



**UNESP – UNIVERSIDADE ESTADUAL PAULISTA**  
**CAMPUS DE PRESIDENTE PRUDENTE**  
**FACULDADE DE CIÊNCIAS E TECNOLOGIA**  
**PROGRAMA DE PÓS GRADUAÇÃO EM CIÊNCIAS CARTOGRÁFICAS**

**LUIZ EDUARDO CHRISTOVAM DE SOUZA**



**ORGANIZAÇÃO E ARMAZENAMENTO DE IMAGENS MULTITEMPORAIS**  
**GEORREFERENCIADAS PARA SUPORTE AO PROCESSO DE DETECÇÃO DE**  
**MUDANÇAS**

**PRESIDENTE PRUDENTE**  
**2018**

**LUIZ EDUARDO CHRISTOVAM DE SOUZA**

**ORGANIZAÇÃO E ARMAZENAMENTO DE IMAGENS MULTITEMPORAIS  
GEORREFERENCIADAS PARA SUPORTE AO PROCESSO DE DETECÇÃO  
DE MUDANÇAS**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Cartográficas da Faculdade de Ciências e Tecnologia da Universidade Estadual Paulista, para obtenção do título de Mestre em Ciências Cartográficas.

Orientador:

Dra. Maria de Lourdes B. T. Galo

Coorientador:

Dr. Milton Hirokazu Shimabukuro

S729o

Souza, Luiz Eduardo Christovam de

Organização e armazenamento de imagens multitemporais georreferenciadas para suporte ao processo de detecção de mudanças / Luiz Eduardo Christovam de Souza. -- Presidente Prudente, 2018  
90 p.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências e Tecnologia, Presidente Prudente

Orientadora: Maria de Lourdes Bueno Trindade Galo

Coorientador: Milton Hirokazu Shimabukuro

1. Banco de dados híbrido. 2. NoSQL. 3. Séries temporais de imagens. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências e Tecnologia, Presidente Prudente. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

**CERTIFICADO DE APROVAÇÃO**

TÍTULO DA DISSERTAÇÃO: ORGANIZAÇÃO E ARMAZENAMENTO DE IMAGENS MULTITEMPORAIS  
GEORREFERENCIADAS PARA SUPORTE AO PROCESSO DE DETECÇÃO  
DE MUDANÇAS

**AUTOR: LUIZ EDUARDO CHRISTOVAM DE SOUZA**

**ORIENTADORA: MARIA DE LOURDES BUENO TRINDADE GALO**

**COORIENTADOR: MILTON HIROKAZU SHIMABUKURO**

Aprovado como parte das exigências para obtenção do Título de Mestre em CIÊNCIAS  
CARTOGRÁFICAS, área: AQUISIÇÃO, ANÁLISE E REPRESENTAÇÃO DE INFORMAÇÕES  
ESPACIAIS pela Comissão Examinadora:



Profa. Dra. MARIA DE LOURDES BUENO TRINDADE GALO  
Departamento de Cartografia / Faculdade de Ciências e Tecnologia de Presidente Prudente



Prof. Dr. JEFERSSON ALEX DOS SANTOS  
Departamento de Ciência da Computação / Universidade Federal de Minas Gerais



Prof. Dr. NILTON NOBUHIRO IMAI  
Departamento de Cartografia / Faculdade de Ciências e Tecnologia de Presidente Prudente

Presidente Prudente, 01 de março de 2018

## **DADOS CURRICULARES**

### **Luiz Eduardo Christovam de Souza**

<b>Nascimento</b>	10/04/1991 – Presidente Prudente – SP
<b>Filiação</b>	Luiz Antônio de Souza Rosemarie Christovam de Souza
<b>2010 - 2015</b>	Graduação Bacharelado em Engenharia Cartográfica Faculdade de Ciências e Tecnologia – UNESP
<b>2016 - 2018</b>	Pós-Graduação Mestrado em Ciências Cartográficas Faculdade de Ciências e Tecnologia - UNESP

*À minha amada família por todo amor e  
dedicação transmitidos ao longo de minha  
vida.*

## AGRADECIMENTOS

Agradeço à Deus por me proporcionar a dádiva da vida, e por me conduzir ao longo dessa caminhada, nem sempre pelos caminhos mais fáceis, mas sempre pelos caminhos de maior aprendizado.

À minha orientadora doutora Maria de Lourdes e coorientador doutor Milton Shimabukuro, pela confiança, paciência e por todos os conhecimentos compartilhados nesse período de trabalho conjunto.

Aos meus pais, Rose e Luiz, por toda dedicação e sacrifício realizados em prol da minha criação. À minha avó Cecília pelos conselhos e sacrifícios, que foram fundamentais para minha formação profissional e pessoal. À minha avó Luzia (*in memoriam*) por me mostrar que a bondade habita no coração das pessoas mais simples. A minha irmã, Ana Lúcia, por me apoiar em todos os momentos. A minha tia Rejiane e a minha madrinha/tia Célia, por todos os conselhos e puxões de orelha.

A minha amada namorada, Isis, pela compreensão, incentivo e conselhos durante esse período, e principalmente pelo exemplo de força e generosidade. A sua família, Dona Lourdes, Sandra, Zé (*in memoriam*) e Gabriel, por me acolherem como um membro da família.

Aos amigos de longa data, Smash, Silas, Paulinho e Flávio, por toda a parceria ao longo de vários anos. Aos amigos que fiz na Unesp, Rapha, Gelo, Sato, Victor Kaneshiro, Bazan e Maurício. E aos amigos que continuaram ao meu lado na pós-graduação, Gabriel, Lucas, Tobias, Alemão, Carioca e Emerson, obrigado pelos momentos de diversão.

Aos companheiros de pós, Carol Campus, Carol Pifer, Sarah, Mari, Prol, Gabi, Renan, Samara, Fran, Ju e outros, pelas discussões e risadas que compartilhamos juntos na cozinha e nos churrascos. Agradeço em especial a minha primeira vizinha de baia, Raquel, pelas conversas matinais e por toda ajuda durante esse período.

À Cida e a Zilda, por manter o ambiente da pós-graduação sempre agradável, e em especial pela amizade construída ao longo desse período.

Aos doutores Nilton Imai e Jeferson Santos, pelas valiosas contribuições. À UNESP e ao PPGCC, pela formação e estrutura oferecida. À CAPES e ao CNPq pelo

financiamento por meio de bolsa. E a todos os demais professores e funcionários do Departamento de Cartografia.

Por fim, gostaria de agradecer a todos que de forma direta ou indireta, contribuíram para a realização desta pesquisa. Meus sinceros agradecimentos!

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.



*“The true sign of intelligence is not knowledge  
but imagination.”*

*Albert Einstein*

## RESUMO

Atualmente o volume de dados produzidos tem atingido patamares nunca imaginados, sobretudo em decorrência da multiplicação do número de sensores e da popularização da internet, com a *web* 2.0 e as redes sociais. Dentre os diversos tipos de sensores existentes, os de imageamento, transportados principalmente por satélites, produzem vastos conjuntos de observações da superfície da Terra. A observação contínua da Terra por satélites possibilita o monitoramento de mudanças no uso e cobertura da terra. Contudo, em diversas pesquisas relacionadas a mudanças no planeta, são utilizados apenas pequenos fragmentos do imenso conjunto de dados existente, essencialmente devido a ainda haver uma lacuna científica-tecnológica relacionada aos procedimentos de organização, armazenamento, análise e representação de grandes conjuntos de dados. Portanto, nessa pesquisa foi definida uma estrutura para organização, armazenamento e recuperação de dados espaço-temporais, com o propósito de fornecer suporte a detecção de mudanças na cobertura da terra. Para tanto, foi definida como aplicação a análise de séries temporais de *Normalized Difference Vegetation Index* (NDVI) derivadas de imagens adquiridas desde 1984 até 2017, pelos sensores *Thematic Mapper* (TM), *Enhanced Thematic Mapper Plus* (ETM+) e *Operational Land Imager* (OLI) para a região de Porto Velho, Rondônia. Foi construída uma série temporal de NDVI para a posição de cada pixel presente na área de estudo. Regiões de referência foram definidas para recuperar séries temporais de referência que descrevem os tipos de cobertura da terra e classes de mudanças (antrópicas ou naturais). O algoritmo *Fast Dynamic Time Warping* (FastDTW), foi adotado para medir a similaridade entre séries temporais, a serem classificadas e de referência. Para determinar a classe da série temporal, as distâncias obtidas com o FastDTW foram classificadas com o algoritmo *k-Nearest Neighbor* (kNN). A classificação de várias séries permitiu a construção do mapa de mudanças para uma área teste inserida na área de estudo. A respeito do banco de dados, esse foi definido com uma abordagem híbrida, foram utilizados o modelo relacional, que permitiu armazenar dados rigidamente estruturados, e o modelo de dados não relacional (NoSQL), que possibilitou armazenar dados semiestruturados. Essa estrutura de armazenamento foi fundamental para a automatização da aplicação, permitindo a organização, armazenamento e recuperação de um extenso conjunto de dados heterogêneos de forma consistente e segura. Com relação às séries temporais foi possível observar que de maneira geral o perfil temporal de NDVI pode descrever o tipo de cobertura da terra, entretanto efeitos atmosféricos, nuvens, fumaça, entre outros, exercem influência no comportamento da série temporal. Foram realizadas duas classificações, com tamanhos de raio iguais a 1 e 20 para o FastDTW. A concordância global dos mapas produzidos foi de 58,06% e 55,02%, respectivamente. Apesar desses valores serem relativamente baixos, esses resultados são justificados pelo caráter subjetivo presente na caracterização de mudanças temporais, pela não filtragem das séries temporais e pela elevada variação no tamanho das séries.

**Palavras chave:** Banco de Dados Híbrido, NoSQL, Séries Temporais de Imagens, LUCC, DTW.

## ABSTRACT

Nowadays the size of datasets has been reaching levels never seen before, mainly due to new sensors and the widespread of the internet, with web 2.0 and social media. Among the various types of sensors, the imaging sensors, mainly carried by satellites, have produced big Earth observations datasets. The regular Earth observation by satellites enable to monitor Land Use/Cover Change (LUCC). However, in many researches related to LUCC, only small parts of the big Earth Observation datasets are normally used, because there is still a scientific-technological gap related to the organization, storage, analysis and representation of big Earth Observations data. Therefore, in this research was defined a database for the organization, storage and retrieval of spatio-temporal data, to support a LUCC task. Therefore, the time series analysis of Normalized Difference Vegetation Index (NDVI) of images acquired from 1984 to 2017 by Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+) and Operational Land Imager (OLI) for the region of Porto Velho, Rondônia was defined as the application. To the position each of the pixel in the study area was built a NDVI time series. Reference areas were defined to retrieve reference time series that describe the land cover types and the change classes (anthropic and natural). The Fast Dynamic Time Warping (FastDTW) algorithm was used to measure the similarity between the time series, to be classified and reference ones. To find the time series class, the distances previously found by FastDTW were classified by k-Nearest Neighbor (kNN) algorithm. Classifying several time series allowed the construction of LUCC map in a test area. Regarding the database, this was defined with a hybrid approach, the relational data model allowed to store highly structured data, and the non-relational data model (NoSQL) allowed store semi-structured data. The database was important in automating the application, enabling the organization, storage and retrieval of an extensive heterogeneous dataset in a consistent and secure way. Regarding the time series, it was possible to observe that in general the NDVI time series can describe land cover types, however atmospheric effects, clouds, smoke, among others, can influence the way of the time series. Two classifications were performed using the FastDTW, the first one with radio size equal to 1 and for the second the radio was 20. The overall accuracy of the maps was 58.06% and 55.02%, respectively. Although these are relatively low values, they are justified by the subjective presented in the temporal changes characterization, by the use of the raw time series and by the high variation in the time series size.

**Keywords:** Hybrid database, NoSQL, Remote Sensing Image Time series, LUCC, DTW.

## LISTA DE FIGURAS

Figura 01 - Imagens de sensoriamento remoto e série temporal para o pixel de posição (x,y) em destaque nas imagens em destaque no conjunto de imagens adquiridas em diferentes tempos .....	23
Figura 02 - Alinhamento realizado para medida de similaridade entre séries temporais por meio dos métodos Distância Euclidiana e DTW, respectivamente .....	24
Figura 03 - Alinhamento e caminho de menor distância entre as séries temporais determinados com o DTW .....	25
Figura 04 - Exemplo de modelo relacional: Tabela ALUNO .....	29
Figura 05 - Representação de banco de dados família de colunas .....	34
Figura 06 - Representação de banco de dados chave-valor .....	35
Figura 07 - Representação de banco de dados orientado a documentos .....	35
Figura 08 - Representação de banco de dados grafos com uso em rede social .....	36
Figura 09 - Inserção da área de estudo no território brasileiro e sua delimitação em composição falsa-cor com as bandas espectrais do vermelho, infravermelho próximo e médio (4B5G6R) de imagem OLI/Landsat 8 .....	42
Figura 10 - Mosaico de Imagens OLI 4B5G6R - Cenas 232/066 do dia 02/07/2016 e 233/066 do dia 09/07/2016 .....	44
Figura 11 - (A) Quantidade de cenas por sensor (B) Volume de espaço ocupado pelas cenas em gigabytes por sensor .....	45
Figura 12 - Fluxograma do desenvolvimento geral da pesquisa .....	47
Figura 13 - Fluxograma para o desenvolvimento do projeto do banco de dados .....	50
Figura 14 - Diagrama entidade relacionamento simplificado .....	51
Figura 15 - Tiles (quadrículas) da Banda 5 (infravermelho próximo) de uma imagem OLI/Landsat 8, órbita/ponto 232/066 do dia 14/06/2015 .....	52
Figura 16 - Diagrama OMT-G para o banco de dados híbrido .....	53
Figura 17 - Estruturas genéricas dos atributos metadata, bands e data armazenados no formato de documentos .....	56
Figura 18 - Etapas desenvolvidas no de pré-processamento dos dados .....	57
Figura 19 - Ilustração de imagem com 80% de cobertura de nuvens, referente a uma cena TM/Landsat , da banda 4 (infravermelho próximo), órbita/ponto 232/066 do dia 03/01/1991 .....	59

Figura 20 - Órbitas/ponto 233/066 e 232/066 dos dias 09/07/2016 e 02/07/2016 respectivamente, da imagem OLI/Landsat 8, banda do infravermelho próximo, destacando a área de estudo.....	59
Figura 21 - Ilustração da aplicação de máscara de nuvens e sombra de nuvens na banda do infravermelho próxima de imagem OLI/Landsat 8, órbita/ponto 232/066, do dia 15/05/2016 - cena com alta cobertura de nuvens.....	61
Figura 22 - Ilustração da aplicação de máscara de nuvens e sombra de nuvens na banda do infravermelho próxima de imagem OLI/Landsat 8, órbita/ponto 232/066, do dia 14/06/2015 -cena com pequena cobertura de nuvens.....	61
Figura 23 - Fluxograma das atividades relacionadas com a geração de séries temporais com valores de NDVI para cada pixel da imagem, corrigidos para reflectância TOA .	63
Figura 24 - Representação das séries temporais de NDVI obtidas para os pixels com id 819700, 6040477 e 10813717.....	65
Figura 25 - Etapas relacionadas com a identificação de áreas de referência.....	66
Figura 26 - Distribuição das regiões de referência na área de estudo .....	69
Figura 27 - Fases para a classificação de um pixel por meio da análise de séries temporais...	70
Figura 28 - Séries Temporais de NDVI para os tipos de cobertura do solo inseridos na classe Sem Mudanças.....	75
Figura 29 - Séries Temporais de NDVI indicando as transições entre tipos de cobertura da terra observadas na classe Mudanças Antrópicas.....	77
Figura 30 - Séries Temporais de NDVI para alterações nos tipos de cobertura do solo inseridos na classe Mudanças Naturais.....	78
Figura 31 - Classificação FastDTW com raio igual a 1 e com 154 séries temporais de referência .....	79
Figura 32 - Classificação FastDTW com raio igual a 20 e com 154 séries temporais de referência .....	79

## LISTA DE TABELAS

Tabela 01 - Tempo médio (em segundos) para a classificação de um pixel, variando o algoritmo, tamanho dos conjuntos de séries temporais de referência e número de núcleos de processamento utilizados.....	72
Tabela 02 - Matriz de Concordância para a Classificação FastDTW com raio igual a 1 e com 154 séries temporais de referência .....	80
Tabela 03 - Matriz de Concordância para a Classificação FastDTW com raio igual a 20 e com 154 séries temporais de referência .....	80

## LISTA DE QUADROS

Quadro 01 - Exemplos de SGBD relacionais .....	28
Quadro 02 - SGBD não relacionais .....	33
Quadro 03 - Características das imagens adquiridas pelos sensores TM, ETM+, OLI. NIR (near infrared) se refere à banda do infravermelho próximo, SWIR (shortwave infrared) é o infravermelho médio e TIR (thermal infrared sensor) é o sensor do infravermelho termal do Landsat 8 .....	46
Quadro 04 - Regiões de referência dos tipos de cobertura da terra que caracterizam áreas sem mudança, áreas de mudança antrópica e de mudança natural, identificadas nos anos de 1985, 1990, 1995, 2000, 2005, 2010 e 2015 .....	67

## LISTA DE SIGLAS

ACID	Atomicidade, Consistência, Isolamento e Durabilidade
BASE	<i>Basically Available, Soft State and Eventual Consistency</i>
DER	Diagrama Entidade-Relacionamento
DTW	<i>Dynamic Time Warping</i>
ETM+	<i>Enhanced Thematic Mapper Plus</i>
FastDTW	<i>Fast Dynamic Time Warping</i>
FDS	<i>Flexible Schema Data</i>
GDAL	<i>Geospatial Data Abstraction Library</i>
GNU	<i>General Public License</i>
JSON	<i>JavaScript Object Notation</i>
JSONB	<i>Binary JavaScript Object Notation</i>
kNN	<i>K-Nearest Neighbors</i>
L1T	<i>Pre-Collection Level 1</i>
LGAC	<i>Landsat Global Archive Consolidation</i>
LUCC	<i>Land Use/Cover Change</i>
ND	Número Digital
NDVI	<i>Normalized Difference Vegetation Index</i>
NIR	<i>Near Infrared</i>
NoSQL	Modelo de Dados Não Relacional
OLI	<i>Operational Land Imager</i>
OMT-G	<i>Object Modeling Technique for Geographic Applications</i>
OSGeo	<i>Open Source Geospatial Foundation</i>
SGBD	Sistema Gerenciador de Banco de Dados
SGBDR	Sistema Gerenciador de Banco de Dados Relacional
SIG	Sistemas de Informações Geográficas
SQL	<i>Structured Query Language</i>
STC	<i>Space Time Composite</i>
SWIR	<i>Shortwave Infrared</i>
TI	Tecnologia da Informação
TIRS	<i>Thermal Infrared Sensor</i>
TM	<i>Thematic Mapper</i>



TOA	<i>Top Of Atmosphere</i>
UCB	Universidade da Califórnia em Berkelei
UHE	Usina Hidrelétrica
USGS	<i>United States Geological Survey</i>
UTM	Universal Transversa de Mercator
WGS 84	<i>World Geodetic System 1984</i>
XML	<i>eXtensible Markup Language</i>

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>19</b>
1.1	Contextualização.....	19
1.2	Objetivos.....	21
1.3	Organização do texto.....	21
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>22</b>
2.1	Monitoramento da Cobertura da Terra por Sensoriamento Remoto.....	22
2.1.1	<i>Dynamic Time Warping (DTW)</i> .....	24
2.2	Bancos de dados.....	27
2.2.1	Modelo de dados relacionais.....	28
2.2.2	Modelo de dados não relacionais.....	31
2.2.3	Multimodelo de dados.....	36
2.2.4	Modelo de dados espaço-temporais.....	38
2.2.5	PostgreSQL.....	39
<b>3</b>	<b>MATERIAIS E MÉTODOS.....</b>	<b>42</b>
3.1	Área de estudo.....	42
3.2	Materiais.....	43
3.2.1	Imagens Landsat.....	43
3.3	Metodologia.....	47
3.3.1	Projeto do banco de dados.....	49
3.3.2	Pré-processamento dos dados.....	56
3.3.3	Construção das Séries Temporais.....	62
3.3.4	Identificação de Regiões de Referência.....	65
3.3.5	Discriminação de mudanças nos tipos de cobertura da terra.....	70

<b>4</b>	<b>APRESENTAÇÃO E ANÁLISE DOS RESULTADOS.....</b>	<b>74</b>
4.1	Banco de dados com abordagem híbrida .....	74
4.2	Séries Temporais de Referência.....	75
4.3	Classificação das mudanças na área teste .....	79
<b>5</b>	<b>CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES.....</b>	<b>82</b>
	<b>REFERÊNCIAS .....</b>	<b>85</b>

## 1 INTRODUÇÃO

### 1.1 Contextualização

O desenvolvimento de novos sensores e a popularização da internet a partir dos anos 2000, com a *web 2.0* e as redes sociais, iniciaram a era *big data*, período em que conjuntos de dados têm crescido a altas taxas de velocidade e elevada variabilidade. A informação existente nesses conjuntos de dados é potencialmente proporcional ao volume disponível para análise. Contudo, organizar, armazenar e processar esses dados com o propósito de extrair informação, são atividades que se tornam complexas, à medida que o volume de observações cresce e integra dados provenientes de diferentes sensores com tipos e formatos distintos (LI *et al.*, 2016).

Plataformas orbitais transportam sensores, muitos deles de imageamento, que têm contribuído na geração de extensos conjuntos de dados geoespaciais. Satélites do programa Landsat carregam sensores que vêm adquirindo imagens multiespectrais há mais de 40 anos, fazendo dessa a mais longa missão de observação por satélite da Terra (LOVELAND E IRONS, 2016). A política de acesso gratuito aos dados adotada pela *United States Geological Survey* (USGS) tem disponibilizado para pesquisa e para atividades operacionais um grande conjunto de imagens. Segundo Wulder *et al.* (2016), em 1 de janeiro de 2015, o centro de armazenamento de dados da USGS, o *Landsat Global Archive Consolidation* (LGAC), compreendia 5.532.454 cenas da superfície terrestre, totalizando 4.134 *terabytes* de dados armazenados.

Imagens adquiridas pelas missões Landsat compõe um conjunto de dados de grandes proporções e que seguirá crescendo, visto que os satélites Landsat 7 e 8 continuam em operação e o Landsat 9 está em desenvolvimento. A continuidade do programa Landsat e a existência de outros satélites de imageamento disponibilizam observações da superfície terrestre de forma contínua, e que permitem o entendimento da dinâmica do planeta por meio do monitoramento de mudanças no uso e cobertura do solo no mesmo. Todavia, apenas pequenos fragmentos desse grande volume de observações têm sido empregados em atividades relacionadas ao monitoramento da superfície terrestre, em razão da capacidade científica-tecnológica de extrair informação de grandes conjuntos de dados ainda ser inferior quando comparada com a capacidade para produzir observações.

Dentre as limitações que dificultam a análise completa das observações da superfície terrestre, está a organização e o armazenamento de grandes conjuntos de dados, que apresentam grande heterogeneidade. Imagens obtidas por sensores embarcados nos satélites Landsat, assim como as derivadas de outros sensores, embarcados em satélites desenvolvidos pelos Estados Unidos, União Europeia, China, Brasil e Índia tem motivado a comunidade científica a investigar questões relacionadas à organização e ao armazenamento dessas observações, com o objetivo de definir soluções estáveis e eficientes que deem suporte aos procedimentos de análise desses conjuntos de dados (CAMARA *et al.*, 2016). No entanto, até o presente momento, não há uma forma amplamente difundida de como integrar todo o conjunto de dados em uma estrutura de armazenamento, operação que é desafiada pelo volume de dados e pelas características distintas das imagens provenientes de diferentes sensores.

Krčál e Ho (2015) e Camara *et al.* (2016) sugerem o emprego de bancos de dados matriciais, como o SciDB, como alternativa à organização e ao armazenamento de dados matriciais, como imagens de sensoriamento remoto, em razão da natureza das imagens e do banco de dados coincidirem. Contudo, esse tipo de banco de dados tem como limitações, a impossibilidade de associar um sistema de referência espacial às matrizes, as células das matrizes não possuem informação de resolução, não é possível armazenar outros tipos de dados espaciais, assim como realizar operações espaciais.

Bancos de dados não relacionais (NoSQL) surgiram como a principal alternativa de armazenamento para aplicações *big data*, pelo fato de serem escaláveis e flexíveis, o que permite o armazenamento distribuído de dados semiestruturados ou sem estrutura (CHEN *et al.*, 2014). Entretanto, para proporcionarem essas características, os bancos de dados NoSQL abdicam da forte consistência, própria dos modelos de dados relacionais. Por essa razão, bancos de dados relacionais e não relacionais devem ser vistos como modelos de armazenamento complementares (AMIRIAN *et al.*, 2014). Logo, combinar diferentes modelos de dados em um único sistema de banco de dados pode resultar em uma estrutura de armazenamento mais adequada e eficiente.

A série histórica de dados Landsat oportuniza estudos relacionados com a detecção de modificações na superfície terrestre. Entretanto, a utilização de todo conjunto de dados, obtidos ao longo de mais de 40 anos de operação do sistema Landsat, é de grande complexidade, pelo fato das imagens serem adquiridas por sensores com características distintas e por constituírem um extenso volume de dados a ser analisado. Diante do exposto, considera-se que o monitoramento de mudanças temporais a partir das imagens que compõe a

série histórica Landsat é possível por meio da organização e do armazenamento dessas imagens em um banco de dados espaço-temporal concebido com base em uma abordagem híbrida (relacional e não relacional). Com o propósito de avaliar a questão proposta, foi definida como aplicação a detecção de mudanças por meio da classificação de séries temporais de imagens de satélite.

## 1.2 Objetivos

O objetivo geral desta pesquisa consiste em definir uma estrutura de armazenamento e recuperação de dados espaço-temporais, com o propósito de detectar mudanças na cobertura da terra por meio da análise de um conjunto de séries temporais derivadas de imagens multiespectrais georreferenciadas. Para isso, foram definidos os seguintes objetivos específicos:

- Conceber e implementar a estrutura de armazenamento semi-estruturada, baseada em banco de dados espaço-temporal com abordagem híbrida, relacional e não relacional (NoSQL);
- Identificar mudanças na cobertura do solo na área de estudo, por meio da análise de similaridade entre séries temporais.

## 1.3 Organização do texto

A dissertação está estruturada em 5 capítulos. No Capítulo 2 é apresentada a fundamentação teórica, abordando conceitos relacionados ao monitoramento de mudanças na cobertura da terra e bancos de dados. Com relação aos bancos de dados, é discorrido principalmente sobre os modelos de dados (relacional, não relacional, multimodelo e espaço-temporal) que geram diferentes possibilidades para a organização e armazenamento de dados. No Capítulo 3, materiais e métodos, é indicada a área de estudo, são caracterizadas as imagens Landsat, suas particularidades e é apontada a metodologia desenvolvida. Essa é desmembrada nas etapas de projeto do banco de dados, pré-processamento, construção das séries temporais, identificação de áreas de referência e discriminação de mudanças. No Capítulo 4, são apresentados os resultados e é realizada a análise dos mesmos, e por fim, no Capítulo 5 são realizadas as considerações finais e as recomendações para estudos futuros.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Monitoramento da Cobertura da Terra por Sensoriamento Remoto

Mapear e monitorar mudanças no uso e na cobertura da terra (*Land Use/Cover Change* - LUCC) é reconhecido como um importante objetivo da comunidade científica (ZHU E WOODCOCK, 2014). A cobertura do planeta tem influência no balanço energético, nos créditos de carbono e nos ciclos hidrológicos, portanto LUCC estão diretamente relacionados com o desenvolvimento sustentável (XIUBIN, 1996). É possível citar como exemplos de estudos relacionados a LUCC, as mudanças climáticas associadas ao desmatamento (GUIMBERTEAU *et al.*, 2016) e as alterações na qualidade do ar em consequência da expansão urbana (SUN *et al.*, 2016).

O monitoramento de mudanças na superfície terrestre proporciona um melhor entendimento da dinâmica de alterações ambientais no planeta, causados por atividade antrópica ou mesmo pela dinâmica natural da Terra. Compreender o processo de mudanças no uso e ocupação da superfície terrestre, mesmo após a ocorrência de eventos antrópicos, como por exemplo o desmatamento, possibilita entender as motivações, que os desencadearam e desenvolver políticas que coíbam essa prática (MAUS, 2014). Portanto, a detecção de mudanças fornece suporte e permite orientar o gerenciamento dos recursos terrestres (ALQURASHI E KUMAR, 2013).

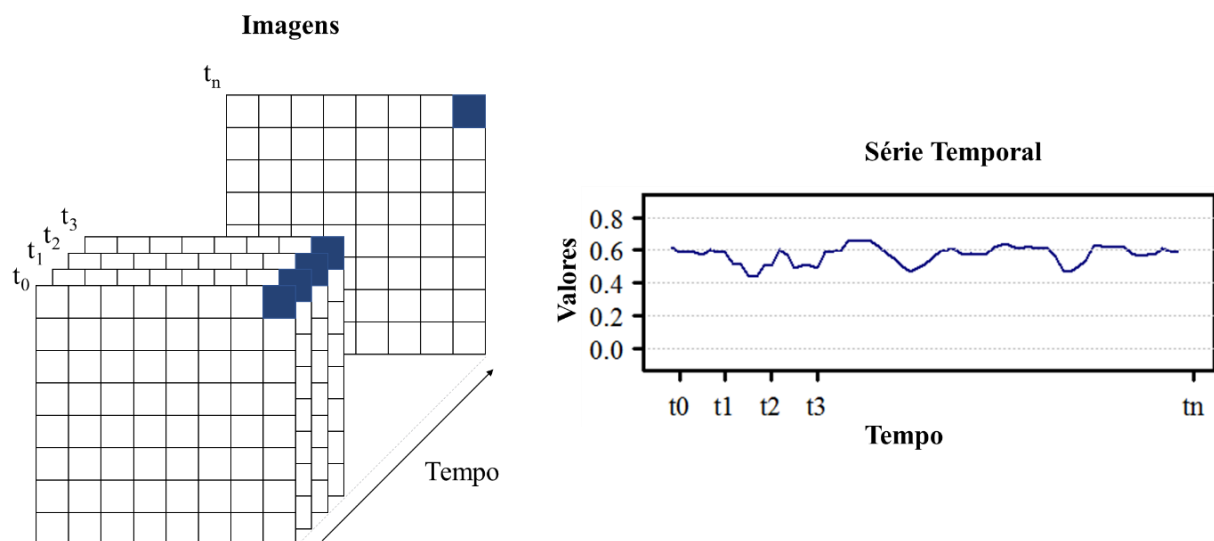
Imagens adquiridas por satélites têm sido utilizadas em estudos relacionados com a cobertura da superfície do planeta, principalmente em razão das características desses dados, como o imageamento de grandes áreas, a resolução temporal que possibilita o monitoramento de diversos fenômenos e a resolução espectral que permite a obtenção de informações de um mesmo alvo em distintas regiões do espectro.

Em aplicações de LUCC, a análise é realizada a partir de imagens multitemporais, visto que o objetivo nessas atividades é avaliar como os objetos ou fenômenos investigados se comportam ao longo do tempo (LU *et al.*, 2004). Contudo, até pouco tempo atrás, a maior parte dos estudos de detecção de mudanças empregavam poucas imagens na análise, em razão da dificuldade de se obter cenas livres de nuvens, da escassez de técnicas para extrair a informação a partir de grandes conjuntos de imagens, ou mesmo devido às limitações computacionais relacionadas com o armazenamento e com a extração dessas informações.

Recentemente, com o desenvolvimento de novos satélites e o aprimoramento na qualidade geométrica e nas resoluções espacial e temporal das imagens de sensoriamento remoto, maior atenção tem sido direcionada para o desenvolvimento de métodos de LUCC para conjuntos de dados maiores, como por exemplo, os baseados em séries temporais de imagens.

A aquisição sequencial de imagens a partir de sensores embarcados em satélites de sensoriamento remoto e a alta qualidade do seu georreferenciamento, permitem a composição de séries temporais para cada pixel da cena, como indicado na Figura 1. Para cada pixel da imagem, localizado em uma posição  $(x, y)$  específica, são adquiridas observações em datas consecutivas  $t_0, t_1, \dots, t_n$ , as quais indicam a variação ao longo do tempo de um determinado atributo, como por exemplo os valores de reflectância em uma determinada banda ou valores derivados de um índice espectral (MAUS, 2014; WANG *et al.*, 2017).

Figura 1 - Imagens de sensoriamento remoto e série temporal para o pixel de posição  $(x,y)$  em destaque nas imagens em destaque no conjunto de imagens adquiridas em diferentes tempos



Diversos estudos, (XIAO *et al.*, 2005; CONRAD *et al.*, 2011; MAUS, 2014; ZHU E WOODCOCK, 2014; MAUS *et al.*, 2016; WANG *et al.*, 2017), tem utilizado séries temporais de imagens de sensoriamento remoto para extrair informações relacionadas ao uso e cobertura da superfície terrestre. De maneira geral, parâmetros como média, desvio padrão, número de picos extraídos a partir dos períodos da série temporal, entre outros, orientam a especificação do tipo de uso e cobertura da terra (MAUS, 2014). Contudo, a realização de medidas de similaridade entre séries temporais é a principal estratégia para a detecção de



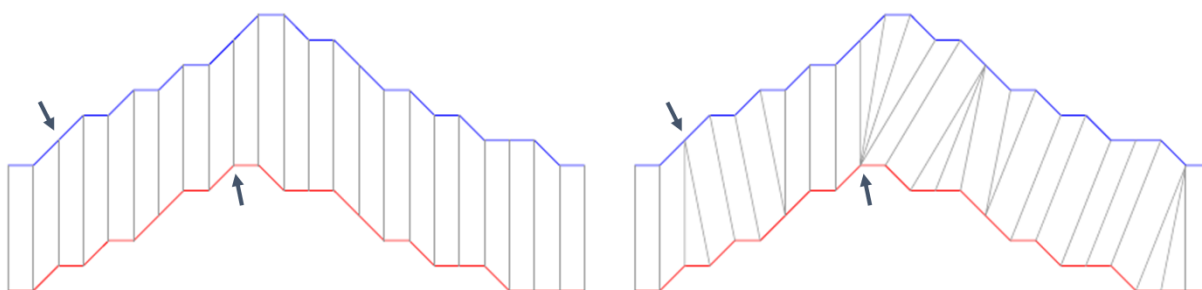
mudanças na superfície terrestre, e o algoritmo *Dynamic Time Warping* (DTW) tem sido indicado como o principal método para a medição de similaridade entre séries temporais (BERNDT E CLIFFORD, 1994; WANG *et al.*, 2017).

### 2.1.1 *Dynamic Time Warping* (DTW)

DTW é um algoritmo capaz de encontrar o melhor alinhamento entre duas séries temporais. É frequentemente utilizado para a determinação de similaridade, para a classificação e para indicação de regiões correspondentes entre séries temporais (SALVADOR E CHAN, 2007).

Ao contrário da distância euclidiana, em que *i-ésima* observação de uma série temporal é alinhada exatamente com a *i-ésima* observação de outra série, o DTW busca alinhar as observações mais similares nas duas séries temporais, o que pode provocar uma deformação no eixo do tempo. Desse modo, o DTW proporciona uma medida de distância confiável, independente das séries temporais apresentarem tamanhos distintos ou de estarem em fases diferentes no eixo do tempo (KEOGH E RATANAMAHATANA, 2005). A correspondência entre as observações no alinhamento entre as séries temporais para os métodos, distância euclidiana e DTW, podem ser observados na Figura 2, respectivamente. As setas na figura exemplificam duas diferenças na correspondência, mostrando que na distância euclidiana a observação de uma série temporal é alinhada exatamente com a observação na mesma posição na outra série temporal, enquanto no DTW a observação é alinhada com as mais similares a ela, podendo ser alinhada com uma ou mais observações.

Figura 2 - Alinhamento realizado para medida de similaridade entre séries temporais por meio dos métodos Distância Euclidiana e DTW, respectivamente

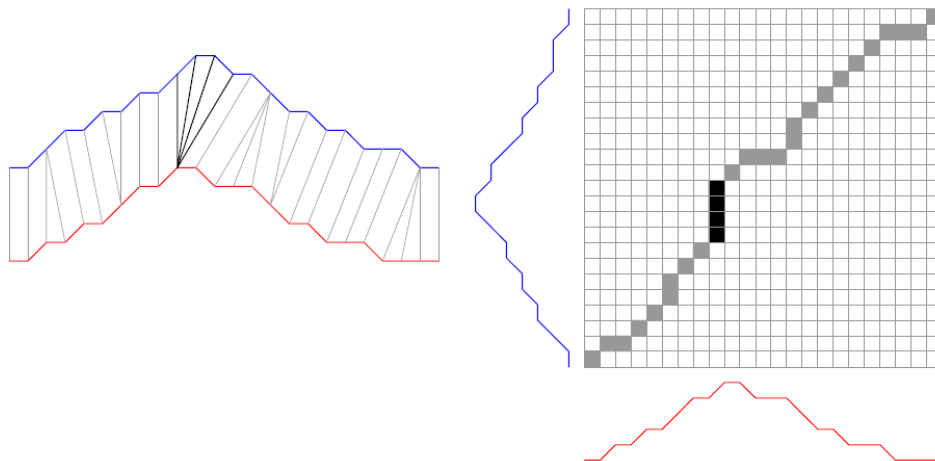


Considerando duas séries temporais  $T = (t_1, t_2, \dots, t_n)$  e  $U = (u_1, u_2, \dots, u_m)$ , a medida similaridade entre ambas utilizando o DTW é realizada por meio da construção da matriz  $D$ , denominada matriz de custo acumulado, a partir da qual é encontrado o caminho de menor distância (*minimum-distance warp path*). Para a determinação desse caminho, inicialmente, é computada a matriz  $V_{n,m}$  em que os elementos  $v_{i,j}$  são as distâncias entre  $t_i \in T \forall i = 1, \dots, n$  e  $u_j \in U \forall j = 1, \dots, m$ . A partir desse resultado é computada a matriz  $D$ , sendo os elementos dessa matriz calculados por meio da Equação 1, que corresponde a soma recursiva das distâncias mínimas em  $D$ .

$$d_{i,j} = v_{i,j} + \min\{v_{i-1,j}, v_{i-1,j-1}, v_{i,j-1}\} \quad (1)$$

Para o cálculo do menor caminho pela Equação 1, devem ser atendidas as condições de borda, continuidade e monotonicidade. A condição de borda demanda que o caminho de menor distância deve iniciar no primeiro elemento da matriz e terminar no último elemento. A restrição de continuidade impõe que o caminho de menor distância deve ser composto por elementos adjacentes para garantir a continuidade. Por fim, a condição de monotonicidade determina que as posições seguintes no caminho de menor distância devem ter sua base na posição anterior, independente da direção da coluna ou linha (WANG *et al.*, 2017). Na Figura 3, é apresentado o menor caminho entre duas séries temporais, com destaque para o comportamento no caminho quando uma observação tem correspondência com quatro observações.

Figura 3 - Alinhamento e caminho de menor distância entre as séries temporais determinados com o DTW



Como já mencionado, o DTW pode ser utilizado na classificação de séries temporais. Nesse processo, é feita a comparação entre a série temporal de um evento conhecido (referência) com a série temporal de um evento desconhecido (a ser classificado). Em aplicações de LUCC, em que existem diferentes tipos de cobertura da terra e possibilidades de mudanças entre essas coberturas, a classificação da série temporal de um pixel é realizada por meio da sua comparação com cada uma das séries temporais de referência, sendo a série temporal de interesse indicada para a classe que definir uma série temporal de menor distância, com base no DTW (PETITJEAN et al., 2012). Com isso, esse processo pode levar a uma grande quantidade de comparações, dependendo do número de pixels a serem classificados, bem como da quantidade de séries temporais de referência, o que pode levar a um processo computacionalmente lento.

Outro ponto que faz com que o DTW aplicado na classificação de séries temporais seja um processo lento é sua complexidade computacional, que segundo Salvador e Chan (2007) é quadrática,  $O(N^2)$ , para uma matriz de custo ( $D$ ) com dimensões  $(X, Y)$  se  $N = |X| = |Y|$ . Com o propósito de aumentar a velocidade de execução do DTW, foram desenvolvidas técnicas que podem ser categorizadas como de Restrições, de Abstração de dados e de Indexação.

Os métodos baseados em Restrições limitam o número de células (vizinhas) que são avaliados na matriz de custo, nos métodos que empregam Abstração de dados o DTW é executado em uma representação de dados reduzida e nos de Indexação são empregadas funções de limite inferior a fim de reduzir o número de vezes de execução do DTW durante a classificação e agrupamento de séries temporais.

Em Salvador e Chan (2007) é apresentado o algoritmo FastDTW que consiste em uma adaptação do DTW, que combina os métodos baseados em Restrições e Abstração de dados a fim de melhorar o desempenho do algoritmo. Um parâmetro de raio é inserido no FastDTW, o qual indica o número de células utilizadas na determinação do caminho de menor custo. Com essas adaptações a complexidade do algoritmo passa a ser linear  $O(N)$ , tornando o algoritmo mais veloz que a implementação original do DTW. O FastDTW apresenta erro médio de 8,6% com raio igual a 1 e erro inferior a 1% com raio igual a 20.

## 2.2 Bancos de dados

Banco de dados pode ser definido como uma coleção de dados relacionados, que representa um universo de discurso definido, uma coleção logicamente coerente de dados com o significado próprio e é desenvolvido com uma finalidade específica, com usuários e aplicações pré-definidos (ELMASRI E NAVATHE, 2011).

Bancos de dados são indicados a contextos em que o volume de dados é extenso, de modo que viabilize a organização e gerenciamento da informação armazenada, com o propósito de assessorar os usuários em operações de recuperação e atualização dos dados, para isso é utilizado um Sistema Gerenciador de Banco de Dados (SGBD) (SILBERSCHATZ *et al.*, 2012).

Segundo Elmasri e Navathe (2011), um SGBD é um conjunto de programas que possibilita definir, construir, manipular e compartilhar o banco de dados entre diversos usuários e aplicações. Definir um banco de dados equivale a determinar os tipos, estruturas e restrições aos dados que serão armazenados, sendo essa definição memorizada pelo SGBD em um dicionário de dados. A construção corresponde ao armazenamento da informação no banco de dados, para tanto o dicionário de dados é consultado com o propósito de verificar se as informações a serem inseridas estão em concordância com o que foi definido. A manipulação é feita por meio de funções de consulta que podem atualizar ou recuperar as informações desejadas. Finalmente, o compartilhamento permite que diversos usuários tenham acesso concomitante aos dados armazenados. Além disso, o SGBD também proporciona integridade e segurança aos dados armazenados (RAMAKRISHNAN, 2000).

Os bancos de dados e os SGBD podem ser classificados de acordo com o modelo de dados. Segundo Silberschatz *et al.* (2012) o modelo de dados é um conjunto de ferramentas conceituais próprias para descrever os dados, relações de dados, semântica de dados e restrições de consistência. Os modelos de dados definem a estrutura lógica do banco de dados, ou seja, como os dados se conectam, como são armazenados e processados pelo sistema.

Dentre os modelos existentes, o relacional é o principal modelo de dados, visto que é altamente eficiente e bem sucedido (GANIEE E BHARGAVA, 2015). Em consequência disto é o modelo de dados mais frequente em projetos desenvolvidos por empresas de tecnologia de informação (TI) (ARORA E GUPTA, 2012). Outros modelos de

dados existentes são o modelo hierárquico, modelo de rede, modelo de objeto, modelo objeto-relacional, modelo não relacional e os multimodelo.

A eficiência de armazenamento e gerenciamento de dados por meio de um sistema de informação para uma determinada aplicação está diretamente relacionada com o modelo de dados escolhido para o projeto de banco de dados (YUAN, 1996). Segundo Ramapriyan (2004), as características das imagens de sensoriamento remoto, o volume dessas e os processamentos a que serão submetidas, são elementos que impõem requisitos as atividades de armazenamento e recuperação, por conseguinte devem ser avaliados na escolha do modelo de dados, com o propósito de melhorar o desempenho da aplicação.

### 2.2.1 Modelo de dados relacionais

O modelo de dados relacional foi apresentado em Codd (1970), e ganhou grande popularidade em virtude de sua simplicidade, baseada nos princípios matemáticos da álgebra relacional (ELMASRI E NAVATHE, 2011). Esse modelo de dados é o mais difundido, e, portanto, implementado em diversos SGBD, comerciais e de código aberto. No Quadro 1 são apresentados alguns dos principais SGBD relacionais (SGBDR).

Quadro 1 - Exemplos de SGBD relacionais

SGBDR	DESENVOLVEDOR	TIPO
DB2	IBM	Comercial
Informix Dynamic Server		
Oracle <sup>1</sup>	Oracle Corporation	
SQL Server	Microsoft	
PostgreSQL <sup>2</sup>	PostgreSQL Global Development Group	Código Aberto
MySQL	Oracle Corporation	

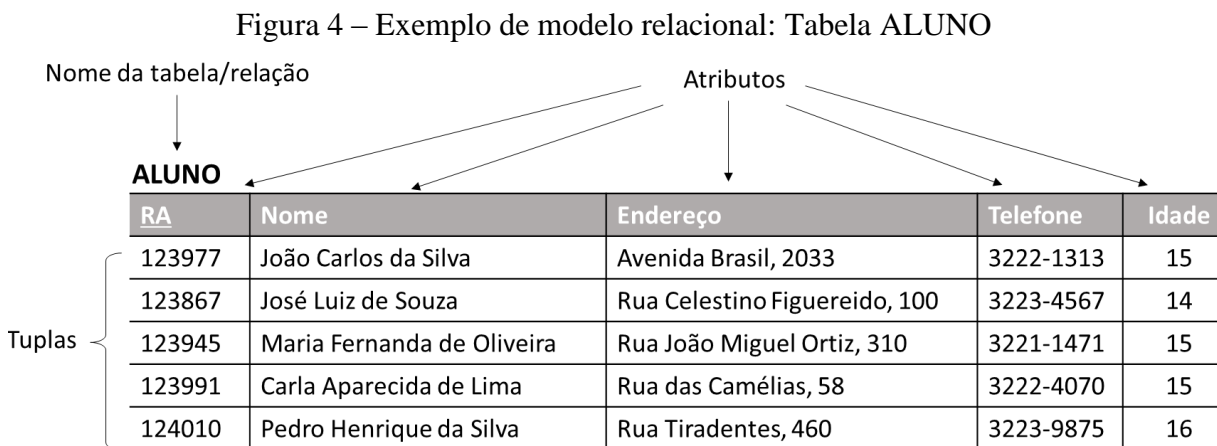
Fonte: Adaptado de (Silberschatz *et al.*, 2012)

O modelo de dados relacional é definido como um modelo estruturado, que requer o delineamento do esquema de armazenamento antes de se realizar o armazenamento

<sup>1,2</sup> Banco de dados objeto-relacional, classificado por diversos autores como relacional

propriamente dito, essa abordagem é denominada por Liu e Gawlick (2015) como ‘*schema first, data later*’.

A estrutura do modelo de dados relacional é constituída de coleções de tabelas, também conhecidas como relações. Cada relação representa uma entidade, que pode ser um objeto de existência física ou conceitual, onde são armazenadas informações logicamente relacionadas e tratadas como uma unidade. Cada relação possui atributos (campos), representados pelas colunas das tabelas. Os dados são armazenados nas linhas da tabela, também intituladas tuplas. Os valores registrados em cada linha devem estar de acordo com o domínio definido para o atributo (ELMASRI E NAVATHE, 2011). Na Figura 4 é apresentado como exemplo a tabela ALUNO com seus atributos e tuplas.



Fonte: Adaptado de Elmasri e Navathe (2011)

Codd (1970) definiu as seguintes propriedades para o modelo relacional:

- Cada linha da tabela representa um registro;
- A ordem das linhas não é significante;
- Todas as linhas são distintas;
- A ordem das colunas é significante;
- O significado de cada coluna é parcialmente transmitido rotulando-a com o nome do domínio correspondente.

No modelo relacional as tabelas podem apresentar relações entre elas, que podem ser dos tipos: um para um (1:1), um para muitos (1:n) ou muitos para muitos (m:n). As relações são implementadas por meio de chaves, que são denominadas chave primária e chave estrangeira. A chave primária garante a unicidade de cada tupla, essa pode ser definida por um

único atributo ou por uma combinação de atributos. A chave estrangeira define a relação de um atributo da tabela de origem com uma chave primária em outra tabela e é utilizada na recuperação de registros. O conceito de chaves do modelo relacional promove a possibilidade de controlar a redundância de dados através da normalização dos dados (ELMASRI E NAVATHE, 2011).

A redundância de dados refere-se ao armazenamento dos mesmos dados repetidamente, fato comum em aplicações que utilizam arquivos como sistema de armazenamento. A redundância de dados provoca desperdício de espaço de armazenamento e pode causar problemas de inconsistência dos dados. Com o propósito extinguir esses inconvenientes é realizado o processo de normalização (CODD, 1970).

A normalização dos dados consiste em armazenar cada item lógico em um lugar do banco de dados, garantindo economia de espaço de armazenamento, consistência aos dados, otimização e flexibilidade a consultas (ALAM, 2015).

Outro conceito importante do modelo de dados relacional é que SGBDR garantem que as transações executadas pelos usuários sejam seguras e confiáveis, uma vez que apresentam as propriedades ACID acrônimo de Atomicidade, Consistência, Isolamento e Durabilidade (SILBERSCHATZ *et al.*, 2012).

Segundo Elmasri e Navathe (2011) uma transação pode ser definida como um programa em execução que forma uma unidade lógica de processamento de banco de dados. Trata-se de uma ou mais operações que acessam o banco de dados com o propósito de inserir, atualizar, modificar ou recuperar dados. As transações podem ser implementadas em programas de aplicação ou diretamente através da linguagem de consulta de alto nível, como a *Structured Query Language* (SQL), padrão em SGBDR.

As propriedades ACID são indispensáveis para que transações realizadas em um banco de dados sejam seguras, confiáveis e consistentes. As propriedades ACID são definidas como:

- Atomicidade: essa propriedade define que uma transação é uma unidade indivisível, o que significa que todas as operações pertencentes à transação devem ser confirmadas, ou caso ocorra algum erro a transação é desfeita completamente, ou seja nenhuma operação é realizada e o banco de dados retorna ao seu estado anterior.

- **Consistência:** essa propriedade determina que após qualquer transação o banco de dados deve permanecer consistente, portanto as transações devem respeitar as regras de integridade dos dados.
- **Isolamento:** em um sistema de banco de dados multiusuário, onde transações são realizadas em paralelo, essa propriedade garante o controle de concorrência, com o propósito de assegurar que os resultados de transações realizadas concorrentemente sejam compatíveis. Portanto, nenhuma transação é afetada pela existência de outra transação.
- **Durabilidade:** essa propriedade define que os dados devem ser armazenados em uma memória permanente, de modo que independente de ocorrer alguma falha de *hardware* ou *software* os dados ainda estarão armazenados. Caso ocorra uma falha durante uma transação antes da confirmação dessa os dados permaneceram inalterados, caso a falha ocorra após a confirmação o banco de dados armazenará os dados modificados.

SGBDR proporcionam mecanismos para armazenar e recuperar dados estruturados com forte consistência e segurança, fato que proporcionou aos bancos de dados relacionais alcançarem estabilidade e alta confiança entre os profissionais de TI. Contudo, o volume e a heterogeneidade dos dados em determinadas aplicações vêm crescendo de maneira contundente, e devido a essas novas características dos dados o uso de bancos de dados relacionais pode resultar em uma solução de baixo desempenho (GESSERT *et al.*, 2016).

## 2.2.2 Modelo de dados não relacionais

O crescimento exponencial do volume de dados e a variabilidade dos dados, deu origem a era *big data*, principalmente em razão da multiplicação da quantidade de sensores, da popularização da internet e do surgimento das redes sociais. Essas características da era *big data* têm desafiado as atividades de armazenamento e manipulação de dados (GESSERT *et al.*, 2016).

Em virtude do armazenamento estruturado, próprio do modelo relacional, não atender as necessidades que emergem nesse novo período, surgiram os modelos de dados não relacionais, também denominados como NoSQL, esses proporcionam a flexibilidade e a escalabilidade requerida pelas aplicações *big data* (VITOLLO *et al.*, 2015).



A flexibilidade é garantida uma vez que os modelos NoSQL permitem o armazenamento de dados semiestruturados ou sem estrutura, através de modelos maleáveis, que seguem o paradigma *'data first, schema later or never'*, denominados por Liu e Gawlick (2015) como *Flexible Schema Data* (FDS). Já a escalabilidade é possível de ser realizada uma vez que SGBD NoSQL dão suporte ao armazenamento distribuído dos dados.

Não há uma definição amplamente aceita para bancos de dados NoSQL, entretanto algumas características são comuns a diversos SGBD que fazem o uso dos modelos de dados não relacionais. Em Edlich *et al.* (2010) *apud* Schmid *et al.* (2015) são apontadas sete características que bancos de dados *NoSQL* compartilham.

- Não são baseados em relações;
- São escaláveis horizontalmente;
- Geralmente são de código aberto, apesar de existirem sistemas comerciais;
- Não apresentam um esquema definido;
- Apresentam aplicativos para integração com outros *softwares*;
- Apresentam arquitetura descentralizada;
- Seguem o princípio BASE, que é acrônimo de *Basically Available, Soft State and Eventual Consistency*.

Bancos de dados NoSQL apresentam melhor desempenho em operações de leitura e gravação de dados, em razão da possibilidade de armazenar os dados em diferentes servidores, contudo para atingir a escalabilidade almejada, as propriedades ACID que garantem a consistência do banco de dados são renunciadas. O teorema CAP define que em sistemas de dados compartilhados em rede, é possível ter apenas duas das três propriedades: consistência, disponibilidade e tolerância a partição (BREWER, 2012). O que implica que em sistemas de bancos de dados distribuídos deve-se escolher entre consistência e disponibilidade. Portanto, em bancos de dados NoSQL as propriedades ACID são substituídas pelas propriedades BASE.

*Basically Available* significa que o sistema se encontra sempre disponível, tal fato é possível devido ao sistema suportar falhas parciais, sem falhar totalmente. Isso significa que se um dos servidores que compõe o sistema apresentar problemas, apenas parte dos usuários serão atingidos, portanto de maneira geral o sistema continuará disponível. *Soft State* significa que o sistema não precisa estar consistente o tempo todo. Por fim, *Eventual Consistency*

manifesta que o sistema pode eventualmente estar em estado de consistência (PRITCHETT, 2008).

Os bancos de dados NoSQL tem em comum as características apresentadas, contudo podem ser diferenciados em quatro modelos de dados não relacionais: famílias de colunas, chave-valor, orientados a documentos e grafos (Sadhalage e Fowler, 2013). Atualmente existem diversos SGBD NoSQL, alguns dos principais e o tipo de modelo de dados adotado por esses, são apresentados no Quadro 2.

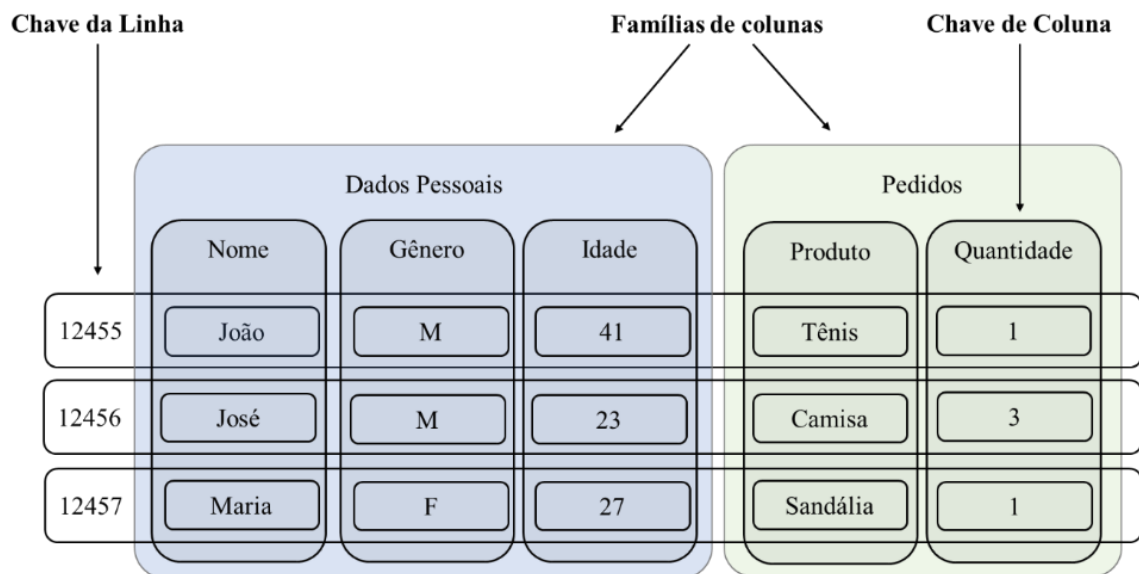
Quadro 2 - SGBD não relacionais

<b>SGBD</b>	<b>Desenvolvedor</b>	<b>Modelo de dados</b>
MongoDB	MongoDB Inc.	Orientados a Documentos
CouchDB	Apache Software Foundation	
BigTable	Google Inc.	Família de Colunas
DynamoDB	Amazon	
HBase	Apache Software Foundation	
Accumulo	Apache Software Foundation	
Riak	Basho Technologies	Chave-valor
Redis	Salvatore Sanfilippo	
Neo4j	Neo Technology	Grafos

Fonte: Adaptado de Moniruzzaman e Hossain (2013)

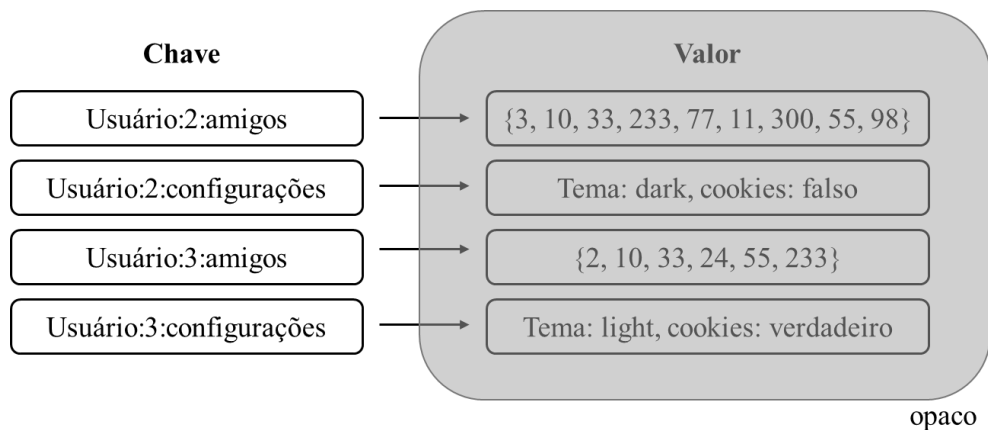
Tal como no modelo de dados relacional, bancos de dados família de colunas tem o armazenamento baseado no conceito de linhas e colunas, como é apresentado na Figura 5. Porém, os dados são armazenados nas colunas e não nas linhas como ocorre no modelo relacional (ALAM, 2015). Segundo Gessert *et al.* (2016) bancos de dados família de colunas podem ser entendidos como mapas ordenados de vários níveis, o primeiro nível funciona como o identificador para as linhas e é denominado de chaves das linhas. No segundo nível há chaves de colunas que são utilizadas para acessar as diferentes famílias de colunas. É possível haver um terceiro nível, como no banco de dados BigTable, que tem um identificador para o tempo. As famílias de colunas são utilizadas para agrupar no disco, dados que normalmente são acessados ao mesmo tempo com o propósito de otimizar a recuperação desses dados.

Figura 5 - Representação de banco de dados família de colunas



No modelo de dados chave-valor os dados são organizados através de um conjunto de chaves, sendo que cada chave é única e tem valores associados a essa, os valores são agregados de dados. A chave funciona como identificador e é utilizada para recuperar os dados armazenados. Em consequência da simplicidade desse modelo, ele suporta apenas operações de recuperação e inserção. Outra característica de bancos de dados chave-valor é que a estrutura dos valores associados às chaves é desconhecida e pode variar de chave para chave, portanto é responsabilidade da aplicação que acessa os dados ter o conhecimento de quais dados podem estar armazenados em cada chave. A vantagem desse modelo é sua simplicidade, que facilita o particionamento e a consulta aos dados, fazendo o sistema ter baixa taxa de latência e alta taxa de transferência (GESSERT *et al.*, 2016). Contudo, o modelo de chave-valor tem como desvantagem a impossibilidade de acessar os valores de forma direta (os valores são opacos a aplicação), portanto não é possível realizar consultas complexas aos dados. Na Figura 6 é apresentada uma representação de banco de dados chave-valor.

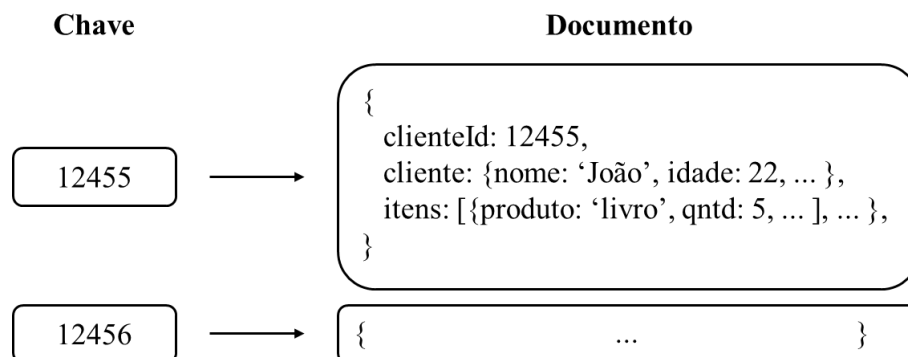
Figura 6 - Representação de banco de dados chave-valor



Fonte: Adaptado de Gessert *et al.* (2016)

O armazenamento de dados utilizando o modelo orientado a documentos é semelhante ao modelo chave-valor, com a distinção que os valores são documentos e esses podem ser dos tipos *JavaScript Object Notation* (JSON), *Binary JavaScript Object Notation* (JSONB) e *eXtensible Markup Language* (XML). Cada documento é independente e sem um esquema definido, havendo a possibilidade de armazenar subdocumentos dentro de um documento. Essa diferença permite realizar o acesso ao documento inteiro pelo identificador, assim como recuperar partes do documento, logo é possível fazer consultas complexas nesse modelo. A Figura 7 apresenta um exemplo de representação de banco de dados orientado a documentos.

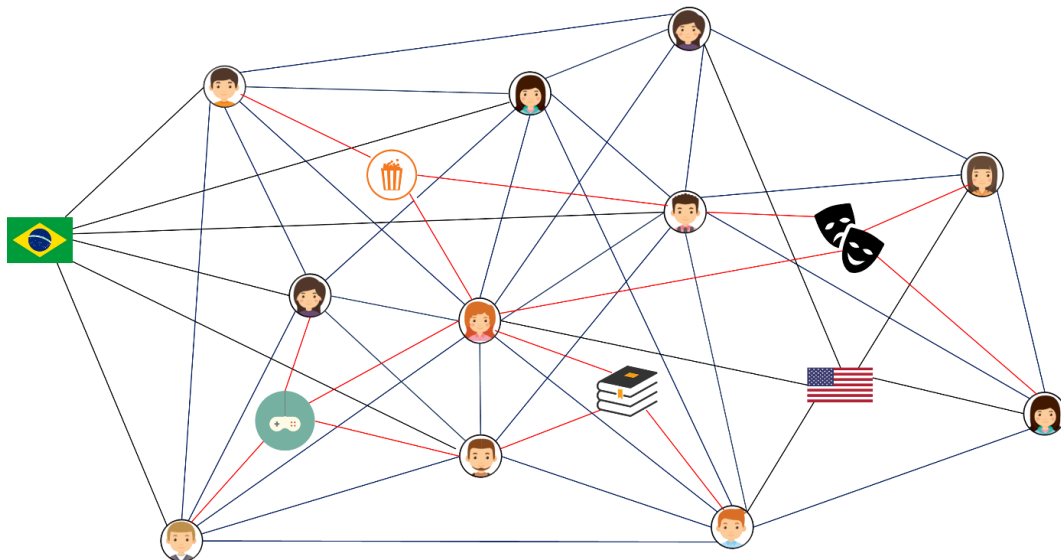
Figura 7 - Representação de banco de dados orientado a documentos



Fonte: Adaptado de Gessert *et al.* (2016)

O modelo de dados baseado em grafos armazena e representa os dados a partir de três estruturas: nós, arestas e propriedades. Os nós são empregados para modelar objetos, as ligações existentes entre os objetos são representadas pelas arestas, e as propriedades são utilizadas para descrever os atributos dos nós e das arestas. Assim como nos modelos descritos anteriormente, esse modelo é livre de esquema. Bancos de dados grafos são favoráveis em aplicações em que as relações entre os dados é tão importante quanto os dados, como no caso das redes sociais, que requerem o armazenamento da ligação entre os usuários (MONIRUZZAMAN E HOSSAIN, 2013). Na Figura 8 é possível observar a representação de um banco de dados baseado em grafos em uma aplicação de rede social. Nesse exemplo, os nós são representados pelos ícones, já as linhas coloridas são as arestas, a linha azul indica que as pessoas se conhecem, a preta aponta a nacionalidade da pessoa e a vermelha assinala em que a pessoa tem interesse.

Figura 8 - Representação de banco de dados grafos com uso em rede social



### 2.2.3 Multimodelo de dados

Diversos modelos de dados foram desenvolvidos com o propósito de satisfazer as diferentes exigências que emergem das variadas aplicações. O procedimento tradicional no desenvolvimento de sistemas de informação foi, por muito tempo, definir o SGBD que apresentasse o modelo de dados que atendesse, de maneira geral, as especificidades do projeto. Entretanto, em aplicações complexas que combinam diferentes tipos de problemas, empregar um único mecanismo de banco de dados pode resultar em uma solução de baixo

desempenho (SADALAGE E FOWLER, 2013). Com o propósito de minimizar essa adversidade, desenvolvedores passaram a adotar o que é conhecido como persistência poliglota.

A persistência poliglota tem sua origem fundamentada na programação poliglota, apresentada por Ford (2006), e se refere a atividade de empregar várias tecnologias de armazenamento de dados para uma ou várias aplicações (SADALAGE E FOWLER, 2013). A persistência poliglota pode ser entendida como precursora dos bancos de dados multimodelo (DATASTAX, 2016), que, segundo Lu e Holubová (2017) pode ser dividida em *multi-database* e *single-database*.

A proposta multimodelo *multi-database* apresenta a mesma abordagem da persistência poliglota, portanto são empregados múltiplos SGBD, de modo que a informação seja distribuída em um conjunto de bancos de dados. Cada um dos SGBD pode diferir no modelo de dados, nas restrições, na linguagem de consulta e na forma de gerenciamento das transações. Em razão disso, há a necessidade de realizar a integração dos dados pelo sistema, o que eleva a complexidade operacional dessa proposta. Outra desvantagem dessa estratégia é que não há um mecanismo automático para manter a consistência dos diferentes bancos de dados.

Na abordagem multimodelo *single-database* é adotado um único SGBD que possibilite realizar o armazenamento dos dados através de diferentes modelos, conseqüentemente toda informação fica restrita a apenas um banco de dados. Essa estratégia elimina a necessidade de realizar a integração dos dados, atividade mandatória na proposta multimodelo *multi-database* e a consistência pode ser gerenciada pelo SGBD adotado. Logo essa estratégia simplifica as atividades de integração, migração, desenvolvimento, manutenção e operação dos dados.

Apesar da abordagem multimodelo ter ganhado força nos últimos anos, principalmente com os bancos de dados NoSQL que frequentemente adotam mais de um dos modelos de dados descritos na Seção 2.2.2, esse comportamento não pode ser considerado inovador, visto que SGBD objeto-relacionais, que apresentam os modelos objeto e relacional surgiram na década de 1990 (LU E HOLUBOVÁ, 2017). Como exemplo de SGBD multimodelo é possível citar o PostgreSQL, o Microsoft SQL que são originalmente classificados como (objeto-)relacional e atualmente aceitam o armazenamento nos modelos de dados chave-valor (HBase) e orientado a documentos (JSON, JSONB e XML). ArangoDB,

OrientDB, Datastax e CouchDB são bancos de dados NoSQL que fazem uso de mais de um modelo de dados.

### 2.2.4 Modelo de dados espaço-temporais

Bancos de dados são classificados como espaciais quando apresentam a capacidade de armazenar e consultar de maneira eficiente dados n-dimensionais, como pontos, linhas, polígonos e dados matriciais (*rasters*). A eficiência está relacionada com definição dos tipos de dados e com a indexação espacial presente nesses sistemas, e que é utilizada para otimizar a velocidade das consultas, visto que possibilita recuperar a informação de interesse sem a necessidade de percorrer todos os objetos armazenados (CASANOVA *et al.*, 2005).

Os bancos de dados espaciais podem ser denominados bancos de dados geográficos quando o espaço de interesse é a superfície terrestre. Nesse caso, são empregados dados geográficos que representam a extensão e posição de entidades do mundo real na superfície terrestre. Outra característica dos bancos de dados espaciais é que possibilitam realizar diversas operações espaciais, como determinar a distância entre diferentes objetos, calcular a área de um polígono, definir uma área de *buffer* ao redor de uma geometria, verificar quais objetos estão contidos em uma área de *buffer*, verificar se objetos se interceptam, entre outras.

Em bancos de dados espaciais é comum utilizar representações estáticas dos fenômenos espaciais, isso é decorrente do fato de que a principal abstração utilizada pelos Sistemas de Informações Geográficas (SIG) é o mapa. Entretanto, em aplicações de monitoramento de fenômenos espaciais há a necessidade incluir o tempo na representação, o que origina os modelos de dados espaço-temporais. É possível afirmar que bancos de dados espaço-temporais são modelados com o propósito de representar o comportamento espacial de objetos ao longo do tempo (CASANOVA *et al.*, 2005).

A representação de dados matriciais espaço-temporais pode ser realizada por meio dos modelos conceituais: *snapshot*, *space-time composite* (STC) e *spatiotemporal object* (YUAN, 1996; NADI e DELAVAR, 2003).

O modelo *snapshot* representa um tema de interesse ao longo do tempo sobrepondo as camadas de representação espacial desse tema obtidas em diferentes datas. Portanto, a representação temporal dos dados matriciais pelo modelo *snapshot* é feita

empilhando as imagens de acordo com a ordem de aquisição dessas (ARMSTRONG, 1988). Uma deficiência desse método é que as datas às quais os dados estão associados, ou o intervalo de tempo de aquisição dos dados, são representados apenas de forma implícita, uma vez que geralmente são gravados em arquivos externos de metadados em vez de diretamente nos dados (SONG et al., 2016), o que inviabiliza realizar consultas para determinadas datas ou períodos. Outro inconveniente desse modelo é que é frequente a presença de dados redundantes, sem modificações espaciais (YUAN, 1996).

O modelo *STC* representa os dados espaço-temporais através de uma única camada bidimensional, composição, espacialmente homogênea e temporalmente uniforme. A composição é construída a partir de um conjunto de camadas adquiridas regularmente em um espaço de tempo e que, segundo Nadi e Delavar (2003), é adequada para representar mudanças em apenas uma feição espacial. A representação das mudanças espaciais pode ser realizada por meio do atributo data de modificação, que é adicionado aos objetos espaciais, portanto é possível classificar o objeto e a partir dessa propriedade retratar as mudanças espaço-temporais.

O modelo *spatiotemporal object* representa o mundo real como um conjunto de objetos discretos que consistem em átomos espaço-temporais. Aos dados espaciais é incorporado o tempo como uma terceira dimensão. Os átomos espaço-temporais são as maiores unidades homogêneas que determinado fenômeno apresenta no espaço e no tempo. Dessa forma o modelo pode armazenar mudanças no espaço e no tempo (YUAN, 1996).

### 2.2.5 PostgreSQL

O SGBD PostgreSQL teve seu desenvolvimento iniciado com o nome de Postgres, em 1986 na Universidade da Califórnia em Berkeley (UCB) pelo professor de Ciências da Computação Michael Stonebraker. Esse sistema surgiu com o propósito de explorar novos conceitos da época sobre bancos de dados, como o modelo objeto-relacional, visto que até então SGBDs eram desenvolvidos a partir da teoria clássica de bancos de dados relacionais. Inicialmente, o sistema fazia uso da linguagem de consulta denominada POSTQUEL, que foi substituída pelo SQL em 1995 pelos alunos Andrew Yu e Jolly Chen. A partir de 1996, o Postgres deixou de ser desenvolvido apenas na UCB, e passou a ser um sistema de código aberto que recebe contribuições de desenvolvedores espalhados pelo mundo, atualmente a comunidade de desenvolvedores é denominada por PostgreSQL Global Development Group. Com a



contribuição assídua de diversas pessoas, o sistema adquiriu a reputação de ser altamente estável e passou a ser denominado como PostgreSQL.

O PostgreSQL é conhecido como o mais avançado SGBD gratuito, devido ao seu desenvolvimento contínuo que possibilita ao sistema ser multiplataforma, compatível com as propriedades ACID, dar suporte a chaves estrangeiras, *joins*, *views*, *triggers* e armazenar procedimentos em diversas linguagens. O PostgreSQL também aceita o armazenamento de diversos tipos de dados como: integer, numeric, boolean, char, varchar, date, interval, timestamp, entre outros. Apresenta, ainda, uma série de extensões e diversos outros recursos avançados.

Dentre os SGBD existentes, o PostgreSQL é um dos mais empregados em Geociências, já que é gratuito e apresenta a extensão espacial PostGIS (Zhang e Yi, 2010). O PostGIS faz parte dos projetos desenvolvidos pela OSGeo, e permite adicionar ao PostgreSQL os tipos de dados espaciais, além de funções e indexadores espaciais empregados na manipulação dos dados espaciais.

A primeira versão do PostGIS foi disponibilizada em 2001, contudo a combinação PostgreSQL/PostGIS, até abril de 2012, era incapaz de armazenar e manipular dados matriciais georreferenciados, como imagens de sensoriamento remoto. Essa lacuna foi preenchida com o projeto PostGIS Raster, que foi integrado ao PostGIS a partir de sua versão 2.0 e passou a possibilitar o armazenamento do tipo matriz como geometria e oferecer um conjunto de funções SQL para manipular esses dados (AGOSTO, 2013).

Outros avanços significantes e relativamente novos no desenvolvimento do PostgreSQL foram disponibilizados na versão 9.4, liberada em dezembro de 2014, na qual são aceitos dados dos tipos HStore, JSON e JSONB, além das funções implementadas com a finalidade de manipular dados desses tipos. Essas funcionalidades fizeram do PostgreSQL um SGBD multimodelo, uma vez que continua sendo objeto-relacional, com dados organizados por meio de tabelas e relacionamentos, mas com a particularidade adicional de aceitar conjuntos de dados semi-estruturados ou sem estrutura, característica de SGBD NoSQL (LERNER, 2015). O tipo de dado HStore segue o modelo de dados chave-valor e os tipos JSON e JSONB seguem o modelo de dados orientado a documentos, apresentados na Seção 2.2.2.

A flexibilidade dos bancos de dados NoSQL é uma propriedade vantajosa quando se trata de dados heterogêneos e em grande volume (XIAO e LIU, 2011), uma vez que apresentam melhor performance quando comparados com o modelo baseado em esquema. Todavia, os bancos de dados relacionais, proporcionam maior integridade e segurança aos dados armazenados, além do fato de que, de maneira geral, bancos de dados relacionais possuem mais funcionalidades para gerir dados espaciais (SCHMID *et al.*, 2015).

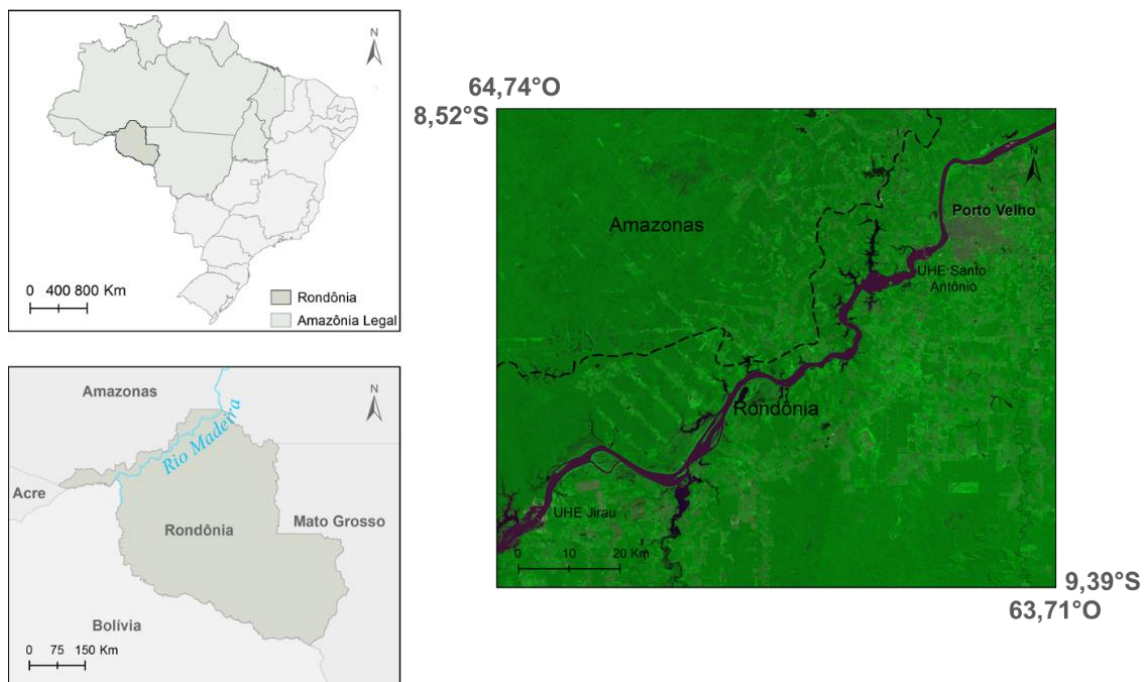
Essas características foram determinantes na escolha do PostgreSQL/PostGIS para o presente projeto, uma vez que o armazenamento de séries temporais geradas a partir de imagens requer flexibilidade, o que é proporcionado pelos tipos de dados JSON e JSONB. Além disso, o PostgreSQL/PostGIS dispõe de um grande conjunto de ferramentas para armazenamento e análise de dados espaciais.

### 3 MATERIAIS E MÉTODOS

#### 3.1 Área de estudo

Com o propósito de organizar e armazenar as imagens da série histórica Landsat e realizar a detecção de mudanças a partir desse conjunto de dados, foi definida como área de estudo uma região localizada na Amazônia legal brasileira. Essa região inclui a cidade de Porto Velho, na divisa entre os estados de Rondônia e Amazonas, englobando o trecho do rio Madeira desde UHE de Jirau até a de Santo Antônio. A área de estudo é delimitada pelas coordenadas geográficas  $8,52^{\circ}$  à  $9,39^{\circ}$  de latitude a Sul e  $63,71^{\circ}$  à  $64,74^{\circ}$  de longitude a Oeste, apresentadas na Figura 9.

Figura 9 - Inserção da área de estudo no território brasileiro e sua delimitação em composição falsa-cor com as bandas espectrais do vermelho, infravermelho próximo e médio (4B5G6R) de imagem OLI/Landsat 8



Essa região foi definida como área de estudo em razão haver uma intensa preocupação relacionada ao monitoramento e controle do desmatamento, prática que pode causar alterações no regime hidrológico e climático local e global, além de representar um risco a diminuição da biodiversidade da floresta amazônica (SHUKLA *et al.*, 1990; GUIMBERTEAU *et al.*, 2016).

## 3.2 Materiais

Para o desenvolvimento da metodologia proposta nesta pesquisa foram implementadas rotinas computacionais utilizando a linguagem de programação *python*, e módulos dessa linguagem. Também foi empregada a biblioteca de abstração de dados geoespaciais (*Geospatial Data Abstraction Library - GDAL*), própria para manipular dados espaciais no formato matricial e vetorial.

Os dados utilizados referem-se a imagens multiespectrais adquiridas pelos sensores TM, ETM+ e OLI. As imagens foram adquiridas a partir março de 1984 até abril de 2017. Informações relacionadas ao conjunto de imagens, assim como as características dos sensores são detalhados na seção 3.2.1.

O armazenamento e gerenciamento dos dados foi realizado por meio do SGBD PostgreSQL versão 9.6.3, juntamente com sua extensão espacial PostGIS versão 2.3.2. O PostgreSQL é classificado como um SGBD objeto-relacional gratuito e de código aberto, outras informações sobre o PostgreSQL, tal como a justificativa para o uso desse, foram esclarecidos na seção 2.2.5.

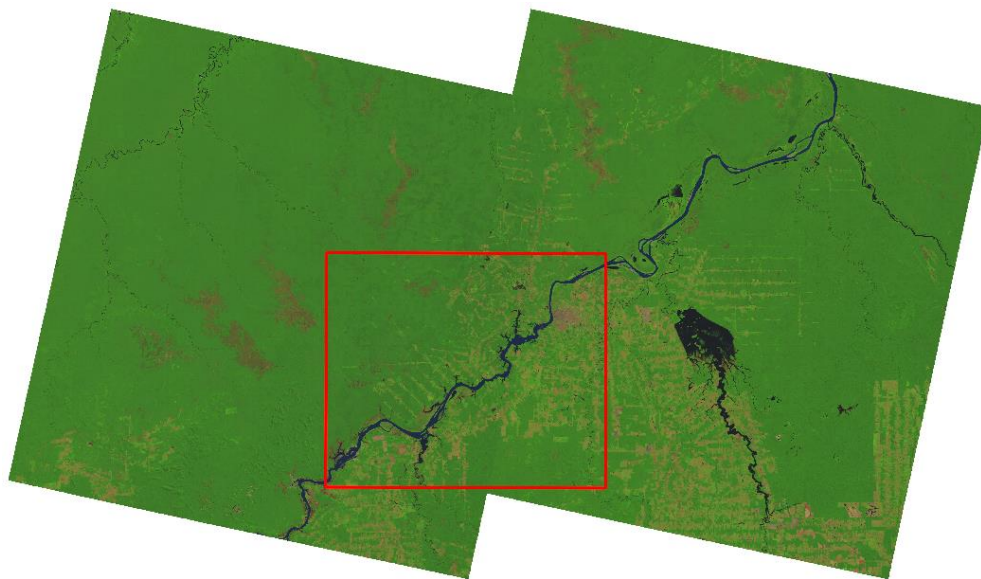
Para a visualização da informação extraída do conjunto de dados, foi selecionado o SIG gratuito e de código aberto QGIS versão 2.18.9. Esse é um sistema multiplataforma que suporta a visualização, edição e análise de dados espaciais e integra o conjunto de projetos desenvolvidos pela *Open Source Geospatial Foundation (OSGeo)*. O uso desse sistema é justificado por ser permissivo a conexão direta com o PostgreSQL/PostGIS, além de ser amplamente utilizado pela comunidade das Geociências.

### 3.2.1 Imagens Landsat

Foram selecionadas para o presente estudo imagens adquiridas pelos sensores TM, ETM+ e OLI, embarcados nos satélites Landsat 5, 7 e 8, respectivamente. Essas são imagens multiespectrais e foram adquiridas nos períodos de março de 1984 até janeiro de 2013 pelo sensor TM, de abril de 1999 até maio de 2003 pelo sensor ETM+ e de fevereiro de 2013 até abril de 2017 pelo sensor OLI.

A área de estudo está inserida em duas cenas adquiridas em órbitas diferentes pelos satélites Landsat, cuja tomada ocorre em um intervalo de 7 dias, correspondentes às órbitas/ponto 233/066 e 232/066. O mosaico dessas cenas, com a área de estudo em destaque, é apresentado na Figura 10.

Figura 10 - Mosaico de Imagens OLI 4B5G6R - Cenas 232/066 do dia 02/07/2016 e 233/066 do dia 09/07/2016

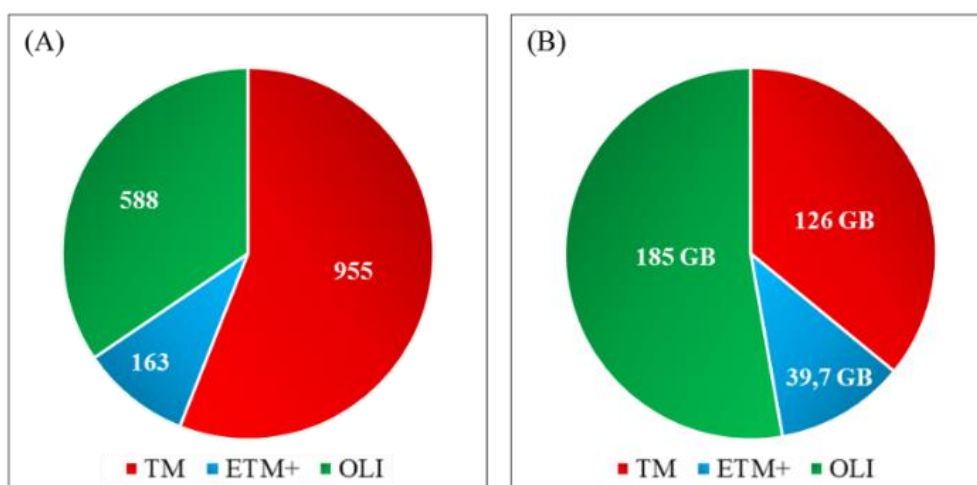


O conjunto de cenas empregado neste estudo refere-se ao nível de processamento *Pre-Collection Level-1* (L1T), no qual as imagens são submetidas a calibração radiométrica e ortorretificação. Com isso, as imagens são disponibilizadas com desvio padrão de até 12 metros na posição do pixel e, devido a essas correções os produtos pertencentes a essa categoria de processamento são adequados a análise de séries temporais em nível de pixel (USGS).

Para o período de aquisição completo foram baixadas 1294 cenas, o que totalizou 353 gigabytes de dados. Na Figura 11 (A) é apresentada a quantidade de cenas por sensor e na Figura 11 (B) o volume em gigabytes que essas cenas ocupam. É possível verificar que, apesar do conjunto de dados conter um número maior de cenas adquiridas pelo sensor TM (955) o maior volume de dados corresponde a imagens tomadas pelo sensor OLI (185 GB). Isso ocorre em razão de cenas oriundas de diferentes sensores apresentarem características distintas, como é o caso do sensor OLI que dispõe de 11 bandas espectrais, enquanto que o sensor TM têm 7 bandas espectrais. Além disso, as bandas da imagem TM são quantizadas em 256 níveis

de brilho, e as imagens OLI são representadas em 65536 níveis. Outras características como resolução espectral, espacial e temporal para os sensores TM, ETM+ e OLI são fornecidas no Quadro 3. Esse contexto de crescente volume de dados e do requisito de compatibilização entre eles reforça a necessidade de buscar formas adequadas de armazenamento e de recuperação dos dados, visando a eficiência da organização e do processamento.

Figura 11 - (A) Quantidade de cenas por sensor (B) Volume de espaço ocupado pelas cenas em gigabytes por sensor



As cenas são disponibilizadas gratuitamente pela USGS, e podem ser adquiridas realizando o *download* diretamente do site da USGS, via Earth Explorer. A cena é fornecida em um arquivo compactado, que tem como conteúdo um conjunto de bandas no formato GeoTIFF e um arquivo de texto com os metadados da cena.

Cada imagem contém um conjunto de bandas. O número de bandas e seus respectivos intervalos espectrais dependem do sensor que adquire a cena (Quadro 3). O formato GeoTIFF permite armazenar na banda, o sistema de referência e a projeção cartográfica. No caso das imagens Landsat, o sistema de referência é o *World Geodetic System 1984* (WGS84), a projeção é Universal Transversa de Mercator (UTM), o fuso e o hemisfério da projeção dependem da localização da cena. A área de estudo está inserida no fuso 20 e no hemisfério sul (S), porém imagens Landsat são disponibilizadas referenciadas ao hemisfério norte (N), de modo que as cenas utilizadas nesse estudo devem ser reprojctadas.

Quadro 3 - Características das imagens adquiridas pelos sensores TM, ETM+, OLI. NIR (*near infrared*) se refere à banda do infravermelho próximo, SWIR (*shortwave infrared*) é o infravermelho médio e TIR (*thermal infrared sensor*) é o sensor do infravermelho termal do Landsat 8

(Esp. = espectral, Rad. = radiométrica, Temp. = temporal)

Bandas	Landsat 5 - TM					Landsat 7 - ETM+					Landsat 8 - OLI					
	Intervalo Espectral ( $\mu\text{m}$ )		Resolução			Intervalo Espectral ( $\mu\text{m}$ )		Resolução			Intervalo Espectral ( $\mu\text{m}$ )		Resolução			
			Esp. (m)	Rad. (bits)	Temp. (dias)			Esp. (m)	Rad. (bits)	Temp. (dias)			Esp. (m)	Rad. (bits)	Temp. (dias)	
1	Azul	0,45 – 0,52	30	8	16	Azul	0,45 – 0,52	30	8	16	Ultra Azul	0,435 – 0,451	30	16 <sup>2</sup>	16	
2	Verde	0,52 – 0,60	30			Verde	0,52 – 0,60	30			Azul	0,452 – 0,512	30			
3	Vermelho	0,63 – 0,69	30			Vermelho	0,63 – 0,69	30			Verde	0,533 – 0,590	30			
4	NIR	0,76 – 0,90	30			NIR	0,77 – 0,90	30			Vermelho	0,636 – 0,673	30			
5	SWIR1	1,55 – 1,75	30			SWIR1	1,55 – 1,75	30			NIR	0,851 – 0,879	30			
6	Termal	10,40 – 12,50	120			Termal	10,40 – 12,50	60 <sup>3</sup>			SWIR1	1,566 – 1,651	30			
7	SWIR2	2,08 – 2,35	30			SWIR2	2,08 – 2,35	30			SWIR2	2,107 – 2,294	30			
8						Pancromática	0,52 – 0,90	15						Pancromática	0,503 – 0,676	15
9											Cirrus	1,363 – 1,384	30			
10											TIRS1	10,60 – 11,19	100*			
11											TIRS2	11,50 – 12,51	100*			

Fonte: Adaptado de USGS

<sup>2</sup> Quantizado em 12 bits e representados em 16 bits

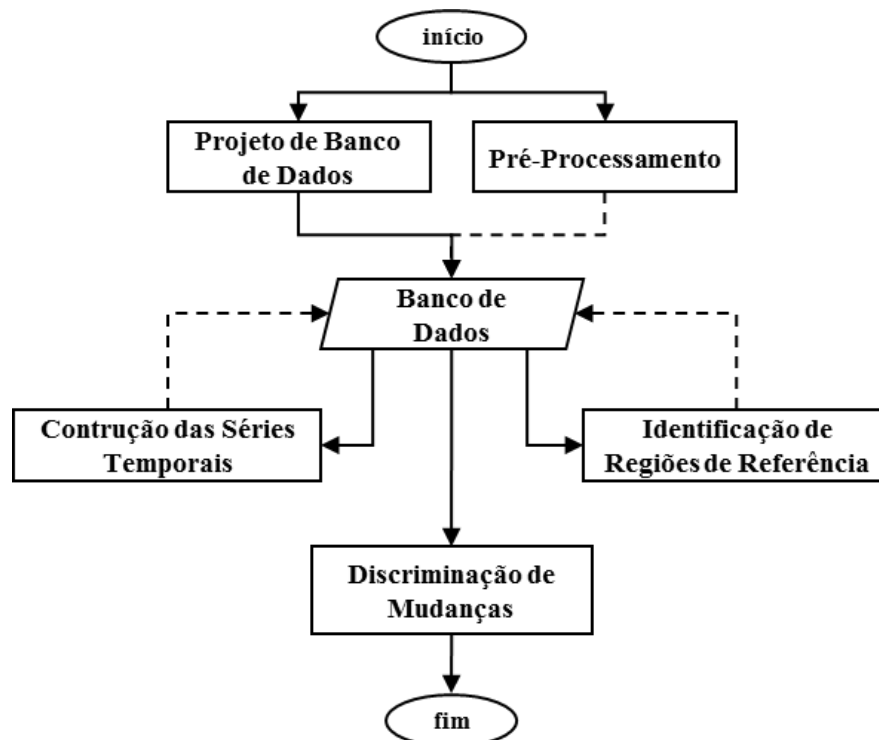
<sup>3</sup> Produtos disponibilizados depois de reamostrados para resolução espacial de 30 metros

O arquivo de metadados da cena possui diversas informações, entre elas: identificador da cena, satélite, sensor, órbita/ponto, data de aquisição, porcentagem de nuvens, azimute solar, elevação solar, distância Terra-Sol e os parâmetros de conversão especificados como ganho e *offset* que são utilizados na transformação dos valores de brilho quantizados (ND) em radiância e reflectância no topo da atmosfera (TOA). Esses parâmetros de conversão são definidos banda a banda, e como o número de bandas é distinto para cada sensor, a quantidade desses parâmetros também é variável

### 3.3 Metodologia

Com expectativa de organizar imagens e séries temporais em um banco de dados híbrido (multimodelo), com o propósito de discriminar mudanças na cobertura da terra na região de estudo, foi definida a metodologia apresentada na Figura 12. As etapas gerais que constituem a metodologia desenvolvida foram: projeto do sistema de banco de dados, pré-processamento dos dados, construção das séries temporais, identificação de regiões de referência para análise de mudanças e discriminação das mudanças. As linhas pontilhadas na Figura 12 indicam a inserção de dados no banco de dados.

Figura 12 - Fluxograma do desenvolvimento geral da pesquisa





O projeto do banco de dados é a etapa na qual foram levantados e analisados os requisitos dos dados e da aplicação. Nessa etapa foi definido o modelo conceitual, o modelo lógico e implementadas as tabelas e relacionamentos, que correspondem ao projeto físico.

O pré-processamento consistiu na aquisição das imagens, filtragem e conversão dos dados de acordo com o modelo definido para o banco de dados. Nessa etapa foram construídas máscaras de nuvens e sombra de nuvens com o algoritmo *Fmask* desenvolvido para detectar nuvens, sombra de nuvens e neve em imagens Landsat e Sentinel. Essa etapa foi definida com o propósito de filtrar os dados, para armazenar e processar o volume de dados úteis de fato.

O projeto do banco de dados e o pré-processamento foram desenvolvidos em paralelo, visto que a análise de requisitos, primeira etapa do projeto do banco de dados, depende do entendimento dos dados, o qual só foi possível após a aquisição das imagens, que é a primeira etapa do pré-processamento. Outra justificativa para essas etapas terem sido desenvolvidas concomitantemente é que para a informação ser armazenada no banco de dados, essa deve estar de acordo com o modelo definido no projeto do banco de dados.

A etapa denominada construção das séries temporais consistiu na produção de uma série temporal para cada pixel, em uma posição (coluna, linha) da cena, considerando imagens adquiridas em várias datas, como ilustrado na Figura 1. Para tanto foram recuperadas informações no banco de dados para gerar imagens Índice de Vegetação da Diferença Normalizada (*Normalized Difference Vegetation Index - NDVI*) e a partir dessas foram construídas as séries temporais. Ao final da construção das séries temporais essa informação foi armazenada no banco de dados no tipo de dados JSON, em razão da flexibilidade que esse tipo de dados possibilita.

Para a caracterização das áreas de mudança e não mudança por meio de medidas de similaridade entre séries temporais, foram identificadas por meio de interpretação visual regiões que correspondiam as classes de interesse. Uma vez que a aplicação proposta para esta pesquisa consiste na discriminação de mudanças, as regiões de referência foram determinadas analisando imagens de diferentes anos. Após a identificação das regiões de referência, foram vetorizados polígonos de delimitação para essas áreas, e os mesmos foram armazenados no banco de dados.

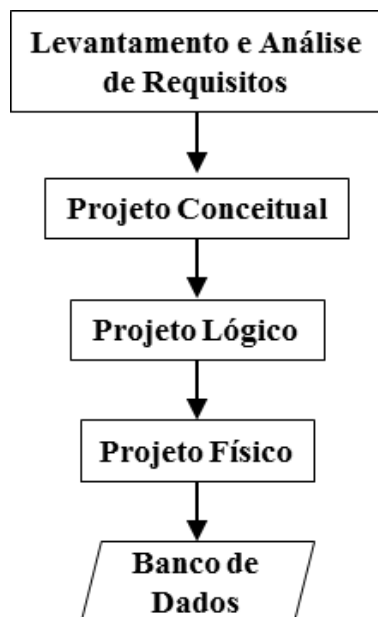
A etapa final da metodologia deste projeto consistiu na discriminação de mudanças na cobertura da terra, por meio da classificação das séries temporais. Para isso foram aplicadas medidas de similaridade entre as séries temporais (as de referência contra a série a ser classificada), utilizando o algoritmo FastDTW.

Uma vez que o principal produto desta pesquisa é o sistema de banco de dados, o qual orienta o desenvolvimento das etapas seguintes, a sua estrutura é apresentada nesta seção de metodologia, assim como os resultados parciais decorrentes da aplicação dos processamentos prévios, e mesmo exemplos de séries temporais produzidas para pixels específicos. As séries temporais produzidas para regiões de referência e a classificação das mudanças são mostrados na Seção 4.

### **3.3.1 Projeto do banco de dados**

O propósito do banco de dados definido para esta pesquisa consistiu em armazenar as imagens adquiridas pelos sensores embarcados nos satélites Landsat, seus metadados, as séries temporais geradas a partir da informação extraída das imagens e regiões de referência, com o objetivo de dar suporte à automatização do processo de discriminação de mudanças. Para tanto foi realizado o projeto do banco de dados, conforme apresentado na Figura 13.

Figura 13 – Fluxograma para o desenvolvimento do projeto do banco de dados



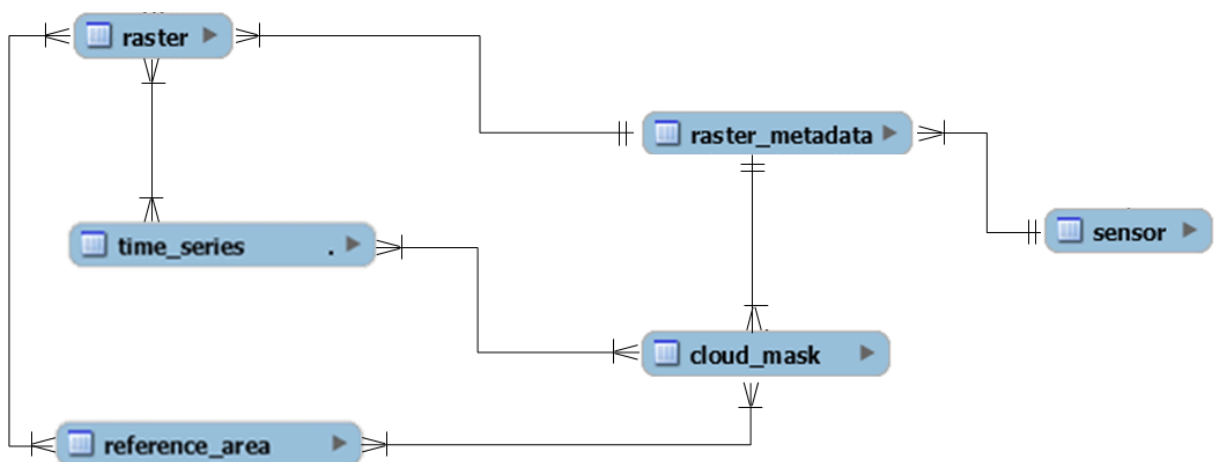
O projeto do banco de dados foi iniciado com o levantamento e análise de requisitos, etapa que consistiu em compreender quais informações deveriam ser armazenadas, quais as particularidades dos dados e operações a serem realizadas. Como resultado, foram definidos os seguintes requisitos:

- Armazenar as imagens adquiridas pelos sensores TM, ETM+ e OLI preservando as características geométricas e radiométricas das cenas;
- Permitir a recuperação das cenas a partir da data de aquisição dessas;
- Armazenar informações (órbita/ponto, banda, data de aquisição, sensor, nome da imagem, satélite, azimute solar, elevação solar, distância Terra-Sol, constantes das bandas termais e parâmetros de transformação (*ganho e offset*) dos ND em radiância e reflectância TOA) das imagens;
- Armazenar os dados (nome, resolução temporal, bandas, intervalo espectral) referentes aos sensores;
- Armazenar máscaras de nuvens e sombra de nuvens a serem geradas na etapa de pré-processamento para cada uma das cenas;
- Permitir a recuperação das bandas das imagens e das máscaras de nuvens;

- Permitir a recuperação dos parâmetros específicos para as bandas, com o propósito de transformar imagens representadas por ND em imagens de reflectância TOA;
- Armazenar séries temporais de NDVI de cada pixel imageado na região de estudo, sendo essas séries temporais de tamanho e frequência irregulares;
- Armazenar regiões de referência, a serem utilizadas na etapa de discriminação de mudanças na cobertura da terra.

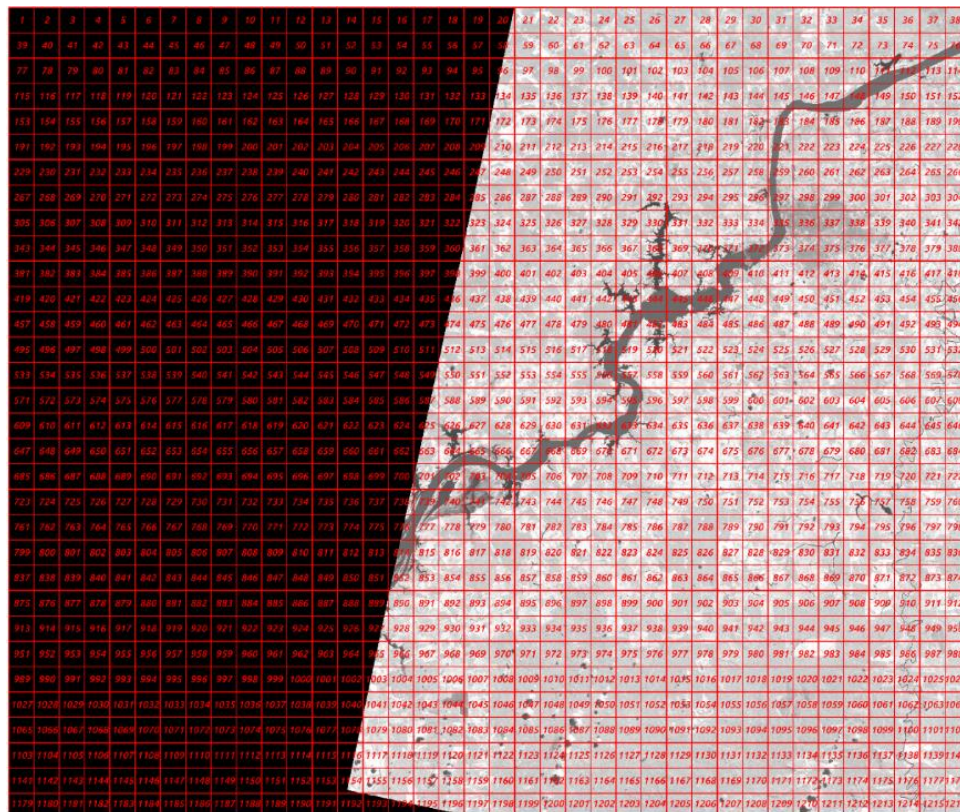
Após a etapa de levantamento e análise de requisitos, foi elaborado o projeto conceitual, no qual foram definidos por meio de um esquema conceitual de alto nível, as entidades, relacionamentos e restrições, sem incluir detalhes sobre o armazenamento ou implementação. Na Figura 14 é apresentado o diagrama entidade-relacionamento (DER) simplificado. As entidades receberam a seguinte nomenclatura no diagrama: imagens como *raster*, metadados das imagens como *raster\_metadata*, informações referentes aos sensores como *sensor*, máscaras de nuvens e sombra de nuvens como *cloud\_mask*, séries temporais nomeadas como *time\_serie* e regiões de referência como *reference\_area*.

Figura 14 - Diagrama entidade relacionamento simplificado



Visando otimizar a performance na recuperação de bandas e de partes específicas da imagem, inicialmente, cada banda foi armazenada de maneira independente e dividida em *tiles* (quadrículas), sendo que um mosaico de *tiles* representa uma banda, como mostrado na Figura 15.

Figura 15 - *Tiles* (quadrículas) da Banda 5 (infravermelho próximo) de uma imagem OLI/Landsat 8, órbita/ponto 232/066 do dia 14/06/2015



Essa estratégia de armazenamento foi selecionada, com base em Song *et al.* (2016), que afirmam que, em termos de velocidade na recuperação das bandas ou recortes dessas, é superior ao método convencional em que as bandas de uma mesma imagem são armazenadas na forma de pilha e sem divisão.

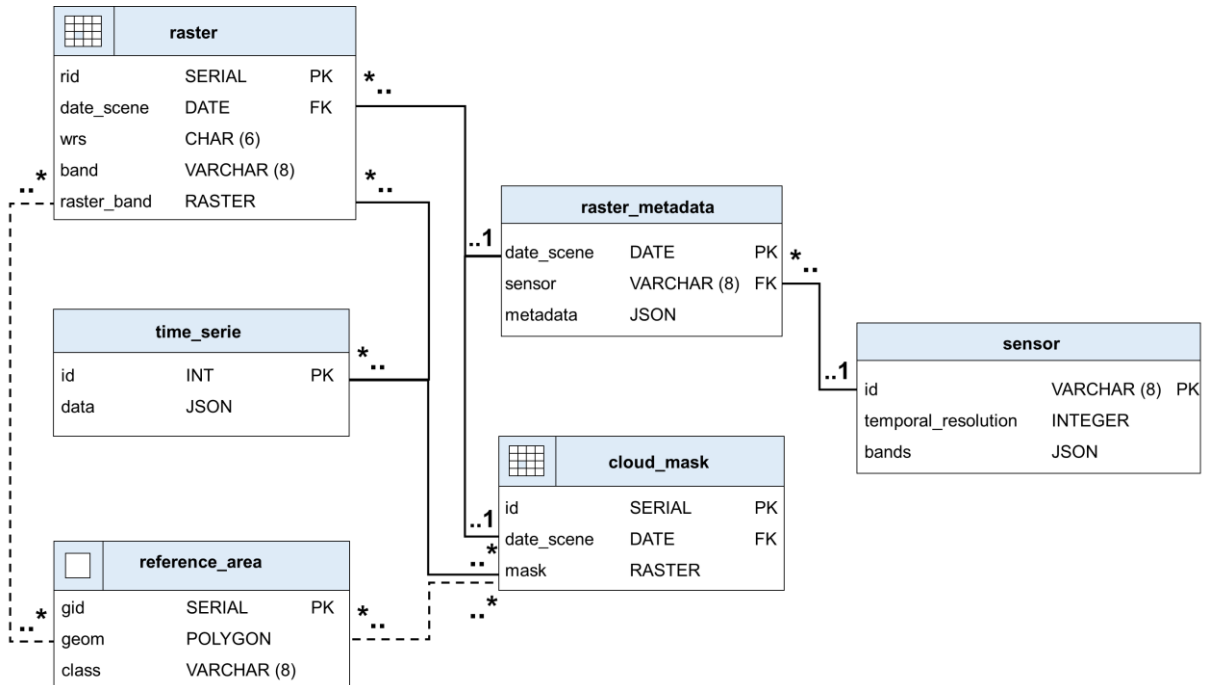
Apesar desses autores apresentarem a eficiência do armazenamento na forma de *tiles*, essa estratégia se mostrou impraticável nesta pesquisa, em razão da dificuldade de reagrupar todos os *tiles* referentes a uma banda espectral, operação que viria a ser realizada em etapas posteriores. A função *ST\_Union* do PostGIS, utilizada para tal tarefa, apresentou como restrição a necessidade de que todos os *tiles* que compõem uma mesma linha ou coluna apresentassem exatamente o mesmo alinhamento o que, possivelmente por questões relacionadas a deformação relativa a projeção das imagens, não ocorreu. Em razão disso, as bandas foram armazenadas sem particionamento e de forma independente, ou seja, cada banda da imagem foi armazenada em uma tupla no banco de dados.

Finalizado o projeto conceitual foi executado o projeto lógico, fase em que o esquema conceitual é mapeado para o modelo de dados empregado pelo SGBD. O produto dessa etapa consiste em um diagrama com detalhes sobre as tabelas, relacionamentos, atributos (domínio, tamanho) do banco de dados no modelo de dados do SGBD.

Posto que informações armazenadas nas entidades *raster\_metadata*, *sensor* e *time\_serie* apresentam variabilidade, propriedade que demanda o armazenamento em um modelo de dados flexível, como o NoSQL, e outras informações se ajustam ao modelo relacional, rigidamente estruturado, foi selecionado o SGBD PostgreSQL para a implementação do banco de dados, pelo fato de permitir uma abordagem híbrida (SQL e NoSQL) de armazenamento.

O produto do projeto lógico é apresentado na Figura 16, por meio do diagrama *Objetct Modeling Technique for Geographic Applications* (OMT-G), o qual exibe entidades, atributos, tipos dos atributos e relacionamentos das entidades.

Figura 16 - Diagrama OMT-G para o banco de dados híbrido



A entidade *raster* apresenta a relação de muitos para um (m:1) com *raster\_metadata*, o que significa que muitos registros da entidade *raster* tem relação com um registro específico de *raster\_metadata*. Isso ocorre devido a cada registro de *raster\_metadata*

armazenar informações referentes a uma cena, todavia uma cena é composta por várias bandas, que são armazenadas de forma independente na entidade *raster*. Já a entidade *cloud\_mask* apresenta relação de um para um (1:1) com *raster\_metadata*, uma vez que cada máscara está relacionada a uma única cena. O relacionamento de *raster* e *cloud\_mask* com *raster\_metadata* é materializado pelo atributo *date\_scene* que é chave estrangeira nas entidades *raster* e *cloud\_mask*.

Na entidade *raster* foi armazenada a informação matricial de uma determinada banda no atributo *raster\_band*. A fim de permitir a identificação da cena pela banda e órbita/ponto em que essa foi adquirida, foram definidos os atributos *band* e *wrs*, respectivamente. O atributo *date\_scene* identifica a data de aquisição da cena. O atributo *rid* tem papel de garantir a unicidade dos registros.

A entidade *cloud\_mask*, referente às máscaras de nuvens e sombra de nuvens, construídas para cada uma das cenas na etapa de pré-processamento, foi definida com os seguintes atributos: *id*, *mask* e *date\_scene*, que correspondem, respectivamente, ao identificador, informação matricial e data a que a máscara se refere.

A entidade *raster\_metadata*, armazena informações das cenas. Em razão de cada cena ser adquirida em uma data única, o atributo *date\_scene* foi definido como chave primária. As demais informações da cena (nome da imagem, satélite, azimute solar, elevação solar, distância Terra-Sol, constantes das bandas termais e parâmetros de transformação) demandam uma estrutura de armazenamento flexível (tipo de dado JSON para o atributo *metadata*), em razão da possibilidade de algumas dessas informações não existirem no arquivo original de metadados e devido ao fato de que a quantidade de constantes das bandas termais e dos parâmetros de transformação diferirem de acordo com o sensor que adquiriu a cena. A relação de uma cena com o sensor é estabelecida pelo atributo (chave estrangeira) *sensor*.

A entidade *sensor* apresenta os atributos *id*, *temporal\_resolution* e *bands*. O *id* corresponde a um identificador exclusivo, *temporal\_resolution* à resolução temporal e *bands* corresponde as informações relacionadas ao espectro eletromagnético e ao intervalo de comprimento de onda em micrometros de cada banda. Em razão dos sensores apresentarem número distinto de bandas espectrais, a informação armazenada no campo *bands* pode ser armazenada de forma semi-estruturada (JSON). A entidade *sensor* apresenta relação de um para muitos (1:m) com *raster\_metadata*, portanto um registro de *sensor* se relaciona com muitos registros de *raster\_metadata*.

A entidade *time\_serie* foi definida com o propósito de armazenar as séries temporais, as quais tem processo de construção apresentado na Seção 3.3.3. Em razão de existir uma série temporal para a posição de cada pixel, foi definido o atributo *id*, determinado a partir da posição do pixel na cena e utilizado para identificar cada série temporal. Já para o armazenamento das séries temporais em si, foi definido o atributo *data*, e especificado como uma estrutura flexível (JSON), uma vez que as séries temporais apresentam tamanhos variados e por permitir o armazenamento de mais séries para o mesmo pixel, como uma série temporal para cada uma das bandas espectrais ou para diferentes índices espectrais.

As regiões de referência se constituem de locais identificados visualmente em um conjunto de imagens e associados a classes específicas de mudança (ou de não mudança), e que são utilizadas para recuperar séries temporais de referência no processo de discriminação de mudanças na cobertura da terra (o processo de identificação de regiões de referência é descrito na seção 3.3.4). Para armazenar essa informação foi definida a entidade *reference\_area* com os atributos *gid*, *geom*, e *class*. O primeiro atributo é um identificador, em *geom* é armazenada a informação vetorial e enquanto *class* é a classe a que a região de referência se refere.

Na Figura 17, podem ser observadas estruturas genéricas que representam os atributos *metadata*, *bands* e *data* das entidades *raster\_metadata*, *sensor* e *time\_serie*, respectivamente, e que foram armazenadas no modelo de dados orientado a documento, no formato JSON.



Figura 17 - Estruturas genéricas dos atributos *metadata*, *bands* e *data* armazenados no formato de documentos

<pre> metadata 01 { 02   "file_name": "LC82330662016239LGN00", 03   "spacecraft": "LANDSAT_8", 04   "image_attributes": { 05     "sun_azimuth": 57.08575716, 06     "sun_elevation": 55.94749481, 07     "earth_sun_distance": 1.0104384 08   }, 09   "rad_rescaling": { 10     "gain": { 11       "B1": 0.012298, 12       ... 13       "B11": 0.0003342 14     }, 15     "offset": { 16       "B1": -61.4884, 17       ... 18       "B11": 0.1 19     } 20   }, 21   "ref_rescaling": { 22     "gain": { 23       "B1": 0.00002, 24       ... 25       "B9": 0.00002 26     }, 27     "offset": { 28       "B1": -0.1, 29       ... 30       "B9": -0.1 31     } 32   }, 33   "thermal_constants": { 34     "k1_B10": 774.8853, 35     "k1_B11": 480.8883, 36     "k2_B10": 1321.0789, 37     "k2_B11": 1201.1442 38   } 39 } </pre>	<pre> bands 01 { 02   "bands": { 03     "band 1": "Ultra Blue - coastal/aerosol", 04     ... 05     "band n": "Thermal Infrared - TIRS2" 06   }, 07   "wavelength-micrometers": { 08     "band 1": "0.43-0.45", 09     ... 10     "band 11": "11.50-12.51" 11   } 12 } </pre>
	<pre> data 01 { 02   "date": ["1984-07-01", ..., "2016-08-26"], 03   "ndvi": [0.673835, ..., 0.166061] 04 } </pre>

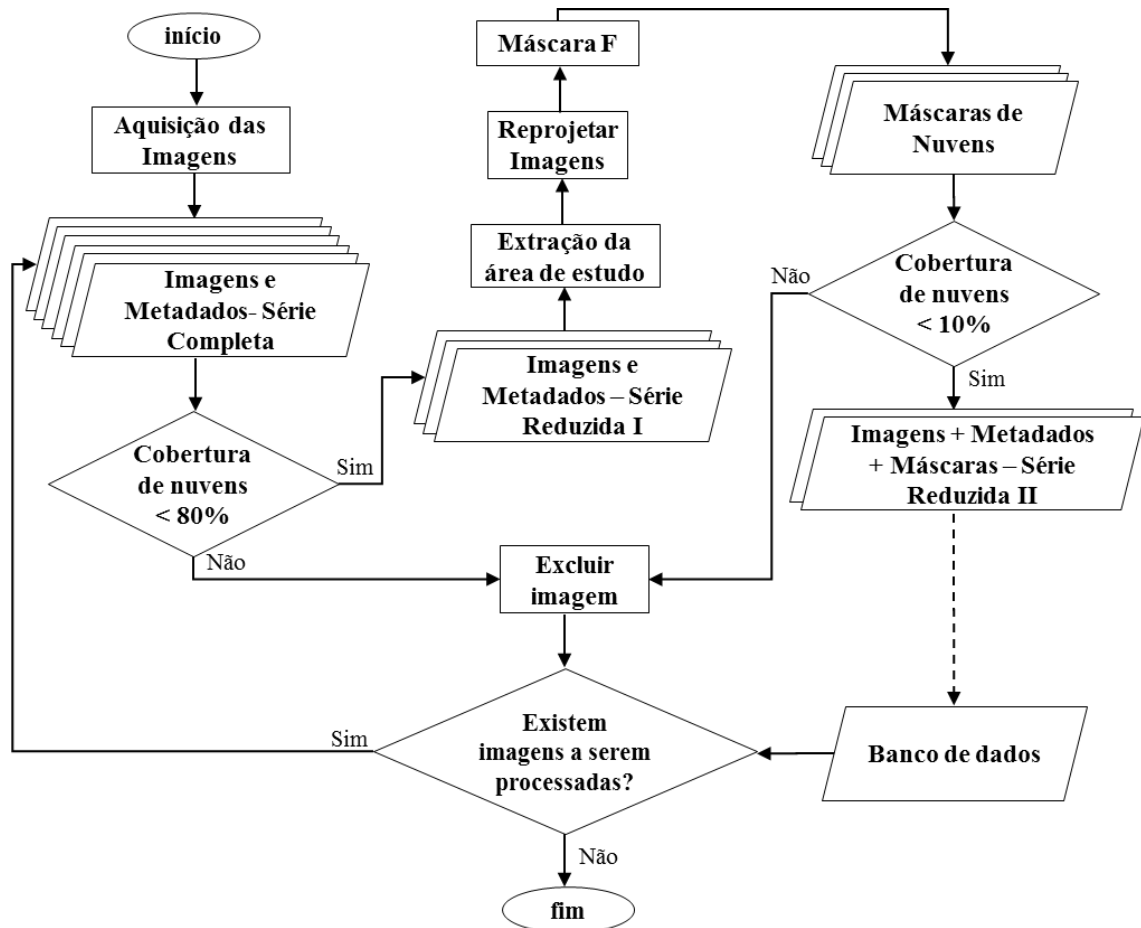
O projeto físico, etapa final do projeto de banco de dados, foi a fase em que o projeto lógico foi implementado. Nessa fase foram elaborados os *scripts*, em linguagem SQL, para a criação do banco de dados como definido nas etapas anteriores.

### 3.3.2 Pré-processamento dos dados

O pré-processamento dos dados, esquematizado na Figura 18, consistiu no processo de aquisição das imagens, aplicação de filtragem no conjunto de dados, conversão dos

dados para o formato definido no projeto do banco de dados e, por fim, inserção no banco de dados.

Figura 18 - Etapas desenvolvidas no de pré-processamento dos dados



A aquisição das imagens foi executada de forma automática com o *script* disponibilizado sob licença *General Public License* (GNU) pelo *Centre d'Etudes Spatiales de la BIOSphère* localizado em Toulouse, França. Esse *script* permite realizar o *download* de cenas adquiridas pelos sensores TM, ETM+ e OLI diretamente do site USGS Earth Explorer.

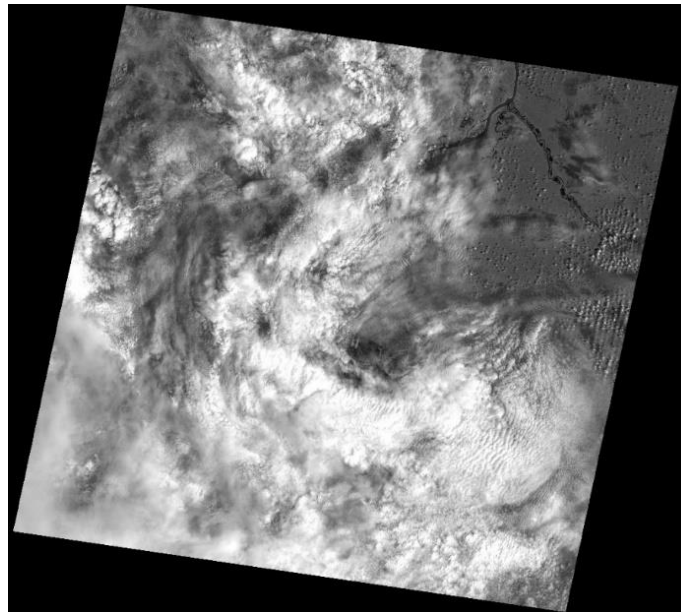
No total foram adquiridas 1294 cenas que totalizam 352 GB, na Figura 18 esse conjunto de dados é denominado como *Imagens e Metadados – Série Completa*. Apesar do volume de espaço ocupado pelas cenas ser pequeno quando comparado com a quantidade de dados de aplicações consideradas *big data*, esse volume de dados é relativamente grande quando confrontado com aplicações convencionais de sensoriamento remoto. Portanto o *cleaning* (limpeza), citada por Chen *et al.*, (2014) como uma técnica convencional de pré-processamento e que tem como propósito otimizar a quantidade de dados, foi aplicada sobre o conjunto de dados com o objetivo de favorecer as análises posteriores.

*Cleaning* pode ser entendido como uma filtragem em que são identificados e eliminados os dados que apresentam informações incorretas, incompletas ou não apropriadas, de modo a aumentar a qualidade do conjunto de dados (CHEN *et al.*, 2014). Na metodologia apresentada, esse processo está relacionado com a extração da área de estudo das imagens e com a eliminação das cenas com cobertura de nuvens e sombra de nuvens na região de interesse.

A eliminação das cenas em razão da presença de nuvens, é decorrente da necessidade de acesso a resposta espectral dos alvos nas imagens. Na Figura 18 é possível verificar que a exclusão das cenas do conjunto de dados em razão da cobertura de nuvens é realizada em duas etapas (*Cobertura de nuvens menor < 80%* e *Cobertura de nuvens menor < 10%*), a primeira antes de extrair a área de estudo e a segunda após o recorte dessa área.

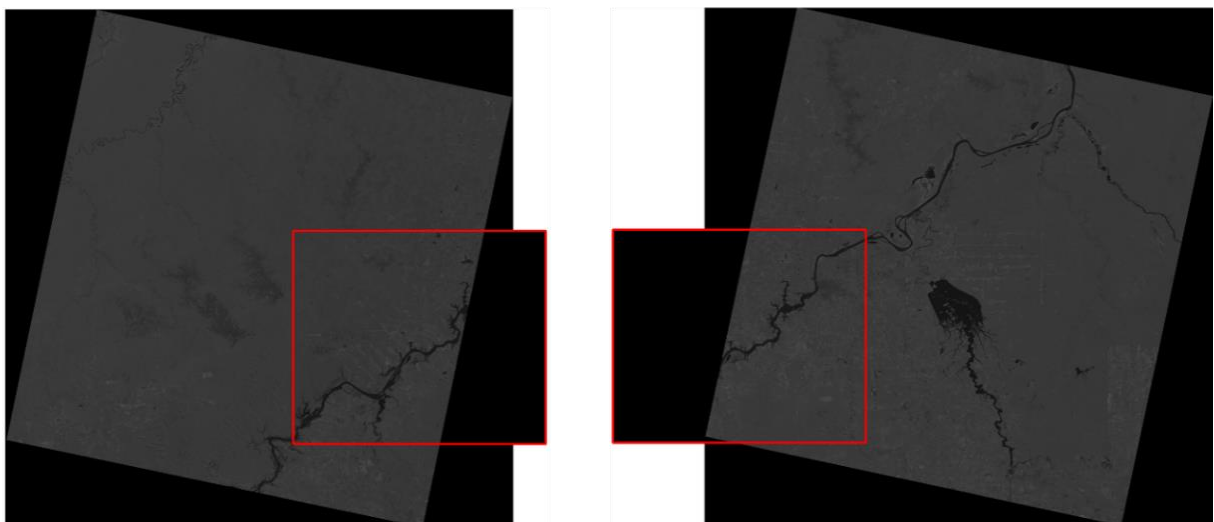
A primeira etapa de filtragem foi realizada considerando a porcentagem de área coberta por nuvens em toda cena, informação presente nos metadados da imagem. Uma vez que a extensão da área estudada corresponde a aproximadamente 20% da área total de cada cena, foi definido como limiar de exclusão o valor de 80% da área da cena coberta por nuvens, visto que, com grande probabilidade, nessas cenas haveriam nuvens na região de interesse, como pode ser observado na Figura 19, a qual tem 80% de sua área coberta por nuvens. O primeiro procedimento de filtragem resulta no conjunto de dados denominado na Figura 18 como *Imagens e Metadados – Série Reduzida I*.

Figura 19 – Ilustração de imagem com 80% de cobertura de nuvens, referente a uma cena TM/Landsat , da banda 4 (infravermelho próximo), órbita/ponto 232/066 do dia 03/01/1991



O pré-processamento seguinte refere-se a extração da área de estudo nas imagens, já que não havia necessidade de armazenar as cenas inteiras. Na Figura 20 são apresentadas as órbitas/ponto 233/066 e 232/066 da imagem OLI/Landsat 8, banda espectral do infravermelho próximo, com destaque para o recorte das áreas de interesse em cada uma das imagens.

Figura 20 - Órbitas/ponto 233/066 e 232/066 dos dias 09/07/2016 e 02/07/2016 respectivamente, da imagem OLI/Landsat 8, banda do infravermelho próximo, destacando a área de estudo



Na sequência os recortes das imagens foram reprojctados para o hemisfério Sul com o propósito de corrigir o georeferenciamento das cenas e atualizar o código EPSG para 32720 que corresponde ao WGS84/UTM 20S, a fim de possibilitar o reconhecimento da posição correta das imagens quando forem armazenadas no banco de dados.

Visando minimizar o volume de dados foi realizada uma nova filtragem a partir de uma nova avaliação da cobertura de nuvens na área de estudo. Para tanto, foi empregado o algoritmo *Fmask* para a construção de máscaras de nuvens e sombra de nuvens para a área de estudo de cada cena. O algoritmo *Fmask* foi desenvolvido com o pressuposto de detectar nuvens, sombra de nuvens e neve em imagens dos satélites Landsat 4, 5 e 7 e atualizado para detectar os mesmos alvos em imagens do satélite Landsat 8 e Sentinel 2, alcançando uma acurácia global de 96,4% para imagens Landsat 4-7 e de 89% para imagens adquiridas a partir do Landsat 8 (ZHU e WOODCOCK, 2012; ZHU *et al.*, 2015).

Verificando a máscara de nuvens e sombra de nuvens gerada para cada cena, foram eliminadas as imagens com mais de 10% dos pixels não nulos cobertos por nuvens ou sombra de nuvens. As máscaras, assim como as imagens e os metadados não descartados, compuseram o conjunto denominado na Figura 18 como *Imagens, Metadados e Máscaras – Série Reduzida II*.

Na Figura 21 é apresentada uma cena que não foi eliminada na primeira etapa de filtragem, devido a cobertura de nuvens de toda cena corresponder a apenas 22,04% de sua área. Entretanto na filtragem seguinte, na qual foi avaliada a máscara de nuvens gerada apenas para a área de interesse, verificou-se que a ocorrência de nuvens e sombra de nuvens era superior ao critério de 10% estabelecido, de modo que, essa cena foi eliminada do conjunto de dados. Na Figura 22, é apresentada uma cena que atende ao critério de cobertura de nuvens e sombra de nuvens menor que 10% e não foi excluída do conjunto de dados.

Figura 21 - Ilustração da aplicação de máscara de nuvens e sombra de nuvens na banda do infravermelho próxima de imagem OLI/Landsat 8, órbita/ponto 232/066, do dia 15/05/2016 - cena com alta cobertura de nuvens

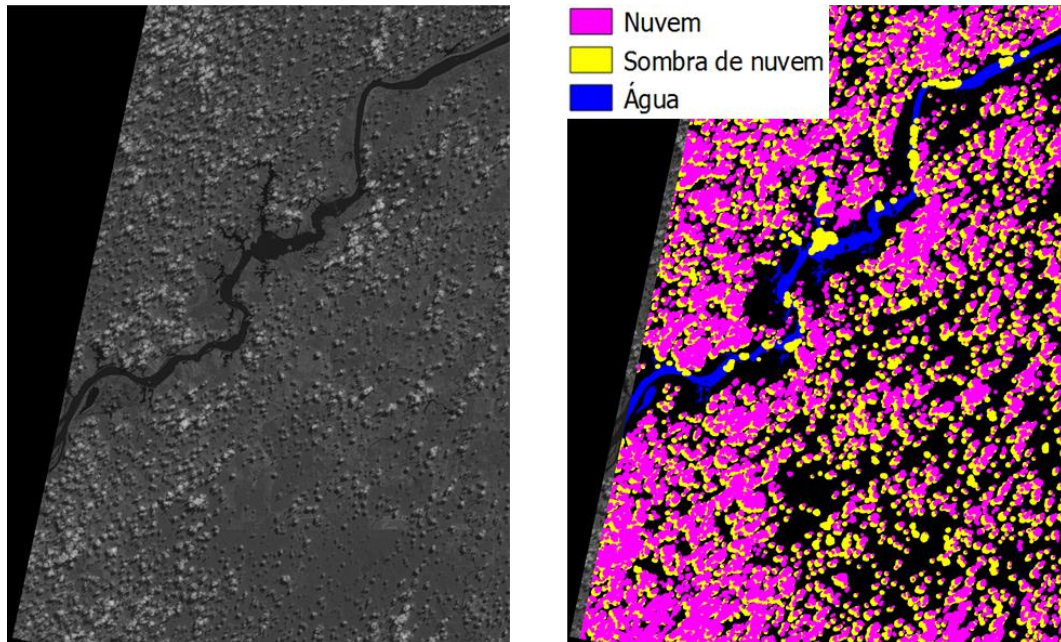
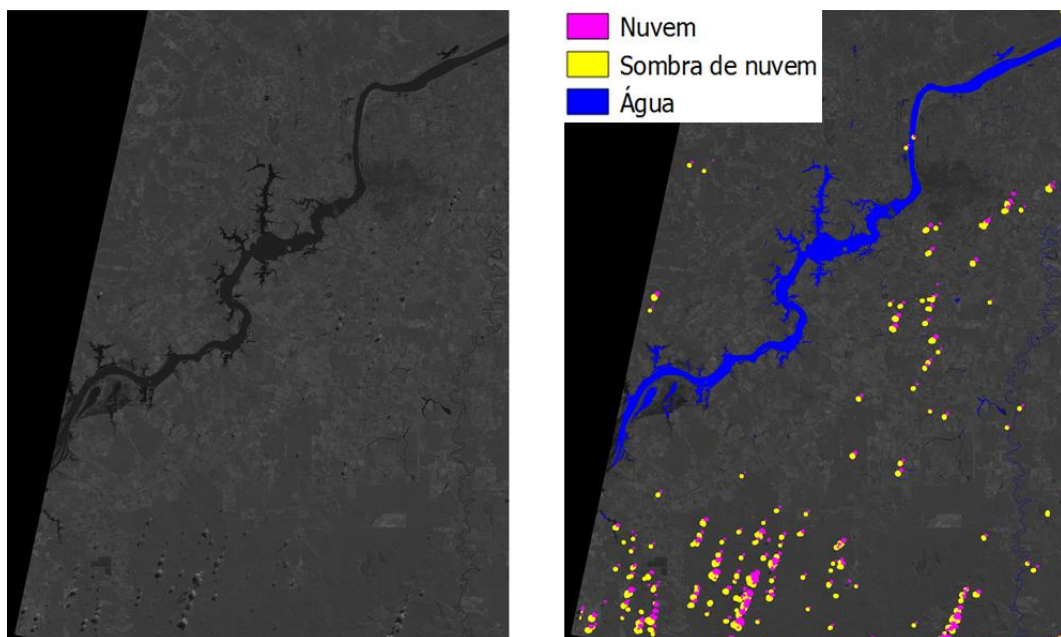


Figura 22 -- Ilustração da aplicação de máscara de nuvens e sombra de nuvens na banda do infravermelho próxima de imagem OLI/Landsat 8, órbita/ponto 232/066, do dia 14/06/2015 - cena com pequena cobertura de nuvens



Na sequência as imagens, os metadados e as máscaras de nuvens foram ajustados aos tipos, estruturas e restrições definidos no projeto do banco de dados, uma vez que, para sua inserção no banco, o SGBD verifica se os dados estão em concordância com o dicionário de dados, antes de aceitar a inserção no banco de dados. Realizada essa operação, a última etapa do pré-processamento foi a inserir as imagens, os metadados e máscaras de nuvens no banco de dados.

### 3.3.3 Construção das Séries Temporais

Essa etapa desenvolvida consistiu em gerar séries temporais com valores de NDVI, para cada pixel da imagem. Para tanto foi definido um identificador (*id*) para cada posição de pixel (linha, coluna) na cena a partir da Equação 2, a fim de possibilitar a recuperação da série temporal no banco de dados com base na posição do pixel na cena.

$$id = lin * ncol + col + 1 \quad (2)$$

Em que:

*id*: Corresponde ao identificador do pixel;

*lin*: a linha do pixel que o identificador será calculado;

*ncol*: a quantidade de colunas da imagem;

*col*: a coluna do pixel que o identificador será calculado.

O NDVI, definido pela equação 3, foi selecionado como o índice espectral a ser utilizado, devido a sua correlação com propriedades biofísicas da vegetação, tais como condições fitossanitárias, índice de área foliar e biomassa (Carlson e Ripley, 1997). Considerando as características da área de estudos e a expectativa de discriminar mudanças na cobertura da terra, como desmatamento, sua aplicação é particularmente apropriada.

$$NDVI = \frac{(IVP - Verm)}{(IVP + Verm)} \quad (3)$$

Em que:

NDVI: Imagem índice de vegetação por diferença normalizada;

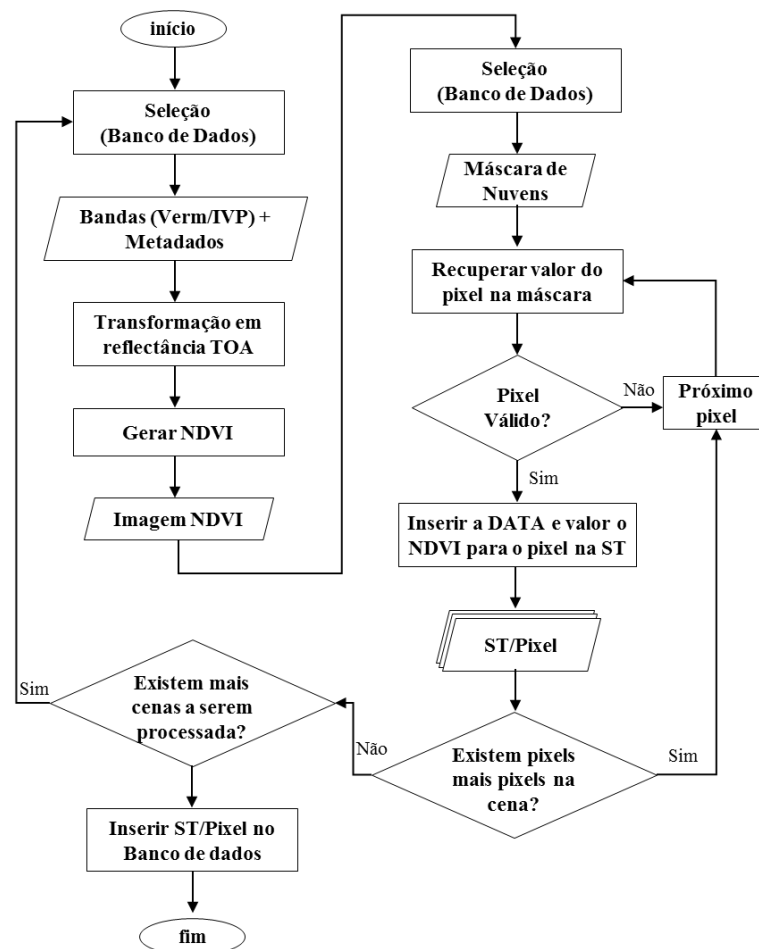
IVP: Imagem adquirida no intervalo espectral do infravermelho próximo;

Verm: Imagem adquirida no intervalo espectral do vermelho visível.

A capacidade do NDVI em detectar variações na cobertura vegetal decorre das bandas utilizadas em sua construção, que são adquiridas nos intervalos espectrais correspondentes ao infravermelho próximo, onde a vegetação apresenta a maior taxa de reflectância, e do vermelho onde ocorre a maior taxa de absorção. Os valores de NDVI variam de -1 a +1, sendo que, valores negativos ou próximos de 0, indicam a presença de água, solo exposto ou área construída, enquanto valores positivos ou mesmo próximos de 1, indicam a presença de vegetação, de forma que quanto maior o valor, maior a quantidade ou melhor a condição dessa vegetação.

Sendo assim, para a construção das séries temporais de NDVI, foram realizados os procedimentos apresentados na Figura 23. Essa sequência de procedimentos foi realizada de forma iterativa para todas as imagens armazenadas no banco de dados.

Figura 23 - Fluxograma das atividades relacionadas com a geração de séries temporais com valores de NDVI para cada pixel da imagem, corrigidos para reflectância TOA





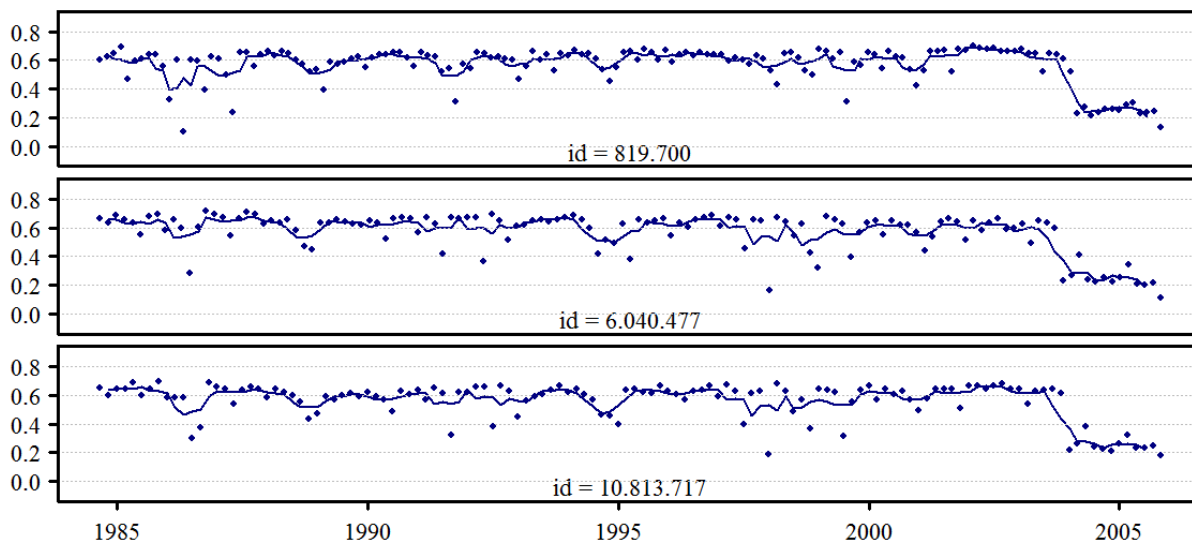
A primeira etapa tratou-se da seleção das bandas correspondentes aos intervalos espectrais do vermelho e infravermelho próximo no banco de dados. Na sequência, os valores de ND das bandas originais foram convertidos para valores de reflectância TOA. Essa transformação foi realizada para a calibração radiométrica dos sensores e a fim de minimizar as variações existentes entre as imagens, devido a efeitos relativos a diferenças na geometria de iluminação, que são provocados pelos diferentes ângulos solares e distâncias Terra-Sol.

Para a conversão dos ND dos pixels das bandas originais em valores de reflectância TOA, foi implementada uma função de transformação diretamente no banco de dados, os parâmetros passados para a função são os valores de ganho (*ref\_rescaling->gain*) e *offset* (*ref\_rescaling->offset*) armazenados na tabela *raster\_metadata* ou valores de ganho e *offset* calculados a partir de atributos do sensor e da imagem armazenados na mesma tabela. As bandas convertidas em reflectância TOA, foram usadas no cálculo do NDVI, a função utilizada também foi definida e executada no banco de dados.

Na sequência foi realizada a seleção, no banco de dados, da máscara de nuvens correspondente a imagem NDVI da data sendo processada. Todos os pixels da máscara de nuvens foram percorridos de forma iterativa, a fim de verificar se esse pixel apresentava valor válido (não é coberto por nuvem ou sombra de nuvem). Caso essa verificação fosse positiva, a data correspondente a imagem sendo processada e o valor de NDVI do pixel avaliado, eram inseridos na série temporal definida pelo identificador do pixel, mais precisamente nos objetos *date* e *ndvi* respectivamente, no atributo *data* da tabela *time\_serie*.

Ao final do processamento foram construídas e armazenadas 12.015.232 séries temporais no banco de dados. A fim de ilustrar a representação de algumas séries, indica-se na Figura 24, as séries temporais correspondentes aos pixels nas posições (307, 217), (2652, 1599) e (3028, 2863), identificados pelos *id* 819700, 6040477 e 10813717, respectivamente. Nessa figura os pontos indicam a série temporal bruta e a linha contínua trata-se da série temporal suavizada pelo filtro da média móvel com janela de tamanho três.

Figura 24 - Representação das séries temporais de NDVI obtidas para os pixels com *ids* 819700, 6040477 e 10813717

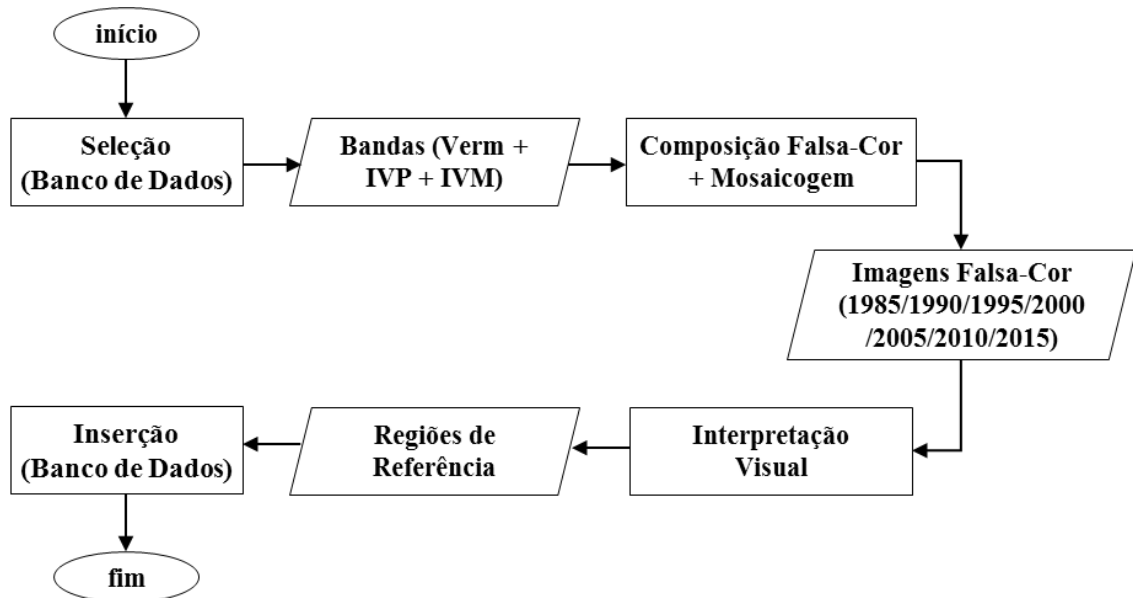


Em razão da existência da área de sobreposição entre as duas cenas utilizadas no estudo e da filtragem realizada na construção das séries, com o propósito de não incluir os pixels cobertos por nuvens ou sombra de nuvens, as séries temporais apresentam tamanhos irregulares, com número máximo de 279 datas válidas, visto que essa foi quantidade de cenas armazenadas no banco de dados.

### 3.3.4 Identificação de Regiões de Referência

A identificação de regiões de referência para as classes de mudança e não mudança, tem seu esquema apresentado na Figura 25. Esse procedimento, foi desenvolvido com o propósito de permitir o reconhecimento dos tipos de cobertura da terra propriamente dito (floresta, vegetação secundária, solo exposto, área urbana e corpo d'água) e classes correspondentes a mudanças antrópicas (desmatamento, cuja trajetória parte de floresta para qualquer uma das outras classes) e mudanças não antrópicas (reflorestamento e inundação), a partir da série temporal de NDVI.

Figura 25 – Etapas relacionadas com a identificação de áreas de referência

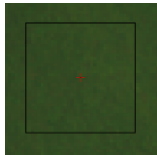
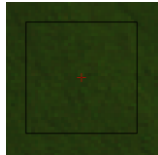

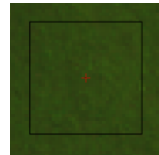
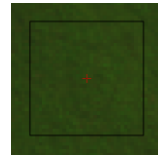
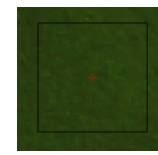
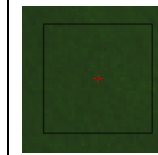

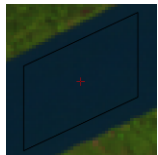
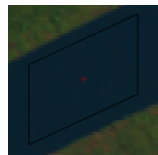
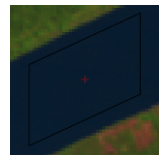

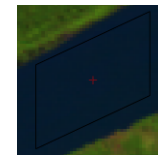
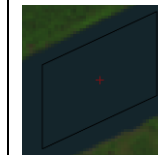




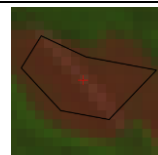


















Em razão de ser impraticável pesquisar todas as 279 cenas armazenadas no banco de dados a fim de identificar as regiões de referência para a classificação, foram selecionadas imagens tomadas em sete anos específicos (1985, 1990, 1995, 2000, 2005, 2010, 2015) para essa avaliação. Para cada um desses anos foi escolhida uma única cena para cada uma das duas órbitas/ponto que compõe a área.

Especificadas as cenas a serem utilizadas na identificação de regiões de referência, foram recuperadas no banco de dados as bandas espectrais para gerar composições falsa-cor e realizar a mosaicagem entre as composições do mesmo ano.

Por fim, realizando inspeção visual sobre esse conjunto de imagens, denominado no fluxograma como *Imagens Falsa-Cor (1985/1990/1995/2000/2005/2010/2015)*, foram identificadas e vetorizadas regiões de referência para cada uma das classes especificadas como de mudança ou não mudança. No Quadro 4 são apresentados exemplos de regiões de referência para as classes definidas, agrupadas em três categorias genéricas: sem mudanças, mudanças antrópicas e mudanças naturais. Por meio desse quadro é possível compreender o processo de identificação de regiões de referência, em que o analista avaliou de forma cronológica a existência de mudanças para os diferentes tipos de cobertura do terra presentes na área de estudo.

Quadro 4 - Regiões de referência dos tipos de cobertura da terra que caracterizam áreas sem mudança, áreas de mudança antrópica e de mudança natural, identificadas nos anos de 1985, 1990, 1995, 2000, 2005, 2010 e 2015

Classe		Anos de Referência						
		1985	1990	1995	2000	2005	2010	2015
Sem Mudanças	Floresta							
	Água							
	Solo Exposto							
	Área Urbana							
	Veg. Secundária							
							continua	

Classe		Anos de Referência						
		1985	1990	1995	2000	2005	2010	2015
Mudanças Antrópicas	Desmatamento (Floresta-Área Urb.)							
	Desmatamento (Floresta-Veg. Sec.)							
	Desmatamento. (Floresta-Solo Exp.)							
	Desmatamento (Floresta-Água)							
Mudanças Naturais	Reflorestamento							
	Inundação							

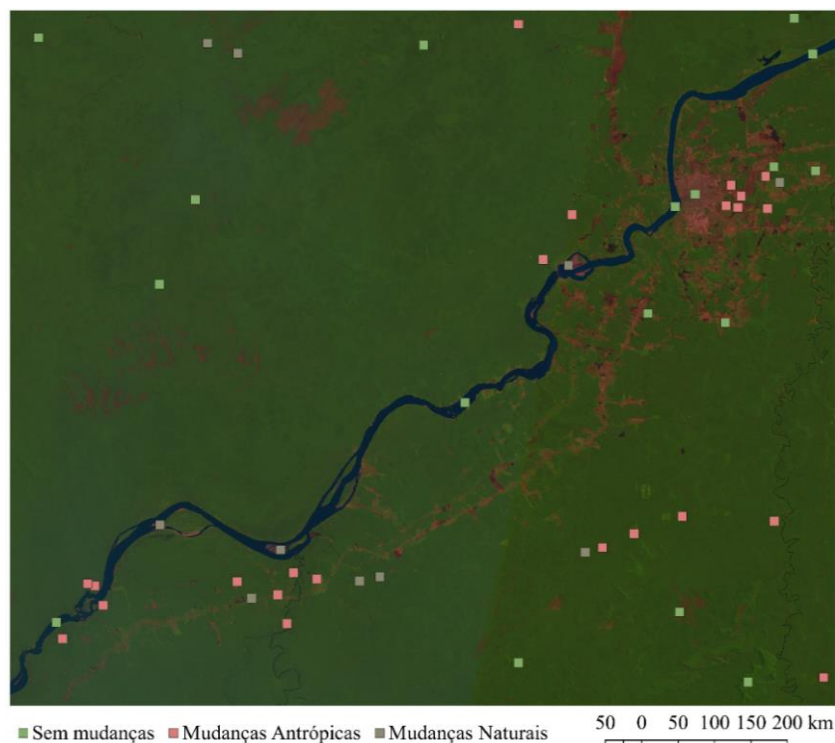
Como pode ser observado no Quadro 4, a categoria denominada sem mudanças compreende os tipos de cobertura do solo que não sofreram alterações desde 1985 até 2015, e inclui as classes: floresta, água, solo exposto, área urbana e vegetação secundária.

Nas mudanças categorizadas como antrópicas foram vetorizados polígonos em locais onde foi observada, em algum momento ao longo da linha temporal, mudança na cobertura da terra em decorrência de intervenção humana, incluindo: mudança de floresta para área urbana ou, vegetação secundária ou, solo exposto ou água.

As mudanças naturais englobaram as alterações de solo exposto ou vegetação secundária para floresta denominadas como reflorestamento. Outra mudança no tipo de cobertura do solo inserida nessa categoria foram as inundações (variações no nível do rio), que são frequentes e dependem do regime de chuvas.

Na Figura 26 é apresentada a disposição das regiões de referência na área de estudo. No total foram vetorizados 50 polígonos, sendo 17 da classe sem mudanças, 23 da classe mudanças antrópicas e 10 da classe mudanças naturais. As áreas de referência foram distribuídas na região de estudo procurando inserir o máximo possível de variações espectrais decorrentes da localização geográfica, inclusive entre os mesmos tipos de cobertura do solo.

Figura 26 - Distribuição das regiões de referência na área de estudo

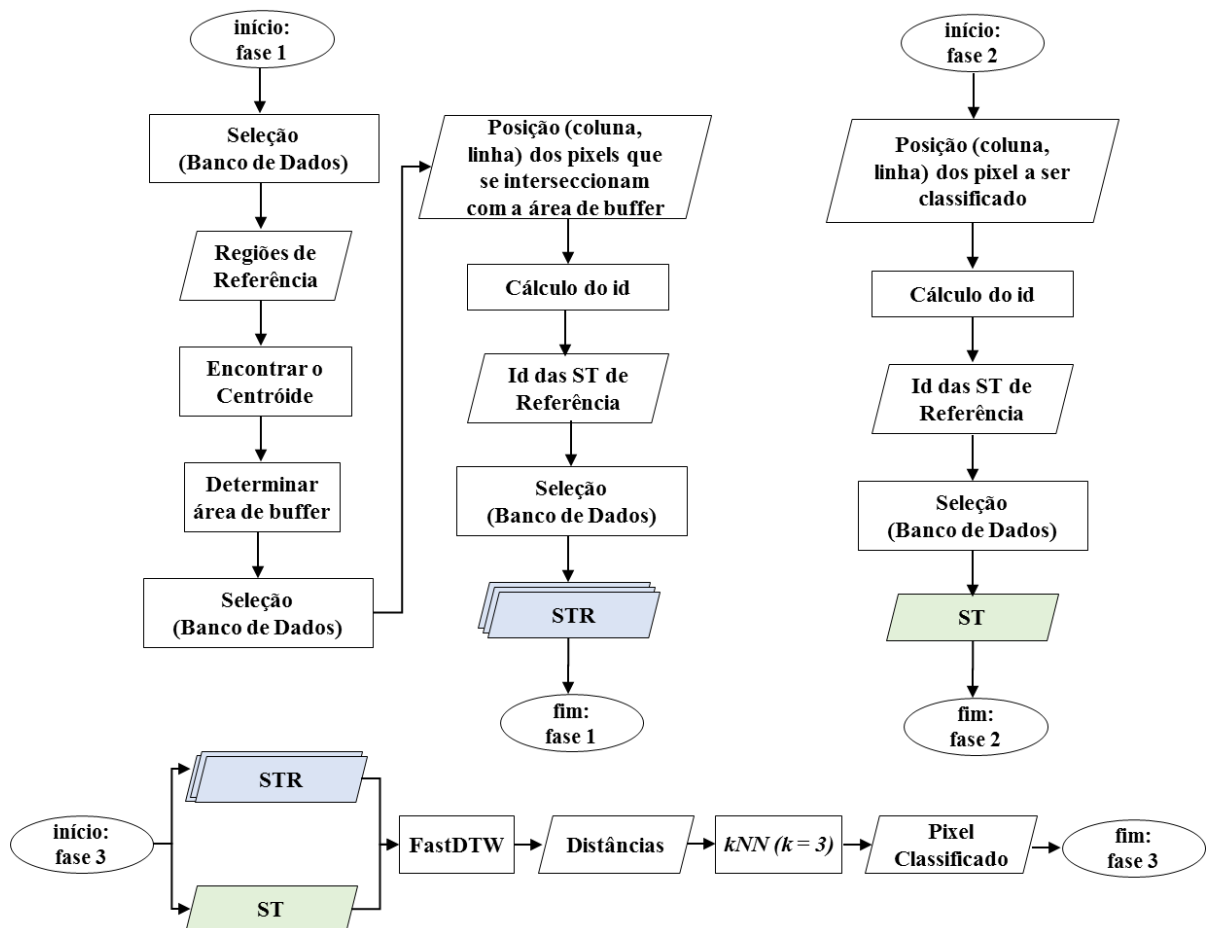


Identificadas e vetorizadas as regiões de referência essa etapa foi finalizada com a inserção dos polígonos no banco de dados, sendo essa informação alocada na tabela *reference\_area*.

### 3.3.5 Discriminação de mudanças nos tipos de cobertura da terra

A etapa de discriminação de mudanças, consistiu na classificação das séries temporais, com o propósito de construir o mapa de mudanças na cobertura da terra. Seu desenvolvimento, incluiu uma estratégia para dispor de um número significativo de séries temporais de referência, das áreas de mudança e não mudança, além da classificação propriamente dita. Na Figura 27 são apresentadas as três fases em que a etapa de discriminação de mudanças pode ser dividida, a fase 1 consiste na recuperação das séries temporais de referência, a fase 2 na recuperação da série temporal do pixel a ser classificado e a fase 3 na para a determinação da classe dessa série temporal utilizando como dado de entrada os resultados da fase 1 e fase 2.

Figura 27 - Fases para a classificação de um pixel por meio da análise de séries temporais



Na fase 1 o primeiro procedimento realizado foi a seleção das regiões de referência, armazenadas no banco de dados na tabela *reference\_area*. Para cada região de referência foi determinado o seu centroide, e a partir dele foi calculada uma área de *buffer* com o objetivo de recuperar a posição (coluna, linha) dos pixels incluídos nessa região. A partir da posição recuperada para cada um dos pixels foi possível calcular, por meio da equação 2, o identificador do pixel. Os identificadores apontam para as séries temporais armazenadas na tabela *time\_serie*, uma vez que correspondem ao atributo *id*. Portanto, o identificador do pixel e sua respectiva classe, definida pela região de referência em que o mesmo foi inserido, possibilitou recuperar as séries temporais denominadas *STR* na Figura 27. As séries temporais de referência indicam perfis de NDVI correspondentes aos tipos de cobertura da terra presentes nas categorias Sem Mudanças, Mudanças Antrópicas e Mudanças Naturais.

A fase 2 consiste na recuperação da série temporal de um pixel específico, essa fase é iniciada a partir da posição do pixel a ser classificado, a qual é utilizada para o cálculo, por meio da Equação 2, do *id* do pixel. Na sequência, a série temporal desse pixel é recuperada no banco de dados. O produto dessa fase está em destaque na Figura 27 e é denominado como *ST*.

Após a etapa de recuperação das séries temporais de referência e do pixel a ser classificado, é realizada a fase 3 que consiste na classificação de uma única série temporal. Para a classificação de uma região específica da imagem (vários pixels) a fase 2 e 3 são realizadas iterativamente para todos os pixels incluídos na região definida.

A fase 3, é realizada por meio da comparação entre a série temporal a ser classificada com todas as séries temporais de referência, logo os resultados das fases 1 e 2 são utilizados como dados de entrada. De forma iterativa é realizada a medida de similaridade entre essa série temporal e todas as séries temporais de referência, para tanto foi empregado o algoritmo FastDTW. Conhecendo as distâncias entre a série temporal a ser classificada e todas as séries temporais de referência, a classificação foi realizada com algoritmo *K-Nearest Neighbors* (kNN), para  $k = 3$ , de modo a associar essa série temporal de interesse a classe mais votada, entre as 3 mais similares a essa.

Como mencionado na seção 2.1.1 o DTW apresenta elevado custo computacional, o que torna a classificação lenta, visto que para definir a classe de um pixel o algoritmo é aplicado na comparação da série temporal do mesmo com todas as séries temporais de referência. Para avaliar o tempo despendido na classificação, foram utilizadas as distâncias



de 30, 35 e 60 metros como parâmetro para a determinação das áreas de *buffer* a partir do centroide das regiões de referência, foram recuperadas conjuntos com 154, 218 e 634 pixels, respectivamente. Na Tabela 1 é apresentado o tempo médio para a classificação de 1 pixel para os algoritmos DTW e FastDTW, levando em conta os diferentes tamanhos dos conjuntos de séries temporais de referência, a quantidade de núcleos utilizados para o processamento paralelo, considerando um servidor com *clock* de 3.4 GHz e 16 GB de memória e para, o FastDTW, valores de raio igual a 1 e 20.

Tabela 1 - Tempo médio (em segundos) para a classificação de um pixel, variando o algoritmo, tamanho dos conjuntos de séries temporais de referência e número de núcleos de processamento utilizados

		<b>Média de Tempo (segundos)</b>				
	<b>Núcleos Tamanho</b>	<b>1</b>	<b>2</b>	<b>4</b>	<b>6</b>	<b>8</b>
<b>DTW</b>	<b>154</b>	382,112	188,851	106,746	105,252	104,659
	<b>218</b>	544,673	284,391	150,909	148,618	146,374
	<b>634</b>	1528,988	806,717	436,692	416,623	410,883
<b>FastDTW r = 1</b>	<b>154</b>	1,504	0,804	0,505	0,508	0,414
	<b>218</b>	2,107	1,204	0,608	0,607	0,610
	<b>634</b>	5,912	3,010	1,708	1,712	1,713
<b>FastDTW r = 20</b>	<b>154</b>	25,339	13,624	7,216	7,122	7,124
	<b>218</b>	36,955	10,622	10,622	10,525	10,328
	<b>634</b>	105,147	53,581	28,550	28,557	29,160

Devido ao elevado tempo de execução, tanto do DTW como do FastDTW (com  $r$  igual a 1 ou a 20), foi impraticável classificar os 12.015.232 pixels da imagem, de modo que foi selecionada uma sub-área (área teste) na cena com 100.000 pixels para executar esse processamento. Essa área teste foi classificada com duas configurações, em ambas foi utilizado o conjunto de referência com 154 séries temporais. O processamento foi realizado em paralelo nos oito núcleos de um servidor, empregando o algoritmo *FastDTW*, com raio de tamanho um ( $r = 1$ ) em uma primeira configuração (I), e raio igual a vinte ( $r = 20$ ) na segunda configuração (II).

A fim de validar a classificação das séries temporais, foram definidas regiões de referência na área teste de 100.000 pixels e, a partir da interpretação visual das imagens adquiridas em 1985, 1990, 1995, 2005, 2010 e 2015, foram estabelecidas as classes de mudança e não mudança, em novas regiões de referência. Com base na comparação do resultado da classificação com essas regiões de referência foram construídas as matrizes de concordância para avaliar o desempenho da classificação.

## 4 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

### 4.1 Banco de dados com abordagem híbrida

A estrutura de armazenamento definida com a propósito de organizar e armazenar imagens multitemporais georreferenciadas, consistiu em um banco de dados espaço-temporal híbrido que combinou características dos modelos de dados relacional e não relacional. O banco de dados foi fundamental para a automatização da aplicação, visto que sem esse, o procedimento convencional seria armazenar os dados no formato de arquivos o que torna o gerenciamento e a extração de informações uma atividade complexa, tornando muitas vezes impraticáveis aplicações com grandes conjuntos de dados, em consequência da dificuldade de se organizar um grande volume de arquivos e devido à alta probabilidade de falhas no gerenciamento manual desses dados. O SGBD utilizado proporcionou mecanismos para o armazenamento e recuperação dos dados de forma consistente e segura, o que consequentemente garantiu a automatização das diversas operações realizadas sobre o conjunto de dados.

Com relação ao armazenamento das imagens na forma de *tiles* foi possível verificar que apesar de estudos indicarem as vantagens dessa abordagem, essa forma de armazenamento mostrou problemas neste projeto, devido à dificuldade de se automatizar a recomposição dos *tiles* no formato original da banda, isso decorreu do não alinhamento de todos os *tiles* pertencentes as mesmas linhas e colunas, em razão disso as imagens foram armazenadas em seu formato original no banco de dados, sem comprometer os procedimentos seguintes, visto que as operações realizadas sobre as imagens foram executadas em toda cena.

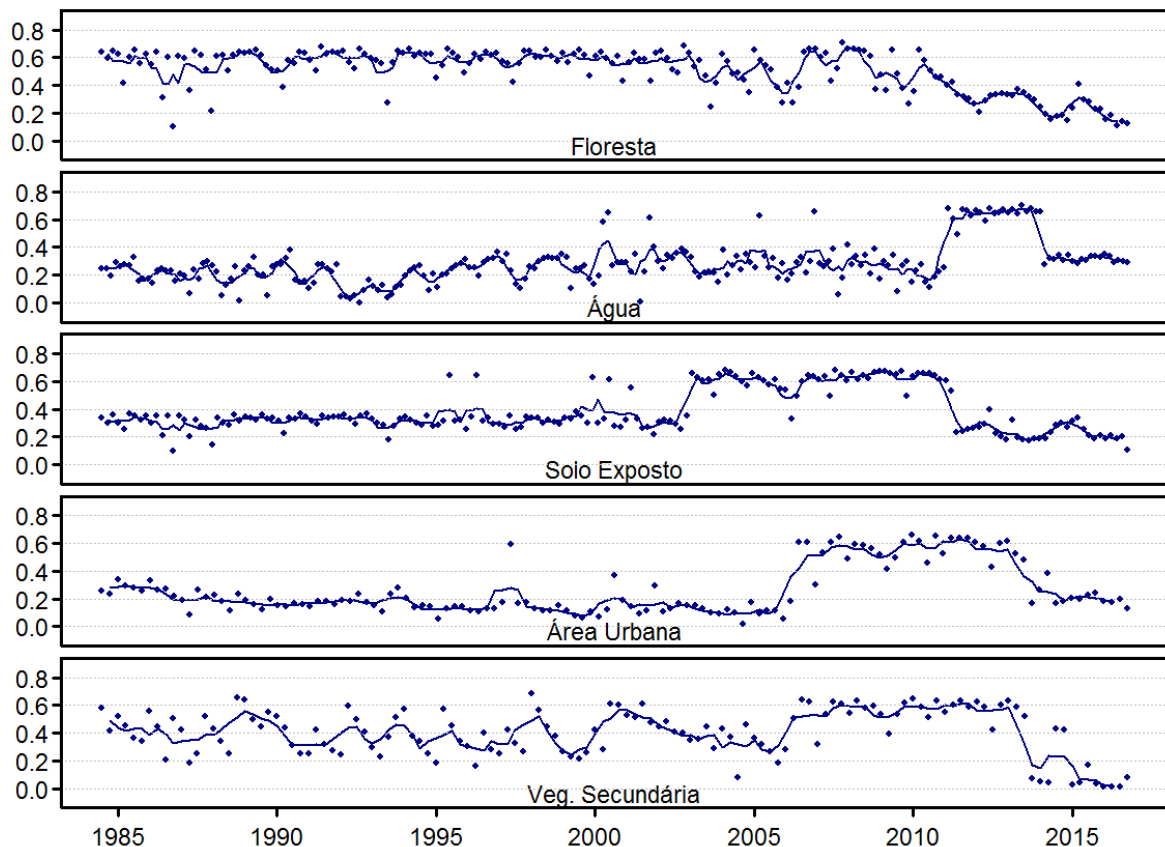
No armazenamento dos metadados das imagens, de informações relativas aos sensores e das séries temporais foi empregado o modelo orientado a documento (NoSQL). Esse modelo de armazenamento foi escolhido devido aos dados demandarem uma estrutura de armazenamento flexível, visto que as informações foram adquiridas por três sensores distintos e devido a isso apresentam variabilidade. Com relação as séries temporais essa forma de armazenamento se mostrou adequada em razão desses dados apresentarem tamanhos variáveis, e em consequência desse modelo favorecer a extração de informação na aplicação adotada, em que a vizinhança temporal entre as observações tem papel fundamental na descrição dos fenômenos de interesse.

## 4.2 Séries Temporais de Referência

Para cada uma das áreas de referência ilustradas no Quadro 4 foi selecionado um pixel e gerada a série temporal de NDVI correspondente. Nas Figuras Figura 28, Figura 29 e Figura 30 são apresentados os perfis de NDVI para os tipos de cobertura da terra incluídos nas classes Sem Mudanças, Mudanças Antrópicas e Mudanças Naturais, respectivamente. Nesses gráficos, os pontos correspondem aos valores de NDVI e a linha contínua indica a série temporal suavizada.

Figura 28 - Séries Temporais de NDVI para os tipos de cobertura do solo inseridos na classe

Sem Mudanças



Na Figura 28 são apresentados os perfis de NDVI para os tipos de cobertura da terra Floresta, Água, Solo Exposto, Área Urbana e Vegetação Secundária. Apesar de cada um desses tipos de cobertura não se alterarem ao longo da série temporal (inseridos na classe Sem Mudanças), os valores de NDVI apresentam oscilações naturais, como pode ser observado

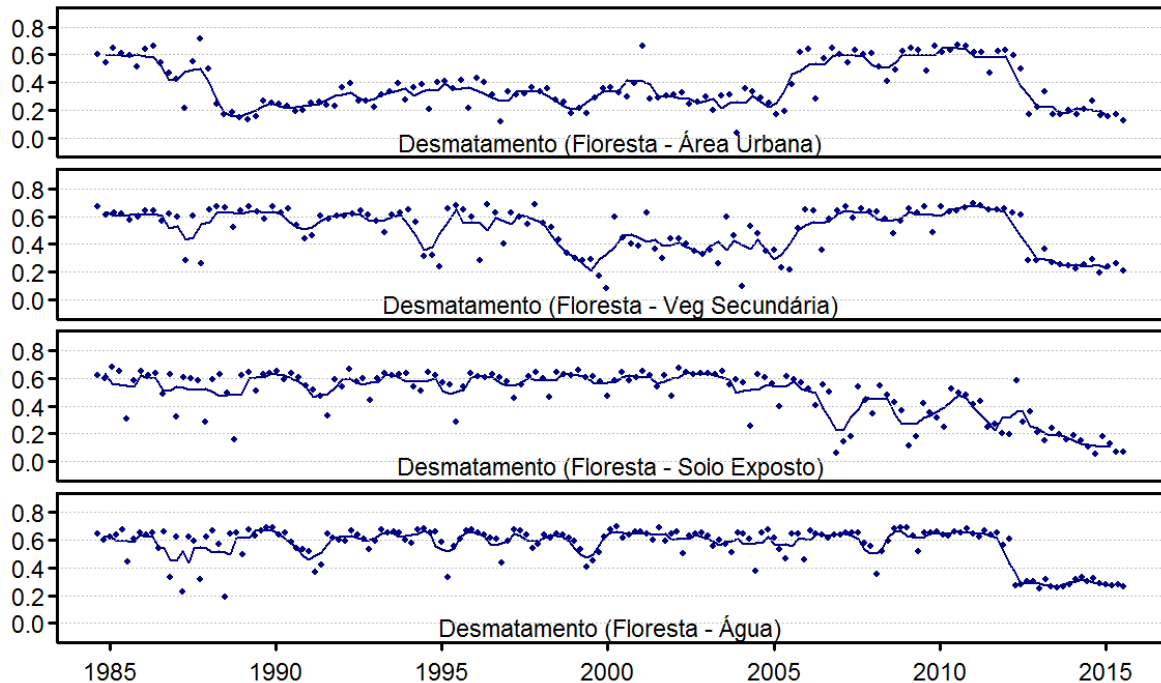
nos gráficos. Essas variações são decorrentes das diferentes condições atmosféricas na aquisição das cenas e da fenologia da vegetação (em Floresta e Vegetação Secundária), nesse caso uma cobertura natural, que é influenciada por fatores intrínsecos do alvo e do ambiente. De maneira geral, as variações nos tipos de cobertura da terra dessa classe são suaves, contudo é possível observar uma variação abrupta próxima dos anos de 2011/2012 para a Água, caracterizada pelo aumento nos valores de NDVI na série temporal, essa alteração coincide com o período em que a UHE de Santo Antônio entrou em operação, o que aponta para a possibilidade desse aumento nos valores de NDVI estar relacionado com o enchimento do reservatório, o que pode ter provocado alguma anomalia na resposta espectral da água.

Além das variações mencionadas, é possível observar uma mudança no comportamento da série temporal (em maior ou menor grau) a partir do ano de 2013, período em que o satélite Landsat 8 entrou em operação e as imagens passaram a ser adquiridas pelo sensor OLI. Embora o NDVI tenha sido calculado a partir de imagens com ND convertido para reflectância TOA, a diferença nos intervalos espectrais das bandas do vermelho e do infravermelho próximo do TM, ETM+ e OLI podem não ter sido totalmente compensadas, assim como a variação na resolução radiométrica dos sensores (TM e ETM+ é de 8 bits e OLI é 12 bits).

Finalmente, é necessário considerar que na seleção das regiões de referência para os tipos de cobertura foi verificada a estabilidade temporal a cada cinco anos e por meio de inspeção visual, o que pode ser insuficiente para detectar alterações de curta duração, ou mesmo muito sutis nas áreas de referência selecionadas. As inconsistências observadas nos valores de NDVI principalmente para Solo Exposto (aproximadamente entre 2002 e 2013) e Área Urbana (2006-2013) pode estar associado à não estabilidade temporal do alvo.

As possibilidades de alteração na cobertura da terra incluídas na classe Mudanças Antrópicas estão indicadas nas quatro séries presentes na Figura 29.

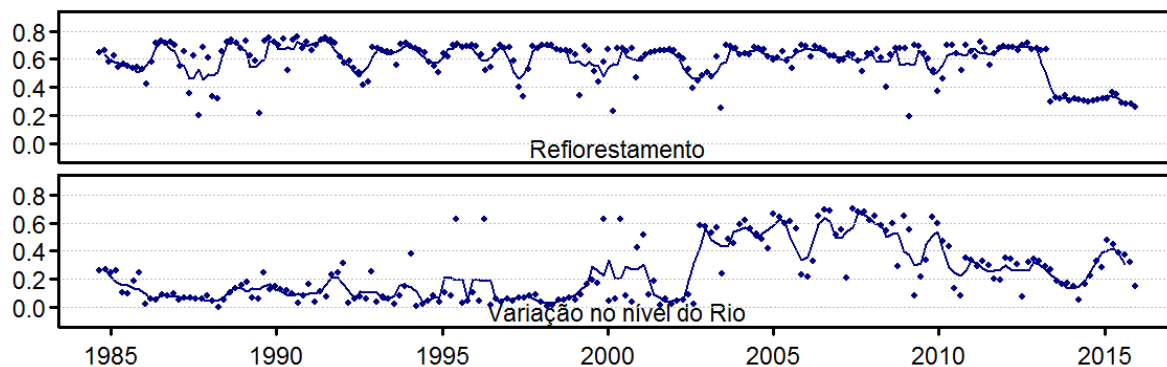
Figura 29 - Séries Temporais de NDVI indicando as transições entre tipos de cobertura da terra observadas na classe Mudanças Antrópicas



Na Figura 29 que inclui as transições nos tipos de cobertura da terra da classe Mudanças Antrópicas, também é possível observar a presença de oscilações decorrentes de variações nas condições atmosféricas e padrões sazonais, porém o que caracteriza esses perfis são as mudanças abruptas causadas pelas alterações no tipo de cobertura da terra. No perfil temporal do pixel referente a transição Floresta-Área Urbana, sugere que o desmatamento ocorreu a partir de 2006 e em 2012, o perfil de NDVI do pixel Floresta-Água aponta para o alagamento de parte da mata decorrente do enchimento da barragem da UHE de Santo Antônio.

A classe Mudanças Naturais, que inclui alterações nos tipos de cobertura da terra relacionadas com Reflorestamento e Inundação têm os perfis temporais do NDVI dessas transições, indicados na Figura 30.

Figura 30 - Séries Temporais de NDVI para alterações nos tipos de cobertura do solo inseridos na classe Mudanças Naturais



Na Figura 30, foram mostradas as séries temporais inseridas na classe Mudanças Naturais as quais, assim como para as séries temporais descritos anteriormente, apresentam variações decorrentes das condições atmosféricas e ambientais. Além disso, a série temporal de Reflorestamento é caracterizada por oscilações acentuadas no NDVI até o ano de 2004, a partir do qual esses valores se tornam mais estáveis, enquanto que as oscilações no perfil do pixel que se propõe a representar a Variação no Nível do Rio (localizado na margem do rio Madeira) indicam alternância inclusive nos tipos de cobertura da terra.

Em todos os gráficos das três figuras pode ser observada presença de valores espúrios. A presença desses ruídos no conjunto de dados pode estar associada a variações nas condições atmosféricas, existência de fumaça proveniente de queimadas, problemas técnicos no sensor no momento de tomada da imagem, ou mesmo pela não detecção de nuvens e sombra de nuvens pelo algoritmo *Fmask*.

A mudança no comportamento geral da série, já relatada anteriormente, atribuída a diferenças nas características do sensor OLI em relação aos seus antecessores, é perceptível em todas as séries temporais de tipos de cobertura temporalmente estáveis (não mudanças) ou características de alterações na cobertura da terra (mudanças antrópicas ou naturais).

### 4.3 Classificação das mudanças na área teste

As imagens resultantes da classificação das mudanças na área teste, a partir 154 séries temporais de referência, utilizando o algoritmo *FastDTW*, com raios de tamanhos 1 e 20, são apresentadas nas Figuras 30 e 31, respectivamente.

Figura 31 - Classificação FastDTW com raio igual a 1 e com 154 séries temporais de referência

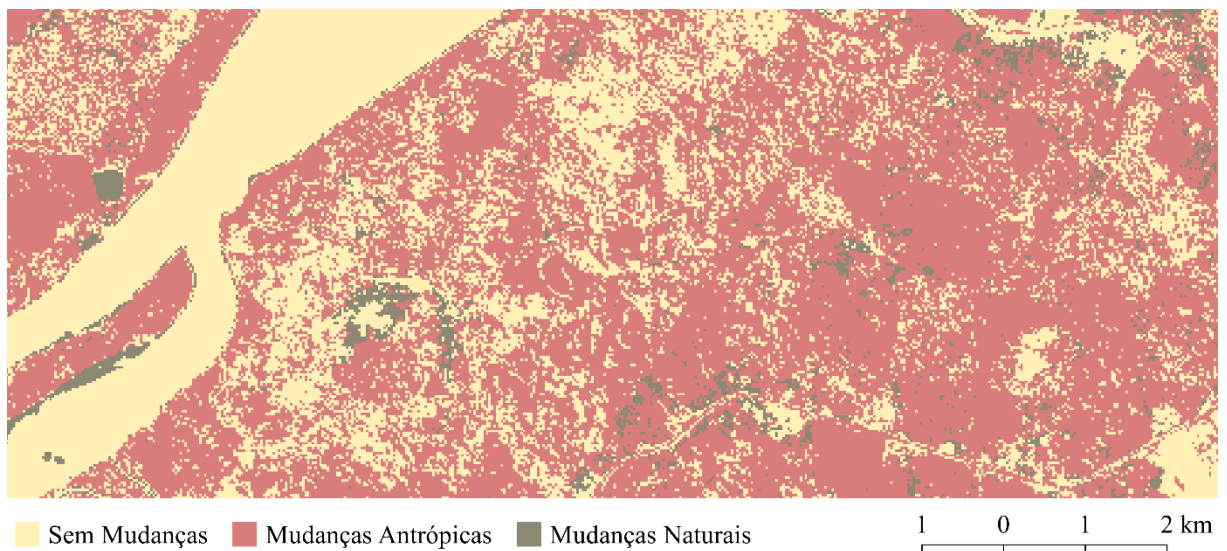
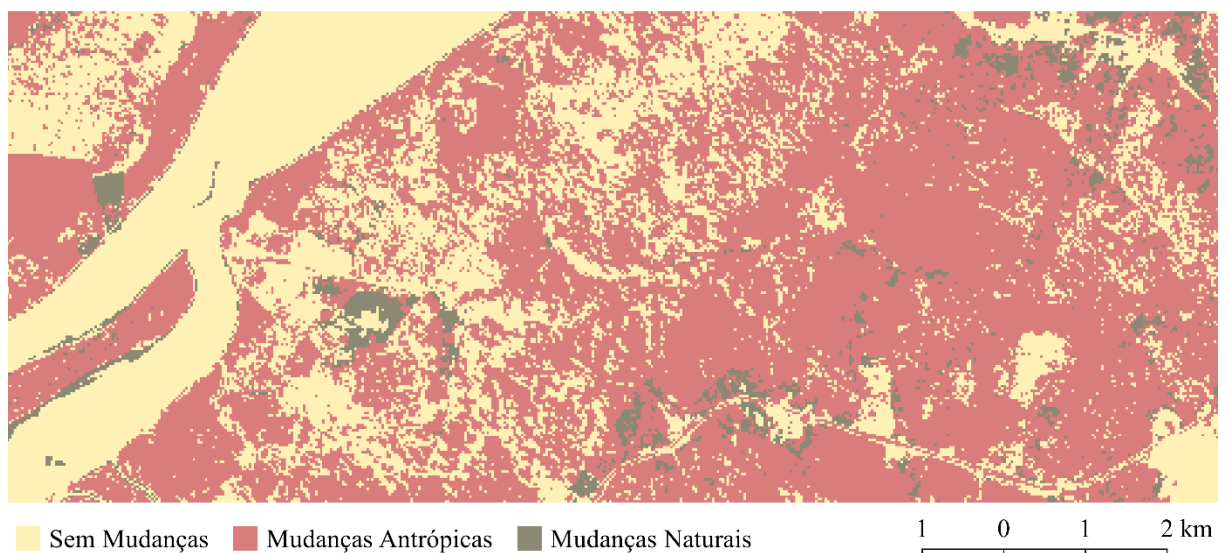


Figura 32 - Classificação FastDTW com raio igual a 20 e com 154 séries temporais de referência





Com a utilização do algoritmo *FastDTW* com tamanho de raio diferente para a classificação das séries temporais da área teste, verificou-se que mesmo utilizando a configuração a mais veloz (raio igual a 1), o tempo necessário para realizar a classificação dos 100 mil pixels foi de aproximadamente 11,5 horas. Na configuração com raio igual a 20 o tempo de classificação foi equivalente a 8,24 dias.

As matrizes de concordância, construídas a partir da comparação do resultado das classificações com as regiões de referência especificadas na área teste, são apresentadas nas Tabelas 2 e 3, respectivamente para o resultado das classificações mostrados nas Figuras 30 ( $r = 1$ ) e 31 ( $r = 20$ ).

Tabela 2 - Matriz de Concordância para a Classificação *FastDTW* com raio igual a 1 e com 154 séries temporais de referência

		REFERÊNCIA			
CLASSIFICAÇÃO		Sem Mudanças	Mud. Antrópicas	Mud. Naturais	Total
	Sem Mudanças	6642	5066	138	11846
	Mud. Antrópicas	8063	13105	79	21247
	Mud. naturais	233	722	58	1013
	Total	14938	18893	275	34106
<b>Concordância Global = 58,06%</b>					

Tabela 3 - Matriz de Concordância para a Classificação *FastDTW* com raio igual a 20 e com 154 séries temporais de referência

		REFERÊNCIA			
CLASSIFICAÇÃO		Sem Mudanças	Mud. Antrópicas	Mud. Naturais	Total
	Sem Mudanças	6246	5727	154	12127
	Mud. Antrópicas	8435	12453	53	20941
	Mud. naturais	257	713	68	1038
	Total	14938	18893	275	34106
<b>Concordância Global = 55,02%</b>					

As matrizes de concordância mostradas nas Tabela 2 e Tabela 3 indicam valores relativamente baixos de concordância global para as classificações de mudanças utilizando as configurações I e a II, (58,06% e 55,05% respectivamente), quando comparados com outros estudos de LUCC usando análises de séries temporais. Outro ponto que contrasta com o que é encontrado na literatura é a menor concordância global para a configuração II, visto que o maior valor de raio aumenta a chance da série temporal a ser classificada encontrar o melhor alinhamento com a série temporal de referência. De modo geral, as discrepâncias encontradas na discriminação de mudanças por meio da classificação de séries temporais podem ser explicadas pela presença de ruídos nos dados, visto que esses podem afetar de maneira significativa o comportamento da série temporal. Outra questão que pode ter afetado o desempenho do classificador foi a diferença nos valores de NDVI para as imagens adquiridas com o sensor OLI para as imagens adquiridas com os sensores TM e ETM+, o que provocou uma mudança abrupta no comportamento das séries temporais a partir de 2013, e provavelmente fez com que áreas de não mudança fossem classificadas como mudança antrópica.

## 5 CONSIDERAÇÕES FINAIS E RECOMENDAÇÕES

O expressivo crescimento de dados geoespaciais adquiridos por diferentes satélites de sensoriamento remoto tem desafiado a comunidade científica em questões relacionadas a organização, armazenamento, processamento e análise de grandes conjuntos de observações da Terra. Diante desses desafios, este projeto foi concebido com a proposta de construir uma estrutura de armazenamento em que fossem organizadas e armazenadas imagens multitemporais georreferenciadas e que, além disso, desse suporte ao processo de detecção de mudanças.

Nesse contexto, o presente projeto buscou apresentar brevemente conceitos relativos a relevância do monitoramento da cobertura da terra por sensoriamento remoto, e a possibilidade de extrair informações de LUCC a partir de séries temporais. Seguindo essa temática foi exposto o algoritmo DTW, em razão desse ser considerado o estado da arte na determinação de similaridade entre séries temporais. Também foi apresentada uma revisão de literatura relacionada a bancos de dados, em que foram tratados os tópicos modelos de dados relacional, modelo de dados não relacional, multimodelo de dados e modelos espaço-temporais, com o propósito de dar suporte à definição da estrutura de armazenamento.

O desenvolvimento da pesquisa se deu por meio da metodologia composta pelas etapas: projeto do banco de dados, pré-processamento, construção das séries temporais, identificação de regiões de referência e discriminação de mudanças. A primeira etapa, consistiu na definição e construção da estrutura de armazenamento. No pré-processamento, foram adquiridas as cenas, geradas máscaras de nuvens e realizadas filtragens nos dados. A construção das séries temporais consistiu na recuperação de bandas das imagens, sua transformação em reflectância TOA, geração de imagens NDVI e na construção da série temporal de cada pixel da imagem. A etapa de identificação das regiões de referência compreendeu a vetorização de regiões que incluíssem as classes de interesse, definidas a partir da interpretação visual de um conjunto multitemporal de sete imagens. A etapa final, de discriminação de mudanças, consistiu na classificação de séries temporais a partir da comparação com séries temporais de referência realizada por meio do algoritmo FastDTW.

A estrutura de armazenamento construída consistiu em um banco de dados espaço-temporal, concebido a partir de uma abordagem híbrida, em que foi empregado tanto o modelo de dados relacional como o não relacional. A abordagem escolhida coincide com a

persistência poliglota no formato *single-database* em que são adotados diferentes modelos de dados na construção de um único banco de dados. Essa abordagem foi utilizada com o propósito de aproveitar as características de cada modelo, a fim de melhorar a adequação geral do sistema de armazenamento, permitindo que o banco de dados construído armazene dados rigidamente estruturados, que se ajustam ao modelo relacional, e dados semi-estruturados que, preferivelmente devem ser armazenados em bancos de dados que adotam os modelos NoSQL.

Outro ponto importante relacionado ao banco de dados foi sua concepção baseada na necessidade de automatização da aplicação, visto que sem o banco de dados o armazenamento e gerenciamento das informações seriam realizados por meio de arquivos, dificultando a utilização de uma grande quantidade de imagens e a adoção do método de classificação adotado para este estudo. O emprego do PostgreSQL como SGBD possibilitou o desenvolvimento de um banco de dados multimodelo, que permitiu o armazenamento de dados heterogêneos. A extensão espacial PostGIS do PostgreSQL viabilizou o armazenamento dos dados geoespaciais, como as imagens, as máscaras de nuvens e as áreas de referência. O modelo de representação temporal adotado para as imagens foi o *snapshot*, materializado pela inserção da data de aquisição das cenas como atributo na tabela de armazenamento, a representação temporal das imagens permitiu a recuperação dessas e a construção das séries temporais. Sendo assim, a definição do modelo espaço-temporal para o banco de dados construído no projeto contribuiu para a aplicação, do ponto de vista que esse permitiu a construção das séries temporais e a classificação dessas a fim de construir mapa de mudanças na região selecionada.

Com relação às séries temporais é possível observar que de maneira geral o perfil temporal de NDVI pode descrever o tipo de cobertura do solo, contudo efeitos atmosféricos, nuvens, sombras, fumaça, problemas na tomada da imagem, entre outros, podem inserir dados espúrios na série temporal que podem alterar a forma da assinatura temporal do tipo de cobertura da terra. Outra questão que tem influência no perfil temporal dos tipos de cobertura da terra é a frequência temporal das observações, sendo que a não regularidade pode implicar na impossibilidade de identificar o padrão do tipo de cobertura da terra a que a série temporal descreve.

Considerando a capacidade de classificar mudanças por meio das séries temporais foram obtidos valores de concordância global de 58,06% e 55,02%, considerados baixos no que diz respeito a classificação de séries temporais por meio de reconhecimento de padrões com o algoritmo DTW. Todavia, esses resultados são justificados pelo caráter subjetivo embutido na caracterização de mudança temporal, pela não filtragem das séries temporais e

pela alta irregularidade que apresentam, visto que a região de estudo apresenta alta taxa de cobertura de nuvens.

Para projetos futuros relacionados a organização e armazenamento de dados geoespaciais recomenda-se armazenar juntamente com cada série temporal a coordenada central do pixel a que essa corresponde, com o propósito de facilitar o processo de reconstrução dos mapas de classificação e de possibilitar ampliar a área de estudo. Com relação a classificação das séries temporais, recomenda-se analisar a influência do tamanho do conjunto de séries temporais de referência, da filtragem e da irregularidade dessas na qualidade da classificação.

**REFERÊNCIAS**

AGOSTO, E. Vector–raster server-side analysis: a PostGIS benchmark. **Applied Geomatics**, v. 5, n. 2, p. 177-184, 2013. ISSN 1866-9298.

ALAM, F. DATA MIGRATION: RELATIONAL RDBMS TO NON-RELATIONAL NOSQL. 2015.

ALQURASHI, A. F.; KUMAR, L. Investigating the use of remote sensing and GIS techniques to detect land use and land cover change: A review. **Advances in Remote Sensing**, v. 2, n. 02, p. 193, 2013.

AMIRIAN, P.; BASIRI, A.; WINSTANLEY, A. Evaluation of data management systems for geospatial big data. International Conference on Computational Science and Its Applications, 2014, Springer. p.678-690.

ARMSTRONG, M. P. Temporality in spatial databases. 1988.

ARORA, I.; GUPTA, A. Cloud databases: a paradigm shift in databases. **International Journal of Computer Science Issues**, v. 9, n. 4, p. 77-83, 2012.

BERNDT, D. J.; CLIFFORD, J. Using dynamic time warping to find patterns in time series. KDD workshop, 1994, Seattle, WA. p.359-370.

BREWER, E. CAP twelve years later: How the "rules" have changed. **Computer**, v. 45, n. 2, p. 23-29, 2012. ISSN 0018-9162.

CAMARA, G. et al. Big earth observation data analytics: matching requirements to system architectures. Proceedings of the 5th ACM SIGSPATIAL International Workshop on Analytics for Big Geospatial Data, 2016, ACM. p.1-6.

CARLSON, T. N.; RIPLEY, D. A. On the relation between NDVI, fractional vegetation cover, and leaf area index. **Remote sensing of Environment**, v. 62, n. 3, p. 241-252, 1997. ISSN 0034-4257.

CASANOVA, M. A. et al. **Banco de dados geográficos**. MundoGEO Curitiba, 2005.

CHEN, M.; MAO, S.; LIU, Y. Big data: A survey. **Mobile Networks and Applications**, v. 19, n. 2, p. 171-209, 2014. ISSN 1383-469X.

CODD, E. F. A relational model of data for large shared data banks. **Communications of the ACM**, v. 13, n. 6, p. 377-387, 1970. ISSN 0001-0782.

- CONRAD, C. et al. Temporal segmentation of MODIS time series for improving crop classification in Central Asian irrigation systems. **International Journal of Remote Sensing**, v. 32, n. 23, p. 8763-8778, 2011. ISSN 0143-1161.
- DATASTAX. **The multi-model database: cloud applications in a complex world** 2016.
- ELMASRI, R.; NAVATHE, S. **Database systems**. Pearson Education Boston, MA, 2011.
- FORD, N. Polyglot programming. **Retrieved September**, v. 25, p. 2015, 2006.
- GANIEE, N.; BHARGAVA, R. NOSQL for Interactive Applications. **International Journal of Allied Practice, Research and Review**, Website: [www.ijaprr.com](http://www.ijaprr.com) (ISSN 2350-1294), retrieved on, 2015.
- GESSERT, F. et al. NoSQL database systems: a survey and decision guidance. **Computer Science-Research and Development**, p. 1-13, 2016. ISSN 1865-2034.
- GUIMBERTEAU, M. et al. Impacts of future deforestation and climate change on the hydrology of the Amazon basin: a multi-model analysis with a new set of land-cover change scenarios. **Hydrol. Earth Syst. Sci. Discuss**, 2016.
- KEOGH, E.; RATANAMAHATANA, C. A. Exact indexing of dynamic time warping. **Knowledge and information systems**, v. 7, n. 3, p. 358-386, 2005. ISSN 0219-1377.
- KRČÁL, L.; HO, S.-S. A SciDB-based Framework for Efficient Satellite Data Storage and Query based on Dynamic Atmospheric Event Trajectory. Proceedings of the 4th International ACM SIGSPATIAL Workshop on Analytics for Big Geospatial Data, 2015, ACM. p.7-14.
- LERNER, R. M. PostgreSQL, the NoSQL Database. LINUX JOURNAL, 2015. Disponível em: < <http://www.linuxjournal.com/content/postgresql-nosql-database?page=0.0> >. Acesso em: 24 de Março 2017.
- LI, S. et al. Geospatial big data handling theory and methods: A review and research challenges. **ISPRS Journal of Photogrammetry and Remote Sensing**, v. 115, p. 119-133, 2016. ISSN 0924-2716.
- LIU, Z. H.; GAWLICK, D. Management of Flexible Schema Data in RDBMSs-Opportunities and Limitations for NoSQL. CIDR, 2015.
- LOVELAND, T. R.; IRONS, J. R. Landsat 8: the plans, the reality, and the legacy. **Remote Sensing of Environment**, v. 185, p. 1-6, 2016. ISSN 0034-4257.

- LU, D. et al. Change detection techniques. **International journal of remote sensing**, v. 25, n. 12, p. 2365-2401, 2004. ISSN 0143-1161.
- LU, J.; HOLUBOVÁ, I. Multi-model Data Management: What's New and What's Next? , EDBT, 2017. p.602-605.
- MAUS, V. Satellite time series analysis for land use/cover change detection. 2014.
- MAUS, V. et al. A time-weighted dynamic time warping method for land-use and land-cover mapping. **IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing**, v. 9, n. 8, p. 3729-3739, 2016. ISSN 1939-1404.
- MONIRUZZAMAN, A.; HOSSAIN, S. A. Nosql database: New era of databases for big data analytics-classification, characteristics and comparison. **arXiv preprint arXiv:1307.0191**, 2013.
- NADI, S.; DELAVAR, M. R. Spatio-Temporal Modeling of Dynamic Phenomena in GIS. ScanGIS, 2003. p.215-225.
- PETITJEAN, F.; INGLADA, J.; GANÇARSKI, P. Satellite image time series analysis under time warping. **IEEE transactions on geoscience and remote sensing**, v. 50, n. 8, p. 3081-3095, 2012. ISSN 0196-2892.
- PRITCHETT, D. Base: An acid alternative. **Queue**, v. 6, n. 3, p. 48-55, 2008. ISSN 1542-7730.
- RAMAKRISHNAN, R. Database Management Systems . pdf. 2000.
- RAMAPRIYAN, H. Satellite imagery in earth science applications. **Image Databases**, p. 35-82, 2004.
- SADALAGE, P. J.; FOWLER, M. **NoSQL Essencial: Um guia conciso para o Mundo emergente da persistência poliglota**. Novatec Editora, 2013. ISBN 8575223380.
- SALVADOR, S.; CHAN, P. Toward accurate dynamic time warping in linear time and space. **Intelligent Data Analysis**, v. 11, n. 5, p. 561-580, 2007. ISSN 1088-467X.
- SCHMID, S.; GALICZ, E.; REINHARDT, W. Performance investigation of selected SQL and NoSQL databases. 2015, AGILE.
- SHUKLA, J.; NOBRE, C.; SELLERS, P. Amazon deforestation and climate change. **Science(Washington)**, v. 247, n. 4948, p. 1322-1325, 1990. ISSN 0036-8075.



SILBERSCHATZ, A.; KORTH, H. F.; SUDARSHAN, S. **Sistema de banco de dados**. Rio de Janeiro Elsevier, 2012. ISBN 978-85-352-4535-6.

SONG, M. et al. Spatiotemporal data representation and its effect on the performance of spatial analysis in a cyberinfrastructure environment—A case study with raster zonal analysis. **Computers & Geosciences**, v. 87, p. 11-21, 2016. ISSN 0098-3004.

SUN, L. et al. Impact of Land-Use and Land-Cover Change on urban air quality in representative cities of China. **Journal of Atmospheric and Solar-Terrestrial Physics**, v. 142, p. 43-54, 2016. ISSN 1364-6826.

USGS. Landsat Processing Details. Disponível em: < <https://landsat.usgs.gov/landsat-processing-details> >. Acesso em: 03 de Janeiro de 2018.

\_\_\_\_\_. What are the band designations for the Landsat satellites? , Disponível em: < <https://landsat.usgs.gov/what-are-band-designations-landsat-satellites> >. Acesso em: 22 de Março de 2017.

VITOLO, C. et al. Web technologies for environmental Big Data. **Environmental Modelling & Software**, v. 63, p. 185-198, 2015. ISSN 1364-8152.

WANG, W. et al. Detecting spatio-temporal and typological changes in land use from Landsat image time series. **Journal of Applied Remote Sensing**, v. 11, n. 3, p. 035006, 2017. ISSN 1931-3195.

WULDER, M. A. et al. The global Landsat archive: Status, consolidation, and direction. **Remote Sensing of Environment**, v. 185, p. 271-283, 2016. ISSN 0034-4257.

XIAO, X. et al. Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. **Remote sensing of environment**, v. 95, n. 4, p. 480-492, 2005. ISSN 0034-4257.

XIAO, Z.; LIU, Y. Remote sensing image database based on NOSQL database. Geoinformatics, 2011 19th International Conference on, 2011, IEEE. p.1-5.

XIUBIN, L. A review of the international researches on land use/land cover change [J]. **Acta Geographica Sinica**, v. 6, 1996.

YUAN, M. Temporal GIS and spatio-temporal modeling. Proceedings of Third International Conference Workshop on Integrating GIS and Environment Modeling, Santa Fe, NM, 1996.

ZHANG, L.; YI, J. Management methods of spatial data based on postgis. *Circuits, Communications and System (PACCS)*, 2010 Second Pacific-Asia Conference on, 2010, IEEE. p.410-413.

ZHU, Z.; WANG, S.; WOODCOCK, C. E. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. ***Remote Sensing of Environment***, v. 159, p. 269-277, 2015. ISSN 0034-4257.

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 0034-4257.

\_\_\_\_\_. Continuous change detection and classification of land cover using all available Landsat data. ***Remote sensing of Environment***, v. 144, p. 152-171, 2014. ISSN 0034-4257.