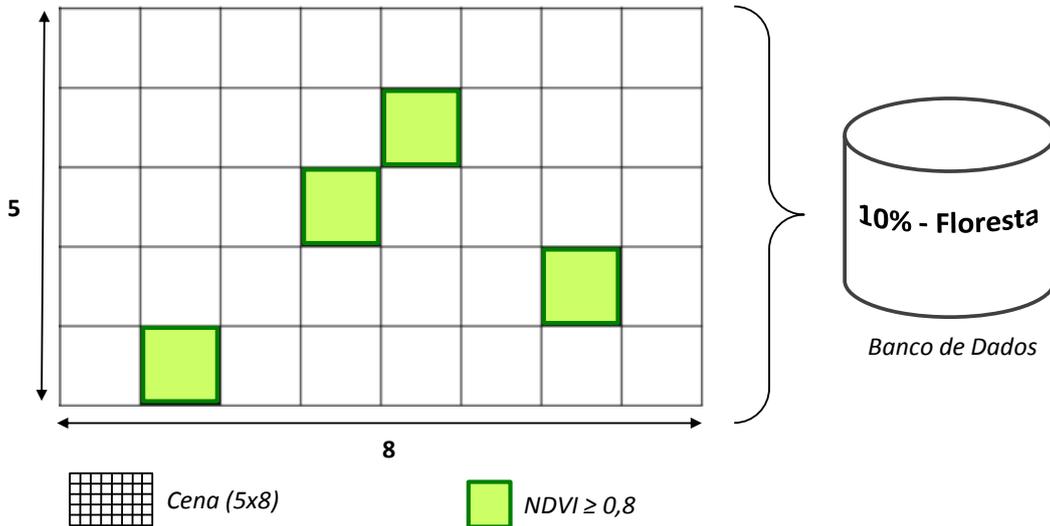
Fonte: Elaborada pela autora.

3.3 Banco de metadados

O banco de metadados de imagens de Sensoriamento Remoto foi idealizado para que o usuário tenha a liberdade de escolher os critérios desejados na etapa de Buscas por Imagens, cujos exemplos de aplicação estão descritos no Subseção 4. Uma ilustração didática de como o banco foi estruturado pode ser analisada na Figura 3.6. Em uma cena (5x8) com 4 pixels classificados pelo ReSIIM como *floresta* ($\text{NDVI} \geq 0,8$, conforme apresentado na Subseção 3.2.3), é armazenado no banco de metadados a

informação de que 10% desta cena é constituída pelo alvo *floresta*. A estrutura do banco de metadados em Python está disponível na Subseção 3.5.

Figura 3.6 - Exemplo do processo de armazenamento no banco de metadados.



Fonte: Elaborada pela autora.

3.4 Buscas por imagens: Avaliação da acurácia do ReSIIM

Considerando que cada usuário possui uma expectativa, não há uma única resposta para a avaliação de um sistema de busca por imagens (KOSKELA et al., 2000). Apesar disso, há abordagens que buscam quantificar a acurácia de sistemas, como a abordagem de Veganzones et al. (2008), a qual envolve três medidas principais: precisão (p), *recall* (r) e erro (e), dentro de um Banco de Metadados (BM). Partindo do pressuposto de que nos resultados de uma busca (A) há tanto imagens relevantes (R) quanto irrelevantes (\bar{R}), a precisão se refere à fração do conjunto de imagens retornadas pelo sistema e que são importantes para a busca (Equação 3.1). O *recall* é definido como a fração de imagens relevantes não retornadas pela busca (Equação 3.2), enquanto que o erro é uma terceira medida que se refere à fração do conjunto de imagens retornadas pelo sistema que são irrelevantes para o usuário (Equação 3.3). Para melhor compreender essa abordagem, foi desenvolvida a Figura 3.7 para exemplificar esses diferentes elementos.

$$p = \frac{|A \cap R|}{|A|} \quad (3.1)$$

$$r = \frac{|R \cap A|}{|R|} \quad (3.2)$$

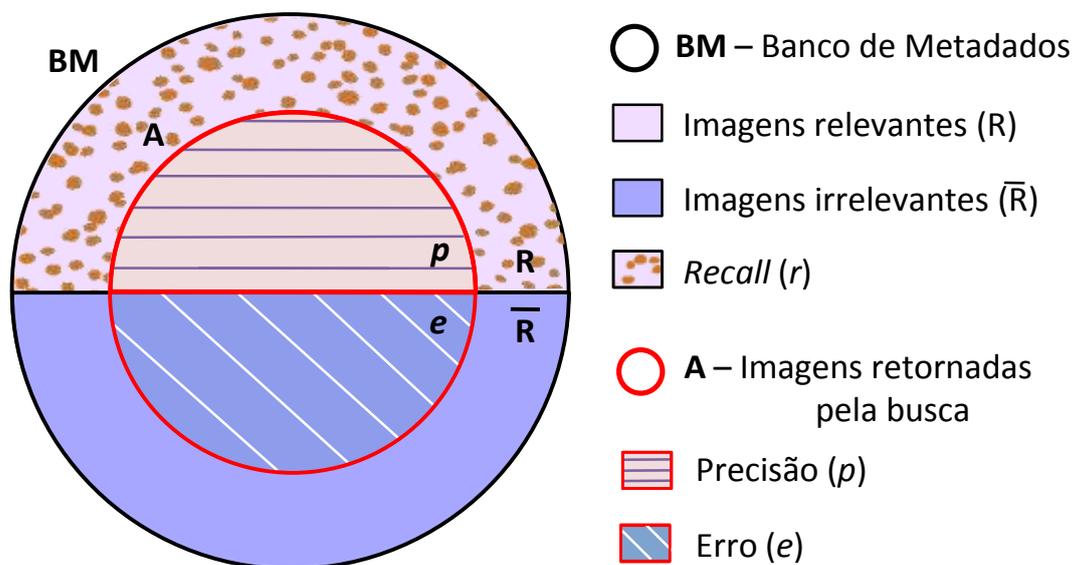
$$e = \frac{|A \cap \bar{R}|}{|A|} \quad (3.3)$$

onde:

- p = precisão;
- r = *recall*;
- e = erro;
- A = imagens retornadas pela busca no Banco de Metadados (BM);
- R = conjunto de imagens relevantes para a busca;
- \bar{R} = conjunto de imagens irrelevantes para a busca.

A precisão pode prejudicar a medida do *recall*, uma vez que para aumentar a precisão há a necessidade de realizar novos *recall*, assim, objetiva-se atingir um equilíbrio entre ambas (LONG et al., 2003). Desta forma, o ideal seria uma alta precisão em uma única busca, retornando todo o conjunto de imagens relevantes, sem a necessidade de realizar *recall*. Embora a avaliação da metodologia ainda seja uma questão em discussão na literatura, grande parte das avaliações utilizadas baseiam-se na busca de informações (do inglês: *Information Retrieval* - IR) (VEGANZONES et al., 2008).

Figura 3.7 - Avaliação da acurácia do ReSIIM.



Fonte: Elaborada pela autora.

3.5 A automatização em *Python*

A linguagem de programação *Python* foi escolhida por ser amplamente difundida na comunidade acadêmica, garantindo assim a característica de extensibilidade ao ReSIIM. Essa linguagem pode agir ainda como uma ferramenta interativa para ligar os diferentes componentes (HUANG et al., 2012), além de oferecer um ambiente com atributos únicos que se apresentam como uma escolha para cientistas a procura de uma linguagem de alto nível para aplicações específicas (OLIPHANT, 2007).

Os principais módulos utilizados no ReSIIM foram o *NumPy*, responsável pelos cálculos de matrizes, e o *GDAL*, que lida com os formatos de imagens de Sensoriamento Remoto. O banco de metadados foi criado com base em *dicionários* no Python, um tipo de objeto que coleciona outros objetos. Entre as características dos dicionários, cita-se aqui a flexibilidade de incluir ou eliminar itens, e a possibilidade de ser integrado a outros tipos de objetos no Python (LUTZ, 2013). Um exemplo da estrutura do armazenamento no banco de metadados pode ser observado na Figura 3.8.

Figura 3.9 - Exemplo da busca combinada dos alvos *nuvem* e AC no ReSIIM.

```
1 import numpy as np
2
3 lista_cenas_nuvem = []
4 lista_cenas_AC = []
5
6 for cena in banco_metadados.items():
7     if 'porcentagem_nuvem' <= 20:
8         lista_cenas_nuvem.append(cena)
9     if 'porcentagem_AC' >= 40:
10        lista_cenas_AC.append(cena)
```

Linhas: 1. Importação do módulo *NumPy*; 3-4. Criação de listas de cenas com as porcentagens desejadas; 6-8. Para cada cena dentro do banco de metadados, se o percentual de *nuvem* for igual ou inferior a 20%, adicione na lista. O mesmo se repete para o alvo área continental (AC) nas linhas 09 e 10.

Fonte: Elaborada pela autora.

Figura 3.8 - Exemplo da estrutura de armazenamento no banco de metadados em Python.

```
1 {'orbita/ponto_AnoMesDia':
2  {
3   'porcentagem_agua': {1.10},
4   'porcentagem_nuvem': {69.46},
5   'porcentagem_solo_exposto': {23.19},
6   'porcentagem_AC': {0.0},
7   'porcentagem_veg': {0.0},
8   'porcentagem_sombra': {6.25}}
9 }
```

Cena (órbita/ponto) e data de aquisição (linha 1), alvos e as respectivas porcentagens (linhas 3-8).

Fonte: Elaborada pela autora.

Devido a estas características, os parâmetros de buscas por imagens podem ser inclusive combinados entre dois ou mais alvos. Um exemplo da estrutura em Python da busca combinada dos alvos *nuvem* e *área continental* (AC) no ReSIIM é apresentado na Figura 3.9. Após adicionar as cenas correspondentes nas listas, linhas 8 e 10 (Figura 3.9), é necessário ainda intersectar (\cap) ambas para obter os resultados desta busca.

4 RESULTADOS E DISCUSSÕES

Esta Seção apresenta o resultado da aplicação do ReSIIM em três perspectivas: i) espacial - ao longo do Brasil no período de seca na Amazônia brasileira (Subseção 4.1); ii) temporal - uma cena ao longo de 1 ano (Subseção 4.2); iii) espectral - como suporte à aplicação do Modelo Linear de Mistura Espectral (MLME), principal abordagem em estudos de cicatrizes de queimadas na Amazônia brasileira (Subseção 4.3).

4.1 Espacial: Aplicação do ReSIIM ao longo do Brasil

Visando apresentar uma aplicação do ReSIIM em larga escala, coletamos dados de todo o Brasil, que é um país de dimensões continentais, para o qual são necessárias pelo menos 375 cenas de resolução *Landsat* para seu mapeamento completo. Em um de nossos estudos de caso, optamos pela época da seca na Amazônia, período entre Agosto e Setembro de 2017 (ARAGÃO et al., 2007). Totalizando quase 13 milhões de km² armazenados em 0,75TB, o processamento durou cerca de três dias. A espacialização dos dados pode ser visualizada na Figura 4.1. Visando explorar o potencial do ReSIIM, buscas combinadas de alvos foram realizadas de acordo com questões que surgiram conforme a análise foi desenvolvida.

As Figuras 4.1-A e 4.1-B são correlacionadas, uma vez que nuvens geralmente são acompanhadas de sombra. As cenas com a maior presença destes elementos são encontradas principalmente no norte e nordeste do país e ao longo da costa brasileira.

Ab'Sáber (2003) entende por domínio morfoclimático e fitogeográfico um conjunto espacial representado por feições do relevo, tipos de solos e formas de vegetação e condições climático-hidrológicas coerentemente esquematizadas. Entre os 6 domínios paisagísticos brasileiros, o Domínio das Terras Baixas Florestadas da Amazônia encontra-se na região norte do país, clima Equatorial (Instituto Brasileiro de Geografia e Estatística, 2002), sendo úmida e geralmente encoberta por nuvens baixas (AB'SÁBER, 2003; THÉRY; MELLO, 2005). Ao longo da costa do Brasil com alta presença do alvo *nuvem*, é encontrado principalmente o Domínio dos Mares de Morros (sudeste e litoral), faixas de transição, parte do Domínio das Araucárias (sul do país), e parte do Domínio das Caatingas (nordeste brasileiro). O Domínio dos Mares de Morros (Serra do Mar, São Paulo) é considerado úmido, enquanto que Domínio das Araucárias, hiperúmido. O Domínio das Caatingas é seco, com 350 a 600 mm anuais (AB'SÁBER, 2003; THÉRY; MELLO, 2005; ROSS, 2006). Entretanto, a distribuição anual de chuvas é atípica no litoral oriental e trecho setentrional do Nordeste, apresentando picos entre março e agosto (CONTI; FURLAN, 2011). A cobertura de nuvens na região é

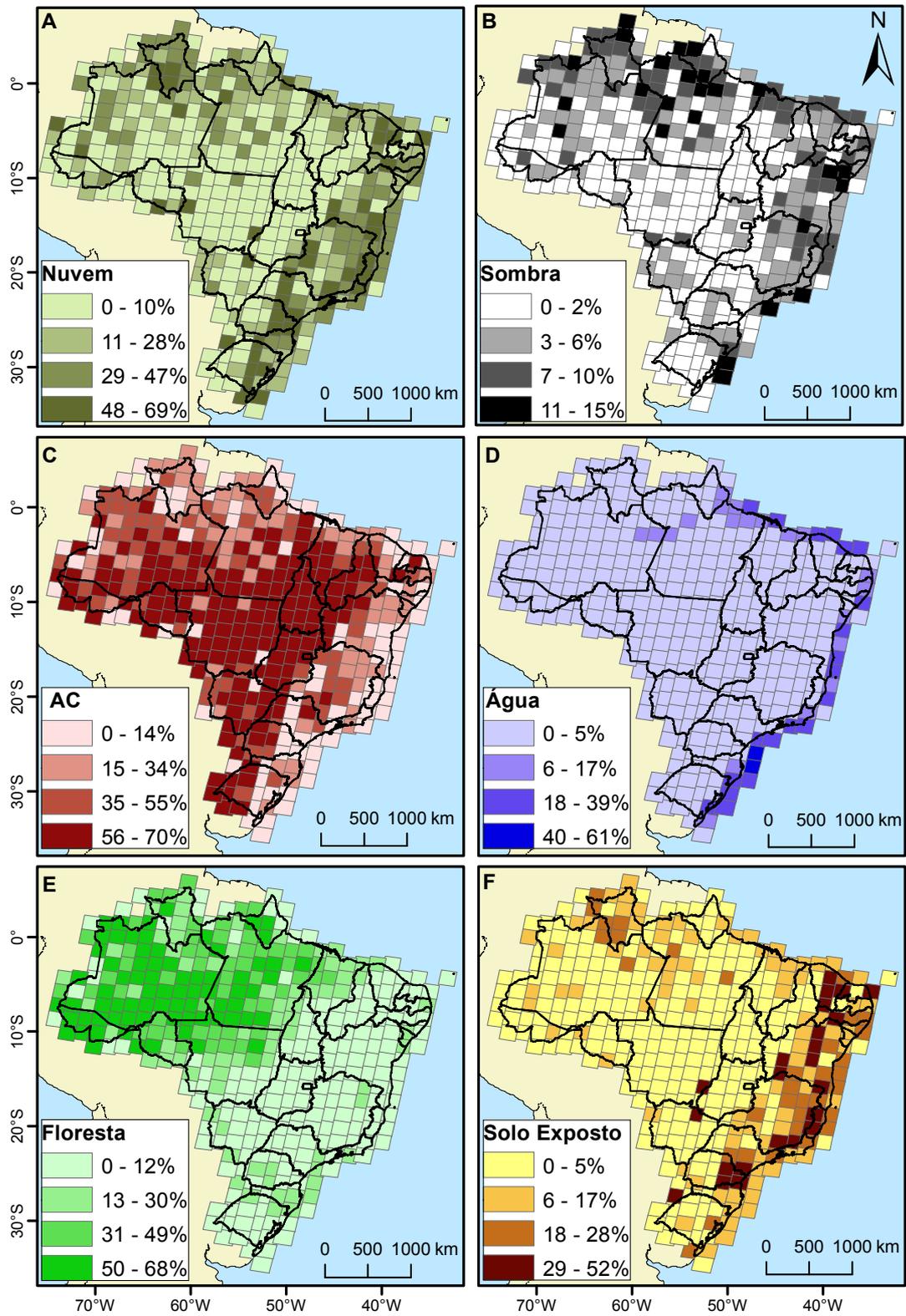
ainda localizada e condicionada por fatores como relevo, umidade e vento (SILVEIRA et al., 2017).

Sendo AC as *áreas continentais*, a Figura 4.1-C indica que sua ocorrência é inversamente proporcional aos alvos *nuvem*, *sombra* e *água*. Embora seja um dado abstrato, esta informação pode suportar a identificação de outros alvos com o desenvolvimento e aperfeiçoamento do ReSIIM. A Figura 4.1-D apresenta as cenas com maior presença de *água* ao longo da costa, assim como A e B. Entretanto, a maior parte das cenas nas bordas do país encontram-se já no Oceano Atlântico, ou seja, não são representativas com relação à quantidade de águas interiores. Algumas cenas com maior porcentagem de *água* também são encontradas no norte do país referentes à drenagem do Rio Amazonas.

A Figura 4.1-E apresenta a espacialização do alvo de *floresta*, sendo encontrado predominantemente no bioma Amazônia (Instituto Brasileiro de Geografia e Estatística, 2004a), devido ao conceito de *floresta* utilizado no ReSIIM. A floresta encontrada na bacia Amazônica é considerada uma floresta tropical úmida, também encontrada no Caribe e na Mata Atlântica no continente americano (CONTI; FURLAN, 2011). Assim, algumas cenas dispersas com 13 a 30% dos pixels classificados como *floresta* podem ainda ser encontradas ao longo do bioma Mata Atlântica, com destaque ao estado de Santa Catarina. Formando uma espécie de telhado verde, as copas das árvores destas florestas se interligam, formando um dossel contínuo (CONTI; FURLAN, 2011). Além disso, ressalta-se aqui também algumas áreas com presença do alvo *floresta* no bioma Pantanal. Entre outras razões, o monitoramento deste alvo é relevante já que o fenômeno da substituição da cobertura florestal por pastagens ou vegetação de campo aberto contribui para a ação de processos erosivos superficiais devido à exposição do solo (MOURA, 1995). Na Amazônia Legal, uma questão é levantada: quais os resultados do ReSIIM frente a uma cena que apresenta mudanças significativas de uso e cobertura da terra? Para responder a essa questão, uma comparação entre os resultados adquiridos e dados do Projeto PRODES está disponível na Seção 4.1.1.

Por fim, a Figura 4.1-F apresenta os metadados do alvo de *solo exposto*. Há maior concentração deste alvo no extremo norte da Amazônia, com destaque ao norte de Roraima, e ao longo dos estados do Ceará, centro e sudoeste da Bahia, sudeste de Minas Gerais, sul de São Paulo e leste do Paraná, compreendendo assim os biomas Amazônia, Cerrado, Caatinga, Mata Atlântica. No norte e leste de Roraima, há a presença de savanas (cerrado), com vegetação gramínea adaptada ao clima

Figura 4.1 - Espacialização dos metadados gerados pelo ReSIIM no Brasil para os meses de Agosto e Setembro de 2017.



A - Nuvem; B - Sombra de Nuvem; C - AC: Área Continental; D - Água; E - Floresta; F - Solo Exposto.

Fonte: Elaborada pela autora.

seco (DANTAS; RODRIGUES, 1982; Instituto Brasileiro de Geografia e Estatística, 2004b; Instituto Brasileiro de Geografia e Estatística, 2012). Apesar do ReSIIM corroborar esse dado demonstrando que essa área possui entre 6 e 17% de *solo exposto*, ao aplicar os limiares para a delimitação deste alvo no desenvolvimento do ReSIIM, houveram algumas confusões principalmente com o alvo de *nuvem*. Como essa região é considerada de clima hiperúmido (THÉRY; MELLO, 2005), e com grande presença de nuvens ao longo do ano, houve a necessidade de verificar a dependência deste alvo com *nuvem* por meio da intersecção dos dois grupos em *buscas combinadas*. O detalhamento deste tópico, bem como os principais resultados está descrito na Seção 4.1.2. Entre outros motivos, a análise deste tipo de alvo é importante, já que poderia nortear a gestão de limites e ameaças, como a desertificação no Nordeste interior, ocasionada entre outros fatores pelo uso inadequado dos recursos naturais (FEARNSIDE, 1979; SANTOS; CÂMARA, 2002; THÉRY; MELLO, 2005).

4.1.1 Aplicação do ReSIIM em imagens com mudanças de cobertura da terra

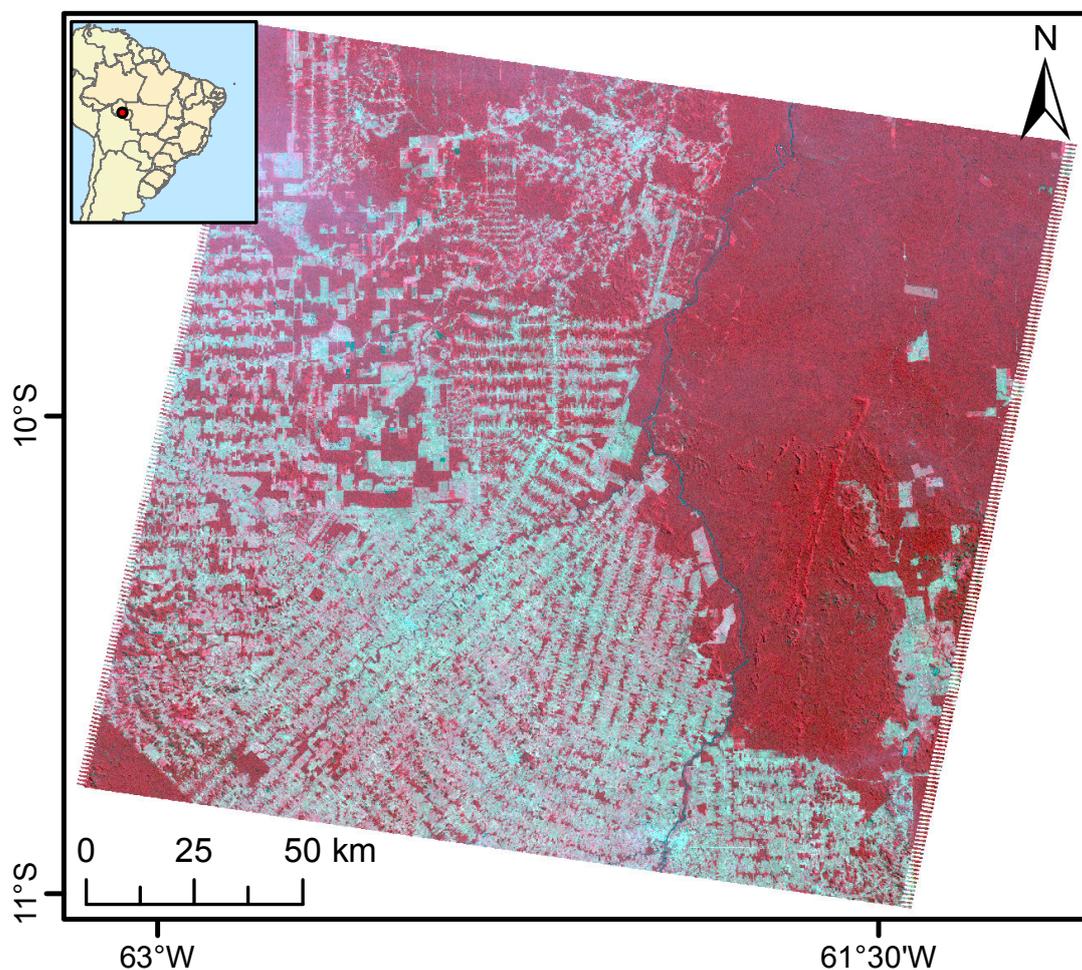
O ReSIIM foi delineado para que seja analisado pontualmente no tempo e no espaço, com a possibilidade de analisar cada cena individualmente e independentemente de seu histórico ou entorno. No entanto, os resultados estáticos podem ser combinados por um especialista para realizar análises mais detalhadas de mudança da cobertura da terra.

Para responder essa questão, o ReSIIM foi aplicado na cena órbita/ponto 231/067 (26/06/2004), Landsat 5 (Figura 4.2). Esta cena foi selecionada por ser a mesma utilizada no Projeto PRODES para o monitoramento do ano de 2004 (Instituto Nacional de Pesquisas Espaciais, 2017d), e foi escolhida por pertencer à porção centro-norte de Rondônia, sul da floresta Amazônica, composta por projetos de colonização do INCRA, e que apresenta mudanças significativas de uso e cobertura da terra entre 1985 e 2000 (ESCADA, 2003). Este ano específico foi selecionado, porque apresenta uma das maiores estimativas anuais de desmatamento (km²) na Amazônia Legal Brasileira. Vale ressaltar ainda que apenas áreas superiores a 6,25 ha são identificadas pelo PRODES como desmatamento por corte raso ¹ (CÂMARA et al., 2006).

Os resultados do ReSIIM estão apresentados na Tabela 4.1. Embora a cena apresente visualmente grandes áreas sem a presença de tons avermelhados, o qual seria um indicativo de cobertura de vegetação densa devido à sua composição RGB, o

¹Desmatamento por corte raso significa a remoção completa da cobertura florestal em um curto intervalo de tempo (Instituto Nacional de Pesquisas Espaciais, 2013)

Figura 4.2 - Cena 231/067 (26/06/2004), Landsat 5.



Composição colorida/infravermelho - R4G3B2, com destaque em vermelho à vegetação densa.

Fonte: Elaborada pela autora.

ReSIIM retornou menos de 1% de pixels com *solo exposto*. Essas áreas já desmatadas apresentam em sua grande parte NDVI entre 0,2 e 0,8, possivelmente devido à presença de agricultura e vegetação rasteira, definições ainda não contempladas na estrutura de detecção do ReSIIM. Além disso, o ReSIIM retornou ainda quase 10% dos pixels com *floresta*. Entretanto, não é possível prever quais áreas não foram caracterizadas com este alvo, uma vez que o ReSIIM não fornece esta informação. Uma pequena área de água também foi identificada pelo ReSIIM, representando quase 1% dos pixels. Os demais alvos retornados pelo ReSIIM não foram expressivos (menos de 2%).

Tabela 4.1 - Resultados do ReSIIM para a cena 231/067 (26/06/2004), Landsat 5.

| Alvos | Porcentagem |
|------------------------|-------------|
| <i>nuvem</i> | 0,18 |
| <i>sombra de nuvem</i> | 0,01 |
| <i>AC</i> | 69,31 |
| <i>água</i> | 0,09 |
| <i>floresta</i> | 9,60 |
| <i>solo exposto</i> | 0,08 |

Vale ressaltar que de acordo com o PRODES, a taxa estimada de corte raso no período de agosto de 2016 a julho de 2017 foi de 6.624 km² (Instituto Nacional de Pesquisas Espaciais, 2017b), corroborando a importância deste tipo de estudo.

4.1.2 O conceito de *solo exposto* precisa de reajustes?

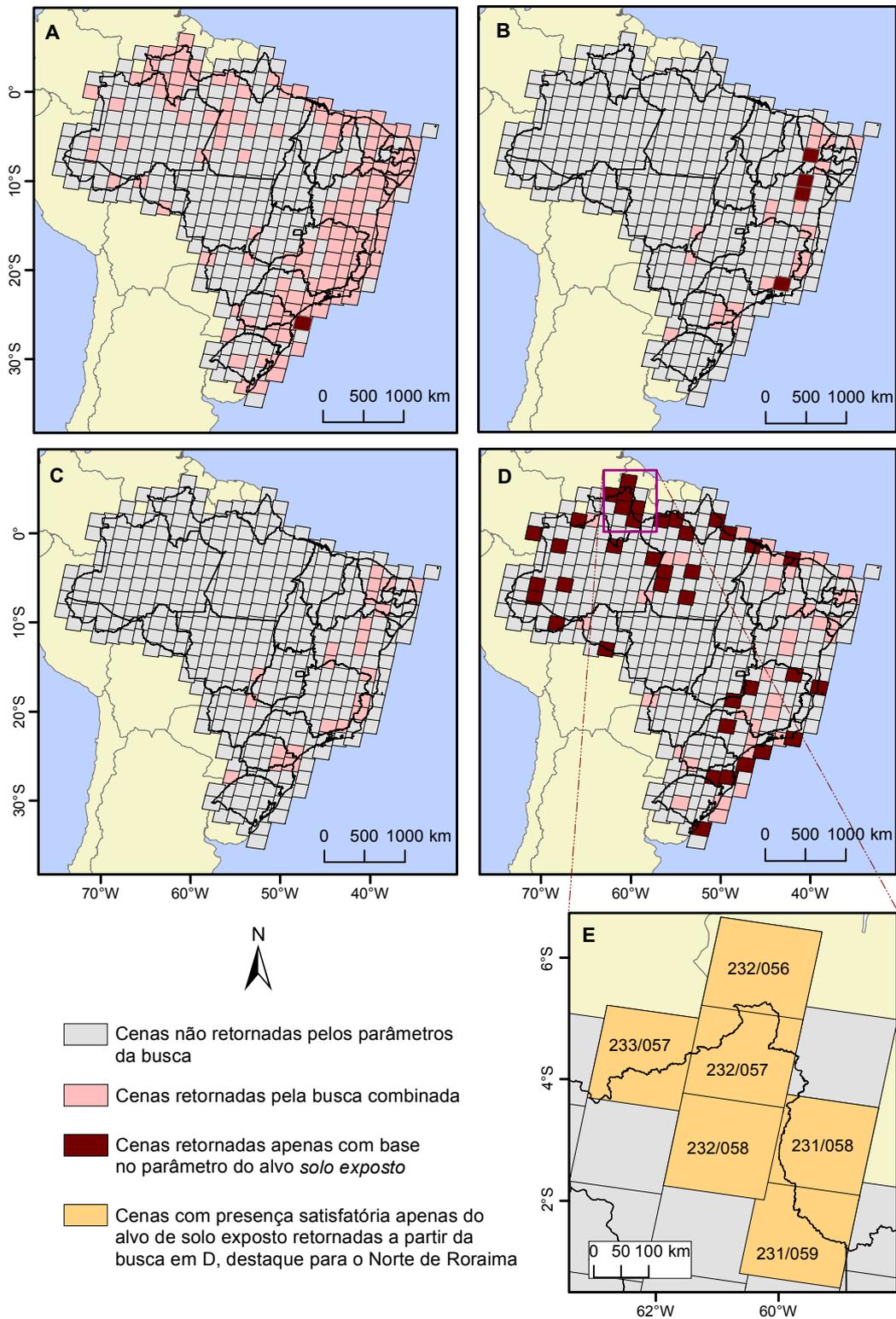
Para responder se o conceito de *solo exposto* aplicado no ReSIIM precisa de reajustes, algumas buscas foram realizadas com base nos alvos de *nuvem* e *solo exposto* no banco de imagens. Assim, foi analisado o retorno das cenas, bem como coeficiente de correlação de Pearson, o qual varia de +1,0 e -1,0, indicando respectivamente máxima e mínima relação entre conjuntos de dados (MITCHEL, 2005).

Os parâmetros de busca foram baseados principalmente nas legendas das Figuras 4.1-A e F, uma vez que foram formulados para ressaltar diferenças entre cada alvo. Um resumo dos parâmetros de busca, bem como os resultados podem ser analisados na Tabela 4.2. A espacialização do dados está disponível na Figura 4.3. Para todo o conjunto de cenas, assumindo que possui uma distribuição normal, o coeficiente de correlação de Pearson é de 0,8.

Tabela 4.2 - Análise da relação entre os alvos *solo exposto* e *nuvem*.

| Figura 4.3 | Parâmetro de Busca (%) | Número de cenas retornadas | | | |
|------------|------------------------|----------------------------|---------------------|--------------|--|
| | <i>solo exposto</i> | <i>nuvem</i> | <i>solo exposto</i> | <i>nuvem</i> | <i>solo exposto</i> \cap <i>nuvem</i> (busca combinada) |
| A | ≥ 6 | ≥ 11 | 137 | 202 | 136 |
| B | ≥ 29 | ≥ 48 | 24 | 47 | 20 |
| C | ≥ 29 | ≥ 29 | 24 | 116 | 24 |
| D | [6;17] | [11;18] | 77 | 86 | 38 |

Figura 4.3 - Espacialização dos resultados da busca combinada dos alvos *solo exposto* e *nuvem*.



A - *solo exposto* $\geq 6\%$ e *nuvem* $\geq 11\%$; B - *solo exposto* $\geq 29\%$ e *nuvem* $\geq 48\%$; C - *solo exposto* $\geq 29\%$ e *nuvem* $\geq 29\%$; D - *solo exposto* entre [6,17] e *nuvem* entre [11,28]; E - Cenas com presença satisfatória apenas do alvo de *solo exposto* retornadas a partir da busca em D, destaque para o Norte de Roraima.

Em um primeiro momento, foi analisada a distribuição das cenas com porcentagem de alvos relevantes, igual ou acima de 6% no caso do *solo exposto* e igual ou acima de 11% para *nuvem*. Essa busca retornou 137 cenas no caso de *solo exposto*, 202 no de *nuvem* e 136 intersecções (\cap) entre ambos os conjuntos (Figura 4.3-A), sendo que a única cena do *solo exposto* não retornada encontra-se na maior parte do Oceano Atlântico (órbita/ponto 219/078). Assim, todas as cenas com maior representatividade do alvo de *solo exposto* dentro destes parâmetros foram retornadas. O coeficiente de Pearson para os dados intersecçados foi de 0,73, indicando uma alta correlação entre os dados. Além disso, é notável que diversas cenas do alvo *solo exposto* acompanham o crescimento da quantidade de pixels com o alvo *nuvem*, conforme pode ser verificado na Figura 4.4.

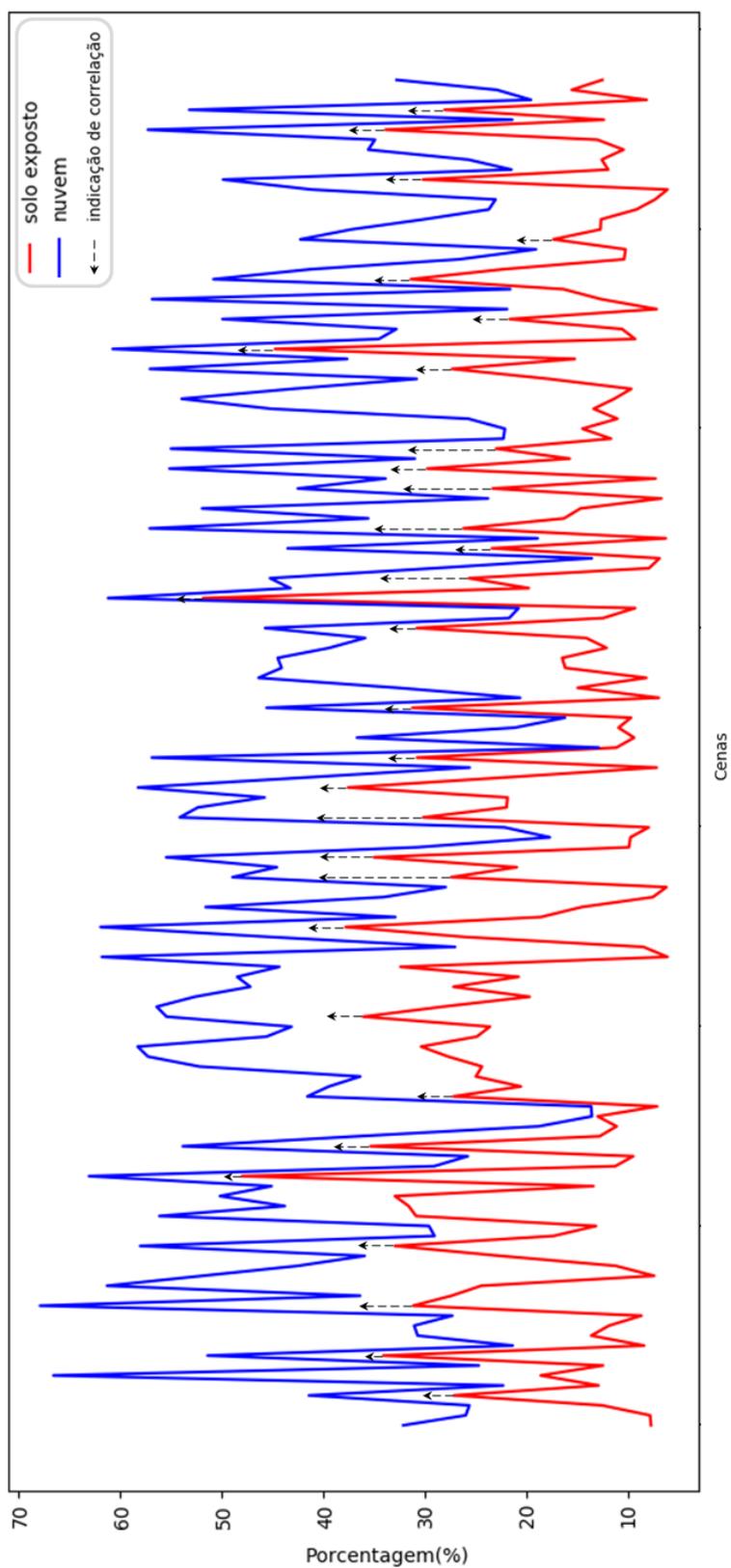
Buscando delimitar os resultados, apenas as cenas com a maior porcentagem de alvos foram selecionadas, sendo igual ou acima de 29% para o alvo de *solo exposto* e 48% para *nuvem*. Desta busca, 24 cenas de *solo exposto*, 47 de *nuvem* e 20 cenas de intersecção foram retornadas. As cenas de *solo exposto* não retornadas foram 217/075, 217/067, 217/065, 217/068 (Figura 4.3-B). A coeficiente de Pearson para os dados intersecçados foi de 0,46 e uma breve análise visual é possível ser realizada na Figura 4.5, denotando uma baixa correlação entre os dados obtidos.

As cenas com 29% ou mais do alvo *solo exposto* totalizaram 24, enquanto que a soma das cenas com 29% ou mais do alvo *nuvem* foi 116. Da busca combinada entre ambos, todas as cenas do alvo *solo exposto* foram retornadas (Figura 4.3-C), e o coeficiente de Pearson foi 0,48 para essa intersecção, sendo levemente superior à correlação do conjunto de dados obtidos na busca anterior. A Figura 4.6 apresenta as porcentagens dos alvos das cenas retornadas e suas possíveis correlações.

Em seguida, as porcentagens nos parâmetros da busca foram ajustadas de acordo com os intervalos do segundo grupo mais recorrente dos alvos, [6;17] de *solo exposto* e [11;18] de *nuvem*, buscando verificar alguma dependência do alvo *solo exposto* dentro destas proporções. O retorno das imagens foi de 77 para *solo exposto*, 86 para o alvo de *nuvem* e intersecção de 38 cenas (Figura 4.3-D). Dentre as buscas, esta foi a com menor correlação, com coeficiente de Pearson de -0,01 e sem correlações visuais (Figura 4.7).

Entre as cenas não retornadas pela intersecção destes intervalos de porcentagem, ressaltam-se aqui as localizadas no norte e leste de Roraima (Figura 4.3-E), já que o retângulo envolvente abrange região de savana (cerrado), também conhecida como *lavrado* ou *campos do Rio Branco*, a qual é encoberta por vegetação de gramínea e

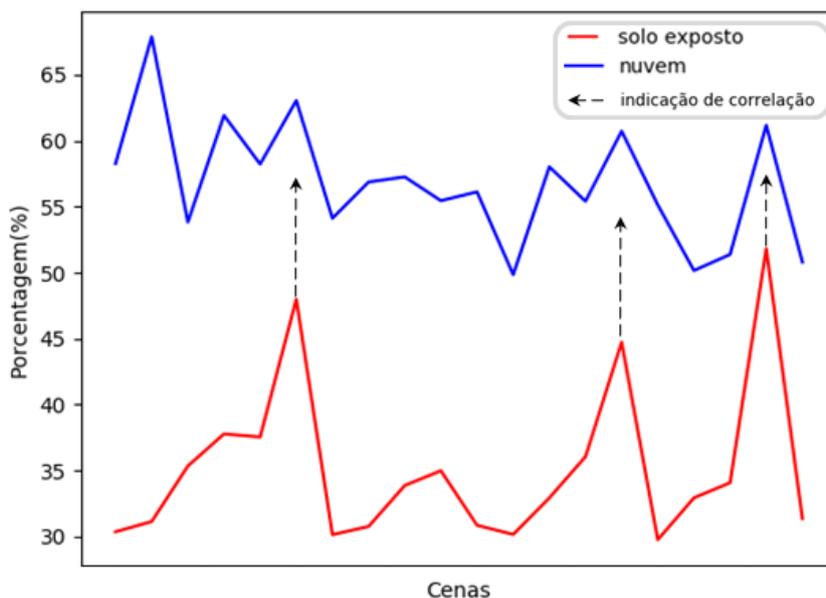
Figura 4.4 - Cenas retornadas a partir da intersecção da busca combinada de *solo exposto* acima de 6% e *nuvem* igual ou acima de 11%.



Porcentagem dos alvos de *solo exposto* e *nuvem* nas cenas retornadas.

Fonte: Elaborada pela autora.

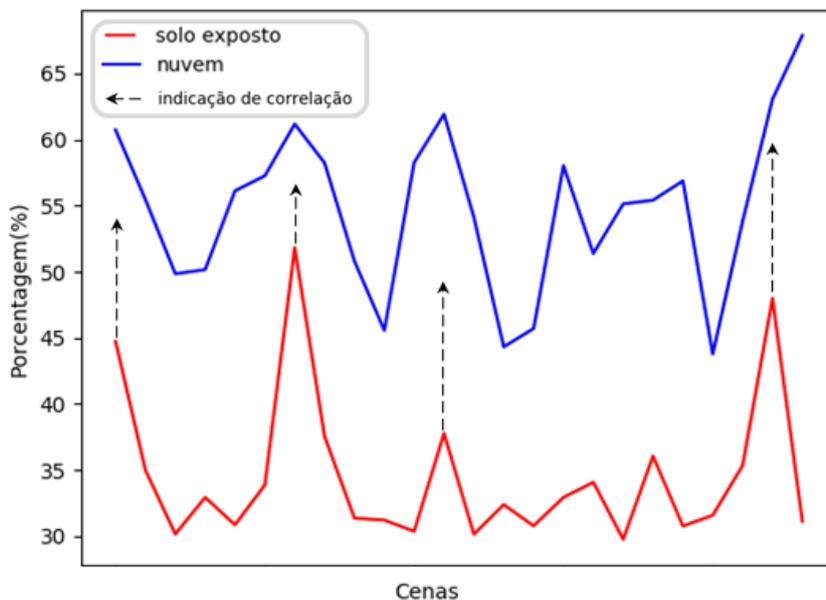
Figura 4.5 - Cenas retornadas a partir da intersecção da busca combinada de *solo exposto* igual ou acima de 29% e *nuvem* igual ou acima de 48%.



Porcentagem dos alvos de *solo exposto* e *nuvem* nas cenas retornadas.

Fonte: Elaborada pela autora.

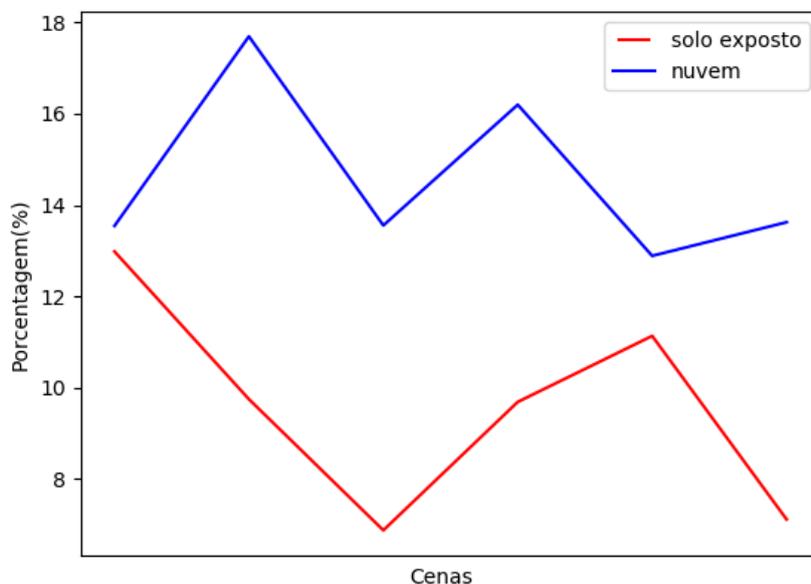
Figura 4.6 - Cenas retornadas a partir da intersecção da busca combinada igual ou acima de 29% para os alvos *solo exposto* e *nuvem*.



Porcentagem dos alvos de *solo exposto* e *nuvem* nas cenas retornadas.

Fonte: Elaborada pela autora.

Figura 4.7 - Cenas retornadas a partir da intersecção da busca combinada de porcentual de [6;17] para o alvo *solo exposto* e de [11;18] para *nuvem*.



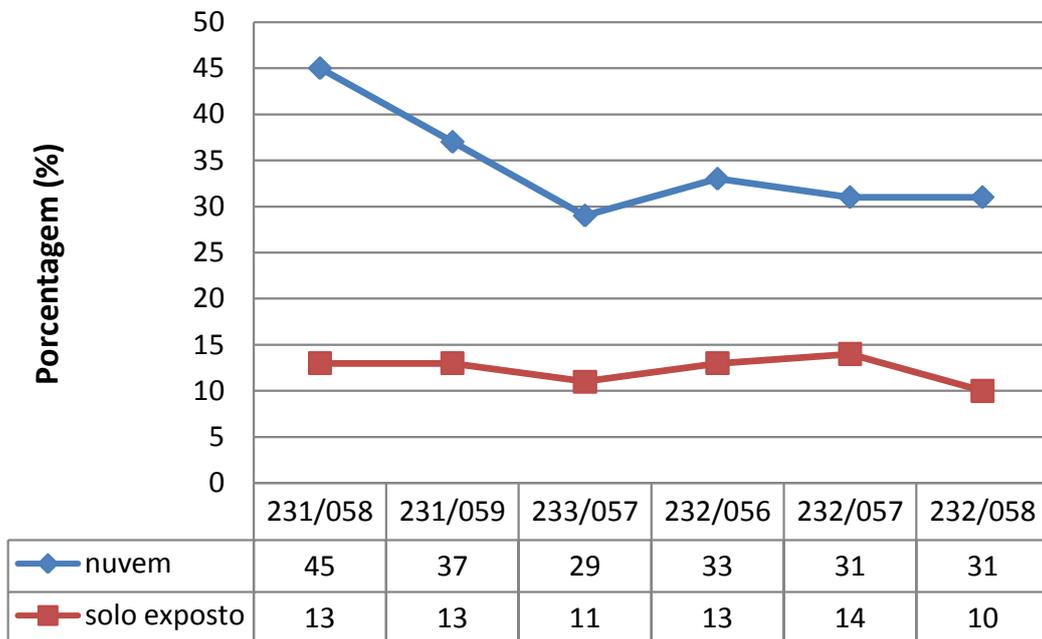
Porcentagem dos alvos de *solo exposto* e *nuvem* nas cenas retornadas.

Fonte: Elaborada pela autora.

adaptada ao clima seco. As áreas de savanas ao longo do estado podem ainda ser sub-divididas conforme as características do local em que são encontradas (DANTAS; RODRIGUES, 1982; Instituto Brasileiro de Geografia e Estatística, 2004b; Instituto Brasileiro de Geografia e Estatística, 2012). As savanas localizadas nas regiões centro-leste e nortordeste de Roraima compreendem o complexo paisagístico Rio Branco-Rupununi, o qual abrange o Brasil, a Guiana e a Venezuela, somando juntas mais de 60.000 km². No Brasil, está inserida no bioma Amazônia e é restrita ao estado de Roraima, com mais de 40.000 km², sendo também conhecida como *Savanas das Guianas* (BARBOSA et al., 2007). Destaca-se aqui ainda que a floresta que limita as savanas é diferente do sul da Amazônia (BRAGA, 1979).

A distribuição das porcentagens dos alvos em cada uma das cenas pode ser analisada na Figura 4.8. Apesar da pouca quantidade de dados, não foram identificados visualmente correlações marcantes entre os alvos nesta região. Além disso, o coeficiente de Pearson para estas seis cenas referente à porcentagem dos alvos *solo exposto* e *nuvem* foi de 0,39, denotando que o ReSIIM, embora apresente certas confusões no alvo *solo exposto* devido a presença de *nuvem*, conseguiu identificar que essa área possui significativa quantidade de pixels com pouca área de vegetação viva.

Figura 4.8 - Cenas não retornadas a partir da intersecção da busca combinada de percentual de [6;17] para o alvo *solo exposto* e de [11;18] para *nuvem*.



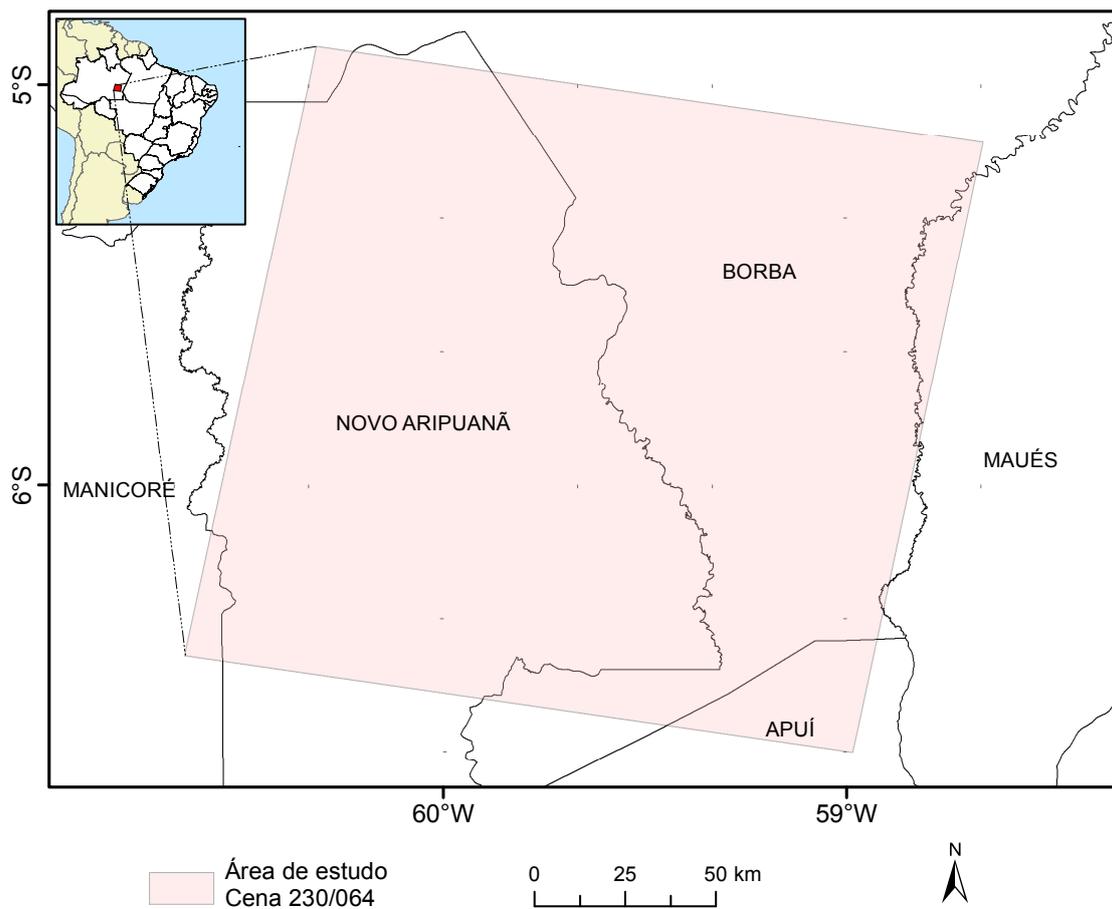
Fonte: Elaborada pela autora.

Levando em consideração os aspectos visuais e o coeficiente de correlação de Pearson, há um indício que o alvo *solo exposto* seja altamente correlacionado com o alvo *nuvem* quando são considerados os dados em uma visão macro (*solo exposto* igual ou acima de 6% e *nuvem* igual ou acima de 11%). Entretanto, ao analisar certos casos mais específicos, há pouca correlação entre os dados de *solo exposto* e *nuvem*. Assim, embora exista a necessidade de refinar a definição do alvo *solo exposto* para a implementação no ReSIIM, a atual definição conseguiu identificar áreas que realmente apresentam *solo exposto* e *rochas expostas*, como no caso do norte de Roraima. Outro exemplo e possível justificativa para a presença marcante do alvo *solo exposto* ao longo da costa são as malhas urbanas que tendem a apresentar baixa porcentagem de vegetação, uma vez que dificilmente mantêm os critérios de áreas verdes conforme previstos na lei (LOCH, 2008), e a dinâmica das vegetações, como no caso da Caatinga, cuja vegetação predominantemente arbustiva perde as folhas no período seco (ROSS, 2006). Para resultados mais concretos sobre o desempenho do alvo *solo exposto*, há a necessidade de analisar um maior conjunto de dados e uma base mais consistente de dados de validação, além de analisar possíveis ajustes no limiar de *solo exposto*.

4.2 Temporal: Aplicação do ReSIIM ao longo de um ano

Com o objetivo de analisar a dispersão dos dados do ReSIIM ao longo de um ano em uma área de baixa mudança de cobertura da terra, a cena 230/064 (Figura 4.9) foi selecionada por estar localizada no centro da região Amazônica. A cena de estudo abrange os municípios de Manicoré, Novo Aripuanã, Borba, Apuí e Maués, no estado do Amazonas. O período de aquisição de cenas do Landsat 8 selecionado no catálogo do INPE para esta análise foi entre 01/01/2017 e 01/01/2018. Totalizando 23 cenas, o processamento dos 45 GB durou cerca de 4 horas. Caso o sistema fosse analisar esta cena com base nos dados da Série Landsat por 40 anos, todo o processamento levaria menos de 7 dias. O resultado da dispersão dos alvos encontrados nesta cena ao longo do tempo pode ser visualizado na (Figura 4.10).

Figura 4.9 - Localização da cena do estudo de caso.



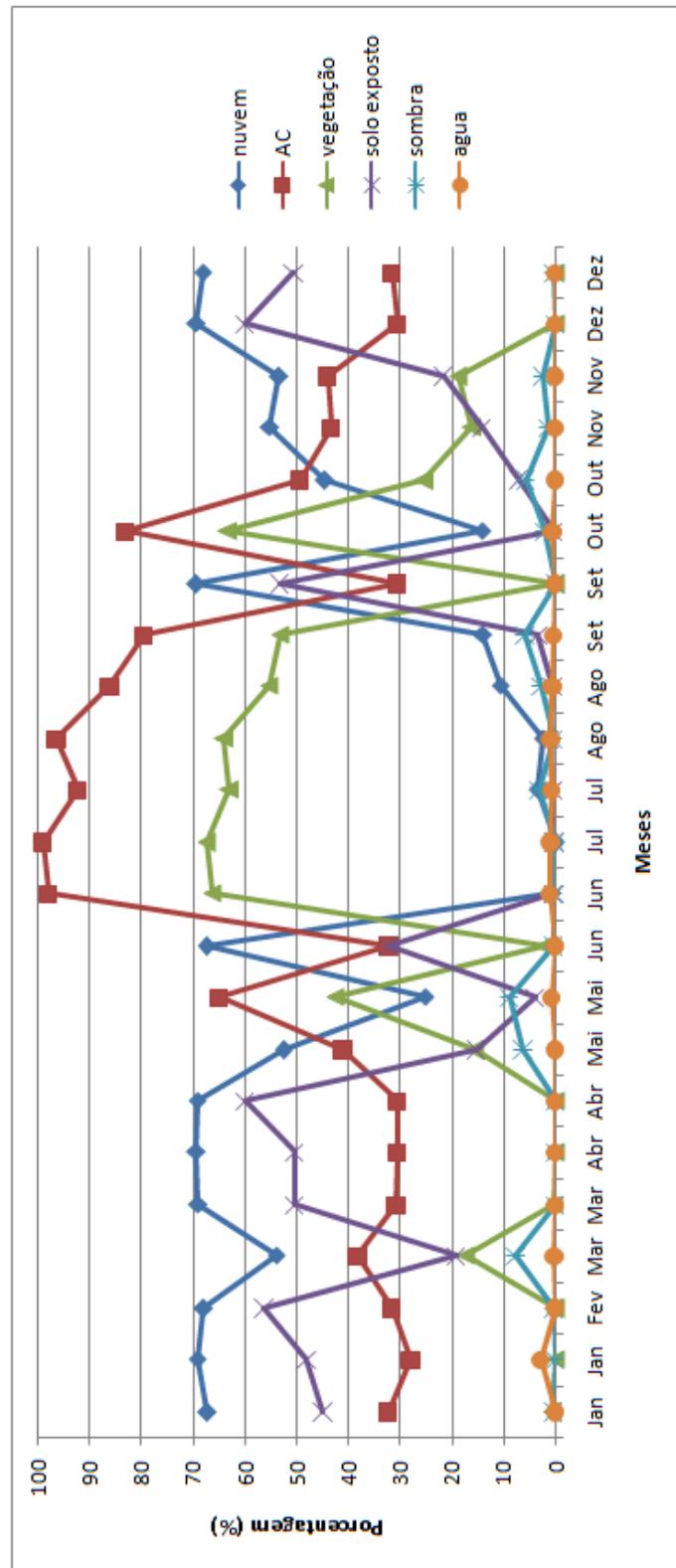
Fonte: Elaborada pela autora.

O período de seca na Amazônia (entre Agosto e Setembro segundo [Aragão et al. \(2007\)](#)), apresenta baixa porcentagem de *nuvem* principalmente entre Junho e Setembro, e no começo do mês de Outubro. Entre Dezembro e Abril, período mais chuvoso, a presença do alvo *nuvem* é significativa, em torno de 70% das cenas. Ao contrário da coerência encontrada no alvo *nuvem*, o alvo *sombra de nuvem* apresenta porcentagens abaixo de 10%, com picos nos meses de Março e Maio. Isso ocorre já que o alvo *sombra de nuvem* não necessariamente é linearmente relacionada com *nuvem*, devido à fatores como o ângulo de aquisição dos dados e período do dia.

Devido à baixa porcentagem principalmente de *nuvem* na época de seca, esse período apresentou maior presença de AC em contraste, o que poderia ser útil mais análise em áreas sem *nuvem*, *sombra de nuvem* e que não apresentam *água*. O alvo de *água* não foi superior a 3% em nenhuma época do ano. Entretanto, destaca-se aqui a sua maior presença na época de chuvas, no mês de Janeiro. O alvo *floresta*, considerando que este é baseado apenas no valor de NDVI, não é possível afirmar que a maior predominância deste alvo ao longo do é no período de seca, e sim que há maior visibilidade da área neste período do ano. Assim, buscas combinadas, um maior conjunto de dados e novos aprimoramentos na estrutura do ReSIIM são requeridos para análises. O mesmo é necessário para o alvo *solo exposto*, já que existe a possibilidade deste alvo ser erroneamente classificado devido à presença de nuvens (mais detalhes na Subseção 4.1.2).

Desta aplicação, foi possível explorar o potencial do ReSIIM em uma perspectiva temporal. Entretanto, caso o ReSIIM fosse utilizado em uma perspectiva de menor escala, como no caso de bairros, poderia surgir a necessidade de limitar a área de interesse, embora era opção ainda não esteja disponível no protótipo.

Figura 4.10 - Variação da porcentagem dos alvos *nuvem*, *sombra de nuvem*, *área continental (AC)*, *água*, *floresta* e *solo exposto* ao longo do ano de 2017.



Fonte: Elaborada pela autora.

4.3 Espectral: ReSIIM como suporte de aplicação do Modelo Linear de Mistura Espectral

Buscando analisar a performance do ReSIIM como uma ferramenta de busca inteligente de imagens tendo como estudo de caso a aplicação do MLME (Subseção 2.4.1), um experimento foi realizado. Nele, o banco de imagens de Sensoriamento Remoto foi formado por 60 cenas Landsat, representando mais de 2 milhões de km² e cerca de 120 GB. Dois principais grupos de dados foram selecionados para compor a base do banco de imagens de Sensoriamento Remoto, sendo metade de dado de referência (DRef), e outra, de dado de ruído (DR), previamente classificados de acordo com um especialista. O objetivo foi identificar se haveria algum critério de busca por imagens que pudesse retornar apenas as imagens de referência (máxima precisão, e mínimo *recall* e erro). Primeiramente, os alvos foram usados no critério de busca por imagens isoladamente, seguido de uma análise integrada dos critérios com melhor desempenho no retorno apenas de imagens de referência.

Através dos testes, foi possível compreender não apenas os dados disponíveis, mas também o papel do ReSIIM dentro da seleção de imagens para a aplicação do MLME. Os principais resultados dos testes de retorno de dados podem analisados na Tabela 4.3.

Os testes 1 e 2 demonstram que o alvo *nuvem* é um dos principais critérios na busca

Tabela 4.3 - Testes realizados na busca por imagens adequadas para a aplicação do MLME (DRef: Dado de Referência; DR: Dado de Ruído). As lacunas em azul representam valores menores ou iguais (\leq), e em laranja, maiores ou iguais (\geq).

| Número do Teste | Critério de Busca | Porcentagem | | | | | | | | | | Imagens Retornadas | | |
|-----------------|-------------------------|-------------|----|----|----|----|----|----|----|----|-----|--------------------|----|----|
| | | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 | DRef | DR | |
| 1 | <i>nuvem</i> | | | | | | | | | | | | 27 | 0 |
| 2 | <i>nuvem</i> | | | | | | | | | | | | 30 | 1 |
| 3 | <i>sombra</i> | | | | | | | | | | | | 30 | 29 |
| 4 | <i>área continental</i> | | | | | | | | | | | | 0 | 30 |
| 5 | <i>área continental</i> | | | | | | | | | | | | 30 | 0 |
| 6 | <i>área continental</i> | | | | | | | | | | | | 26 | 0 |
| 7 | <i>água</i> | | | | | | | | | | | | 30 | 29 |
| 8 | <i>floresta</i> | | | | | | | | | | | | 12 | 17 |
| 9 | <i>solo exposto</i> | | | | | | | | | | | | 30 | 1 |
| 10 | <i>nuvem</i> | | | | | | | | | | | | 30 | 0 |
| | <i>área continental</i> | | | | | | | | | | | | | |

por imagens adequadas para aplicação do MLME. Se o critério para retornar imagens for de até 10% com a presença deste alvo, 27 DRef foram retornadas. Se a busca for pautada em cenas com até 20% deste alvo, 30 DRef são retornadas, juntamente com uma DR. Esses testes indicam que embora o alvo *nuvem* seja importante na busca por imagens, pode haver outro alvo que contribua para que o retorno seja composto apenas por cenas de DRef. Isso pode ser explicado devido a algumas áreas da Amazônia brasileira apresentarem nuvens ao longo de todo o ano, devido a umidade local.

No teste 3, é possível reconhecer que o alvo *sombra* não foi importante nesse tipo de estudo, já que retornou tanto DRef quanto DR. Assim como o alvo de *nuvem*, os testes 4-6 demonstram que o alvo área continental foi importante no retorno de DRef. Cenas com 40% ou mais de áreas continentais retornaram todos os DRef e nenhum DR.

Os alvos de *água* e *floresta*, assim como o alvo *sombra de nuvem*, não apresentam eficácia no retorno apenas de DRef (Testes 7, 8). Vale ressaltar que algumas regiões apresentam água durante todo o período do ano, logo, esse critério de buscas por imagens poderia enviesar o estudo de caso.

No teste 9, o critério de buscas por imagens foi baseado na presença de até 10% de *solo exposto*. Embora o resultado se apresente satisfatório, com retorno de 30 DRef e apenas um DR, foi identificado que os limiares de NDVI utilizados no Re-SIIM apresentaram alta confusão com outros atributos, como nuvens. A discussão pormenorizada do resultado desta possível confusão está descrita na Subseção 4.1.2. Considerando que esta é uma questão ainda em aberto, esse alvo não foi considerado nos demais testes.

Finalmente, com base nos testes descritos, os melhores resultados dos testes dos alvos *nuvem* e *área continental* foram integrados em um critério de busca por imagens, busca combinada. Cenas de até 20% de *nuvem* e igual ou acima de 40% de área continental retornaram todas as cenas DRef e nenhuma DR. Com precisão máxima, e sem a necessidade de *recall* ou apresentação de erros, esse foi o melhor critério de retorno de imagens identificado para este estudo de caso.

Caso os catálogos de imagens oferecessem esse tipo de critério, a pesquisa de detecção e mapeamento de cicatrizes de queimadas na região da Amazônia por meio do MLME poderia ser facilitada pelo uso dos metadados. Lembrando que a busca por imagens adequadas para este tipo de estudo pode representar uma parcela considerável do

esforço necessário para o mapeamento dos alvos e a rapidez destas aplicações podem ser ainda decisivas em tomadas de decisões.

5 CONCLUSÕES

Os resultados desta dissertação foram gerados a partir do processamento de quase 2 mil cenas, representando quase 4 TB de dados, o que no caso das imagens Landsat, representam quase 70 milhões de km² de área, ou seja, mais de 13% da área da superfície terrestre.

A catalogação e a indexação têm como principal função determinar a existência de algum dado específico, bem como a sua localização e uma possível aplicação (FRANK, 1994). Assim, o refinamento da busca por imagens em grandes catálogos tem como principal vantagem a rápida visualização de fenômenos de maior complexidade, já que diferentes metadados e limiares podem ser considerados. Apesar disso, o desafio em tratar esses dados considerando o seu grande volume, velocidade de produção e variedade é alto. Esse conjunto de características dos dados pode sobrecarregar a capacidade das instituições de gerir e buscar o seu conteúdo. Embora exista demanda por ferramentas mais robustas de buscas por imagens, as atuais plataformas limitam-se a apresentar parâmetros espaciais e temporais, como a área e o período desejados.

Nesse contexto, ReSIIM se apresenta como uma rápida e acessível ferramenta alternativa para minerar as informações destas imagens, armazenando e permitindo aos usuários buscar imagens de acordo com o seu conteúdo de *nuvem*, *sombra de nuvem*, *área continental*, *água*, *floresta* e *solo exposto*. Desse modo, a abordagem desenvolvida dá suporte a uma gama de estudos, sendo ainda possível o seu aperfeiçoamento por colaboradores. Apesar disso, a acurácia da metodologia de extração de atributos e geração de metadados presentes no ReSIIM não é absoluta, uma vez que apresenta uma indicação da correlação destes elementos por meio de modelos matemáticos. Além disso, a metodologia empregada no ReSIIM classificou erroneamente o alvo *solo exposto* como *nuvem* em certas áreas, carecendo assim também de atenção.

No estudo de caso, o suporte de ReSIIM na aplicação do MLME superou as expectativas, buscando todas as imagens de referência do banco principal (máxima precisão), sem a necessidade de *recall* ou presença de erros. Os principais alvos identificados na etapa de buscas por imagens foram *nuvem* e *área continental*. Para outras aplicações, há a necessidade de ajustar estes parâmetros.

5.1 Indicação de trabalho futuros

Como trabalhos futuros, indica-se aplicar nas cenas já adquiridas novos índices e modelos matemáticos que possam abarcar uma maior gama de estudos e necessidades de usuários. Estes índices, ou novos algoritmos de detecção, incluem os alvos de áreas queimadas, áreas construídas, ou até mesmo a identificação de feições particulares, como por exemplo retas e círculos. A abordagem empregada no ReSIIM teve como base cenas da série Landsat, assim, indica-se também a aplicação do protótipo com base em outras plataformas de imagens de Sensoriamento Remoto, como o satélite sino-brasileiro CBERS-4, já em órbita há três anos (Instituto Nacional de Pesquisas Espaciais, 2017a) e o Amazonia-1, cujo lançamento está previsto para 2018 (Instituto Nacional de Pesquisas Espaciais, 2018).

Para análises mais detalhadas no ReSIIM, faz-se necessário ainda permitir abordagens, cuja área de interesse possa ser menor que uma cena, como um armazenamento dos dados a partir das dimensões de um pixel ao invés de uma cena. A sinergia e introdução de dados complementares, como dados estatísticos providos de instituições oficiais, também são sugeridos. Assim, buscas combinadas entre distintas fontes seriam possíveis, possibilitando uma maior compreensão do conjunto de dados. Além disso, considerando os aspectos do ReSIIM, há a necessidade de aperfeiçoar a validação dos metadados, por meio da correlação com alvos reais, bem como refinar os dados já retornados pelo ReSIIM, como o alvo *solo exposto*.

Por fim, para aprimorar o ReSIIM, é necessário compreender as necessidades dos usuários visando nortear as próximas etapas de atuação na estrutura do protótipo, utilizando assim o maior potencial do ReSIIM, a extensibilidade e a participação da comunidade interessada.

REFERÊNCIAS BIBLIOGRÁFICAS

- AB'SÁBER, A. N. **Os domínios de natureza no Brasil: potencialidades paisagísticas**. São Paulo: Ateliê Editorial, 2003. 43
- ANDERE, L.; ANDERSON, L. O.; DUARTE, V.; ARAI, E.; ARAGÃO, J. R. L. d.; ARAGÃO, L. E. O. e. C. d. Dados multitemporais do sensor modis para o mapeamento de queimadas na amazônia. In: GHERARDI, D. F. M.; ARAGÃO, L. E. O. e. C. d. (Ed.). **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2015. p. 3534–3541. ISBN 978-85-17-0076-8.
Disponível em:
<<http://urlib.net/sid.inpe.br/marte2/2015/06.15.15.28.29>>. Acesso em: 13 mar. 2018. 27, 28
- ANDERSON, L. O.; ARAGÃO, L. E. O. C.; GLOOR, M.; ARAI, E.; ADAMI, M.; SAATCHI, S. S.; MALHI, Y.; SHIMABUKURO, Y. E.; BARLOW, J.; BERENGUER, E.; DUARTE, V. Disentangling the contribution of multiple land covers to fire-mediated carbon emissions in Amazonia during the 2010 drought. **Global Biogeochemical Cycles**, v. 29, n. 10, p. 1739–1753, 2015. 26, 27
- ANDERSON, L. O.; ARAGÃO, L. E. O. E. C. D.; LIMA, A. D.; SHIMABUKURO, Y. E. Detecção de cicatrizes de áreas queimadas baseada no modelo linear de mistura espectral e imagens índice de vegetação utilizando dados multitemporais do sensor MODIS/TERRA no estado do Mato Grosso, Amazônia brasileira. **Acta Amazonica**, v. 35, n. 4, p. 445–456, 2005. ISSN 0044-5967.
Disponível em: <http://www.scielo.br/scielo.php?script=sci_{_}arttext&pid=S0044-59672005000400009>. Acesso em: 21 fev. 2017. 26, 27, 28
- ANDERSON, L. O.; CHEEK, D.; ARAGAO, L. E.; ANDERE, L.; DUARTE, B.; SALAZAR, N.; LIMA, A.; DUARTE, V.; ARAI, E. Development of a point-based method for map validation and confidence interval estimation: a case study of burned areas in Amazonia. **Journal of Remote Sensing & GIS**, v. 06, n. 01, p. 1–9, 2017. ISSN 24694134. 26
- ANDERSON, L. O.; SHIMABUKURO, Y. E.; DEFRIES, R. S.; MORTON, D. Assessment of deforestation in near real time over the brazilian amazon using multitemporal fraction images derived from terra modis. **IEEE Geoscience and Remote Sensing Letters**, IEEE, v. 2, n. 3, p. 315–318, 2005. 27, 28

ARAGÃO, L. E.; ANDERSON, L. O.; LIMA, A.; ARAI, E. Fires in amazonia. In: NAGY, L.; FORSBERG, B. R.; ARTAXO, P. (Ed.). **Interactions between biosphere, atmosphere and human land use in the Amazon basin**. Berlin, Heidelberg: Springer, 2016. p. 301–329. 4, 26, 28

ARAGÃO, L. E. O. C.; MALHI, Y.; ROMAN-CUESTA, R. M.; SAATCHI, S.; ANDERSON, L. O.; SHIMABUKURO, Y. E. Spatial patterns and fire response of recent amazonian droughts. **Geophysical Research Letters**, Wiley Online Library, v. 34, n. 7, 2007. ISSN 1944-8007. Disponível em: <<http://dx.doi.org/10.1029/2006GL028946>>. Acesso em: 14 mar. 2018. 43, 56

ARAGÃO, L. E. O. C.; POULTER, B.; BARLOW, J. B.; ANDERSON, L. O.; MALHI, Y.; SAATCHI, S.; PHILLIPS, O. L.; GLOOR, E. Environmental change and the carbon balance of amazonian forests. **Biological Reviews**, Blackwell Publishing Ltd, v. 89, n. 4, p. 913–931, 2014. ISSN 1469-185X. Disponível em: <<http://dx.doi.org/10.1111/brv.12088>>. Acesso em: 14 mar. 2018. 26

ARAI, K. Remote sensing satellite image database system allowing image portion retrievals utilizing principal component which consists spectral and spatial features extracted from imagery data. **International Journal of Advanced Research in Artificial Intelligence**, v. 2, n. 5, p. 32–38, 2013. Disponível em: <<http://thesai.org/Publications/ViewPaper?Volume=2&Issue=5&Code=IJARAI&SerialNo=5>>. Acesso em: 10 jun. 2017. 14, 15

BANNARI, A.; MORIN, D.; BONN, F.; HUETE, A. A review of vegetation indices. **Remote sensing reviews**, Taylor & Francis, v. 13, n. 1-2, p. 95–120, 1995. 27

BARBOSA, R. I.; CAMPOS, C.; PINTO, F.; FEARNSIDE, P. M. The 'lavrados' of roraima: biodiversity and conservation of brazil's amazonian savannas. **Functional Ecosystems and Communities**, v. 1, n. 1, p. 29–41, 2007. 53

BIJL-BROUWER, M. van der; DORST, K. Advancing the strategic impact of human-centred design. **Design Studies**, v. 53, p. 1 – 23, 2017. ISSN 0142-694X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0142694X17300455>>. Acesso em: 28 fev. 2017. 15

BOWKER, D. E.; DAVIS, R. E.; MYRICK, D. L.; STACY, K.; JONES, W. T. **Spectral reflectances of natural targets for use in remote sensing studies**. United States: NASA, Jun 1985. Reference Publication 1139. Disponível

em: <<https://ntrs.nasa.gov/search.jsp?R=19850022138>>. Acesso em: 13 mar. 2018. 7

BRAGA, P. I. S. Subdivisão fitogeográfica, tipos de vegetação, conservação e inventário florístico da floresta amazônica. **Acta amazonica**, SciELO Brasil, v. 9, n. 4, p. 53–80, 1979. 53

BUSH, M. B.; SILMAN, M. R.; TOLEDO, M. B. de; LISTOPAD, C.; GOSLING, W. D.; WILLIAMS, C.; OLIVEIRA, P. E. de; KRISSEL, C. Holocene fire and occupation in Amazonia: records from two lake districts. **Philosophical Transactions of the Royal Society B: Biological Sciences**, v. 362, n. 1478, p. 209–218, 2007. ISSN 0962-8436. Disponível em: <<http://rstb.royalsocietypublishing.org/cgi/doi/10.1098/rstb.2006.1980>>. Acesso em: 22 fev. 2017. 26

CÂMARA, G.; VALERIANO, D. d. M.; SOARES, J. V. **Metodologia para o cálculo da taxa anual de desmatamento na Amazônia Legal**. São José dos Campos: INPE, 2006. Disponível em: <<http://www.obt.inpe.br/prodes/metodologia.pdf>>. Acesso em: 13 mar. 2018. 46

CAMPBELL, J. B.; WYNNE, R. H. **Introduction to remote sensing**. 5. ed. United States: Guilford Press, 2011. ISBN 9781609181765. 8

CARDOZO, F. d. S.; PEREIRA, G.; SHIMABUKURO, Y. E.; MORAES, E. C. Análise do uso do modelo linear de mistura espectral (mlme) para o mapeamento das áreas queimadas no estado de Rondônia no ano de 2010. In: EPIPHANIO, J. C. N.; GALVÃO, L. S. (Ed.). **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2013. p. 7265–7272. ISBN 978-85-17-00066-9 (Internet) and 978-85-17-00065-2 (DVD). Disponível em: <<http://urlib.net/dpi.inpe.br/marte2/2013/05.29.00.38.12>>. Acesso em: 13 mar. 2018. 27, 28

CARRARA, P.; PASI, G.; PEPE, M.; RAMPINI, A. An indexing model of remote sensing images. In: ENSER, P.; KOMPATSIARIS, Y.; O'CONNOR, N. E.; SMEATON, A. F.; SMEULDERS, A. W. M. (Ed.). **Proceedings...** Berlin, Heidelberg: Springer Berlin Heidelberg, 2004. p. 517–525. 23, 24, 25

CHAZDON, R. L.; BRANCALION, P. H.; LAESTADIUS, L.; BENNETT-CURRY, A.; BUCKINGHAM, K.; KUMAR, C.; MOLL-ROCEK, J.; VIEIRA, I. C. G.; WILSON, S. J. When is a forest a forest? forest concepts and

definitions in the era of forest and landscape restoration. **Ambio**, Springer, v. 45, n. 5, p. 538–550, 2016. 35

CHI, M.; PLAZA, A.; BENEDIKTSSON, J. A.; SUN, Z.; SHEN, J.; ZHU, Y. Big data for remote sensing: challenges and opportunities. **IEEE**, v. 104, n. 11, p. 2207–2219, 2015. ISSN 0018-9219. Disponível em: <<http://ieeexplore.ieee.org/document/7565634/>>. Acesso em: 15 jan. 2017. 1, 2

CHUVIECO, E.; CONGALTON, R. G. Mapping and inventory of forest fires from digital processing of tm data. **Geocarto International**, Taylor & Francis, v. 3, n. 4, p. 41–53, 1988. 28

COMBER, A. J. The separation of land cover from land use using data primitives. **Journal of Land Use Science**, Taylor & Francis, v. 3, n. 4, p. 215–229, 2008. 18

CONTI, J. B.; FURLAN, S. A. Geocologia: o clima, os solos e a biota. In: _____. **Geografia do Brasil**. 6. ed. São Paulo: Edusp - Editora da Universidade de São Paulo, 2011. cap. 2, p. 67–208. ISBN: 978-85-314-0242-5. 43, 44

DANTAS, M.; RODRIGUES, I. **Estudos fitoecológicos do trópico úmido brasileiro: IV, levantamentos botânicos em campos do Rio Branco**. Belém, Pará: EMBRAPA, Centro de Pesquisa Agropecuária do Trópico Úmido, 1982. Boletim de Pesquisa. 46, 53

DATCU, M.; SEIDEL, K. An innovative concept for image information mining. In: ZAIANE, O. R.; SIMOFF, S. J.; DJERABA, C. (Ed.). **Lecture Notes in Computer Science: Mining multimedia and complex data**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003. p. 84–99. ISBN 978-3-540-39666-6. LNAI 2797. 1, 16

_____. Image information mining: exploration of Earth observation archives. **Geographica Helvetica**, v. 58, n. 2, p. 154–168, 2003. Disponível em: <<http://www.geogr-helv.net/58/154/2003/>>. Acesso em: 10 abr. 2017. 1, 2, 3, 14, 15, 16

DATCU, M.; SEIDEL, K.; D'ELIA, S.; MARCHETTI, P. Knowledge-driven information mining in remote-sensing image archives. **E.S.A. Bulletin**, n. 110, p. 26–33, 2002. ISSN 0376-4265. 1, 2

DATCU, M.; SEIDEL, K.; PELIZZARRI, A.; SCHROEDER, M.; REHRAUER, H.; PALUBINSKAS, G.; WALESSA, M. Image information mining and remote

sensing data interpretation. In: **Proceedings... IEEE**, 2000. v. 7, n. July, p. 3057–3059. ISBN 0-7803-6359-0. Cat. No.00CH37120. Disponível em: <<http://ieeexplore.ieee.org/document/860334/?arnumber=860334>>. Acesso em: 8 jul. 2017. 2, 16

DONG, J.; ZHUANG, D.; HUANG, Y.; FU, J. Advances in multi-sensor data fusion: algorithm and applications. **Sensors**, Molecular Diversity Preservation International, v. 9, n. 10, p. 7771–7784, 2009. 1

EAKINS, J.; GRAHAM, M. E. **Content based image retrieval**: a report to the JISC technology applications programme. Manchester: Joint Information Systems Committee's Technology Applications Programme - JTAP, 1999. 59 p. 14, 15, 17, 18

EHLERS RONALD JANOWSKY, M. G. M. New remote sensing concepts for environmental monitoring. In: **International Symposium on Remote Sensing**. Toulouse, France: Remote Sensing for Environmental Monitoring, GIS Applications, and Geology, 2002. v. 4545, p. 1 – 12. 28

ESCADA, M. I. S. **Evolução de padrões de uso e cobertura da terra na região Centro-Norte de Rondônia**. 264 p. Tese (Doutorado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2003-04-14 2003. Disponível em: <<http://mtc-m12.sid.inpe.br/sid.inpe.br/jeferson/2003/06.30.13.31>>. Acesso em: 14 mar. 2018. 46

FAYYAD, U.; GRINSTEIN, G. G.; WIERSE, A. **Information visualization in data mining and knowledge discovery**. 1st. ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001. ISBN 1558606890. 16

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. The kdd process for extracting useful knowledge from volumes of data. **Communications of the ACM**, ACM, v. 39, n. 11, p. 27–34, 1996. 3, 16

FEARNSIDE, P. M. O processo de desertificação e os riscos de sua ocorrência no brasil. **Acta Amazonica**, SciELO Brasil, v. 9, n. 2, p. 393–400, 1979. 46

FERREIRA, N. C. d. R.; ARCANJO, J. d. S. **Panorama do sistema de aquisição, processamento, catalogação e distribuição de imagens de sensoriamento remoto do INPE**. 13. ed. Florianópolis, SC, Brasil: São José dos Campos: INPE, 2007. 875–877 p. ISBN 978-85-17-00031-7. Disponível em:

<<http://marte.sid.inpe.br/col/dpi.inpe.br/sbsr{%}7B{%}25{%}7D4080/2007/01.31.19.10/doc/{%}7B{%}25{%}7D40sumario.htm>>. Acesso em: 05 jun. 2017. 1

FISHER, P.; COMBER, A. J.; WADSWORTH, R. Land use and land cover: contradiction or complement. **Re-presenting GIS**, Wiley Chichester, p. 85–98, 2005. 18

FLOOD, N.; GILLINGHAM, S. **Python Fmask documentation - Release 0.4.4**. 2017. Disponível em: <<http://pythonfmask.org/en/latest/>>. Acesso em: 15 jan. 2017. 20, 32

Florida Exotic Pest Plant Council, FLEPPC. **List of Florida's invasive species**. 2017. Disponível em: <www.fleppc.org>. Acesso em: 13 mar. 2018. 22

FOGA, S.; SCARAMUZZA, P. L.; GUO, S.; ZHU, Z.; DILLEY, R. D.; BECKMANN, T.; SCHMIDT, G. L.; DWYER, J. L.; HUGHES, M. J.; LAUE, B. Cloud detection algorithm comparison and validation for operational landsat data products. **Remote Sensing of Environment**, Elsevier, v. 194, p. 379 – 390, 2017. ISSN 0034-4257. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0034425717301293>>. Acesso em: 23 fev. 2017. 34

FORMAGGIO, A. R.; SANCHES, I. D. **Sensoriamento Remoto em agricultura**. São Paulo: Oficina de Textos, 2017. ISBN 978-85-7975-277-3. 21

FRANK, S. Cataloging digital geographic data in the information infrastructure: a literature and technology review. **Information Processing and Management**, v. 30, n. 5, p. 587–606, 1994. ISSN 03064573. 18, 19, 61

GARCIA, M. L.; CASELLES, V. Mapping burns and natural reforestation using thematic mapper data. **Geocarto International**, Taylor & Francis, v. 6, n. 1, p. 31–37, 1991. 27

GATTI, L.; GLOOR, M.; MILLER, J.; DOUGHTY, C.; MALHI, Y.; DOMINGUES, L.; BASSO, L.; MARTINEWSKI, A.; CORREIA, C.; BORGES, V.; FREITAS, S.; BRAZ, R.; ANDERSON, L.; ROCHA, H.; GRACE, J.; PHILLIPS, O.; LLOYD, J. Drought sensitivity of amazonian carbon balance revealed by atmospheric measurements. **Nature**, Nature Research, v. 506, n. 7486, p. 76–80, 2014. 26

GIGLIO, L. **MODIS collection 6 active fire product user guide**: revision A. United States: University of Maryland, Department of Geographical Sciences, 2015. Disponível em: <https://cdn.earthdata.nasa.gov/conduit/upload/3865/MODIS_C6_Fire_User_Guide_A.pdf>. Acesso em: 05 jun. 2017. 27

GIGLIO, L.; SCHROEDER, W.; JUSTICE, C. O. The collection 6 modis active fire detection algorithm and fire products. **Remote Sensing of Environment**, Elsevier, v. 178, p. 31–41, 2016. 27

GUPTILL, S. C. Metadata and data catalogues. **Geographical information systems**, John Wiley & Sons, Inc, v. 2, p. 677–692, 1999. 16, 17, 18, 19

_____. Spatial data. In: WRIGHT, J. D. (Ed.). **International Encyclopedia of the Social and Behavioral Sciences**. Second edition. Oxford: Elsevier, 2015. p. 126 – 129. ISBN 978-0-08-097087-5. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9780080970868720599>>. Acesso em: 05 jun. 2017. 18

HAN, J.; PEI, J.; KAMBER, M. **Data mining**: concepts and techniques. 3. ed. Waltham, Mass.: Elsevier, 2011. ISBN 9780123814807. 3, 16

HAWBAKER, T. J.; VANDERHOOF, M. K.; BEAL, Y.-J.; TAKACS, J. D.; SCHMIDT, G. L.; FALGOUT, J. T.; WILLIAMS, B.; FAIRAUX, N. M.; CALDWELL, M. K.; PICOTTE, J. J.; HOWARD, S. M.; STITT, S.; DWYER, J. L. Mapping burned areas using dense time-series of landsat data. **Remote Sensing of Environment**, Elsevier, v. 198, p. 504–522, 2017. ISSN 0034-4257. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0034425717302857>>. Acesso em: 14 mar. 2018. 27

HUANG, J.; GAO, J.; HÖRMANN, G.; MOOIJ, W. M. Integrating three lake models into a phytoplankton prediction system for lake taihu (taihu pps) with python. **Journal of Hydroinformatics**, IWA Publishing, v. 14, n. 2, p. 523–534, 2012. 41

Instituto Brasileiro de Geografia e Estatística. **Introdução ao processamento digital de imagens**. Rio de Janeiro: Manuais Técnicos em Geociências, 2001. GE00017281-9. ISBN 8524007621. 7

_____. Mapa de clima do brasil. **Diretoria de Geociências**, Rio de Janeiro, 2002. Disponível em: <ftp://geoftp.ibge.gov.br/informacoes_ambientais/climatologia/mapas/brasil/clima.pdf>. Acesso em: 13 mar. 2018. 43

_____. Mapa de biomas do Brasil. **Diretoria de Geociências**, Rio de Janeiro, 2004. Disponível em: <ftp://geoftp.ibge.gov.br/informacoes_ambientais/vegetacao/mapas/brasil/biomas.pdf>. Acesso em: 13 mar. 2018. 44

_____. Mapa de vegetação do Brasil. **Diretoria de Geociências**, Rio de Janeiro, 2004. Disponível em: <<https://mapas.ibge.gov.br/tematicos/vegetacao>>. Acesso em: 22 mar. 2018. 46, 53

_____. Manuais técnicos em geociências: manual técnico da vegetação brasileira. **Diretoria de Geociências**, Rio de Janeiro, 2012. 46, 53

Instituto Nacional de Pesquisas Espaciais. **Tutorial de geoprocessamento: modelos de mistura**. 2006. Disponível em: <<http://www.dpi.inpe.br/spring/portugues/tutorial/modelos.html>>. Acesso em: 13 mar. 2018. 28

_____. **Monitoramento da floresta Amazônica Brasileira por satélite**. 2013. Disponível em: <<http://www.obt.inpe.br/prodes/index.php>>. 46

_____. **CBERS-4 completa três anos em órbita**. 2017. A. Disponível em: <http://www.cbears.inpe.br/noticia.php?Cod_Noticia=4670>. 62

_____. **INPE estima 6.624 km² de desmatamento por corte raso na Amazônia em 2017**. 2017. B. Disponível em: <<http://www.obt.inpe.br/OBT/noticias/INPE-estima-desmatamento-por-corte-raso-na-Amazonia-em-2017>>. 48

_____. **Portal de monitoramento de queimadas e incêndios**. 2017. C. Disponível em: <<http://www.inpe.br/queimadas>>. 27

_____. **Projeto PRODES: monitoramento da floresta Amazônica Brasileira por satélite**. 2017. D. Disponível em: <<http://www.obt.inpe.br/prodes/index.php>>. 35, 46

_____. **Missão Amazonia-1**. 2018. Disponível em: <<http://www3.inpe.br/amazonia-1/amazonia.php>>. 62

JENSEN, J. R. **Sensoriamento remoto do ambiente: uma perspectiva em recursos terrestres**. São José dos Campos: Editora Parêntese, 2009. 672 p. ISBN 85605070 and 13: 9788560507061. Acesso em: 14 mar. 2018. 7, 8

JIN, X.; ZHANG, Y.-K.; TANG, Y.; HU, G.; GUO, R. Quantifying bare soil evaporation and its relationship with groundwater depth. **International Journal**

of **Remote Sensing**, v. 35, n. 21, p. 7567–7582, 2014. ISSN 0143-1161. Disponível em:

<<http://www.tandfonline.com/doi/abs/10.1080/01431161.2014.975374>>.

Acesso em: 13 mar. 2018. 21

JUNIOR, C. H. L. S.; ARAGÃO, L. E. O. e. C. d.; ANDERSON, L. O.; PENHA, T. V. Mapeamento de queimadas na amazônia legal brasileira utilizando dados modis e viirs: resultados preliminares. In: GHERARDI, D. F. M.; ARAGÃO, L. E. O. e. C. d. (Ed.). **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2017. p. 3475–3482. ISBN 978-85-17-00088-1. Disponível em: <<http://urlib.net/sid.inpe.br/marte2/2017/10.27.13.11>>. Acesso em: 13 mar. 2018. 27, 28

KEY, C.; BENSON, N. **General technical report: Landscape assessment (LA) - sampling and analysis methods.** United States: US Department of Agriculture, USDA. Forest Service, 2006. RMRS-GTR-164-CD. 27, 28

KEY, C. H.; BENSON, N. C. **The normalized burn ratio (NBR):** a Landsat TM radiometric index of burn severity. United States: US Geological Survey, USGS, 1999. 33 p. Department of the Interior. 27

KHERFI, M. L.; BRAHMI, D.; ZIOU, D. Combining visual features with semantics for a more effective image retrieval. In: **Proceedings...** Cambridge, UK: IEEE, 2004. v. 2, p. 961–964. ISSN 1051-4651. Disponível em: <<http://ieeexplore.ieee.org/document/1334418/>>. Acesso em: 13 mar. 2018. 14

KÖRTING, T. S.; NAMIKAWA, L. M.; FONSECA, L. M. G.; FELGUEIRAS, C. A. How to effectively obtain metadata from remote sensing big data? In: GEOBIA 2016. SOLUTIONS AND SYNERGIES, 14-16 set., Enschede, The Netherlands. **Proceedings...** 2016. Disponível em: <http://https://www.conftool.net/geobia2016/index.php?page=browseSessions\&abstracts=show\&form_session=16\&presentations=show>. Acesso em: 13 mar. 2018. 1

KOSKELA, M.; LAAKSONEN, J.; LAAKSO, S.; OJA, E. The picsom retrieval system: description and evaluations. In: **Proceedings...** Brighton, UK: CIR2000, 2000. Disponível em: <<http://www.cis.hut.fi/picsom/publications.html>>. Acesso em: 13 mar. 2018. 24, 39

LAAKSONEN, J.; KOSKELA, J.; OJA, E. Picsom: a framework for content-based image database retrieval using self-organizing maps. In: **Proceedings...** Kangerlussuaq, Greenland - Denmark: Pattern Recognition Society of Denmark, 1999. v. 1, p. 151–156. 24

LAAKSONEN, J.; KOSKELA, M.; LAAKSO, S.; OJA, E. Picsom: content-based image retrieval with self-organizing maps. **Pattern recognition letters**, Elsevier, v. 21, n. 13, p. 1199–1207, 2000. 24

LANEY, D. 3D data management: controlling data volume, velocity and variety. **Application Delivery Strategies**, v. 949, n. February 2001, p. 4, 2001. ISSN 09505849. Disponível em: <<http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>>. Acesso em: 15 jan. 2017. 1, 2

LARSON, R. R.; FRONTIERA, P. Geographic information retrieval and spatial browsing. In: SMITH, L. C.; GLUCK, M. (Ed.). **Proceedings...** United States: Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, 1996. p. 81–124. ISBN 0878450971. ISSN 0069-4789. Disponível em: <<https://www.ideals.illinois.edu/handle/2142/416>>. Acesso em: 13 mar. 2018. 15

LASHOF, D. The contribution of biomass burning to global warming: an integrated assessment. In: LEVINE, J. S. (Ed.). **Global biomass burning: atmospheric, climatic, and biospheric implications**. MIT Press, 1991. p. 441–444. ISBN 9780262310895. Disponível em: <<http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6302650>>. Acesso em: 22 fev. 2017. 26

LI, J.; NARAYANAN, R. M. Integrated information mining and image retrieval in remote sensing. In: CHANG, C. I. (Ed.). **Recent Advances in Hyperspectral Signal and Image Processing**. 1. ed. Trivandrum, India: Transworld Research Network, 2006. cap. 16, p. 449—478. ISBN 81-7895-218-1. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.100.4884>>. Acesso em: 15 jan. 2017. 2, 3, 16

LI, S.; DRAGICEVIC, S.; CASTRO, F. A.; SESTER, M.; WINTER, S.; COLTEKIN, A.; PETTIT, C.; JIANG, B.; HAWORTH, J.; STEIN, A.; CHENG, T. Geospatial big data handling theory and methods: a review and research

challenges. **ISPRS Journal of Photogrammetry and Remote Sensing**, International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS), v. 115, p. 119–133, 2016. ISSN 09242716. Disponível em: <<http://dx.doi.org/10.1016/j.isprsjprs.2015.10.012>>. Acesso em: 08 jan. 2017. 1, 2, 15

LI, S.; YU, H.; YUAN, L. A novel approach to remote sensing image retrieval with multi-feature vp-tree indexing and online feature selection. In: **Proceedings...** Second International Conference on Multimedia Big Data, BigMM, 2016. p. 133–136. Disponível em: <<http://ieeexplore.ieee.org/document/7545009/>>. Acesso em: 07 fev. 2017. 2

LIANG, S. **Quantitative remote sensing of land surfaces**. United States: John Wiley & Sons, 2005. ISBN 9780471723721. 21

LIANG, S.; ZHANG, X.; XIAO, Z.; CHENG, J.; LIU, Q.; ZHAO, X. Longwave Emissivity. In: **Global LAnd Surface Satellite (GLASS) Products: algorithms, validation and analysis**. Cham: Springer International Publishing, 2014. p. 73–121. ISBN 978-3-319-02588-9. Disponível em: <http://dx.doi.org/10.1007/978-3-319-02588-9_{_}4>. Acesso em: 13 mar. 2018. 21

_____. Shortwave albedo. In: **Global LAnd Surface Satellite (GLASS) products: algorithms, validation and analysis**. Cham: Springer International Publishing, 2014. p. 33–72. ISBN 978-3-319-02588-9. Disponível em: <http://dx.doi.org/10.1007/978-3-319-02588-9_{_}3>. Acesso em: 13 mar. 2018. 21, 38

LIMA, A.; SILVA, T. S. F.; ARAGÃO, L. E. O. e. C. de; FEITAS, R. M. de; ADAMI, M.; FORMAGGIO, A. R.; SHIMABUKURO, Y. E. Land use and land cover changes determine the spatial relationship between fire and deforestation in the Brazilian Amazon. **Applied Geography**, Elsevier Ltd, v. 34, p. 239–246, 2012. ISSN 01436228. Disponível em: <<http://dx.doi.org/10.1016/j.apgeog.2011.10.013>>. Acesso em: 22 fev. 2017. 27, 28

LIMA, A. d. **Influência da cobertura da terra na extensão e configuração espacial de áreas queimadas em anos de seca extrema na Amazônia Oriental**. 110 p. Tese (Doutorado) — Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2013-03-21 2013. Disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m19/2013/03.08.16.30>>. Acesso em: 14 mar. 2018. 27

LOCH, C. **A interpretação de imagens aéreas**. 5. ed. Florianópolis: Editora da UFSC, 2008. 54

LONG, F.; ZHANG, H.; FENG, D. D. Fundamentals of content-based image retrieval. In: _____. **Multimedia Information Retrieval and Management: Technological Fundamentals and Applications**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003. p. 1–26. ISBN 978-3-662-05300-3. Disponível em: <http://dx.doi.org/10.1007/978-3-662-05300-3_{_}1>. Acesso em: 10 jun. 2017. 2, 14, 15, 16, 17, 40

LUTZ, M. **Learning Python: powerful object-oriented programming**. United States: "O'Reilly Media, Inc.", 2013. 41

MARTÍN, M. **Cartografía e inventario de incendios forestales en la Península Ibérica a partir de imágenes NOAA–AVHRR**. Tese (Doutorado) — Universidad de Alcalá, Facultad de Filosofía y Letras, Departamento de Geografía. Alcalá de Henares, Spain, 1998. 27

MENESES, P. R.; ALMEIDA, T. d. **Introdução ao processamento de imagens de sensoriamento remoto**. Brasília: Brasília: UNB/CNPq, 2012. 266 p. 7

MEYER, W. B.; TURNER, I. B. **Changes in land use and land cover: a global perspective**. Cambridge: Cambridge University Press, 1994. 18

MITCHEL, A. The esri guide to gis analysis, volume 2: spartial measurements and statistics. **ESRI Guide to GIS analysis**, ESRI press, 2005. 48

MOLINIER, M.; LAAKSONEN, J.; AHOLA, J.; HÄME. Self-organizing map application for retrieval of man-made structures in remote sensing data. In: **Proceedings...** Frascati, Italy: ESA-EUSC, 2005. 23, 24

MOLINIER, M.; LAAKSONEN, J.; HAME, T. Detecting man-made structures and changes in satellite imagery with a content-based information retrieval system built on self-organizing maps. **IEEE Transactions on Geoscience and Remote Sensing**, IEEE Transactions on Geoscience and Remote Sensing, v. 45, n. 4, p. 861–874, 2007. Special issue on Image Information Mining. 23, 24

MOLINIER, M.; LAAKSONEN, J.; VÄÄTÄINEN, S.; HÄME, T. Self organising maps for change detection and monitoring of human activity in satellite imagery. In: **Proceedings...** Madrid, Spain: ESA-EUSC, 2006. 23, 24

MOREIRA, M. A. **Fundamentos do sensoriamento remoto e metodologias de aplicação**. 3. ed. Viçosa: Editora da Universidade Federal de Viçosa, UFV, 2005. 8

MOURA, J. R. d. S. Geomorfologia do Quaternário. In: _____. **Geomorfologia: uma atualização de bases e conceitos**. 2. ed. Rio de Janeiro: Bertrand Brasil, 1995. cap. 8, p. 335–364. ISBN 85-286-0326-1. 44

MÜLLER, H.; GEISSBUHLER, A.; MARCHAND-MAILLET, S.; CLOUGH, P. Benchmarking image retrieval applications. In: **Proceedings...** San Francisco, USA: Seventh International Conference on Visual Information Systems, 2004.

Disponível em:

<<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.136.1260>>.

Acesso em: 03 fev. 2017. 14

MURA, M. D.; PRASAD, S.; PACIFICI, F.; GAMBA, P.; CHANUSSOT, J.; BENEDIKTSSON, J. A. Challenges and opportunities of multimodality and data fusion in remote sensing. **IEEE**, v. 103, n. 9, Sept 2015. ISSN 0018-9219. 2

NAMIKAWA, L. M.; CASTEJON, E. F. **Mapas de lamina de água para todo o Brasil extraídos do RapidEye**. 2017. Disponível em: <<http://wiki.dpi.inpe.br/doku.php?id\unhbox\voidb@x\bgroup\let\unhbox\voidb@x\setbox\@tempboxa\hbox{m\global\mathchardef\accent@spacefactor\spacefactor}\accent9m\egroup\spacefactor\accent@spacefactorapas:waterbodies>>. 21

NAMIKAWA, L. M.; KÖRTING, T. S.; CASTEJON, E. F. Water body extraction from RapidEye images: an automated methodology based on hue component of color transformation from RGB to HSV model. **Revista Brasileira de Cartografia**, v. 68, n. 6, p. 1097–1111, 2016. Disponível em:

<<http://www.rbc.lsie.unb.br/index.php/rbc/article/view/1662>>. Acesso em: 13 mar. 2018. 20, 34

NAPOLETANO, P. Visual descriptors for content-based retrieval of remote sensing images. **Arxiv**, p. 1–12, 2016. Disponível em:

<<http://arxiv.org/abs/1602.00970>>. Acesso em: 07 fev. 2017. 22

NOAA National Oceanic & Atmospheric Administration. **MGVI**: normalized difference vegetation index. 2017. Disponível em:

<<http://www.ospo.noaa.gov/Products/land/mgvi/NDVI.html>>. 21, 38

NOVO, E. M. d. M. **Sensoriamento remoto: princípios e aplicações**. 2. ed. São Paulo: Edgard Blücher Ltda, 1992. 7

OLIPHANT, T. E. Python for scientific computing. **Computing in Science & Engineering**, IEEE, v. 9, n. 3, 2007. 41

PADILLA, M.; OLOFSSON, P.; STEHMAN, S. V.; TANSEY, K. Stratification and sample allocation for reference burned area data. **Remote Sensing of Environment**, Elsevier, v. 203, p. 240–255, dec 2017. ISSN 0034-4257. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0034425717303000>>.

Acesso em: 21 fev. 2017. 26

PADILLA, M.; STEHMAN, S. V.; RAMO, R.; CORTI, D.; HANTSON, S.; OLIVA, P.; ALONSO-CANAS, I.; BRADLEY, A. V.; TANSEY, K.; MOTA, B. et al. Comparing the accuracies of remote sensing global burned area products using stratified random sampling and estimation. **Remote Sensing of Environment**, Elsevier, v. 160, p. 114–121, 2015. 27

PETRUSHIN, V. A.; KHAN, L. **Multimedia data mining and knowledge discovery**. United States: Springer, 2007. ISBN 978-1-84628-799-2. 16

PLAZA, A. J. Content-based hyperspectral image retrieval using spectral unmixing. In: **Proceedings...** SPIE - The International Society for Optical Engineering, 2011. v. 8180. ISBN 9780819488077. ISSN 0277786X. Disponível em:

<<https://www.spiedigitallibrary.org/conference-proceedings-of-spie/8180/1/Content-based-hyperspectral-image-retrieval-using-spectral-unmixing/10.1117/12.898833.full?SS0=1>>. Acesso em: 21 jun. 2017. 23, 25, 26

PLETSCH, M. A. J. S.; KÖRTING, T. S. Remote sensing image information mining applied to burnt forest detection in the brazilian amazon. In: JR., C. A. U. D.; QUEIROZ, G. R. d. I. (Ed.). **Anais...** São José dos Campos: GeoInfo Series, 2017. p. 322–333. ISSN 2179-4820. Disponível em:

<<http://urlib.net/sid.inpe.br/mtc-m16c/2017/12.01.20.06>>. Acesso em: 13 mar. 2018. 31, 33

PLETSCH, M. A. J. S.; PENHA, T. V.; JUNIOR, C. H. L. S.; KÖRTING, T. S.; ARAGÃO, L. E. O. C.; OIGHENSTEIN, L. Integração do algoritmo fmask ao modelo linear de mistura espectral como subsídio à detecção de áreas queimadas na Amazônia Brasileira. **Revista Brasileira de Cartografia**, 2018. NO PRELO. 28

QUARTULLI, M.; OLAIZOLA, I. G. A review of EO image information mining. **ISPRS Journal of Photogrammetry and Remote Sensing**, International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS), v. 75, p. 11–28, 2013. ISSN 0924-2716. Disponível em: <<http://dx.doi.org/10.1016/j.isprsjprs.2012.09.010>>. Acesso em: 10 jun. 2017. 1, 3, 8

ROSA, R. **Introdução ao sensoriamento remoto**. Uberlândia: Editora da Universidade Federal de Uberlândia, EDEFU, 1990. 7

ROSS, J. L. S. **Ecogeografia do Brasil**: subsídios para planejamento ambiental. São Paulo: Oficina de Textos, 2006. 43, 54

ROUSE, J. W. J.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the Great Plains with ERTS. In: **Proceedings...** Washington, D.C.: NASA. Goddard Space Flight Center, 1974. p. 309–317. Disponível em: <<https://ntrs.nasa.gov/search.jsp?R=19740022614>>. Acesso em: 13 mar. 2018. 21

ROY, D. P.; BOSCHETTI, L.; JUSTICE, C. O.; JU, J. The collection 5 modis burned area product global evaluation by comparison with the modis active fire product. **Remote Sensing of Environment**, Elsevier, v. 112, n. 9, p. 3690–3707, 2008. 27

SABINS, F. F. **Remote sensing**: principles and applications. 3. ed. United States: Waveland Press, 2007. 512 p. 7

SAMAL, A.; BHATIA, S.; VADLAMANI, P.; MARX, D. Searching satellite imagery with integrated measures. **Pattern Recognition**, Elsevier, v. 42, n. 11, p. 2502–2513, 2009. ISSN 0031-3203. Disponível em: <<http://dx.doi.org/10.1016/j.patcog.2009.01.035>>. Acesso em: 05 jun. 2017. 23, 25

SANTOS, T. C.; CÂMARA, J. B. **Geo Brasil 2002**: perspectivas do meio ambiente no Brasil. Brasília: Edições Ibama, 2002. 46

SATO, L. Y. **Tecnologia LIDAR para quantificação dos impactos de inêndios na estrutura**. 98 p. Tese (Doutorado) — Instituto Nacional de Pesquisas Espaciais, 2016. Disponível em: <<http://urlib.net/8JMKD3MGP3W34P/3M8GSQ2>>. Acesso em: 13 out. 2017. 4, 26

SCHRÖDER, M.; DIMAI, A. Texture information in remote sensing images: a case study. In: **Proceedings...** Zürich: Citeseer, 1998. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.20.3528>>. Acesso em: 13 fev. 2018. 2, 22

SCHUH, M. A.; BANDA, J. M.; WYLIE, T.; MCINERNEY, P.; PILLAI, K. G.; ANGRYK, R. A. On visualization techniques for solar data mining. **Astronomy and Computing**, Elsevier B.V., v. 10, p. 32–42, 2015. ISSN 22131337. Disponível em: <<http://dx.doi.org/10.1016/j.ascom.2014.12.003>>. Acesso em: 03 fev. 2017. 14, 15

SEIDEL, K.; SCHRODER, M.; REHRAUER, H.; SCHWARZ, G.; DATCU, M. Query by image content from remote sensing archives. In: **Proceedings...** Seattle, USA: IEEE, 1998. v. 1, p. 393–396. Disponível em: <<http://ieeexplore.ieee.org/document/702916/>>. Acesso em: 08 jan. 2017. 22

SEIDEL, K.; THERRE, J.-P. Remote sensing image archive for national user requirements. In: **Proceedings...** Marino, Rome: Citeseer, 1994. v. 37. Disponível em: <http://www.vision.ee.ethz.ch/publications/pub_readall.cgi?lang=en&authors=Seidel>. Acesso em: 11 jan. 2017. 22

SHIMABUKURO, Y. E.; DUARTE, V.; ARAI, E.; FREITAS, R.; LIMA, A.; VALERIANO, D.; BROWN, I.; MALDONADO, M. Fraction images derived from terra modis data for mapping burnt areas in brazilian amazonia. **International Journal of Remote Sensing**, Taylor & Francis, v. 30, n. 6, p. 1537–1546, 2009. 27, 28

SHIMABUKURO, Y. E.; MIETTINEN, J.; BEUCHLE, R.; GRECCHI, R. C.; SIMONETTI, D.; ACHARD, F. Estimating burned area in Mato Grosso, Brazil, using an object-based classification method on a systematic sample of medium resolution satellite images. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 8, n. 9, p. 4502–4508, 2015. ISSN 21511535. 26

- SHIMABUKURO, Y. E.; SMITH, J. A. The least-squares mixing models to generate fraction images derived from remote sensing multispectral data. **IEEE Transactions on Geoscience and Remote Sensing**, v. 29, n. 1, p. 16–20, 1991. ISSN 15580644. Disponível em: <ieeexplore.ieee.org/document/103288>. Acesso em: 24 fev. 2017. 25, 28
- SHYU, C. R.; KLARIC, M.; SCOTT, G. J.; BARB, A. S.; DAVIS, C. H.; PALANIAPPAN, K. GeoIRIS: geospatial information retrieval and indexing system mdash; content mining, semantics modeling, and complex queries. **IEEE Transactions on Geoscience and Remote Sensing**, v. 45, n. 4, p. 839–852, 2007. ISSN 0196-2892. 15, 16, 23, 25
- SILVEIRA, H. L. F. d.; EBERHARDT, I. D. R.; SANCHES, I. D. A.; GALVÃO, L. S. Análise da cobertura de nuvens no nordeste do brasil e seus impactos no sensoriamento remoto agrícola operacional. In: GHERARDI, D. F. M.; ARAGÃO, L. E. O. e. C. d. (Ed.). **Anais...** São José dos Campos: Instituto Nacional de Pesquisas Espaciais (INPE), 2017. p. 400–407. ISBN 978-85-17-00088-1. Disponível em: <<http://urlib.net/sid.inpe.br/marte2/2017/10.23.17.05>>. Acesso em: 13 mar. 2018. 44
- STEFFEN, C. A.; MORAES, E. C.; GAMA, F. F. Radiometria óptica espectral. **Simpósio Brasileiro de Sensoriamento Remoto**, v. 8, p. 1–1, 1996. 7
- STEPINSKI, T. F.; NETZEL, P.; JASIEWICZ, J. LandEx: a geoweb tool for query and retrieval of spatial patterns in land cover datasets. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 7, n. 1, p. 257–266, 2014. ISSN 19391404. Disponível em: <<http://sil.uc.edu/pdfFiles/jstars2013.pdf>>. Acesso em: 10 jun. 2017. 2
- THÉRY, H.; MELLO, N. D. **Atlas do Brasil**. São Paulo: Edusp, 2005. 43, 46
- USGS, U. S. Geological Survey. **Landsat 8 (L8): data users handbook**. 2016. Disponível em: <<https://landsat.usgs.gov/documents/Landsat8DataUsersHandbook.pdf>>. Acesso em: 07 mar. 2018. 8
- VEGANZONES, M. A.; DATCU, M.; GRAÑA, M. Further results on dissimilarity spaces for hyperspectral images rf-cbir. **Pattern Recognition Letters**, Elsevier, v. 34, n. 14, p. 1659–1668, 2013. 23, 25
- VEGANZONES, M. A.; MALDONADO, J. O.; GRAÑA, M. On content-based image retrieval systems for hyperspectral remote sensing images. In: _____.

Computational Intelligence for Remote Sensing. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008. cap. 5, p. 125–144. ISBN 978-3-540-79353-3. Disponível em: <http://dx.doi.org/10.1007/978-3-540-79353-3_{_}5>. Acesso em: 15 jan. 2017. 14, 15, 17, 22, 39, 40

WULDER, M. A.; MASEK, J. G.; COHEN, W. B.; LOVELAND, T. R.; WOODCOCK, C. E. Opening the archive: How free data has enabled the science and monitoring promise of Landsat. **Remote Sensing of Environment**, Elsevier, v. 122, p. 2–10, 2012. 28

XIE, Z.; ROBERTS, C.; JOHNSON, B. Object-based target search using remotely sensed data: a case study in detecting invasive exotic Australian Pine in south Florida. **ISPRS Journal of Photogrammetry and Remote Sensing**, Elsevier, v. 63, n. 6, p. 647–660, 2008. 22, 23

YANG, Y.; NEWSAM, S. Geographic image retrieval using local invariant features. **IEEE Transactions on Geoscience and Remote Sensing**, v. 51, n. 2, p. 818–832, 2013. ISSN 0196-2892. Disponível em: <<http://ieeexplore.ieee.org/document/6257473/>>. Acesso em: 16 jan. 2017. 1

ZHU, Z.; WANG, S.; WOODCOCK, C. E. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 47, 8, and Sentinel 2 images. **Remote Sensing of Environment**, v. 159, p. 269–277, 2015. ISSN 00344257. Disponível em: <<http://www-sciencedirect-com.ez61.periodicos.capes.gov.br/science/article/pii/S0034425714005069>>. Acesso em: 15 jan. 2017. 20

ZHU, Z.; WOODCOCK, C. E. Object-based cloud and cloud shadow detection in Landsat imagery. **Remote Sensing of Environment**, v. 118, p. 83–94, 2012. ISSN 00344257. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0034425711003853>>. Acesso em: 03 fev. 2017. 20, 28

PUBLICAÇÕES TÉCNICO-CIENTÍFICAS EDITADAS PELO INPE

Teses e Dissertações (TDI)

Teses e Dissertações apresentadas nos Cursos de Pós-Graduação do INPE.

Manuais Técnicos (MAN)

São publicações de caráter técnico que incluem normas, procedimentos, instruções e orientações.

Notas Técnico-Científicas (NTC)

Incluem resultados preliminares de pesquisa, descrição de equipamentos, descrição e ou documentação de programas de computador, descrição de sistemas e experimentos, apresentação de testes, dados, atlas, e documentação de projetos de engenharia.

Relatórios de Pesquisa (RPQ)

Reportam resultados ou progressos de pesquisas tanto de natureza técnica quanto científica, cujo nível seja compatível com o de uma publicação em periódico nacional ou internacional.

Propostas e Relatórios de Projetos (PRP)

São propostas de projetos técnico-científicos e relatórios de acompanhamento de projetos, atividades e convênios.

Publicações Didáticas (PUD)

Incluem apostilas, notas de aula e manuais didáticos.

Publicações Seriadas

São os seriados técnico-científicos: boletins, periódicos, anuários e anais de eventos (simpósios e congressos). Contam destas publicações o Internacional Standard Serial Number (ISSN), que é um código único e definitivo para identificação de títulos de seriados.

Programas de Computador (PDC)

São a seqüência de instruções ou códigos, expressos em uma linguagem de programação compilada ou interpretada, a ser executada por um computador para alcançar um determinado objetivo. Aceitam-se tanto programas fonte quanto os executáveis.

Pré-publicações (PRE)

Todos os artigos publicados em periódicos, anais e como capítulos de livros.