UNIVERSIDADE DE SÃO PAULO

ESCOLA POLITÉCNICA

THAÍS PASSOS CORREIA

**Aplicação de Geotecnologias na Elaboração de Índices
Socioeconômicos Associados às Ocorrências de Desastres Naturais no
Entorno da Ferrovia entre as Estações Caieiras e Francisco Morato,
SP**

São Paulo

2021

THAÍS PASSOS CORREIA

**Aplicação de Geotecnologias na Elaboração de Índices
Socioeconômicos Associados às Ocorrências de Desastres Naturais no
Entorno da Ferrovia entre as Estações Caieiras e Francisco Morato,
SP**

Versão Corrigida

Dissertação apresentada à Escola Politécnica da
Universidade de São Paulo para obtenção do título
de Mestre em Ciências.

Área de Concentração: Engenharia de Transportes

Subárea: Informações Espaciais –
Geoprocessamento

Orientador: Prof. Dr. José Alberto Quintanilha

São Paulo

2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Este exemplar foi revisado e corrigido em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.

São Paulo, _____ de _____ de _____

Assinatura do autor: _____

Assinatura do orientador: _____

Catálogo-na-publicação

Correia, Thaís Passos

Aplicação de Geotecnologias na Elaboração de Índices Socioeconômicos Associados às Ocorrências de Desastres Naturais no Entorno da Ferrovia entre as Estações Caieiras e Francisco Morato, SP / T. P. Correia -- versão corr. -- São Paulo, 2021.
177 p.

Dissertação (Mestrado) - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Transportes.

1.Inundação 2.Escorregamento 3.Ferrovia 4.SIG 5.Big Data I.Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Transportes II.t.

Nome: CORREIA, Thaís Passos

Título: Aplicação de Geotecnologias na Elaboração de Índices Socioeconômicos Associados às Ocorrências de Desastres Naturais no Entorno da Ferrovia entre as Estações Caieiras e Francisco Morato, SP.

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do título de Mestre em Ciências.

Aprovada em: 11/02/2021

Banca Examinadora

Prof. Dr. José Alberto Quintanilha (Presidente)

Instituição: Departamento de Engenharia de Transportes – Escola Politécnica – Universidade de São Paulo

Prof^a. Dr^a. Amarilis Lucia Casteli Figueiredo Gallardo

Instituição: Departamento de Engenharia Hidráulica e Ambiental - Escola Politécnica - Universidade de São Paulo

Dr. Laercio Massaru Namikawa

Instituição: Divisão de Processamento de Imagens - Instituto Nacional de Pesquisas Espaciais

DEDICATÓRIA

Aos moradores e trabalhadores dos municípios de Caieiras, Franco da Rocha e Francisco Morato (SP), por frequentemente sofrerem com ocorrências de desastres naturais na região.

Aos passageiros da Linha 7 – Rubi da Companhia Paulista de Trens Metropolitanos (CPTM), que eventualmente enfrentam transtornos no transporte público devido às inundações e escorregamentos que afetam a linha férrea.

E a todos os pesquisadores brasileiros, que continuam lutando por seus estudos e princípios, em um cenário cada vez mais desmotivador da ciência, com destaque à todas as pesquisadoras mães, que tiveram que aprender a conciliar maternidade, estudos, trabalho e tarefas domésticas em tempo integral, em uma rotina difícil e isolada durante a pandemia da Covid-19.

AGRADECIMENTOS

Meus sinceros agradecimentos:

À Deus por iluminar o meu caminho nos momentos difíceis e por me proporcionar a oportunidade de estudar em uma universidade conceituada.

Ao meu filho, Pedro Henrique Correia, que foi uma linda surpresa durante este Mestrado e por ser a minha maior motivação para seguir em frente.

Aos meus pais, Yara e Luís Correia, e à minha irmã, Amanda Correia, por me apoiarem em minhas decisões, por me compreenderem nas situações difíceis, por me ajudarem a cuidar do meu filho enquanto eu estudava e por sempre me motivarem a nunca desistir dos meus sonhos.

Ao meu orientador, Prof. Dr. José Alberto Quintanilha, por ter visto potencial em mim, pela paciência e por ter aceitado seguir essa jornada comigo.

A todos os docentes do Programa de Pós-graduação em Engenharia de Transportes do Departamento de Engenharia de Transportes, da Escola Politécnica da Universidade de São Paulo, pela paixão em lecionar, pela dedicação nas aulas e pela generosidade de conceder os seus conhecimentos.

À minha coorientadora da Seção de Investigações, Riscos e Gerenciamento Ambiental, da Área de Cidades, Infraestrutura e Meio Ambiente do Instituto de Pesquisas Tecnológicas do Estado de São Paulo, Dr^a. Alessandra Cristina Corsi, pela compreensão, por acreditar em mim e por sempre me motivar.

À toda equipe da Seção de Investigações, Riscos e Gerenciamento Ambiental, da Área de Cidades, Infraestrutura e Meio Ambiente do Instituto de Pesquisas Tecnológicas do Estado de São Paulo, pela receptividade e por se dedicarem a um trabalho tão bonito.

Ao Elcio Niwa e ao Conrado de Souza, da Companhia Paulista de Trens Metropolitanos, pela atenção e pelo fornecimento de dados utilizados nessa pesquisa.

Ao meu colega Lucas Henrique Sandre, pela amizade, pela ajuda nas análises em ambiente *Big Data* e pela parceria no relatório “Do SIG ao *Big Data*: Estruturação dos Dados Espaciais do Entorno das Estações Ferroviárias Caieiras e Francisco Morato, SP”, apresentado no 34º ANPET.

Ao meu colega Leonardo Godoy, do Laboratório de Geoprocessamento do Departamento de Engenharia de Transportes, da Escola Politécnica da Universidade de São Paulo, pela disponibilidade em ajudar com *Big Data* e por ter me auxiliado na geocodificação automática realizada neste estudo.

À minha amiga Amanda Kunath, pela amizade de longa data, por todo carinho, apoio e pelo auxílio com as técnicas do GeoMesa – *Bigtable*.

Ao Marcos Rosa e à equipe da *Solved Scholar*, por terem me auxiliado nas minhas dúvidas em relação ao *Google Earth Engine*.

À colega Michelle Trintinalia por ter elaborado a análise de componente principal e os *boxplots* desta pesquisa.

À Prof.^a Dr.^a Cláudia Machado, do Departamento de Engenharia de Transportes, da Escola Politécnica da Universidade de São Paulo, pelas conversas, apoio e ensinamentos.

À Prof.^a Dr.^a Rúbia Gomes Morato, da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo, pela atenção durante suas aulas e pelas sugestões relacionadas aos índices de qualidade de vida.

Ao colega Caio Azevedo e ao Prof. Dr. Mauro de Mesquita Spinola, do Departamento de Engenharia de Produção, da Escola Politécnica da Universidade de São Paulo, pela parceria no artigo “*A Systematic Review: Big Data Analytics and Landslides*”, submetido no XIII Simpósio Internacional de Deslizamentos.

À Luciane Watanabe, da Seção de Pós-graduação do Departamento de Engenharia de Transportes, da Escola Politécnica da Universidade de São Paulo, por torcer por mim e sempre estar disposta a me ajudar o mais rápido possível.

À Prof.^a Dr.^a Amarilis Lucia Gallardo, do Departamento de Engenharia Hidráulica e Ambiental, da Escola Politécnica da Universidade de São Paulo, e ao Dr. Laércio Namikawa, da Divisão de Processamento de Imagens do Instituto Nacional de Pesquisas Espaciais, pela disposição em relatar o meu projeto na qualificação e na defesa, contribuindo com suas avaliações críticas.

À Fundação de Amparo à Pesquisa do Estado de São Paulo, pela concessão da bolsa de Mestrado e pelo apoio financeiro para a realização desta pesquisa (Processos 2018/21279-7 e 2017/50343-2).

E ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pelo apoio à essa pesquisa.

RESUMO

A crise econômica brasileira, a apropriação das melhores áreas das cidades pelo mercado imobiliário mais a carência de áreas urbanizadas destinadas a moradia popular, favorece que os cidadãos de baixa renda ocupem as áreas desvalorizadas pelo mercado, como áreas ambientalmente frágeis, contribuindo para o aumento do perigo tanto desses locais como das pessoas que vivem neles. As inundações e os escorregamentos estão entre os desastres naturais mais comuns no Brasil, que são responsáveis por prejuízos de ordem social, ambiental e econômica, podendo resultar em perdas de vidas. As geotecnologias tornaram-se uma opção para auxiliar no mapeamento desses eventos e, conseqüentemente, ajudar na previsão de riscos, na avaliação de vulnerabilidades e na quantificação de perdas. As técnicas de *Big Data* também vêm se destacando por sua capacidade de integração de uma grande quantidade de bancos de dados, de diversas fontes e de formatos distintos; e colaborando para a melhoria da tomada de decisão, útil para a prevenção dos desastres ambientais. Este estudo teve como foco o entorno das estações Caieiras e Francisco Morato da Linha 7 - Rubi da CPTM, na qual esses eventos são frequentes e podem afetar a circulação de trens, além de acarretarem perdas econômicas e danos sociais aos moradores e trabalhadores da região. Dois índices foram criados para identificar os locais com uma maior interferência nas ocorrências de desastres naturais no entorno da linha férrea (IISEF) e os lugares que apresentam uma maior concentração de comunidades vulneráveis a esses tipos de eventos (IVSDN) através dos métodos da média aritmética simples, dos *outliers* e da análise de componente principal dos objetos espaciais. Esta pesquisa também identificou as necessidades e dificuldades de se realizar a transferência de dados (estruturados e não-estruturados) em SIG para um ambiente *Big Data*, utilizando as ferramentas de *Big Data* Geoespacial GeoMesa e *Google Earth Engine*. Esta Dissertação contribui para uma melhor compreensão dos desastres naturais, principalmente escorregamentos e inundações, e como eles afetam o sistema ferroviário, seus passageiros e moradores da região.

Palavras-chave: Inundação. Escorregamento. Ferrovia. Sistemas de Informações Geográficas. *Big Data*.

ABSTRACT

The Brazilian economic crisis, the appropriation of the best areas of the cities by the real estate market plus the lack of urbanized areas destined for popular housing, favors that low-income citizens occupy areas devalued by the market, as environmentally fragile areas, contributing to the increase of the danger both from these places and from the people who live in them. Floods and landslides are among the most common natural disasters in Brazil, which are responsible for social, environmental and economic damage, which can result in loss of life. Geotechnologies have become an option to assist the mapping of these events and, consequently, to assist in risk prediction, vulnerability assessment and loss quantification. Big Data techniques have also stood out for their ability to integrate a large number of databases from different sources and in different formats; and it collaborates to improve decision making, useful for the prevention of environmental disasters. This study focused on the surroundings of the Caieiras and Francisco Morato stations on Line 7 - Rubi of CPTM, in which these events are frequent and can affect the circulation of trains, in addition to causing economic losses and social damage to residents and workers in the region. Two indexes were created to identify the places with the greatest interference in the occurrence of natural disasters around the railway line (IISEF) and the places that have the highest concentration of communities vulnerable to these types of events (IVSDN) using the simple arithmetic mean, outliers and main component analysis of space objects methods. This research also identified the needs and difficulties of transferring data (structured and unstructured) in GIS to a Big Data environment, using the tools of Big Data Geospatial: GeoMesa and Google Earth Engine. This master thesis contributes to a better understanding of natural disasters, especially landslides and floods, and how they affect the rail system, its passengers and residents of the region.

Keywords: Flood. Landslide. Railway. Geographic Information System. Big Data.

LISTA DE ABREVIATURAS E SIGLAS

ACP	Análise de Componentes Principais
ANA	Agência Nacional de Águas
ANEEL	Agência Nacional de Energia Elétrica
ANTT	Agência Nacional de Transportes Terrestres
API	<i>Application Programming Interface</i>
CBTU	Companhia Brasileira de Trens Urbanos
CEM	Centro de Estudos da Metrópole
CEMADEN	Centro Nacional de Monitoramento e Alertas de Desastres Naturais
CENAD	Centro Nacional de Gerenciamento de Riscos e Desastres
CETESB	Companhia Ambiental do Estado de São Paulo
CP	Componente Principal
CPRM	Companhia de Pesquisa de Recursos Minerais
CPTM	Companhia Paulista de Trens Metropolitanos
DAEE	Departamento de Águas e Energia Elétrica
DER	Departamento de Estradas e Rodagem do Estado de São Paulo
DNIT	Departamento Nacional de Infraestrutura de Transportes
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
EMPLASA	Empresa Paulista de Planejamento Metropolitano S.A
EPUSP	Escola Politécnica da Universidade de São Paulo
ESA	Agência Espacial Europeia
ESA	<i>European Space Agency</i>
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FEPASA	Ferrovias Paulista S/A
FF	Fundação para a Conservação e a Produção Florestal do Estado de São Paulo
GEE	<i>Google Earth Engine</i>
GIS	<i>Geographic Information System</i>
HDFS	<i>Hadoop Distributed File System</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
ICMBio	Instituto Chico Mendes de Conservação da Biodiversidade
IDE	Infraestrutura de Dados Espaciais
IF	Instituto Florestal

IGC-SP	Instituto Geográfico e Cartográfico de São Paulo
IG-SP	Instituto Geológico do Estado de São Paulo
IISEF	Índice de Interferência Social Externa à Ferrovia
INPE	Instituto Nacional de Pesquisas Espaciais
IoT	<i>Internet of Things</i>
IPT	Instituto de Pesquisas Tecnológicas do Estado de São Paulo
IPVS	Índice Paulista de Vulnerabilidade Social
IQVU	Índice de Qualidade de Vida Urbana
IVSDN	Índice de Vulnerabilidade Social à Ocorrências de Desastres Naturais
JTS	<i>Java Topology Suit</i>
LabGeo	Laboratório de Geoprocessamento
MMA	Ministério do Meio Ambiente
NDVI	<i>Normalized Difference Vegetation Index</i>
ODS	Objetivos de Desenvolvimento Sustentável
PRISMA	<i>Preferred Reporting Items for Systematic Reviews and Meta-Analyzes</i>
PTR	Departamento de Engenharia de Transportes
RDDs	<i>Resilient Distributed Datasets</i>
RMSP	Região Metropolitana de São Paulo
SAEs	Sistemas Analíticos Espaciais
SEADE	Fundação Sistema Estadual de Análise de Dados
SIG	Sistema de Informação Geográfica
SQL	<i>Structured Query Language</i>
SR	Sensoriamento Remoto
SVI	<i>Social Vulnerability Index</i>
TWI	<i>Topographic Wetness Index</i>
UNDRR	<i>United Nations Office for Disaster Risk Reduction</i>
USGS	<i>United States Geological Survey</i>
USP	Universidade de São Paulo

LISTA DE ILUSTRAÇÕES

Figura 1 - Desastres naturais no mundo em 2017.	9
Figura 2 - Diferença entre inundação e enchente.	12
Figura 3 - Fluxograma do método da pesquisa.	40
Figura 4 - Fluxograma PRISMA.	42
Figura 5 - Mapa da Área de Estudo.	46
Figura 6 - Exemplo de ferrovia afetada por uma inundação: Linha 7 - Rubi da CPTM inundada entre as estações Caieiras e Franco da Rocha.	50
Figura 7 - Caminhões presos em inundação em Caieiras, próximo a Linha Rubi da CPTM.	51
Figura 8 - Deslizamento de terra interrompe circulação de trens na Linha 7 da CPTM.	51
Figura 9 - Sabesp abriu comportas da represa em 11 de março de 2016.	53
Figura 10 - Mapa de Cobertura e Uso do Solo do Entorno da Linha 7 – Rubi.	55
Figura 11 - Método da passagem de dados de um ambiente SIG para o GeoMesa e GEE.	78
Figura 12 - Mapa de suscetibilidade e de ocorrências de escorregamentos na área de estudo.	83
Figura 13 - Mapa de suscetibilidade e de ocorrências de inundações na área de estudo.	85
Figura 14 - Relação de cobertura e uso do solo com ocorrências de escorregamentos e inundações na área de estudo.	89
Figura 15 - Mapa do IPVS da área de estudo.	92
Figura 16 - Indicadores de Lixo Acumulado nos Logradouros e de Esgoto à Céu Aberto na área de estudo.	95
Figura 17 - Indicadores de Não-Abastecimento por Água Encanada e de Baixa Renda na área de estudo.	98
Figura 18 - Indicadores de Idosos e de Crianças na área de estudo.	101
Figura 19 - Indicadores de Analfabetismo e de Residentes Sem Moradia Própria na área de estudo.	104
Figura 20 - Boxplots dos indicadores que compõem o ISEF.	106
Figura 21 - Exemplo do ISEF, método dos outliers, no GEE.	109
Figura 22 - Comparação do ISEF para cada método.	110
Figura 23 - Boxplots dos indicadores que compõem o IVSDN.	112

Figura 24 - Exemplo do IVSDN, método da média aritmética, no GEE.	115
Figura 25 - Comparação do IVSDN para cada método.....	116
Figura 26 - Visualização de parte dos dados em SIG no GeoMesa/GeoServer.	117
Figura 27 - Visualização de alguns dados em SIG no GEE.	118

LISTA DE GRÁFICOS

Gráfico 1 - Proporção de Suscetibilidade de Escorregamentos e Inundações entorno da Linha 7 - Rubi.....	45
Gráfico 2 - Ocorrências associadas a inundações na Linha 7 – Rubi.....	48
Gráfico 3 - Número de publicações encontradas entre 2008 e 2019.	80
Gráfico 4 - Formatos primários dos dados levantados.	81
Gráfico 5 - Tipos de formatos dos dados selecionados.	82

LISTA DE QUADROS

Quadro 1 - Variáveis da qualidade de vida urbana.....	32
Quadro 2 - Proporção de uso do solo e de ocorrências na área de estudo (buffer de 500 m).....	87
Quadro 3 - Relação IPVS com escorregamentos e inundações na área de estudo.	90
Quadro 4 - Matriz das componentes principais das variáveis do ISEF.....	107
Quadro 5 - Matriz das componentes principais das variáveis do IVSDN.....	113

LISTA DE TABELAS

Tabela 1 - Levantamento de ocorrências de desastres naturais que afetaram a Linha 7 – Rubi da CPTM.....	49
Tabela 2 - Número de registros sobre o tema de estudo por país.....	80

SUMÁRIO

1	INTRODUÇÃO	1
1.1	Problemática	4
1.2	Hipóteses de pesquisa	4
1.3	Objetivos	5
1.3.1	Objetivo geral	5
1.3.2	Objetivos específicos	5
1.4	Contribuições deste estudo	5
2	REVISÃO BIBLIOGRÁFICA	7
2.1	Desastres: eventos naturais ou instigados?	7
2.1.1	Desastres naturais ao redor do mundo	8
2.1.2	Desastres naturais no Brasil	10
2.2	Inundações: características, causas e consequências	11
2.2.1	Inundações e análise espacial	12
2.2.2	Inundações e transporte	14
2.3	Escorregamentos: conceitos, tipos e causas	15
2.3.1	Escorregamentos e geoprocessamento	16
2.3.2	Escorregamentos e transporte	17
2.4	<i>Big Data</i> : definição e características	18
2.4.1	<i>Big Data</i> e transporte	20
2.4.2	<i>Big Data</i> e análise espacial	21
2.4.2.1	<i>Big Data Geoespacial e suas ferramentas</i>	24
2.4.2.2	<i>GIS Tools for Hadoop (ArcGIS)</i>	25
2.4.2.3	<i>Hadoop-GIS</i>	26
2.4.2.4	<i>SpatialHadoop</i>	26
2.4.2.5	<i>SpatialSpark</i>	27
2.4.2.6	<i>GeoSpark</i>	27
2.4.2.7	<i>GeoMesa</i>	28
2.4.2.8	<i>LocationSpark</i>	29
2.4.2.9	<i>SparkGIS</i>	29
2.4.2.10	<i>Elcano</i>	29
2.4.2.11	<i>Manifold</i>	30
2.4.2.12	<i>Google Earth Engine (GEE)</i>	30

2.5	Indicadores sociais	31
2.5.1	Índices de qualidade de vida	31
2.5.2	Índices de vulnerabilidade social	33
2.5.3	Indicadores utilizados nesta pesquisa	37
2.5.3.1	Indicador de Lixo	37
2.5.3.2	Indicador de Esgoto	37
2.5.3.3	Indicador de Abastecimento de Água	38
2.5.3.4	Indicador de Renda	38
2.5.3.5	Indicador de Idosos	38
2.5.3.6	Indicador de Crianças	39
2.5.3.7	Indicador de Alfabetização	39
2.5.3.8	Indicador de Residentes Sem Moradia Própria	40
3	MATERIAIS E MÉTODOS	40
3.1	Análise bibliométrica	41
3.2	Área de estudo	44
3.2.1	Delimitação da área de estudo	44
3.2.2	Ocorrências de desastres naturais na área de estudo	47
3.2.3	Descrição da área de estudo	52
3.2.3.1	Recursos hídricos	52
3.2.3.2	Cobertura e uso do solo	53
3.2.3.3	Relevo	56
3.2.3.4	Transporte	56
3.3	Levantamento de fontes/dados oficiais na área de estudo	57
3.4	Ocorrência de eventos/Geocodificação automática	59
3.5	Relação de ocorrências de desastres naturais com cobertura e uso do solo	60
3.6	Relação das ocorrências de inundações e escorregamentos com o Índice Paulista de Vulnerabilidade Social (IPVS – SEADE)	61
3.7	Indicador de Lixo Acumulado	61
3.8	Indicador de Esgoto à Céu Aberto	63
3.9	Indicador de Abastecimento por Água Não-Encanada	64
3.10	Indicador de Baixa Renda	65
3.11	Indicador de Idosos	66
3.12	Indicador de Crianças	67

3.13	Indicador de Analfabetismo	68
3.14	Indicador dos Residentes Sem Moradia Própria	69
3.15	Índice de Interferência Socioeconômica à Ferrovia	70
3.15.1	Média Aritmética Simples	71
3.15.2	<i>Outliers</i>	72
3.15.3	Análise de Componente Principal de Objetos Espaciais	72
3.16	Índice de Vulnerabilidade Social à Ocorrências de Desastres Naturais ...	73
3.16.1	Média Aritmética Simples	74
3.16.2	<i>Outliers</i>	74
3.16.3	Análise de Componente Principal de Objetos Espaciais	74
3.17	Inserção e visualização dos dados de um ambiente SIG para um ambiente <i>Big Data</i>	75
3.17.1	GeoMesa: implementação e passagem dos dados	75
3.17.2	<i>Google Earth Engine</i> : inserção dos dados	76
4	RESULTADOS E DISCUSSÕES	79
4.1	Análise bibliométrica	79
4.2	Levantamento de fontes/dados oficiais na área de estudo	81
4.3	Ocorrência de eventos/Geocodificação automática	82
4.3.1	Escorregamentos.....	82
4.3.2	Inundações	84
4.4	Relação de ocorrências de desastres naturais com cobertura e uso do solo 86	
4.4.1	Escorregamentos.....	86
4.4.2	Inundações	87
4.5	Relação das ocorrências de inundações e escorregamentos com o Índice Paulista de Vulnerabilidade Social (IPVS do SEADE)	90
4.6	Indicador de Lixo Acumulado	93
4.7	Indicador de Esgoto à Céu Aberto	94
4.8	Indicador de Abastecimento por Água Não-Encanada	96
4.9	Indicador de Baixa Renda	96
4.10	Indicador de Idosos	99
4.11	Indicador de Crianças	99
4.12	Indicador de Analfabetismo	102
4.13	Indicador de Residentes Sem Moradia Própria	102

4.14	Índice de Interferência Socioeconômica Externa à Ferrovia (IISEF)	105
4.14.1	Média Aritmética Simples	105
4.14.2	<i>Outliers</i>	105
4.14.3	Análise de Componente Principal de Objetos Espaciais	107
4.14.4	Comparação entre os métodos	108
4.15	Índice de Vulnerabilidade Social à Ocorrência de Desastres Naturais (IVSDN)	111
4.15.1	Média Aritmética Simples	111
4.15.2	<i>Outliers</i>	111
4.15.3	Análise de Componente Principal de Objetos Espaciais	112
4.15.4	Comparação entre os métodos	114
4.16	Inserção e visualização dos dados de um ambiente SIG para um ambiente <i>Big Data</i>	117
4.16.1	GeoMesa: visualização dos dados e dificuldades encontradas.....	117
4.16.2	GEE: visualização dos dados e dificuldades encontradas	118
5	CONCLUSÕES E RECOMENDAÇÕES	119
6	REFERÊNCIAS	122
	ANEXOS	143

1 INTRODUÇÃO

No Brasil, a crise econômica, a apropriação das melhores áreas das cidades pelo mercado imobiliário, mais a carência de áreas destinadas a moradia popular, favorecem que os cidadãos de baixa renda ocupem as áreas desvalorizadas pelo mercado, como áreas suscetíveis e ambientalmente frágeis (por exemplo, rios, mangues, encostas íngremes etc.), contribuindo para o aumento da vulnerabilidade (das pessoas que vivem neles) e do perigo devido à precariedade da ocupação. Essa ocupação precária em áreas ambientalmente frágeis pode contribuir para a ocorrência de alguns tipos de desastres naturais, como escorregamentos e inundações (MINISTÉRIO DAS CIDADES/IPT, 2007).

Para construir o seu espaço, o homem não consegue enxergar que deveria haver uma harmonização entre o desenvolvimento econômico e social e, portanto, atende apenas as suas próprias exigências (ROBAINA, 2008). A rápida urbanização, alterações no uso do solo e o acelerado processo de desenvolvimento socioeconômico são fatores potenciais para provocar inundação, principalmente em grandes cidades (CHAN *et al.*, 2018). Projetos de construção civil e outras obras de engenharia envolvem movimentação de terra, capazes de gerar taludes íngremes e sem vegetação, que podem acarretar movimentos de massa (FERNANDES *et al.*, 2009).

O aumento de eventos climáticos extremos (como chuvas intensas, temporais, tornados, estiagens, entre outros) em maior intensidade favorecem os desastres naturais (TOMINAGA, 2009). Segundo Maffra e Mazzola (2007), os fatores que potencializam a ocorrência de desastres ambientais no Brasil são o desmatamento, a ocupação de áreas com maior suscetibilidade natural a esses fenômenos e a utilização imprópria dos recursos naturais.

De acordo com Tobin e Montz (1997, p. 5), há diferença entre os termos “perigo natural” e “desastre”: “Um perigo natural representa a interação potencial entre humanos e eventos naturais extremos. Representa o potencial ou a probabilidade de um evento (não é o evento em si)” enquanto que “um desastre é geralmente definido como um evento que tem um grande impacto na sociedade [...] pode ser definido qualitativamente como um evento perigoso que perturba significativamente o funcionamento da sociedade”. Os desastres naturais mais comuns no país são: inundações, seca, erosão e escorregamentos

(movimentos de massa), que são responsáveis por muitas vítimas e por perdas materiais anualmente (MAFFRA; MAZZOLA, 2007).

As inundações e os escorregamentos estão entre os desastres naturais que mais afetaram a população, incluindo óbitos, no Brasil e ao redor do mundo. Além das perdas humanas, os desastres são responsáveis por prejuízos econômicos bilionários anuais em todo o mundo. Esses fenômenos também podem estar relacionados com o aumento de problemas de saúde, como doenças potencialmente epidêmicas que surgem no período de enchentes e são transmitidas através de água e alimentos contaminados (DE FREITAS; XIMENES, 2012).

A preparação para um desastre se refere a medidas que devem ser tomadas para preparar e reduzir seus efeitos; ou seja, prever e, quando possível, prevenir desastres, diminuir seus impactos sobre populações vulneráveis, responder e lidar com suas consequências. Assim, é muito importante estudar os eventos de desastres naturais no sentido de melhorar sua previsão, prevenção e para amenizar suas consequências (LUKIĆ, 2013).

Wisner *et al.* (2004) busca as conexões entre os riscos que as pessoas enfrentam e as razões de sua vulnerabilidade aos perigos. O estudo mostra como os desastres podem ser percebidos dentro dos padrões mais amplos da sociedade e, de fato, como analisá-los para poder fornecer uma forma de construir políticas que podem ajudar a reduzir desastres e mitigar perigos, enquanto ao mesmo tempo, está melhorando os padrões de vida e as oportunidades em geral.

As alternativas de prevenção dos desastres naturais ou minimização dos seus efeitos podem ser realizadas através da criação de um sistema de informações de mapeamento de emergência de desastres naturais (LI; ZHAO, 2017), como a identificação do risco, que é essencial para o gerenciamento desses eventos. A gestão dos desastres pode ser melhorada via integração de dados e a utilização de técnicas de *Big Data*. *Big Data* é um termo recente que se refere à conjuntos de dados muito grandes e complexos (VENKATRAMAN; VENKATRAMAN, 2019). Segundo Horita (2017), “os tomadores de decisão possuem acesso à diversas fontes de dados geoespaciais, de diferentes formatos e em grande quantidade, o que dificulta a visualização e a integração dessas informações” e a geração de meios para a integração desses dados, seja via bancos

de dados ou via infraestrutura de dados espaciais, facilitará aos tomadores de decisão a escolha das melhores alternativas à luz dos dados disponíveis.

Portanto, são atividades fundamentais: a identificação de áreas suscetíveis aos desastres naturais para prevenir ou mitigar os prejuízos provocados por esses eventos; o uso de dados georreferenciados e suas técnicas de análise em ambiente de *Big Data*, em função do grande número de informações e diversidade de formatos. A mitigação dos prejuízos provocados por eventos de desastres naturais é uma das finalidades dos Objetivos de Desenvolvimento Sustentável - ODS, que são uma agenda mundial adotada durante a Cúpula das Nações Unidas sobre o Desenvolvimento Sustentável em setembro de 2015, constituída por 17 objetivos e 169 metas a serem atingidos até 2030 (ESTRATÉGIA ODS, 2020).

Nos ODS estão previstas ações nas áreas de erradicação da pobreza, segurança alimentar, agricultura, saúde, educação, igualdade de gênero, redução das desigualdades, energia, água e saneamento, padrões sustentáveis de produção e de consumo, mudança do clima, cidades sustentáveis, proteção e uso sustentável dos oceanos e dos ecossistemas terrestres, crescimento econômico inclusivo, infraestrutura, industrialização, entre outros objetivos (ESTRATÉGIA ODS, 2020). Dentre os ODS, cabe destacar que o ODS1 e o ODS11 estão relacionadas com a presente pesquisa.

O ODS1 refere-se à Erradicação da Pobreza, busca erradicar a pobreza até 2030. A pobreza é a maior ameaça global para alcançar o desenvolvimento sustentável. Erradicar a pobreza é, portanto, fundamental para o sucesso dos ODS (ROY *et al.*, 2018). Segundo Echendu (2020) existe uma relação de duas vias entre inundações e pobreza: inundações agravam a pobreza existente e a pobreza torna o impacto de inundações muito mais graves, aumentando a vulnerabilidade. A relação entre inundações e pobreza, portanto, merece mais atenção. Nesta Dissertação, ocorrências de inundações e escorregamentos no entorno da via férrea, serão relacionados às condições de pobreza e vulnerabilidade dos moradores da região.

Quanto ao ODS 11 - Cidades e Comunidades Sustentáveis, visa tornar cidades e assentamentos humanos inclusivos, seguros, resilientes e sustentáveis (IPEA, 2019). A meta até 2030 é reduzir significativamente o número de mortos e o número de pessoas impactadas por desastres, incluindo desastres relacionados à água e escorregamentos (ECHENDU, 2020). Ainda segundo o IPEA *op. cit.*, o Brasil, até 2030, “deve garantir o acesso de todos a moradia digna, adequada e a preço acessível; aos serviços básicos e

urbanizar assentos precários de acordo com as metas assumidas no Plano Nacional de Habitação, com especial atenção para os grupos em situação de vulnerabilidade.”

1.1 Problemática

A problemática desta Dissertação é o uso de dados espaciais públicos e disponíveis para executar a passagem de dados de um ambiente de Sistema de Informação Geográfica - SIG a um ambiente *Big Data*, e utilizar dados espaciais e não-espaciais para apresentar exemplos de aplicações utilizando índices socioeconômicos, educacionais e habitacionais para elaboração de um índice de interferência socioeconômica externa à ferrovia e de um índice de vulnerabilidade social dos moradores/usuários da Companhia Paulista de Trens Metropolitanos - CPTM aos eventos de desastres naturais que afetam a Linha 7 – Rubi.

1.2 Hipóteses de pesquisa

- I. Dados diversos, gratuitos, públicos e acessíveis, em grande volume, gerados e mantidos por diferentes fontes, são passíveis de serem tratados por SIG e transferidos para ambiente *Big Data*, de forma a serem processados de maneira mais rápida, eficiente e fidedigna;
- II. Os índices socioeconômicos e as técnicas utilizadas permitem a geração do índice de interferência socioeconômica externa à ferrovia e do índice de vulnerabilidade social dos moradores/usuários da CPTM às ocorrências de desastres naturais que afetam a linha férrea, que mostram os locais mais afetados pelos eventos de desastres.
- III. As ocorrências de escorregamentos e inundações coincidem com áreas de classe alta de suscetibilidade do relevo a esses tipos de desastres naturais e com locais de alta vulnerabilidade social e também correspondem com as classes de altos índices de interferência socioeconômica externa à ferrovia
- IV. As classes com valores altos do índice de vulnerabilidade aos desastres naturais dos moradores do entorno da ferrovia estão relacionadas aos locais com maior vulnerabilidade social e também com as ocorrências desses eventos na região.

1.3 Objetivos

1.3.1 Objetivo geral

O objetivo deste estudo é identificar as necessidades e as dificuldades de realizar a passagem de um dado espacial em ambiente SIG para um ambiente *Big Data*, analisando dados estruturados e não-estruturados de diferentes bancos de dados espaciais, além de elaborar índices socioeconômicos e de vulnerabilidade, para uma melhor compreensão das ocorrências de escorregamentos e inundações e como elas afetam o sistema ferroviário e os moradores da região/passageiros da CPTM, no trecho considerado.

1.3.2 Objetivos específicos

- Obter um diagnóstico, em termos de disponibilidade e viabilidade de uso dos dados existentes da região de estudo, sobre variáveis de interesse para a tomada de decisão com foco nos escorregamentos e nas inundações;
- Realizar a transferência de dados espaciais tradicionalmente utilizados em ambiente SIG para dentro de um ambiente de análise de *Big Data*, comparando vantagens e desvantagens de ambos os ambientes;
- Gerar um índice de interferência socioeconômica externa à ferrovia para entender como a situação socioeconômica dos moradores da área de estudo pode contribuir para ocorrência de escorregamentos e inundações na ferrovia;
- Elaborar um índice de vulnerabilidade social às ocorrências de desastres naturais dos moradores do entorno da via férrea/usuários da CPTM para verificar quem é a população mais afetada por esses eventos;

1.4 Contribuições deste estudo

As contribuições deste estudo seriam a identificação dos problemas e limitações atuais, neste momento, da conversão do ambiente SIG para *Big Data*; a apresentação do contexto atual das pesquisas sobre uso de *Big Data*, sobre dados necessários e técnicas de migração de SIG para *Big Data*; uma atualização do inventário relacionado a escorregamentos e inundações na área estudada; a proposição de um índice de interferência socioeconômica externa à ferrovia e um índice de vulnerabilidade social à eventos de desastres naturais dos usuários da Linha 7 – Rubi, que propiciará à CPTM, neste caso particular, e às empresas do setor de transporte ferroviário, priorizarem

possíveis intervenções de engenharia no sentido de evitar problemas que causem o interrompimento dos transportes de passageiros e de carga, minimizando paralisação dos trens, atrasos de viagens, pagamento de multas devido aos minutos atrasados, acidentes ferroviários, prejuízos financeiros e até a perda de vidas humanas.

A presente Dissertação foi desenvolvida no âmbito do “Plano de Desenvolvimento Institucional de Pesquisa na Área de Transformação Digital: Manufatura Avançada e Cidades Inteligentes e Sustentáveis (PDIP)” do Instituto de Pesquisas Tecnológicas do Estado de São Paulo – IPT, financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP (Processo 2017/50343-2), na linha de pesquisa “Sistema de Gestão Ambiental” e sublinha “Sistema de predição e gestão de desastres naturais”. Sua realização estava prevista na proposta PDIP-IPT submetida à FAPESP (Processo 2018/21279-7), conforme plano de bolsa intitulado “Estruturação de informações geoespaciais em ambientes *Big Data*”. Parte das atividades previstas no projeto de pesquisa foram conduzidas nas instalações do IPT, mais precisamente na Seção de Investigações, Riscos e Gerenciamento Ambiental (SIRGA), sob a supervisão da pesquisadora-líder Zehbour Panossian com o apoio da equipe técnica formada pelos pesquisadores Alessandra C. Corsi, Eduardo S. de Macedo e Marcela P. P. Guimarães. Houve também o apoio do Laboratório de Geoprocessamento (LabGeo) do Departamento de Engenharia de Transportes da Escola Politécnica da Universidade de São Paulo (PTR – EPUSP).

O presente estudo está estruturado da seguinte maneira: além desta introdução, que mostra também a justificativa, o problema, os objetivos e contribuições desta pesquisa; a Dissertação apresenta no capítulo 2 uma revisão bibliográfica sobre desastres naturais, principalmente escorregamentos e inundações, geoprocessamento, *Big Data* e tipos de índices de qualidade de vida e de vulnerabilidade social; a seção 3 dos materiais e métodos caracteriza a área de estudo e aponta os métodos aplicados nesta pesquisa; o capítulo 4 mostra os resultados e as discussões, como análises em ambiente SIG da área de estudo, levantamento de dados e fontes importantes para esta pesquisa, análise bibliométrica sobre o tema, a implementação de dados em duas ferramentas de *Big Data* Geoespacial, a criação dos índices de interferência socioeconômica externa da ferrovia e de vulnerabilidade social dos moradores da região; e, por fim a seção 5 explica as conclusões e considerações deste estudo.

2 REVISÃO BIBLIOGRÁFICA

Nessa revisão bibliográfica serão mostrados os conceitos de desastres, em seguida os desastres naturais no Brasil e no mundo; após será explicado sobre as inundações e os escorregamentos (eventos tratados nesta Dissertação), seus conceitos, sua relação com análise espacial e com transporte ferroviário; depois será demonstrado os conceitos de *Big Data*, sua relação com transporte e com análise espacial (destacando ferramentas de *Big Data* Geoespacial); e, por fim, serão apontados índices sociais que auxiliam na compreensão da qualidade de vida e de vulnerabilidade social de uma região.

2.1 Desastres: eventos naturais ou instigados?

De acordo com o Centro Nacional de Gerenciamento de Riscos e Desastres – CENAD (BRASIL, 2014, p. 15):

Desastre é o resultado de eventos adversos, naturais ou provocados pelo homem, sobre um cenário vulnerável, causando grave perturbação ao funcionamento de uma comunidade ou sociedade, envolvendo extensivas perdas e danos humanos, materiais, econômicos ou ambientais, que excede a sua capacidade de lidar com o problema usando meios próprios.

Pode-se distinguir os desastres quanto à sua origem e intensidade. Em relação à origem, os desastres são classificados como (ALCÂNTARA-AYALA, 2002; MARCELINO, 2008; TOMINAGA, 2009):

- I. Naturais: aqueles provocados por fenômenos e desequilíbrios da natureza que ocorrem de forma independente da ação humana, como chuvas intensas causadoras de inundações, escorregamentos, tornados etc.;
- II. Humanos: aqueles provocados pela interferência do homem ou por sua omissão, como exemplo a contaminação de rios, acidentes de trânsito, rompimentos de barragens, entre outros.

Já em relação à intensidade, os desastres podem ser distinguidos como (KOBIYAMA *et al.*, 2006; TOMINAGA, 2009):

- I. De pequeno porte: onde os impactos são pouco importantes e os prejuízos pequenos;
- II. De média intensidade: em que os impactos têm alguma importância e os prejuízos são significativos, mas não vultosos;

- III. De grande intensidade: na qual há danos relevantes e consideráveis prejuízos;
- IV. De muito grande intensidade: com impactos muito significativos e prejuízos muito grandes.

Há diferentes tipos de desastres naturais: movimentos de massa, erosão, seca e estiagem, alagamento, enxurradas, inundação, granizo, vendaval, chuvas intensas e incêndios florestais, além de desastres com menor número de registros: baixa umidade do ar, doenças infecciosas bacterianas, doenças infecciosas virais, frentes frias e zonas de convergência, friagem, geada, infestações de animais, marés de tempestade (ressacas), ondas de calor, outras infestações, tempestades de raios, tornados, tremor de terra e tsunamis (BRASIL, 2014).

2.1.1 Desastres naturais ao redor do mundo

Os desastres naturais ocorrem em diversas partes do mundo há séculos: há registros de que a cidade de Lima, no Peru, foi afetada por um intenso terremoto em 1746, matando cerca de 1.000 pessoas em uma população de 50.000. Nos últimos anos, terremotos continuam acontecendo, como na costa leste do Japão em 2011, onde esse fenômeno provocou um tsunami que, juntos, podem ter levado a óbito em torno de 20.000 cidadãos, além de comprometer a rede elétrica da área (UNISDR, 2011).

Os incêndios florestais são outro problema, como o que teve no oeste da Rússia em 2010: podem ser provocados pela combinação de temperaturas altas, falta de chuvas ou tempestades com raios, e campos secos, florestas e turfeiras que apresentam alto risco para queimadas. Em 2005, no município de Mumbai, na Índia, 24 horas de precipitação resultaram em uma grande inundação que sobrecarregou o sistema de esgoto e danificou as infraestruturas físicas e os sistemas de comunicação da região (UNISDR, 2011).

Em 2017 ocorreram 335 desastres naturais ao redor do mundo, sendo a maioria tempestades (127) e inundações (126), seguidos por deslizamentos (25), terremotos (22), incêndios florestais (15), temperatura extrema (10), seca (7), atividades vulcânicas (2) e movimento de massa (1) (Figura 1). Esses fenômenos afetaram mais de 95,6 milhões de pessoas, sendo a maioria na Ásia (69,5%); a maior parte da população (55 milhões) foi afetada por inundações, principalmente na Índia, no Nepal, em Bangladesh e na China (CRED/USAID, 2018).

Esses eventos levaram a óbito 9.697 pessoas, sendo o local mais impactado a Ásia (58%). As inundações causaram mais mortes (3.331), seguidas pelas tempestades (2.510),

deslizamentos (2.312), terremotos (1.012), temperatura extrema (359), incêndios florestais (165) e movimento de massa (8). Os desastres custaram uma perda econômica de 335 bilhões de dólares em 2017, no qual as Américas tiveram um maior prejuízo (88%); o evento que mais causou perda financeira foram as tempestades (284,7 bilhões de dólares), sobretudo furacões nos Estados Unidos e no Caribe (CRED/USAID, 2018).

Figura 1 - Desastres naturais no mundo em 2017.

Modificado de: CRED/USAID, 2018.

Os desastres naturais atingem as pessoas de forma negativa e, geralmente, são as pessoas mais pobres, aquelas que menos fizeram para provocar essas alterações significativas. A Estratégia Internacional das Nações Unidas para a Redução de Desastres – UNDRR (*United Nations Office for Disaster Risk Reduction*) é o ponto focal das Nações Unidas, que trabalha na intersecção entre a compreensão do risco e o impacto do risco, reduzindo as perdas por desastres e evitando o surgimento de novos riscos, de forma a proteger os ganhos obtidos em relação aos Objetivos de Desenvolvimento Sustentável (UNDRR, 2020).

Para a implementação da redução de riscos de desastres, de 2005 a 2015, foi adotado o Marco de Ação de Hyogo (MAH), adotado pelos Estados Membros das Nações Unidas. O MAH possuía a finalidade de “aumentar a resiliência das nações e das comunidades frente aos desastres ao alcançar, para o ano de 2015, uma redução considerável das perdas que ocasionaram os desastres, tanto em termos de vidas humanas

quanto aos bens sociais, econômicos e ambientais das comunidades e dos países” (EIRD, p. 02, 2005).

Atualmente, a UNDRR supervisiona a implementação do Marco de Sendai para a Redução do Risco de Desastres 2015 – 2030, monitorando e revisando o progresso. O Marco de Sendai foi o primeiro grande acordo da agenda de desenvolvimento pós-2015 e oferece aos Estados Membros ações concretas para proteger os ganhos de desenvolvimento do risco de desastres (UNDRR, 2020).

2.1.2 Desastres naturais no Brasil

Segundo Tominaga (2009), no Brasil, os principais desastres naturais derivam da dinâmica externa da Terra, ou seja, aqueles que são orientados pelo clima e por processos atmosféricos. Os eventos mais comuns são inundações e enchentes, escorregamentos de solo e rochas, e tempestades, que geralmente estão relacionados a eventos pluviométricos intensos e/ou prolongados. Também ocorrem no país desastres originados da dinâmica interna terrestre, como alguns tremores de baixa magnitude (variando entre 2° e 4° na escala Richter).

De acordo com o CEMADEN - Centro Nacional de Monitoramento e Alertas de Desastres Naturais, ocorreram 174 desastres naturais de natureza hidrogeológica (deslizamentos de terra) e 198 desastres naturais de natureza hidrometeorológica (enchentes e enxurradas) em 2019 no Brasil (RAMOS *et al.*, 2020). E no primeiro semestre de 2017, os desastres naturais decorrentes de seca e de excesso de chuvas afetaram 2.462 municípios brasileiros, 40, 1 milhões de pessoas e provocaram um prejuízo econômico de R\$39,4 bilhões (CNM, 2018).

O acelerado processo de urbanização, como o crescimento desordenado das cidades em locais impróprios à ocupação, proporcionou um aumento na ocorrência de desastres naturais no Brasil. Desmatamentos, cortes, aterros, alterações nas drenagens, lançamento de lixo e a construção de moradias sem uma infraestrutura adequada são atividades antrópicas que aumentam o perigo desses desastres (TOMINAGA, 2009). Informações sobre esses fenômenos que acontecem no país são essenciais “para se traçar os perfis das ocorrências e planejar o gerenciamento desses desastres, principalmente para se adotar medidas de prevenção, de modo a evitá-los ou a diminuir os impactos causados por eles” (BRASIL, 2014, p. 45).

O Centro Nacional de Gerenciamento de Riscos e Desastres (Cenad) é um dos órgãos brasileiros responsáveis por consolidar as informações sobre riscos no Brasil, através de mapas de áreas de risco de deslizamentos e inundações, além dos dados relativos à ocorrência de desastres naturais e tecnológicos e os impactos negativos associados (MINISTÉRIO DO DESENVOLVIMENTO REGIONAL, 2019). Outro órgão relevante no gerenciamento de desastres naturais é o Centro Nacional de Monitoramento e Alertas de Desastres Naturais (Cemaden), que possui a finalidade de realizar o monitoramento das ameaças naturais em áreas de risco nos municípios brasileiros suscetíveis à ocorrência de desastres naturais; o Cemaden também realiza pesquisas e inovações que podem melhorar seu sistema de alerta antecipado, visando diminuir o número de vítimas fatais e prejuízos materiais (CEMADEN, 2016).

A Companhia de Pesquisa de Recursos Minerais (CPRM) do Serviço Geológico do Brasil também colabora na setorização de áreas de risco geológico, que visa apontar para a Defesa Civil e aos gestores municipais, estaduais e federais, quais são as áreas prioritárias para a implementação de ações de gerenciamento, mitigação, monitoramento e resposta aos desastres naturais (CPRM, 2020). Outro órgão de destaque é a Defesa Civil, uma organização de toda a sociedade para autodefesa através de ações de prevenção, mitigação, preparação, resposta e recuperação destinadas a evitar desastres e reduzir os impactos sobre a população, promovendo o retorno à normalidade social, econômica ou ambiental (ASSEMBLEIA LEGISLATIVA DO ESTADO DE SÃO PAULO, 2019). A Seção de Investigações, Riscos e Desastres Naturais do Instituto de Pesquisas Tecnológicas do Estado de São Paulo (Sirden – IPT) oferece serviços tecnológicos (como exemplo, a identificação, análise e gestão de riscos e desastres naturais) e desenvolve projetos de pesquisa para apoiar políticas públicas e também ao setor privado do país, com ênfase em prevenção, atendimento de emergência e recuperação (IPT, 2020).

2.2 Inundações: características, causas e consequências

Inundações, enchentes, alagamentos e enxurradas: esses termos são comumente usados pela população com um mesmo significado, porém representam eventos distintos. Segundo o Ministério das Cidades/IPT (2007, p. 90, 91), enchente - ou cheia - é a “elevação temporária do nível d’água em um canal de drenagem devido ao aumento da vazão ou descarga” enquanto inundação representa o “processo de extravasamento das águas do canal de drenagem para as áreas marginais (planície de inundação, várzea ou

leito maior do rio) quando a enchente atinge a cota acima do nível máximo da calha principal do rio” (Figura 2). Já alagamentos são o “acúmulo momentâneo de águas em uma dada área decorrente de deficiência do sistema de drenagem” e as enxurradas são definidas como o “escoamento superficial concentrado e com alta energia de transporte” (MINISTÉRIO DAS CIDADES/IPT, 2007, p. 94).

Figura 2 - Diferença entre inundação e enchente.

Modificado de: Ministério das Cidades/IPT, 2007.

As inundações, enchentes e alagamentos afetam grande parte da população – principalmente as regiões metropolitanas das cidades – e seus efeitos podem ser classificados em diretos e indiretos. Os efeitos diretos se referem a mortes por afogamento, a destruição de moradias, danos materiais e gastos com recuperação enquanto os indiretos são aqueles relacionados às doenças transmitidas através da água contaminada, como a leptospirose, a febre tifoide, a hepatite e a cólera (MINISTÉRIO DAS CIDADES/IPT, 2007).

As inundações são um dos desastres naturais mais comuns ao redor do mundo e é aquele que mais causa óbitos, principalmente em países como Índia, Nepal e Bangladesh. Também é o desastre natural que mais afeta pessoas (12 milhões somente na China) e foi responsável por uma perda econômica de 20,3 bilhões de dólares em 2017 (CRED/USAID, 2018).

2.2.1 Inundações e análise espacial

Várias técnicas e métodos são utilizados para identificar áreas suscetíveis a eventos hidrológicos extremos; muitas delas são baseadas no uso de geotecnologias, como o Sensoriamento Remoto (SR) e os Sistemas de Informações Geográficas (SIGs)

(DE OLIVEIRA *et al.*, 2018). Os desastres possuem um componente espacial, assim, são indispensáveis informações geográficas adequadas sobre os perigos e as áreas vulneráveis a estes para poder mitigar os danos dos desastres (OUMA; TATEISHI, 2014).

Geotecnologias, dados fluviométricos e Modelos Digitais de Terreno (MDT) em conjunto são muito utilizados em estudos de monitoramento, previsão e alertas a inundações (CORRÊA, 2020). Os mapas de risco de inundação devem ser elaborados porque fornecem uma base para o gerenciamento de risco de inundação, sendo que esses planos têm que ser efetivamente comunicados aos grupos-alvo – como os tomadores de decisão, as unidades de resposta a emergências e a sociedade – para diminuir o risco de inundação, integrando interesses, potenciais e conflitos distintos sobre espaço e o uso da terra em um município (OUMA; TATEISHI, 2014). O SIG é capaz de oferecer mais e melhores informações sobre as situações de tomada de decisão, possibilitando que o gestor “identifique e liste um conjunto pré-definido de critérios com o processo de sobreposição” (OUMA; TATEISHI, 2014, p. 1.520). A principal vantagem do uso do SIG para o mapeamento de inundação é que ele cria potencial para analisar esse fenômeno a fim de estimar prováveis danos, além de gerar uma visualização do desastre (OUMA; TATEISHI, 2014).

Por outro lado, dados obtidos pelo SR podem ser utilizados para aperfeiçoar o desempenho de modelos hidrológicos e também possibilitam o acesso de informações de áreas que são fisicamente inacessíveis (OPOLOT, 2013). Elevação do terreno, extensão de uma inundação, largura do rio, nível de água e cobertura do solo são algumas das variáveis de SR que, quando integradas com uma modelagem de inundações, permitem uma melhora considerável nos processos de compreensão e previsão (DOMENEGHETTI; SCHUMANN; TARPANELLI, 2019).

Apesar de seus potenciais, o SIG e o SR apresentam algumas limitações. Sanyal e Lu (2005) relatam a presença de nuvens durante o período de enchentes como um dos grandes desafios no uso do sensoriamento remoto óptico (*Synthetic Aperture Radar - SAR*) para o monitoramento de inundações; além disso, há os problemas de disponibilidade limitada de imagens, variações sazonais, limitações técnicas e, principalmente, de baixa resolução temporal, pois a maioria das imagens de radar são obtidas algum tempo antes ou depois da inundação, sendo que na maior parte dos casos o ápice da inundação não é capturado (OPOLOT, 2013). Inundações e transporte ferroviário

2.2.2 Inundações e transporte

De acordo com Kellerman *et al.* (2015), a infraestrutura ferroviária é de suma importância para o transporte de bens e pessoas, contribuindo para o bem-estar econômico e social. No entanto, esse tipo de transporte está sujeito a perigos naturais frequentes, como inundações, terremotos e fluxo de detritos (HONG *et al.*, 2015).

A mudança climática provocou o aumento de chuvas curtas nos últimos anos, gerando preocupação de que tornados e enchentes, por exemplo, também aumentem (HONG *et al.*, 2015). Segundo Ota (2018), na elaboração de projetos para estruturas de engenharia civil deve-se verificar a capacidade de resistência dessas estruturas em relação a forças externas, como riscos naturais.

Rios podem transbordar dependendo da duração da chuva e, assim, podem fluir para estradas e ferrovias (OTA, 2018). As inundações podem afetar negativamente a estrutura e as operações de uma rede ferroviária, como atrasos de viagens, cancelamento de serviços de trens e grandes multas para o transporte ferroviário devido aos minutos atrasados (MURPHY *et al.*, 2016; CHEETHAM; CHIROUZE; BREDIER, 2016).

Em 2006, uma grande inundação ocorreu no entorno do Rio March, localizado na fronteira entre a Áustria e a Eslováquia, e afetou a Ferrovia do Norte, uma relevante linha de conexão da Ferrovia Federal Austríaca entre Viena e a República Tcheca. Este evento permitiu mostrar a alta vulnerabilidade da infraestrutura ferroviária às inundações.

Diante desses fenômenos, é necessária uma tecnologia preditiva mais detalhada. Alguns autores estudam técnicas para prever e mitigar os danos das inundações nas ferrovias. Ozaki, Watanabe e Fukusawa (2018), através da utilização de resultados de avaliações de mapeamentos de perigos, desenvolveram um algoritmo baseado em previsões de vento e precipitação que permite identificar quais trens serão afetados pela inundação, onde os trens devem parar e qual a rota que os passageiros devem tomar para evacuar, caso seja necessário. Kellerman *et al.* (2015) apresentaram uma abordagem de modelagem empírica para estimar danos diretos estruturais causados por inundações à infraestrutura ferroviária e às perdas financeiras associadas; em contrapartida. Hong *et al.* (2015) propuseram uma metodologia para avaliar quantitativamente a vulnerabilidade do sistema ferroviário aos eventos de inundação utilizando dados históricos disponíveis e tecnologia SIG.

Dos Santos *et al.* (2016) estudaram o risco de inundações nas linhas férreas e ferroviárias do município de São Paulo e, para isso, consideraram características topográficas e hidrológicas das áreas de estudo, além da análise de interferências desse tipo de desastre na operação ótima do sistema ferroviário. Enquanto Ota (2018) elaborou um sistema de mapeamento de risco de inundações e indica os melhores locais para os trens pararem, além de rotas de abrigo para os passageiros.

2.3 Escorregamentos: conceitos, tipos e causas

De acordo com Fornasari Filho e Faria (2018, p. 214) “os escorregamentos consistem no movimento rápido de massas de solo ou rocha, geralmente bem definidas quanto ao seu volume, cujo centro de gravidade se desloca para baixo e para fora de um talude (natural, de corte ou aterro)”. Esse fenômeno pode estar associado a períodos de intensa pluviosidade (FORNASARI FILHO; FARIA, 2018). Localidades situadas em regiões serranas – como áreas de intensa urbanização – e locais onde foram instaladas obras de engenharia de grande porte – como rodovias, ferrovias, dutovias e linhas de transmissão – são ambientes favoráveis à ocorrência desse tipo de evento (MINISTÉRIO DAS CIDADES/IPT, 2007).

Os escorregamentos podem ser classificados como (FORNASARI FILHO; FARIA, 2018):

- I. Translacionais ou planares: envolvem solos superficiais; são processos muito frequentes nas encostas serranas brasileiras e em taludes, condicionados por estruturas planares desfavoráveis à estabilidade;
- II. Circulares ou rotacionais: possuem superfície de escorregamentos curvas; estão associados a aterros, pacotes de solo ou depósitos mais espessos, rochas sedimentares ou cristalinas intensamente fraturadas;
- III. Em cunha: estão relacionados a saprolitos e maciços rochosos, na qual há duas estruturas planares que provocam o deslocamento de um prisma ao longo do eixo de interseção desses planos

Os escorregamentos podem ser previstos, pois é possível monitorar seus agentes condicionantes ou potencializadores e os mecanismos relacionados com a movimentação dos materiais nos diferentes tipos de escorregamentos, ou seja, pode-se identificar as áreas mais suscetíveis a ocorrência de escorregamentos, além da sua tipologia e magnitude. Desse modo, permite-se definir e implantar mecanismos e ações distintas para evitar e/ou

minimizar as consequências negativas provocadas por esse fenômeno em certa área. Para isso, a gestão do território deve considerar três fatores: evitar o surgimento de áreas e situações de risco, identificar e caracterizar essas regiões já existentes, e resolver ou diminuir os riscos que existem (VEDOVELLO; MACEDO, 2007).

2.3.1 Escorregamentos e geoprocessamento

Os SIGs permitem um melhor aproveitamento de dados existentes e, a partir destes, produzem novas informações, possibilitando um ordenamento de ações mais eficiente, que é relevante para o planejamento do território, como para o gerenciamento de risco (VARANDA; MAHLER; OLIVEIRA, 2010). As geotecnologias representam possibilidades de análise de risco, incluindo monitoramento, modelagem e gestão dos desastres naturais (SIMÕES *et al.*, 2016).

Um outro avanço das geotecnologias nesse aspecto é que há cada vez mais uma maior integração entre SIGs e modelos computacionais de diferentes naturezas (como de escorregamentos), permitindo que as análises de gestão de dados espaciais e de modelagem possam ser executadas em um mesmo ambiente. O problema é que a análise para compreender as incertezas dos dados de entrada em SIG e os modelos computacionais são pouco considerados na gestão de desastres naturais. Por exemplo, nos modelos de escorregamentos, os dados de entrada incluem dados topográficos, hidrológicos, com alto grau de incerteza na sua distribuição espacial (SIMÕES *et al.*, 2016).

Por outro lado, o sensoriamento remoto apresenta uma excelente aplicabilidade em estudos de detecção, mapeamento e monitoramento dos escorregamentos (LOPES, 2006). O SR também permite o mapeamento das cicatrizes de escorregamentos, que contribuem para o entendimento e distribuição espaço-temporal desse fenômeno (FERNANDES; AMARAL, 1996). Lopes (2006) ressalta que as fotos aéreas possibilitam o mapeamento das formas e dimensões das cicatrizes dos escorregamentos, enquanto as imagens de satélite permitem identificar apenas as cicatrizes maiores, sem precisão dos detalhes (CORRÊA, 2020).

Varanda, Mahler e Oliveira (2010) apresentam um modelo para análise quantitativa de risco com a aplicação de SIG, utilizando a teoria bayesiana, na integração de mapas temáticos que representou a suscetibilidade aos escorregamentos considerados os aspectos do meio ambiente através do mapa de perigo que integrado com mapas e

dados representativos das características urbanas resultou em um mapa de risco; e Simões *et al.* (2016) discutiram métodos de abordagens para tratar com incerteza de dados de entrada em modelos de escorregamento e também analisaram os processos de interação entre SIGs e modelagem, enfatizando os modelos de escorregamento, Zhang, Wang e Wu (2012) monitoraram a dinâmica de movimento dos deslizamentos de terra utilizando imagens de SR de alta resolução e propuseram uma aplicação do método de regressão ótimo baseado na análise de regressão multivariável no monitoramento dos deslizamentos; Tan *et al.* (2021) estudaram a avaliação de risco de perigos geológicos dos escorregamentos baseados em SR e SIG; Zhang *et al.* (2021) usaram imagens ALOS-2 (banda L) e Sentinel-1 (banda C) para mapear potenciais locais sujeitos à ocorrência de deslizamentos de terra, além de técnicas multitemporais; Manfré, Albuquerque e Quintanilha (2018) apresentaram procedimentos de processamento digital de imagens e de geoprocessamento para áreas de risco e de cicatrizes de deslizamento vinculadas às vias de acesso da Central Nuclear Almirante Álvaro Alberto (CNAAA) em Angra dos Reis (RJ), como modelos digitais de elevação - MDE, imagens Landsat e classificação baseada em objeto; entre outros estudos.

2.3.2 Escorregamentos e transporte

Um dos principais problemas socioambientais que as concessionárias brasileiras enfrentam são as ocorrências de processos erosivos acelerados e associados à implantação de sistemas de transportes (SOUZA *et al.*, 2014). As linhas de transporte, como estradas e ferrovias, principalmente em terrenos montanhosos, são frequentemente propensas a escorregamentos que ocorrem em taludes ou declives naturais (JAISWAL *et al.*, 2011). Alguns autores estudaram a ocorrência desse tipo de desastre nas proximidades de vias férreas. Pomante e Di Felice (2018) buscaram integrar Rede de Sensores Sem Fio (*Wireless Sensor Network* – WSN) e Sistema de Informação Geográfica para um monitoramento em tempo real de escorregamento em estações e linhas ferroviárias da Região Abruzzo da Itália; Bednarik *et al.* (2010) avaliaram a suscetibilidade de escorregamentos na ferrovia entre os municípios Kral'ovany e Liptovský Mikuláš na Eslováquia; e Jaiswal *et al.* (2011) mostraram um modelo quantitativo de risco de escorregamentos para vias de transporte, com um exemplo de alinhamento de estradas e ferrovias, em partes das colinas de Nilgiri no sul da Índia.

Esse tipo de desastre provoca danos diretos ou indiretos às ferrovias e para estimar esses riscos é importante considerar o número esperado de escorregamentos e sua

probabilidade anual de ocorrência. Desse modo, é possível estimar a probabilidade de um escorregamento atingir um veículo em movimento ou um passageiro. Para estabelecer essa relação estatística deve-se ter um inventário dos eventos, o que pode ser concomitantemente vantajoso e desvantajoso: o benefício do inventário é que ele é preparado logo depois da ocorrência, permitindo registrar a maioria desses eventos em uma área, enquanto sua limitação é a indisponibilidade de um registro contínuo e a dificuldade de acesso aos dados (JAISWAL *et al.*, 2011).

Bednarik *et al.* (2010) utilizaram os seguintes parâmetros de entrada para avaliar a suscetibilidade à escorregamentos entre a seção ferroviária Kral'ovany e Liptovský Mikuláš na Eslováquia: interpretação das condições geológicas; parâmetros morfométricos (MDE, ângulo de inclinação e aspecto da inclinação); estrutura real da paisagem; interpretação de escorregamentos; análise estatística bivariada (baseia-se na comparação entre o mapa de inventário como uma variável dependente com cada mapa paramétrico de entrada, possibilitando o cálculo do peso para cada variável de entrada); processo de ponderação; mapa prognóstico de suscetibilidade a escorregamentos e, finalmente, o processo de verificação. “A avaliação da suscetibilidade a escorregamentos e o processamento estatístico são baseados em um axioma do atualismo”, ou seja, supõe-se que os eventos futuros ocorrerão nas mesmas condições do passado (BEDNARIK *et al.*, p. 162, 2010).

2.4 Big Data: definição e características

De acordo com Allam e Dhunny (2019), os municípios estão cada vez mais preocupados com tecnologias especializadas para tratar problemas relacionados à sociedade, ecologia, morfologia, entre outras áreas. O conceito emergente “Cidades Inteligentes” (*Smart Cities*) incentiva essa perspectiva e promove a incorporação de sensores de *Big Data* através da “Internet das Coisas” (*Internet of Things – IoT*).

As cidades inteligentes são caracterizadas por componentes como mobilidade, governança, meio ambiente e pessoas, saúde, transporte, educação inteligente e energia (KHAN; ANJUM; KIANI, 2013). Uma das mais recentes tecnologias com um grande potencial para aprimorar os serviços das cidades inteligentes é a análise de *Big Data* (AL NUAIMI *et al.*, 2015).

Os dados cresceram em larga escala e em vários campos; estão em toda a parte, organizações, governos, servidores web e até mesmo em nosso corpo (AL-SAI; ABDULLAH, 2019). Atualmente existem muitas definições e teorias distintas sobre *Big Data* (MALIK, 2013). De modo geral, “*Big Data* é um termo que descreve grandes volumes de dados complexos, alta velocidade de processamento e variáveis que requerem técnicas avançadas e tecnologias para permitir a captura, armazenamento, distribuição, gerenciamento e análise da informação” (TECHAMERICA FOUNDATION, 2012, p. 10).

Como o *Big Data* apresenta um grande volume de dados, é necessário que ele consiga atuar com uma ampla variedade de dados. Desse modo, a variedade do *Big Data* é caracterizada em (JAHN, 2017):

- I. Dados Estruturados: são dados que têm forma definida, de forma rígida, como aqueles extraídos de planilhas ou banco de dados no formato *Structured Query Language* (SQL);
- II. Dados Semiestruturados: são dados semelhantes aos estruturados, mas não obedientes na totalidade quanto à forma, como os registros de linguagens baseadas em HTML e XML;
- III. Dados Não Estruturados (ou NoSQL): são dados que não possuem um formato específico; são coletados na sua forma primária, como um texto, um vídeo, um fragmento de e-mail ou uma foto.

Essa nova tecnologia é caracterizada e derivada principalmente do *Business Intelligence* e *Business Analytics* (BI&BA). A maioria dos principais *players* de tecnologia, como IBM, Microsoft, Google, Facebook, Amazon, etc, começaram a incluir técnicas de *Big Data* em seus planos nos últimos anos (AL-SAI; ABDULLAH, 2019).

O crescimento de recursos de *Big Data* está aumentando exponencialmente, trazendo novos desafios críticos para sua implementação. Quanto as tecnologias, a análise de *Big Data* necessita de novas técnicas que atendam dados massivos, que requer habilidades e capacidade de coletar, armazenar e analisar dados usando ferramentas de gerenciamento de dados que utilizam a tecnologia de *Big Data*, como o *Hadoop* e o *Spark*, por exemplo (CHEN; MAO; LIU, 2014).

O *Hadoop* foi criado para processamento em lote, permitindo escalabilidade e tolerância a falhas, possibilitando que os aplicativos sejam executados em milhares de

nós com *petabytes* de dados. Ele gerencia uma grande quantidade de dados, dividindo os arquivos em blocos e os distribui aos nós do *cluster* do *Hadoop* (MAVRIDIS; KARATZA, 2017).

O *Hadoop* está associado ao *Hadoop Distributed File System* (HDFS) e ao *MapReduce*. O HDFS armazena arquivos em blocos de tamanho fixo em diferentes nós do *cluster Hadoop*, dividindo os arquivos em blocos menores, permitindo armazenar arquivos muito maiores do que a capacidade de cada nós do disco. Já o *MapReduce* é um modelo de computação distribuído, baseado em lote, que consiste em uma Fase de Mapa e em uma Fase de Redução (MAVRIDIS; KARATZA, 2017).

Em relação ao *Spark*, ele estende o modelo *MapReduce* e suporta a combinação de várias técnicas de processamento de dados, além de possuir *APIs Python*, Java, Scala, SQL e muitas bibliotecas incorporadas. Uma das principais vantagens do *Spark* é que ele usa mais efetivamente a memória principal e minimiza as transferências de dados para o disco, realizando cálculos na memória. Em suma, a grande diferença entre o *Hadoop* e o *Spark* é que o *Hadoop* processa dados no disco enquanto o *Spark* processa os dados na memória e tenta minimizar o uso do disco (MAVRIDIS; KARATZA, 2017).

2.4.1 *Big Data* e transporte

A rápida urbanização modernizou a vida de várias pessoas, porém causou grandes problemas, como poluição, consumo de energia, congestionamento de tráfego etc. A fim de minimizar esses problemas, a computação urbana utiliza dados gerados nas cidades, por exemplo, dados geográficos, fluxo de tráfego e mobilidade urbana. Este tipo de computação conecta sensoriamento remoto, gerenciamento de dados, analisa dados e prestação de serviços de forma a melhorar a vida dos indivíduos, dos sistemas operacionais dos municípios e do meio ambiente (ZHENG *et al.*, 2014).

O setor de transporte público foi um dos primeiros a utilizar e implementar *Analytics* e *Big Data*, como em previsão de viagens a operações de trânsito. Há também estudos sobre aplicações de *Big Data* no setor de ferrovias. Os sistemas de transporte ferroviário tendem a ser receptivos às vantagens de *Analytics* e *Big Data* porque estão relacionados aos sistemas fechados, que envolvem processamento sofisticado de grandes volumes de dados (THADURI; GALAR; KUMAR, 2015).

Existem diversas fontes de dados geradas a partir das ferrovias, como horários, velocidade, número de passageiros etc., que constituem um *Big Data* (THADURI;

GALAR; KUMAR, 2015). A indústria ferroviária foi revolucionada pela Análise de *Big Data* (BDA), que auxilia nos processos de tomada de decisão das empresas ferroviárias. Há estudos na literatura que demonstram as várias vantagens da aplicação de BDA na redução de custos e de atrasos nos sistemas de transporte ferroviário (RTS), além da manutenção de elevados padrões de segurança, confiabilidade e satisfação do cliente (GHOFRANI *et al.*, 2018).

2.4.2 *Big Data* e análise espacial

O avanço da tecnologia da informação proporcionou o grande aumento das demandas de processamento, análise e visualização de *Big Data* (QINGQUAN; DEREN, 2014). O processamento das informações geográficas está integrado ao processo de negócios SIG específico do usuário e inclui os dados espaciais e os algoritmos de processamento (BADEA; BADEA, 2013) na qual a relação entre esses elementos é formada por entrada, processamento e saída (ZHANG *et al.*, 2019).

O sistema de informação geográfica enfrenta novos desafios, tendo que evoluir suas tecnologias para poder superar as dificuldades causadas pelo *Big Data* (LI *et al.*, 2016). Entre os desafios para o SIG estão analisar e processar *Big Data* espaço-temporal, agrupar e distribuir *Big Data* espacial, indexar e gerenciar *Big Data*, e calcular e visualizar *Big Data* no sistema, sobretudo mantendo um alto desempenho (YANG *et al.*, 2017a; YANG *et al.*, 2017b). As plataformas populares de *Big Data* ainda não são capazes de realizar análises espaciais, computação espacial ou mineração de dados espaciais; portanto, é necessário que o SIG integre a tecnologia geral de *Big Data* (WANG; ZHONG; WANG, 2019).

Devido ao aumento no volume e na variedade de dados, o banco de dados tradicional é propenso a baixa eficiência de armazenamento e fraca capacidade de acesso simultâneo (LI *et al.*, 2015; LI *et al.*, 2016). Além disso, o sistema SIG tradicional executa somente as tarefas de computação na fila de trabalhos; portanto é preciso desenvolver novas tecnologias de armazenamento de dados espaciais (WANG; ZHONG; WANG, 2019) e para se adaptar às necessidades da era de *Big Data*, também é importante um SIG baseado em nuvem; mas o processo de negócios de um SIG na nuvem é mais complicado em relação ao SIG tradicional (NG *et al.*, 2018), já que as extensões de processos de negócios de SIG na nuvem apresentam deficiências significativas em versatilidade e flexibilidade (ZHANG *et al.*, 2019).

Com o crescimento da utilização de smartphones, plataformas de mídia social, entre outros tipos de tecnologia, ocorreu um aumento na quantidade de dados disponíveis e também uma proliferação de novas fontes de dados (HORITA *et al.* 2017). Os recursos de *Big Data* fornecidos pelas redes móveis e mídias sociais oferecem uma oportunidade para a extração de informações valiosas, que são aproveitadas pelas entidades comerciais para medir o nível de satisfação do cliente (ALLAM; DHUNNY, 2019).

Com o surgimento do termo *Big Data* apareceu o desafio de como alinhar os processos de tomada de decisão dentro de uma organização de maneira que não haja falta de informações ou dados inúteis para os tomadores de decisão (HORITA *et al.*, 2017). Na área das geociências, séries temporais muito grandes ou imagens aéreas e de satélite de alta resolução são fontes de informações valiosas; o problema é que nem sempre é possível extrair conhecimento desses dados através de técnicas estatísticas padrão. É necessário utilizar métodos automatizados de aprendizado de máquina para obter padrões relevantes, já que a computação de alto desempenho e a visualização de dados são amplamente aplicados com sucesso aos dados relacionados a desastres naturais (MARTÍNEZ-ÁLVAREZ; MORALES-ESTEBAN, 2019).

Os dados de sensoriamento remoto são uma das principais fontes de *Big Data* para monitoramento e detecção de desastres, pois um gerenciamento eficaz de desastres pode ser feito através de imagens multitemporais detectadas remotamente; o que ajuda os tomadores de decisão na implementação de estratégias de mitigação. A segunda principal fonte de *Big Data* para detecção de desastres são os dados gerados pelos usuários, como sites de mídia social e informações geográficas voluntárias (*volunteered geographic information* – VGI) (YU; YANG; LI, 2018).

Para avaliar os danos dos desastres, a interpretação visual de imagens de satélite pode ser demorada, imprecisa e cara. Em contrapartida, os algoritmos de aprendizado de máquina se adaptam e aprendem o problema independente das suposições estatísticas sobre a distribuição de dados, apresentando uma melhor precisão em relação aos métodos tradicionais de detecção e classificação de alterações (YU; YANG; LI, 2018).

A integração de diversas fontes de dados pode contribuir para a melhoria da qualidade e integridade dos dados, porém antes dessa integração deve-se haver a validação dos dados. Quando diferentes fontes de *Big Data* são integradas cria-se uma incerteza para resposta de desastres; a interpretação e análise manuais dos dados não são

mais adequadas, sendo necessários métodos sofisticados de análise automática, como o aprendizado de máquina, para eliminar dados não relacionados e acelerar a análise, contribuindo para uma rápida previsão dos desastres e na identificação de estratégias de resposta ideais (YU; YANG; LI, 2018).

Há alguns estudos que utilizaram ferramentas *Big Data* para previsão ou monitoramento de desastres naturais. Horita *et al.* (2017) aplicaram em sua pesquisa uma estrutura de modelagem a um cenário prático com o objetivo de conectar tarefas de tomadores de decisão às fontes de dados através do oDMN + (*observation-aware Decision Model and Notation* – Modelo e Notação de Decisão com reconhecimento de observação), que é um modelo e notação integrados estendidos, e um processo de modelagem para empregar o oDMN + na prática; Ragini, Anand e Bhaskar (2018) classificaram tweets em seu estudo durante um desastre, ajudando na construção de um modelo de sentimentos sobre as diversas necessidades das pessoas; esse modelo proposto permitiu a equipe de resgate a entender a situação do desastre e agir em conformidade.

Em relação à aplicação de recursos *Big Data* no mapeamento de inundações, Chen *et al.* (2014) projetaram uma abordagem chamada RFim (modelo de inundação em tempo real) que foi proposta para fazer simulação e previsão da extensão de uma inundação; o RFim procura encontrar uma relação entre *Big Data* de sensoriamento remoto e dados *in situ* para construir o modelo de previsão em tempo real, aproveitando a vasta cobertura espacial e a alta resolução de imagens de sensoriamento remoto, além da cobertura temporal contínua e a fácil acessibilidade de observações *in situ*. Wang *et al.* (2019) utilizaram as mídias sociais para melhorar o controle de inundações, coletando e analisando dados de um aplicativo por meio de processamento de linguagem natural e técnicas de versão em computador. Também focado na previsão de inundações.

Restrepo-Estrada *et al.* (2018) propuseram um método para acionar alarmes de emergência para inundação usando dados de mídia social e medições de chuvas de fontes autorizadas (MARTÍNEZ-ÁLVAREZ; MORALES-ESTEBAN, 2019), já Cian, Marconcini e Ceccato (2018) apresentaram uma técnica de mapeamento de inundação baseada em um índice que é calculado usando estatísticas multitemporais de imagens de radar, comparando uma grande quantidade de cenas de referência às adquiridas durante a enchente investigada, facilitando a categorização das áreas inundadas: áreas exclusivamente temporariamente cobertas por água ou áreas com água e vegetação misturadas.

Quanto aos escorregamentos, Intrieri *et al.* (2017) mostraram em seu artigo como os problemas logísticos vinculados a técnicas avançadas de monitoramento, como transferência e armazenamento de *Big Data*, podem ser tratados de forma compatível com um sistema de alerta de escorregamento, focando na interação entre uma ferramenta de monitoramento da área (um radar interferométrico no solo) e um DCPC (*Data Collecting and Processing Center* - Centro de Coleta e Processamento de Dados). Zhou *et al.* (2018) desenvolveram estratégias para prevenção e mitigação de riscos para escorregamento: usaram dois modelos de aprendizado de máquina (*vector machine* e redes neurais artificiais) e um modelo multivariado (regressão logística) em dados coletados da China (MARTÍNEZ-ÁLVAREZ; MORALES-ESTEBAN, 2019).

Há uma carência de estudos que tratam sobre a passagem do dado espacial para o ambiente *Big Data*, principalmente em relação à desastres naturais como inundações e escorregamentos. Portanto, é necessário descobrir essa lacuna de informação através de uma revisão bibliográfica tanto de nível internacional como nacional.

2.4.2.1 *Big Data* Geoespacial e suas ferramentas

O *Big Geospatial Data*, que se refere ao crescente volume de dados geográficos e espaciais, contribuiu para o desenvolvimento de diversas ferramentas com a finalidade de processar eficientemente esses dados (BACHIEGA *et al.*, 2018). Os Sistemas Analíticos Espaciais (SAEs) representam uma tecnologia nova que permite gerenciar um grande volume de dados espaciais através de *frameworks* de processamento paralelo e distribuído de dados; a sua maioria possui como framework base o *Apache Hadoop* ou o *Apache Spark* (CASTRO, 2019).

O *Apache Hadoop* é uma implementação de código aberto do *MapReduce*, que processa um grande volume de dados de maneira paralela/distribuída em um cluster (HADOOP, 2020); sua maior vantagem é que ele pode ser executado em um hardware comum, não necessitando de nenhum supercomputador para gerenciar um grande banco de dados (SINGH; BAWA, 2016). Por outro lado, o *Apache Spark* é um sistema de computação em um cluster de memória, que fornece uma nova abstração de dados chamada *resilient distributed datasets* - RDDs (conjuntos de dados distribuídos resilientes) (ZAHARIA *et al.*, 2012). Os RDDs são coleções de objetos particionados em um cluster de máquinas, que permitem que o *Spark* supere modelos existentes, como o *MapReduce*. A desvantagem do *Spark* é que ele não fornece suporte para dados e operações espaciais; isso significa que seus usuários necessitam programar seus próprios

trabalhos de processamento de dados espaciais sobre o *Spark* (YU; WU; SARWAT, 2015).

2.4.2.2 GIS Tools for Hadoop (ArcGIS)

O ArcGIS, que é um software pago, que consiste em uma coleção de *frameworks*, dentre eles há o ArcGIS Desktop, uma estrutura básica para compilação, mapeamento, edição, análise e visualização de dados espaciais. É um conjunto integrado de aplicativos de *desktop*, como ArcMap, ArcCatalog, ArcToolbox e ArcGlobe (ESRI, 2015; SINGH; BAWA, 2016).

A nova tecnologia do ArcGIS, encontrada na versão Pro do *software*, o *GIS Tools for Hadoop* é um kit de ferramentas de código aberto que proporciona análise espacial para grandes dados. As ferramentas GIS para o *Hadoop* ampliam a plataforma do *Hadoop*, que não apresenta uma funcionalidade para o processamento e visualização de grandes volumes de dados em mapas (LEONARDI, 2013).

Essa ferramenta possibilita que os usuários do ArcGIS exportem dados dos mapas em arquivos de extensão HDFS (formato do sistema *Hadoop*) e cruzem com uma grande quantidade de registros armazenados na plataforma, sendo que os resultados podem ser salvos diretamente para o banco de dados do sistema ou reimportados para o ArcGIS, permitindo um melhor nível de geoprocessamento e visualização (LEONARDI, 2013). O GIS para o *Hadoop* apresenta uma série de bibliotecas e utilitários que conectam o ArcGIS com o ambiente *Hadoop*. O *GIS Tools for Hadoop* é composto por quatro projetos (ESRI, 2013):

- I. Os blocos de construção: refere-se a API (*application programming interface*) da *Esri Geometry* para Java, que inclui objetos de geometria, operações espaciais e indexação espacial;
- II. O *framework*: é a estrutura espacial para o *Hadoop*, que inclui funções definidas pelo usuário (UDFs), construídas com base nos recursos da API *Esri Geometry*, permitindo evitar algoritmos complicados do MapReduce e manter um fluxo de trabalho mais familiar;
- III. O conector: são as ferramentas de geoprocessamento para *Hadoop*, que são baixadas como uma caixa de ferramentas e aplicadas no ArcMap; permitem conectar

dados entre o *Hadoop* e o ArcGIS, enviar tarefas de fluxo de trabalho e transportar resultados do *Hadoop* para o ArcGIS para visualização;

IV. O kit de ferramentas: refere-se as ferramentas GIS para *Hadoop*, que sintetiza os três projetos anteriores.

2.4.2.3 *Hadoop-GIS*

O *Hadoop-GIS* é um sistema de armazenamento de dados escalável e de alto desempenho para executar consultas espaciais em grande escala no *Hadoop*. Esse sistema suporta vários tipos de consultas espaciais no *MapReduce* e está disponível como um conjunto de bibliotecas para o processamento de consultas espaciais e como um pacote de software integrado no Hive, uma interface de consulta declarativa (AJI *et al.*, 2013). Esse sistema trabalha com linguagem de programação C++ e Java e fornece suporte tanto para dados espaciais simples como complexos, porém não emprega uma biblioteca espacial externa e oferece apenas um módulo básico de visualização espacial que permite que seus usuários vejam fronteiras de partições (CASTRO, 2019).

O *Hadoop-GIS* tem como finalidade oferecer um sistema de consulta espacial expressivo, escalonável e eficiente para apoiar consultas analíticas em dados espaciais em grande escala e para fornecer uma solução viável que possa ser oferecida para operações diárias. O sistema suporta consultas espaciais como ponto, contenção, junção e consultas complexas, como correspondência espacial (junção espacial em larga escala) e consultas de vizinhos mais próximos. O *Hadoop-GIS* oferece particionamento de dados espaciais para obter uma paralelização de tarefas, porém o particionamento pode gerar problemas de distorção de dados espaciais - o que pode causar um desequilíbrio de carga das tarefas distribuídas e um maior tempo de resposta – e problemas de limites de objetos – que pode provocar resultados incorretos da consulta, se não forem tratados adequadamente (AJI *et al.*, 2013).

2.4.2.4 *SpatialHadoop*

O *SpatialHadoop* consiste em quatro camadas: camada de linguagem, camada de operações, camada *MapReduce* e camada de armazenamento (ELDAWY; MOKBEL, 2015). O *SpatialHadoop* fornece índices espaciais e permite visualização de dados espaciais (ELDAWY; MOKBEL; JONATHAN, 2016), porém apenas objetos espaciais simples. Esse sistema utiliza exclusivamente a linguagem de programação Java (CASTRO, 2019) e suporta vários tipos de geometria, como polígono, ponto, sequência

de linhas, multiponto, além de várias técnicas de particionamento espacial (ELDAWAY; ALARABI; MOKBEL, 2015; YU; ZHANG; SARWAT, 2018).

2.4.2.5 *SpatialSpark*

O *SpatialSpark* é escrito na linguagem de programação Scala, permitindo a utilização da maioria das bibliotecas Java sem nenhuma alteração. Embora esse sistema tenha implementado várias técnicas de indexação espacial e filtragem espacial, ele reutiliza a biblioteca espacial JTS (*Java Topology Suit*) para refinamento espacial, ou seja, ele testa se dois objetos geométricos satisfazem uma certa relação espacial ou calcula uma determinada medida entre dois elementos geométricos (YOU; ZHANG; GRUENWALD, 2015).

As desvantagens é que o *SpatialSpark*, em alguns casos, pode ter um pior desempenho em vários nós de computação do que em um único nó, apresentando uma baixa escalabilidade, principalmente para experimentos que consomem mais dados (YOU; ZHANG; GRUENWALD, 2015), e não disponibiliza nenhum módulo para a visualização distribuída de dados espaciais, sobrecarregando o usuário com a busca de *softwares* para este fim. Em relação ao suporte provido aos tipos de dados espaciais, o *SpatialSpark* tem a vantagem de aceitar todos os tipos de dados espaciais (simples e complexos, ponto, linha e polígono) (CASTRO, 2019).

2.4.2.6 *GeoSpark*

O *GeoSpark* é um sistema de computação em *cluster* para processamento de dados espaciais em larga escala (GEOSPARK, 2020). O *GeoSpark* estende o mecanismo principal do *Apache Spark* e *SparkSQL* para suportar tipos de dados espaciais, índices e operações geométricas em escala, ou seja, esse sistema entende o conceito de RDDs (*Resilient Distributed Datasets*) para suportar dados espaciais. Esse sistema também pode produzir planos de consulta espacial otimizados e executar consultas em conjuntos de dados espaciais em larga escala, possibilitando a criação de mapas de alta resolução em paralelo (YU; ZHANG; SARWAT, 2018).

Essa ferramenta utiliza as linguagens de programação Java e Scala e suporta todos os tipos de dados espaciais, tanto os simples como os complexos (linha, ponto e polígono), porém permite a visualização apenas para objetos espaciais simples (CASTRO, 2019). O *GeoSpark* consiste em três camadas principais (YU; WU; SARWAT, 2015):

- I. Camada *Apache Spark*: que consiste em operações regulares suportadas pelo *Apache Spark*; essas funções nativas são responsáveis por carregar e salvar dados para armazenamento;
- II. Camada de conjunto de dados distribuído resiliente espacial (SRDD): estende o *Spark* com RDDs espaciais (SRDDs) que particiona eficientemente elementos de dados SRDD entre máquinas e introduz novas transformações espaciais paralelas e ações que fornecem uma interface mais intuitiva para os usuários escreverem programas de análise de dados espaciais;
- III. Camada de processamento de consultas espaciais: suporta consultas espaciais para conjuntos de dados espaciais em larga escala; depois que os objetos geométricos são armazenados e processados na camada *Spatial RDD*, o usuário pode chamar uma consulta espacial fornecida nessa terceira camada, processando essa consulta e, assim, retornando os resultados finais para o usuário.

2.4.2.7 GeoMesa

O GeoMesa é um conjunto de ferramentas de *software* livre e de código aberto que permite consultas e análises geoespaciais em larga escala em sistemas de computação distribuídos. Esse sistema fornece indexação espaço-temporal sobre os bancos de dados Accumulo, HBase, *Google Bigtable* e Cassandra para armazenamento massivo de dados (simples e complexos) de pontos, linhas e polígonos, além de oferecer processamento de dados espaço-temporais por fluxo quase em tempo real. O GeoMesa suporta o *Apache Spark* para análises geoespaciais personalizadas distribuídas (GEOMESA, 2020a). Esse sistema fornece um módulo chamado GeoMesaSpark, que permite que o *Spark* leia os dados pré-processados e pré-indexados do repositório de dados Accumulo (APACHE ACCUMULO, 2020).

O GeoMesa possui a extensão *GeoMesa Bigtable Data Store*, que é uma implementação da interface do *GeoTools DataStore*, que tem apoio do *Google Cloud Bigtable*, cujo código baseia-se na *API Bigtable HBase* (GEOMESA, 2021). Segundo Chang *et al.* (2008, p. 1), o “*Bigtable* é um sistema de armazenamento distribuído para o gerenciamento de dados estruturados, projetado para ser dimensionado para um tamanho muito grande: *petabytes* de dados em milhares de servidores de *commodities*”. Diversos projetos do *Google* armazenam seus dados no *Bigtable*, como *Google Earth* e *Google Finance* e aplicam demandas muito variadas a esse sistema, desde a utilização de dados

de tamanho muito grandes, como imagens de satélite, até a entrega de informações em tempo real (CHANG *et al.*, 2008).

Esta pesquisa tentou implementar os seus dados no GeoMesa, porém foram encontradas algumas dificuldades como a ausência de algumas informações essenciais sobre o *software*, tanto em documentações disponibilizadas como pela dificuldade de se obter suporte dos canais oficiais de comunicação.

2.4.2.8 *LocationSpark*

O *LocationSpark* é um sistema de processamento de dados espaciais construído sobre o *Apache Spark*. Ele é capaz de rastrear dados espaciais acessados com frequência e libera dinamicamente dados acessados com menos frequência no disco (TANG *et al.*, 2016). Esse sistema não emprega uma biblioteca espacial externa, é desenvolvido com a linguagem de programação Scala e disponibiliza suporte para dados espaciais apenas de pontos simples (CASTRO, 2019).

2.4.2.9 *SparkGIS*

SparkGIS é um sistema orientado na memória para manipulação de consultas espaciais, que utiliza o *Apache Spark*. Esse sistema suporta consultas espaciais básicas e reduz a distorção no processamento distribuído, oferecendo suporte a vários algoritmos de particionamento dinâmico adequados para um rico conjunto de cenários de aplicativos contemporâneos (BAIG *et al.*, 2018).

O *SparkGIS* emprega a reescrita dinâmica de consultas para gerenciar normalmente grandes fluxos de trabalho de consultas espaciais que excedem os recursos distribuídos disponíveis, mas não permite uma visualização distribuída de dados espaciais, sobrecarregando o usuário com a busca de *softwares* para essa finalidade (BAIG *et al.*, 2018; CASTRO, 2019).

2.4.2.10 *Elcano*

O *Elcano* é um novo sistema projetado para o gerenciamento eficiente e interoperável de *Big Data Geoespacial*, mas somente para dados vetoriais. O sistema também permite carregar dados espaciais de maneira simples e genérica, para facilitar a alimentação de dados e se estender para outros formatos; o *Elcano* também visa garantir a persistência dos dados na memória, de maneira compacta, colaborando para um processamento mais rápido do componente geoespacial (ENGÉLINUS; BADARD, 2018). O *Elcano* é desenvolvido no *Apache Spark*, utiliza linguagem de programação

Java e biblioteca JTS, e oferece suporte para todos os tipos de dados espaciais (simples e complexos) (CASTRO, 2019).

2.4.2.11 *Manifold*

De acordo com *Manifold* (2020), o *Manifold Release 9* é um novo SIG mais ágil, oferece recursos de dados superiores e lida com dados maiores e com melhor qualidade do que o ESRI - *Environmental Systems Research Institute*. Suporta dados em tabelas, geometria vetorial, dados rasterizados, desenhos, mapas e imagens. Também comanda diversas fontes ao mesmo tempo para combinar, preparar, extrair, transformar, carregar, analisar, validar, visualizar e explorar seus dados.

O *software* completo é pago, mas sua versão de somente leitura é disponibilizada gratuitamente: o *Manifold Viewer*. Embora o *Viewer* não possa gravar projetos ou salvar dados editados de volta às fontes de dados originais, ele permite a exibição e análise de quase todos os tipos diferentes de dados (MANIFOLD, 2020). Os dados deste estudo foram adicionados no *Manifold Viewer* e identificou-se como pontos positivos do sistema uma implementação simples e velocidade rápida, e se confirmou que seu ponto negativo é a impossibilidade de salvar projetos e dados.

2.4.2.12 *Google Earth Engine* (GEE)

Segundo Gorelick *et al.* (2017, p.18):

O *Google Earth Engine* é uma plataforma baseada em nuvem para análise geoespacial em escala planetária que traz os enormes recursos computacionais do *Google* para lidar com uma variedade de questões sociais de alto impacto, incluindo desmatamento, seca, desastre, doença, segurança alimentar, gestão de água, monitoramento do clima e proteção ambiental.

O GEE é acessado e controlado por meio de uma API (interface de programação de aplicativos) acessível pela internet e um ambiente de desenvolvimento interativo (IDE) baseado na web associado, possibilitando uma rápida criação de protótipos e visualização dos resultados. Seu diferencial é oferecer uma plataforma integrada e projetada para capacitar não apenas os cientistas tradicionais de sensoriamento remoto, mas também um público muito mais amplo que carece da capacidade técnica necessária para utilizar supercomputadores tradicionais ou recursos de computação em nuvem (GORELICK *et al.*, 2017).

Quando um algoritmo é desenvolvido no GEE, seus usuários podem produzir produtos de dados sistemáticos ou implantar aplicativos interativos apoiados pelos recursos do *Earth Engine*, sem a necessidade de ser um especialista em desenvolvimento de aplicativos, programação da Web ou HTML. O catálogo de dados abriga um grande repositório de conjuntos de dados geoespaciais publicamente disponíveis, composto em sua maior parte por imagens de sensoriamento remoto que observam a Terra, incluindo todo o arquivo *Landsat* e arquivos completos de dados do Sentinel-1 e Sentinel-2, além de incluir previsões climáticas, dados de cobertura do solo e muitos outros conjuntos de dados ambientais, geofísicos e socioeconômicos (GORELICK *et al.*, 2017).

2.5 Indicadores sociais

Na literatura há vários estudos sobre indicadores sociais, de qualidade de vida, de vulnerabilidade social, de vulnerabilidade à ocorrência de desastres naturais, de suscetibilidade a escorregamentos e inundações, entre outras variáveis. A seguir, esta pesquisa contextualizará alguns índices que deram uma orientação para os métodos desta Dissertação.

2.5.1 Índices de qualidade de vida

Os estudos sobre qualidade de vida em área urbana têm sido cada vez mais frequentes, principalmente aqueles relacionados aos mapeamentos utilizando dados do Instituto Brasileiro de Geografia e Estatística – IBGE, dados de prefeituras, além de imagens de satélites. As pesquisas sobre a qualidade de vida abordam diversos fatores (Quadro 1); dificilmente um índice de qualidade de vida analisaria todas estas variáveis (MIRANDA; MORATO; KAWAKUBO, 2012).

Nahas (2002) apresenta a ideia de um Índice de Qualidade de Vida Urbana – IQVU, composto por dez dimensões: abastecimento de alimentos, cultura, educação, esportes, habitação, infraestrutura urbana, meio ambiente, saúde, serviços urbanos e segurança urbana. Desse modo, o IQVU procura quantificar a desigualdade espacial de serviços que estão disponíveis e acessíveis à população, podendo ser utilizados para identificar áreas que necessitam de um maior investimento público (SMARZARO; LIMA; DAVIS JR., 2017).

Miranda, Morato e Kawakubo (2012) mapearam a qualidade de vida urbana no município de Pouso Alegre, em Minas Gerais. Para calcular o IQVU, eles combinaram

três índices: um índice de qualidade ambiental (que consistiu nos índices de abastecimento de água, de esgotamento sanitário, de coleta de lixo e de vegetação), um índice socioeconômico (que considerou renda familiar per capita estimada e pessoas por banheiro) e outro índice de educação (utilizando o número de analfabetos com mais de 10 anos, analfabetos que eram responsáveis por domicílios, anos de estudo dos responsáveis por domicílios, e responsáveis por domicílios com menos de 4 anos de estudo).

Quadro 1 - Variáveis da qualidade de vida urbana.

<i>Espaciais</i>	<i>Biológicos</i>	<i>Sociais</i>	<i>Econômicos</i>
Bem-estar vegetação, espaços abertos, tranquilidade	Saúde Física saneamento, insolação, níveis de ruído, qualidade do ar	Organização comunitária, de classe, associações	Oportunidades emprego, trabalho, negócios
Acessibilidade sistema viário, transporte	Saúde Mental estresse, congestionamentos, filas, solidão, reclamações	Realização Pessoal amizade, afeto, reconhecimento	Produtividade economia e deseconomias de aglomeração; trânsito, custo de vida; competição, complementaridade
Desenho Urbano elementos visuais, monotonia, desordem, informação referências orientação, história	Segurança trânsito, edificações, marginalidade	Contatos encontros, privacidade, solidariedade	Diversidade escolhas
Uso e ocupação do solo densidades, conflito de usos, facilidades, permeabilidade, segregação		Atividades de Lazer recreação, cultura, compras	
		Realização Profissional mobilidade, oportunidades	
		Acesso e Opções moradia, trabalho, serviços urbanos, serviços sociais, transporte	

Modificado de: RIBEIRO; VARGAS, 2001.

Um dos objetivos de Fagundes (2017) era aferir um índice de qualidade de vida para o município de Sorocaba, em São Paulo. Os principais indicadores (aqueles referentes às necessidades básicas) utilizados para compor esse índice são: educação (associada à alfabetização dos moradores e do chefe familiar), habitação/entorno (iluminação pública, calçada, arborização, sem lixo acumulado, sem esgoto à céu aberto e pavimentação), renda (renda per capita maior que um salário-mínimo) e saneamento (água encanada, rede de esgoto, coleta de lixo e banheiro).

Smarzaro, Lima e Davis Jr. (2017) relacionaram o índice de qualidade de vida urbana do município de Belo Horizonte, em Minas Gerais, com os dados do Censo 2010 do IBGE. Os indicadores utilizados foram: equipamentos de abastecimento de alimentos

(hiper e supermercado, mercearia e similares), comércio e serviços culturais (equipamento cultural (por exemplo, teatro), livraria e papelaria, locadora de filmes e estande de revistas), educação (alunos da educação infantil, da escola fundamental e da escola secundária), esportes (espaços públicos para recreação), habitação (qualidade de moradia), infraestrutura urbana (segurança residencial, saúde ambiental, fornecimento de energia elétrica, pavimentação e transporte coletivo), ambiente (conforto acústico, qualidade do ar e áreas verdes), saúde (centros de saúde e clínicas de dentista), serviços urbanos (vigilância sanitária, agência bancária, posto de gasolina, drogaria, correios, espaço público de inclusão digital e telefones públicos) e segurança urbana (segurança pessoal, segurança patrimonial e segurança no trânsito). O IQVU foi a base para a seleção dos indicadores utilizados nesta pesquisa.

2.5.2 Índices de vulnerabilidade social

Além do IQVU, há também o Índice de Vulnerabilidade Social (IVS). Enquanto o IQVU apresenta uma visão global da população e delimita políticas públicas, o IVS mensura a relação entre dados populacionais pela vulnerabilidade social nas dimensões renda, escolaridade e ciclo de vida familiar (FUZARO; CARNIELLO, 2020). Por exemplo, o estado de São Paulo possui o Índice Paulista de Vulnerabilidade Social (IPVS), que se baseia em dois pressupostos (SEADE, 2013):

- I. O IPVS considera o conceito de vulnerabilidade social proposto por Katzman (1999) de que “a vulnerabilidade de um indivíduo, família ou grupo social refere-se a sua maior ou menor capacidade de controlar as forças que afetam seu bem-estar”, ou seja, a vulnerabilidade à pobreza não leva em consideração somente a privação de renda, mas também a composição familiar, as condições e o acesso aos serviços de saúde, o acesso e a qualidade do sistema educacional, a possibilidade de obter trabalho com qualidade e remuneração adequadas, a existência de garantias legais e políticas, entre outros fatores;
- II. O IPVS considera que a segregação espacial é um fenômeno presente nos centros urbanos paulistas e que contribui decisivamente para a permanência dos padrões de desigualdade social, ou seja, a segregação espacial é um forte condicionante da própria condição de pobreza e, em sua forma extrema, a marginalização residencial cria “guetos” de famílias pobres ou que apresentam características que as tornam vulneráveis à pobreza enquanto que também produz as áreas que concentram as parcelas da população com altíssimos níveis de riqueza.

A elaboração do IPVS é baseada no conjunto de informações existentes no banco de dados do Censo Demográfico (para esse estudo considerou-se o Censo 2010), que consiste nas informações socioeconômicas (renda domiciliar per capita, rendimento médio da mulher, responsável pelo domicílio, porcentagem de domicílios com renda, domiciliar per capita até 1/2 salário-mínimo, porcentagem de domicílios com renda, domiciliar per capita até 1/4 salário-mínimo, porcentagem de pessoas responsáveis pelo domicílio alfabetizadas) e nas informações demográficas (porcentagem de pessoas responsáveis de 10 a 29 anos, porcentagem de mulheres responsáveis de 10 a 29 anos, idade média das pessoas responsáveis, porcentagem de crianças de 0 a 5 anos de idade) da população brasileira. Os setores censitários com no mínimo 50 domicílios particulares permanentes são classificados em sete grupos (SEADE, 2013):

- I. Grupo 1 – baixíssima vulnerabilidade;
- II. Grupo 2 – vulnerabilidade muito baixa;
- III. Grupo 3 – vulnerabilidade baixa;
- IV. Grupo 4 – vulnerabilidade média;
- V. Grupo 5 – vulnerabilidade alta;
- VI. Grupo 6 – vulnerabilidade muito alta (esse grupo engloba apenas os setores censitários classificados no Censo Demográfico como aglomerados subnormais com concentração de população jovem e de baixa renda);
- VII. Grupo 7 – setores censitários rurais de alta ou muito alta vulnerabilidade (refere-se aos setores censitários rurais de baixa renda e com pelo menos 50 domicílios particulares).

A vulnerabilidade social refere-se aos fatores socioeconômicos e demográficos que afetam a resiliência das comunidades. Estudos têm mostrado que em eventos de desastre, os socialmente vulneráveis são mais propensos a serem adversamente afetados, ou seja, são menos propensos a se recuperar e mais propensos a morrer. Rasch (2016) afirma que com o aumento dos eventos climáticos extremos e com a conhecida desproporcionalidade de risco das populações vulneráveis, tem aumentado muito a demanda por índices de vulnerabilidade compostos. Apresenta um índice de vulnerabilidade a enchentes incorporando dados socioeconômicos, do ambiente construído e ambientais para medir a vulnerabilidade de 1.276 municípios no Brasil. Encontra que 83 municípios (22% da população brasileira) são de risco e alerta para uso desse tipo de índices em áreas com extrema desigualdade.

A mesma autora, Rasch (2017), argumenta que desigualdade de renda (“*income inequality*”) pode também indicar altos níveis de vulnerabilidade de áreas municipais urbanas, em relação às mudanças climáticas. Usando análise fatorial (“*factor analysis*”, do qual a análise de componentes principais utilizada nesta Dissertação é uma das maneiras de calcular) e modelos de regressão multinível testa se esse tipo de desigualdade é um preditor de vulnerabilidade a enchentes (“*flood hazards*”). Seus resultados mostraram que, no caso de áreas urbanas do Brasil, desigualdade de renda é um bom preditor. Conclui que provê evidências empíricas de que desigualdade de renda, segregação espacial e acesso desigual a recursos públicos têm consequências sociais negativas.

Flanagan *et al.* (2011) descreveram o desenvolvimento de um índice de vulnerabilidade social (SVI) para uso na gestão de desastres e examinaram seu valor potencial, explorando o impacto do furacão Katrina em populações locais para ilustração. Os domínios que formam a base do SVI eram: status socioeconômico, composição familiar e deficiência, status de minoria e idioma, habitação e transporte.

O trabalho de Cutter, Boruff e Shirley (2003) é o mais citado quando se trata de índices de vulnerabilidade social a eventos ambientais. Usaram dados socioeconômicos e demográficos para construir um índice de vulnerabilidade social, o *Social Vulnerability Index* (SoVI) para os Estados Unidos, baseados em dados de 1990. Usando a técnica de componentes principais, reduziram a 11 (explicando 76,4% da variabilidade existente entre os estados) o total inicial de 42 variáveis. A técnica é a mesma empregada nesta Dissertação, porém, aqui, desde o início foi utilizado um número pequeno de variáveis e a componente principal foi utilizada para gerar os ponderadores dessa variáveis na composição dos índices finais.

Sung e Liaw (2020) aplicaram uma abordagem baseada em SIG com o índice de vulnerabilidade social (SoVI) para investigar e quantificar a vulnerabilidade social aos riscos ambientais no Condado de Yilan, em Taiwan. O SoVI pode ser aplicado em vários tipos de escala e seleciona diferentes tipos de variáveis socioeconômicas, como idade, densidade populacional, gênero, etnia, suporte financeiro e educação.

Martines *et al.* (2020) correlacionaram um índice de vulnerabilidade socioeconômica com os padrões de uso e ocupação do solo, para investigar a segregação espacial em uma planície de inundação urbana da cidade de Sorocaba, em São Paulo.

Duas dimensões analíticas foram consideradas: socioeconômica (com as variáveis de renda domiciliar per capita, renda familiar média das mulheres, percentual de domicílios com renda per capita até 1/2 salário mínimo, percentual de domicílios com renda per capita até 1/4 salários mínimos, percentual de pessoas alfabetizadas responsáveis pelo domicílio) e demográfica (considera os percentuais de pessoas de 10 a 29 anos e mulheres de 10 a 29 anos, além da idade média dos responsáveis pelo domicílio e o percentual de crianças, de 0 a 5 anos).

Nascimento Júnior e Sant'Anna Neto (2020) apresentam uma aplicação do SoVI, utilizando dados do censo demográfico por setores censitários urbanos da cidade de Santos, na Zona Costeira Paulista, com o interesse de contribuir aos estudos geográficos da vulnerabilidade social à desastres naturais em áreas urbanas. Segundo os autores, “os resultados, indicam que em Santos, a espacialização dos índices apresenta-se como um subproduto de eventos históricos e contextos socioespaciais, que explicam como os processos de urbanização (precária e segregada) incorporaram diferentes níveis de vulnerabilidade social, com privações de direitos sociais, à cidade e à justiça social” (NASCIMENTO JÚNIOR; SANT'ANNA NETO, 2020).

Tapsell *et al.* (2002) propuseram um índice de vulnerabilidade social de inundação (SFVI), que pode ser usado para prever áreas e populações que são suscetíveis de serem mais gravemente afetados em termos de saúde e outros fatores "intangíveis" dos impactos de inundações. O SFVI é um índice aditivo composto baseado em três características sociais (pais solteiros, pessoas com mais de 75 anos de idade e doentes a longo prazo) e quatro indicadores de privação financeira (desempregados, domicílios com superlotação, domicílios sem carro e famílias que não possuem casa própria).

Barreca, Curto e Rolando (2017), investigam a ligação entre vulnerabilidade social e vulnerabilidade territorial que, no caso desta Dissertação, é representada pela possibilidade de ocorrências de escorregamentos e inundações no entorno do trecho da ferrovia considerada. Os resultados mostraram que esses dois indicadores são espacialmente correlacionados.

Não foram identificados na literatura referente aos temas tratados, o uso da técnica exploratória de *boxplot* para a formação de índices na forma empregada nesta Dissertação. Seu uso atualmente está disseminado em várias áreas de aplicação e, recentemente, tem sido utilizado largamente em análises que usam técnicas de inteligência artificial. Uma

descrição curta do método pode ser encontrada em Morettin e Singer (2019). Os passageiros da CPTM e os moradores da região entorno da Linha 7 – Rubi são uma população vulnerável às ocorrências de desastres naturais e, portanto, foram fundamentais para a escolha dos indicadores usados neste estudo.

2.5.3 Indicadores utilizados nesta pesquisa

Uma vez que um dos objetivos é caracterizar, a partir de bases de dados que podem ser exploradas em ambiente *Big Data*, a população da área de estudo e como é a sua distribuição e relação com o entorno da linha férrea, tomando os conceitos supra mencionados e aqueles referenciados à seguir, a caracterização socioeconômica da população, no entorno do trecho de linha férrea considerado, será feito por meio da proposição do Índice de Interferência Social Externa à Ferrovia e do Índice de Vulnerabilidade Social da População aos Eventos de Desastres Naturais, foram estudados e incorporados nas análises, os seguintes índices:

2.5.3.1 Indicador de Lixo

Para elaborar um índice de habitação do município de Sorocaba, Fagundes (2017) procurou calcular uma variável de “sem lixo acumulado”, considerando os moradores em domicílios particulares permanentes com moradia adequada, onde não existia lixo acumulado, dividido pelo total de moradores. Para esta pesquisa, era relevante conhecer quais setores censitários possuíam lixo acumulado nos logradouros e averiguar se estes poderiam estar relacionados a ocorrências de inundações, além de doenças transmissíveis por água contaminada; portanto, analisou-se os setores censitários em que os domicílios particulares possuíam lixo acumulado em seu entorno.

2.5.3.2 Indicador de Esgoto

Para calcular um índice de habitação do município de Sorocaba, Fagundes (2017) criou uma variável de “sem esgoto à céu aberto”, considerando os moradores em domicílios particulares permanentes com moradia adequada, onde não existia esgoto à céu aberto, dividido pelo total de moradores. Para este estudo, era interessante verificar quais setores apresentavam esgoto à céu aberto e confirmar se estes estariam associados a uma provável inundação insalubre, além da capacidade de deixar o terreno mais úmido, favorecendo escorregamentos em taludes; sendo assim, averiguou-se os setores em que os domicílios apresentavam esgoto à céu aberto em seu entorno.

2.5.3.3 Indicador de Abastecimento de Água

A fim de obter um indicador de saneamento para o município de Sorocaba, Fagundes (2017) calculou uma variável de “água encanada”, considerando os moradores em domicílios particulares permanentes com abastecimento de água da rede geral, dividindo pelo total de moradores. Para este estudo, interessava saber quais setores não tinham acesso à um abastecimento de água da rede geral, pois esses locais poderiam estar relacionados à criação de poços e fossas, que podem estar relacionadas a escorregamentos, principalmente em taludes, e potenciais inundações insalubres.

2.5.3.4 Indicador de Renda

De acordo com o inciso IV, do artigo 7º, da Constituição Federal de 1988, está entre os direitos dos trabalhadores um salário-mínimo, capaz de atender suas necessidades vitais básicas, e de suas famílias, como moradia, alimentação, educação, saúde, lazer, vestuário, higiene, transporte e previdência social (BRASIL, 1988). Em sua pesquisa, Fagundes (2017) calculou o rendimento *per capita* maior que 1 salário-mínimo (S.M.), como uma das variáveis para aferir a qualidade de vida do município de Sorocaba, SP. Para isso, calculou os domicílios particulares com rendimento nominal mensal domiciliar *per capita* de mais de 1 S. M. e dividiu pelo número de domicílios particulares permanentes.

Por outro lado, Sung e Liaw (2020), para calcular um índice de vulnerabilidade social aos riscos ambientais em Taiwan, utilizaram como uma das variáveis a falta de recursos financeiros (famílias de baixa renda), que pode estar relacionada a informação insuficiente, mobilidade restrita e capacidade de preparação inadequada. Para esta pesquisa, interessava saber quais setores da área de estudo tinham uma menor renda salarial e verificar se estes estariam mais sujeitos à ocorrência de desastres naturais, e se estes também poderiam contribuir para a ocorrência de escorregamentos e inundações (por exemplo, se a população de menor renda estivesse associada ao acúmulo de lixo entorno da ferrovia); portanto, considerou-se os setores censitários em que os domicílios particulares permanentes recebiam, no máximo, até um salário-mínimo.

2.5.3.5 Indicador de Idosos

Sung e Liaw (2020) combinaram 12 variáveis para calcular um índice de vulnerabilidade social aos riscos ambientais em Taiwan; uma dessas variáveis era um índice de idosos. A idade pode representar a capacidade física de movimento ou mobilidade; assim, residentes idosos têm a possibilidade de aumentar a dificuldade de

evacuação em caso de desastre ambiental. Portanto, os autores consideraram como variável a população com no mínimo 65 anos de idade (SUNG; LIAW, 2020).

Para este estudo, também era interessante conhecer a proporção de idosos em cada setor censitário da área de estudo, pois quanto mais idosos, maior seria a vulnerabilidade dessa população quando ocorresse um evento de desastre natural. Desse modo, verificou-se o número de pessoas residentes, em domicílios particulares e domicílios coletivos, a partir de 65 anos.

2.5.3.6 Indicador de Crianças

Entre as 12 variáveis que Sung e Liaw (2020) combinaram para calcular um índice de vulnerabilidade social aos riscos ambientais, uma dessas era um índice de crianças. Utilizando a mesma lógica da variável de idosos, as crianças também teriam maiores dificuldades de evacuação caso ocorresse algum evento de desastre natural. Dessa forma, os autores usaram como variável a população com menos de 5 anos de idade (SUNG; LIAW, 2020).

Para esta pesquisa, também era viável descobrir a quantidade de crianças na área de estudo, pois um maior número de crianças estaria associado a uma maior vulnerabilidade social aos eventos de escorregamentos e inundações. Assim, analisou-se o número de pessoas residentes, em domicílios particulares e domicílios coletivos, menores de 5 anos de idade.

2.5.3.7 Indicador de Alfabetização

A situação educacional também desempenha um papel significativo: níveis de educação mais baixos podem significar a informação potencialmente insuficiente, falta de mobilidade e capacidade de preparação inadequada. Dessa maneira, Sung e Liaw (2020) usaram como variável a população sem ensino médio para verificar a vulnerabilidade social aos riscos ambientais. Em contrapartida, para obter um indicador de educação para o município de Sorocaba, para verificar a qualidade de vida urbana da região, Fagundes (2017) calculou o número de pessoas alfabetizadas com 5 anos de idade ou mais em relação ao número de pessoas do setor censitário com 5 anos de idade ou mais.

Para este estudo era necessário conhecer quais locais apresentavam um menor nível de educação para conhecer as áreas com a população mais vulnerável à ocorrência

de desastres naturais. Para tal finalidade, foi calculado o número de pessoas não alfabetizadas de cada setor censitário.

2.5.3.8 Indicador de Residentes Sem Moradia Própria

Uma das variáveis consideradas por Tapsell *et al.* (2002), para calcular um índice de vulnerabilidade social à inundação era a propriedade não residencial, que se referia às famílias que não possuíam casa própria. Essa variável indica um grupo de pessoas que é carente financeiramente e, portanto, contribui para vulnerabilidade aos impactos das inundações (TAPSELL *et al.*, 2020). Baseando-se nessa lógica, esta pesquisa usou os dados dos domicílios que não são próprios, fornecidos pelo Censo 2010 (IBGE), considerando que os moradores desses domicílios seriam mais vulneráveis aos eventos de desastres naturais.

3 MATERIAIS E MÉTODOS

O fluxograma a seguir apresenta as principais etapas do desenvolvimento da presente pesquisa:

Figura 3 - Fluxograma do método da pesquisa.

Fonte: elaboração própria, 2020.

3.1 Análise bibliométrica

Realizou-se uma análise bibliométrica para determinar a frequência dos seguintes termos em publicações: *Big Data*, aprendizado de máquina, desastres naturais e ferrovias. A revisão sistemática seguiu as diretrizes do PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) de acordo com o "Diagrama de fluxo PRISMA" (Figura 4). A declaração PRISMA é um método desenvolvido para revisões sistemáticas e meta-análises de avaliações de intervenções de assistência à saúde (LIBERATI *et al.*, 2009).

Embora a declaração PRISMA tenha sido criada originalmente para revisões de intervenções de saúde, estudos de diferentes áreas também utilizaram esse método para desenvolver revisões sistemáticas. Ao pesquisar o termo "Declaração Prisma" no campo "Article Title, Abstract, Keywords" na base de dados Scopus, o filtro "by subject area" mostra que a maioria dos estudos que utilizaram esse método para revisões sistemáticas pertencem à área da saúde (2.506), mas também podem estar relacionados a outras áreas, como ciências agrárias e biológicas (86), ciências sociais (62), multidisciplinares (60), engenharia (50), ciências ambientais (33), ciências da computação (20), ciência dos materiais (11), artes e humanidades (8), gestão e contabilidade empresarial (8), ciências da decisão (6), economia, econometria e finanças (6), matemática (6), química (4), física e astronomia (3), energia (2) e ciências da terra e planetárias (1) (CORREIA; CORSI; QUINTANILHA, 2020).

A instrução PRISMA consiste em um diagrama de fluxo e um *checklist* de 27 itens. Esse processo de revisão consiste nos quatro estágios do diagrama de fluxo PRISMA: Identificação (onde mostra o número de registros identificados através da pesquisa no banco de dados e também o número de registros adicionais identificados através de outras fontes), Triagem (onde os registros duplicados são removidos e identificam o número de registros rastreados e, ao mesmo tempo, o número de registros excluídos), Elegibilidade (aponta o número de artigos em texto completo avaliados para elegibilidade e, consequentemente, o número de artigos em texto completo excluídos, com motivos) e, finalmente, Inclusão (apresenta o número de estudos incluídos na síntese qualitativa e, posteriormente, o número de estudos na síntese quantitativa) (LIBERATI *et al.*, 2009).

Figura 4 - Fluxograma PRISMA.

Modificado de: Liberati *et al.*, 2009.

A análise para este estudo foi feita através de pesquisas de referências nas seguintes bases de dados: Scopus, *Web of Science*, Scielo e *Google Scholar*. Escolheu-se palavras-chave que se referem ao tema do estudo (*big data*, desastres naturais, deslizamentos/escorregamentos e inundações), técnicas que podem ser usadas para a análise do *Big Data* (aprendizado de máquina, análise espacial, sensoriamento remoto e SIG) e termos colaborativos da área de estudo (ferrovia, via férrea). A busca no banco de dados Scopus utilizou os campos “*Article Title, Abstract, Keywords*”, enquanto com o *Web of Science*, Scielo e *Google Scholar*, apenas o campo “Título” foi utilizado (CORREIA; CORSI; QUINTANILHA, 2020).

A estratégia de busca utilizou um formulário estruturado com os termos "*natural disaster**", "*landslide**" ou "*flood**" no campo "*Title Words*", juntos (ou não) aos termos secundários "*big data*", "*machine learning*", "*spatial analysis*", "*remote sensing*", "*GIS*", "*railway*", "*railroad*" e os operadores booleanos "*AND*" e "*OR*". Na primeira etapa (Identificação), foi obtido o número de registros identificados na pesquisa, além do número de registros adicionais identificados por outras fontes. Depois disso, removeu-se os registros duplicados. Na segunda etapa (Triagem), selecionou-se os registros publicados entre 2008 e 2019, considerando os idiomas português, inglês e espanhol. Os estudos que não atendiam a esses critérios foram excluídos. A terceira etapa (Elegibilidade) foi construída sobre os artigos com textos completos, avaliados por sua elegibilidade. Os critérios de elegibilidade adotados basearam-se na relevância do artigo: primeiro, selecionamos os 20 artigos mais relevantes e, em seguida, selecionamos aqueles que apresentavam resumos, objetivos e conclusões semelhantes ao tema do estudo, que foram consequentemente úteis para a pesquisa (CORREIA; CORSI; QUINTANILHA, 2020).

Os artigos foram classificados de acordo com a relevância e colocados em ordem decrescente do mais para o menos relevante. A opção de classificação de relevância ordena os resultados em ordem decrescente com base em um sistema de classificação que considera quantos dos termos de pesquisa são encontrados em cada registro, permitindo que os estudos com a classificação mais alta apareçam no topo da lista (CLARIVATE ANALYTICS, 2019). A quarta etapa (Inclusão) considerou primeiro o número de estudos incluídos na síntese qualitativa e depois o número de estudos na síntese quantitativa. Os estudos foram incluídos quando uma descrição de técnicas foi encontrada e aplicada à gestão de riscos ou aos estudos de inundações e escorregamento/deslizamentos (CORREIA; CORSI; QUINTANILHA, 2020).

Para a análise quantitativa, o número de registros utilizados referente ao tema deste estudo na base de dados Scopus, foi datado entre 2008 e 2019, e estava em todas as línguas. No campo de pesquisa "*Article Title, Abstract, Keywords*", utilizou-se os termos "*Natural Disasters AND Big Data*", além de "*Natural Disaster AND Machine Learning*", "*Natural Disasters AND Remote Sensing*", "*Natural Disasters AND GIS*", "*Natural Disasters AND Railway*" e "*Natural Disasters AND Railroad*". Também se pesquisou na base de dados Scopus, quais países tiveram publicações sobre o tema de estudo nos últimos anos. O campo "*Article Title, Abstract, Keywords*" foi utilizado novamente com

os termos “*Natural disasters*”, “*Smart cities*” e “*Big data*” através do operador “AND”. O filtro “*Country/Territory*” foi aplicado, que indica os países com mais publicações e considera a localização de todos os autores de cada estudo.

3.2 Área de estudo

3.2.1 Delimitação da área de estudo

O local de estudo escolhido é o entorno da Estação Caieiras a Estação Francisco Morato da Linha 7 - Rubi da CPTM, na escala 1:25.000. Essa região foi escolhida como área de estudo devido à ocorrência de inundações e escorregamentos na área, que muitas vezes afetam a circulação dos trens da CPTM, e em função de vários projetos que o IPT tem no local, possibilitando o trabalho de maneira integrada.

O município de Caieiras encontra-se no estado de São Paulo e localiza-se em torno de 35 km de sua capital (IPT, 2005). O município possui aproximadamente 97,6 km² de área e tem cerca de 94.516 habitantes (PREFEITURA MUNICIPAL DE CAIEIRAS, 2018). Já o município de Franco da Rocha está a 27 km do centro da capital de São Paulo, com território de 132,775 km² (PREFEITURA MUNICIPAL DE FRANCO DA ROCHA, 2017) e possui uma população estimada de 152.433 habitantes (IBGE, 2018). Quanto à Francisco Morato, o município está localizado a cerca de 30,5 km da cidade de São Paulo, possui uma área de 49,2 km², resultando em uma densidade demográfica de 3.470 hab/km² (CÂMARA MUNICIPAL DE FRANCISCO MORATO, 2020).

Para definir o entorno da área de estudo, utilizou-se os dados de suscetibilidade de escorregamentos e de inundação dos municípios de Caieiras, Franco da Rocha e Francisco Morato, na escala de 1:25.000, do IPT/CPRM. As classes de alta suscetibilidade foram associadas a *buffers* – que são mapas de distância; polígonos paralelos ao redor de elementos (DAVIS; CÂMARA, 2001) - calculados a cada 100 m até 1800 m entre as estações Caieiras e Francisco Morato. O *buffer* inicial de 100 m foi escolhido baseado no estudo de Braghirolli (2017), que gerou um *buffer* de 100 m para sua área de estudo para avaliar e propor uma metodologia de mapeamento voltado às áreas de risco a inundação; e realizou-se *buffers* a cada 100 m até 1800 m, pois foi em 1800 m que se percebeu que a proporção de áreas com classes altas de inundação e de escorregamentos se tornaram constantes. Analisou-se a proporção de áreas com classe alta à suscetibilidade de escorregamentos ou de inundações em relação a área de

determinado *buffer*, considerando que as classes de alta suscetibilidade apresentam uma maior probabilidade de ocorrência destes desastres naturais.

Notou-se que, tanto para os escorregamentos quanto para as inundações, a proporção de suscetibilidade começa a se tornar constante a partir do *buffer* de 1800 m da ferrovia. Para os escorregamentos, percebeu-se que a maior taxa de proporção de suscetibilidade de escorregamentos está presente no *buffer* de 500 m (0,0823 de proporção), enquanto notou-se que a maior taxa de proporção de suscetibilidade de inundações encontra-se há apenas 100 m da ferrovia (0,149) (Gráfico 1). Nessa mesma análise, foi identificado que quanto mais próximo da ferrovia, maior era a suscetibilidade natural da região a inundações. Sendo assim, o *buffer* de 500 m foi considerado como o “entorno” da área de estudo, já que ele engloba também o *buffer* de 100 m (Figura 5).

Gráfico 1 - Proporção de Suscetibilidade de Escorregamentos e Inundações entorno da Linha 7 - Rubi.

Fonte: Elaboração própria, 2020.

Figura 5 - Mapa da Área de Estudo.

3.2.2 Ocorrências de desastres naturais na área de estudo

Os escorregamentos e as inundações são causados por condicionantes naturais e antrópicos. Para as inundações, entre os condicionantes naturais destacam-se as formas do relevo; características da rede de drenagem; da bacia hidrográfica; intensidade, quantidade, distribuição e frequência das chuvas (aquelas intensas ou de longa duração saturam os solos, aumentando o escoamento superficial e a concentração de água no local); características do solo e o teor de umidade; e a presença ou ausência da cobertura vegetal (a vegetação colabora para a retenção de água no solo, diminuindo a velocidade do escoamento superficial e, conseqüentemente, a erosão) (DO AMARAL; RIBEIRO, 2009). Quanto aos escorregamentos, os condicionantes naturais dividem-se em agentes predisponentes, “quando determinam as condições iniciais que favorecem ou dificultam o rompimento e a movimentação dos materiais das encostas” (MINISTÉRIO DAS CIDADES/IPT, p. 79, 2007), e em agentes efetivos, que são diretamente responsáveis pelo desencadeamento dos escorregamentos, e que são diferenciados em “preparatórios (pluviosidade, erosão pela água e vento, congelamento e degelo, variação de temperatura e umidade, dissolução química, ação de fontes e mananciais, oscilação do nível dos lagos e marés e do lençol freático, ação de animais e humana, inclusive desflorestamento) e imediatos (chuva intensa, vibrações, fusão do gelo e neves, erosão, terremotos, ondas, vento, ação do homem etc.)” (MINISTÉRIO DAS CIDADES/IPT, p. 41, 2007).

Em contrapartida, os fatores antrópicos das inundações são o uso e a ocupação irregular nas planícies e margens de cursos d’água, a disposição irregular de lixo nas proximidades dos corpos hídricos, alterações nas características da bacia hidrográfica e dos cursos d’água – como impermeabilização do solo, retificação e canalização dos cursos d’água etc. – e intenso processo de erosão dos solos e de assoreamento dos cursos d’água (DO AMARAL; RIBEIRO, 2009). Enquanto que os condicionantes antrópicos dos escorregamentos têm como principais elementos causadores de escorregamentos: desmatamento, lançamento e concentração de águas fluviais ou servidas, vazamento na rede de água ou esgoto, execução de cortes com alturas e inclinações com limites superiores aqueles tecnicamente seguros, execução deficiente de aterros, lançamento de resíduos sólidos nas encostas e taludes, retirada do solo superficial expondo horizontes mais suscetíveis, entre outros fatores (MINISTÉRIO DAS CIDADES/IPT, 2007).

A Linha 7 – Rubi é aquela, de toda a malha ferroviária operada pela CPTM, que apresenta o maior número de ocorrências associadas a processos geodinâmicos, como

inundação, enchente e movimentos de massa. 54 ocorrências associadas a inundações com paralisação total e parciais da circulação dos trens, ou com restrições de velocidade, foram registradas entre 2009 e 2019 (Gráfico 2) (AMORIM *et al.*, 2019).

Gráfico 2 - Ocorrências associadas a inundações na Linha 7 – Rubi.

Modificado de: AMORIM *et al.*, 2019.

Este estudo realizou um levantamento de informações na internet sobre a ocorrência de inundações, alagamentos e escorregamentos na Linha 7 da CPTM, utilizando o Google Notícias, de 2009 a 2020 (já que a primeira ocorrência de desastre natural que afetou a linha férrea e que aparece na internet é referente ao ano de 2009). Os principais resultados são dos sites Estadão, G1 e o Diário da CPTM (Tabela 1 e Anexo A).

Tabela 1 - Levantamento de ocorrências de desastres naturais que afetaram a Linha 7 – Rubi da CPTM.

Data do Evento	Tipo de Desastre	Estações Afetadas	Dados de Chuva
08/12/2009	Inundação	Caieiras a Jundiaí	121,1 mm em Franco da Rocha
03/11/2014	Deslizamento	Franco da Rocha a Baltazar Fidélis	Não informado
22/12/2014	Inundação e alagamento	Luz a Francisco Morato	35,4 mm em Francisco Morato
05/02/2015	Inundação	Caieiras a Baltazar Fidélis	45 mm em Caieiras
11/03/2016	Inundação e alagamento	Caieiras a Jundiaí	68,9 mm em Caieiras
23/01/2017	Deslizamento	Perus a Francisco Morato	1,6 mm em Francisco Morato
01/12/2018	Inundação e alagamento	Caieiras a Francisco Morato	127 mm em Caieiras
19/12/2018	Inundação	Perus	100,2 mm em Caieiras
23/01/2019	Inundação e alagamento	Luz a Francisco Morato	14,6 mm em Francisco Morato
15/02/2019	Inundação e alagamento	Francisco Morato a Jundiaí	27,5 mm em Francisco Morato
17/03/2019	Inundação e alagamento	Francisco Morato a Jundiaí	25,8 mm em Francisco Morato
04/07/2019	Inundação e alagamento	Francisco Morato a Jundiaí	69,5 mm em Francisco Morato
03/02/2020	Alagamento	Caieiras a Francisco Morato	71,19 mm em Caieiras
20/02/2020	Alagamento e deslizamento	Francisco Morato a Jundiaí	39,95 mm em Francisco Morato

Fonte: Google Notícias/ ANA¹/ CEMADEN², 2019.

Existem relatos de inundações na Linha 7 – Rubi, principalmente nas estações Caieiras, Franco da Rocha e Francisco Morato (Figura 6), e alagamentos nessa região desde 2009, ano no qual passageiros ficaram ilhados na estação Caieiras. Nesse tipo de situação, a CPTM aciona o Plano de Apoio entre Empresas em Situação de Emergência (Paese) para garantir o transporte aos usuários, porém, muitas vezes os ônibus do Paese não conseguem circular devido aos alagamentos e, portanto, o serviço é interrompido. Até fevereiro de 2020 foram encontradas notícias relatando sobre inundações e alagamentos que fizeram os trens circularem com velocidade reduzida e maiores intervalos de parada entre as estações da Linha Rubi.

1: Dados pluviométricos disponíveis em: <<http://www.snirh.gov.br/hidroweb/serieshistoricas>>. Acesso em: 18 dez. 2019.

2: Dados pluviométricos disponíveis em: <<http://www.cemaden.gov.br/mapainterativo/>>. Acesso em: 20 dez. 2019.

Figura 6 - Exemplo de ferrovia afetada por uma inundação: Linha 7 - Rubi da CPTM inundada entre as estações Caieiras e Franco da Rocha.

Modificado de: Juliana Cardili/G1, 2011.

Essas são algumas das notícias que confirmam o transtorno que os passageiros passaram durante esse tipo de evento: no dia 11/03/2016 “na linha 7- Rubi, os passageiros passaram a madrugada ilhados dentro das estações ou dormindo dentro dos trens nas estações Francisco Morato, Caieiras, Perus e Franco da Rocha [...] A cidade de Caieiras também ficou debaixo d’água e as pessoas se deitaram no chão porque os trens deixaram de circular desde 2h40 devido alagamento na linha férrea” (G1, 2016). E no dia 01/12/2018 “os municípios de Caieiras, Franco da Rocha e Francisco Morato estão alagados nesta manhã. Franco da Rocha foi uma das mais atingidas pela chuva [...] A principal alternativa de mobilidade da cidade é a estação da CPTM, que também fica no Centro, e o passageiro tem dificuldades para acessar o local devido aos alagamentos [...] Na cidade de Caieiras pelo menos nove caminhões presos na enchente. Os motoristas estão presos no local e terão que esperar a água baixar para se locomoverem” (G1, 2018) (Figura 7).

Figura 7 - Caminhões presos em inundação em Caieiras, próximo a Linha Rubi da CPTM.

Modificado de: Reprodução TV Globo/G1, 2018.

Destaca-se também a ocorrência de escorregamentos em 2014 entre as estações Franco da Rocha e Baltazar Fidélis e em 2017 entre as estações Perus e Caieiras (Figura 8). Ambos interromperam a circulação dos trens. Além de prejudicar a viagem dos passageiros, esses eventos também podem causar danos ao transporte de carga que há nessa linha férrea.

Figura 8 - Deslizamento de terra interrompe circulação de trens na Linha 7 da CPTM.

Modificado de: Reprodução/Twitter BandNews FM, 2017.

Figura 7 disponível em: <<https://g1.globo.com/sp/sao-paulo/noticia/2018/12/01/temporal-provoca-alagamentos-em-franco-da-rocha-caieiras-e-francisco-morato.ghtml>>. Acesso em: 18 ago. 2019.

Figura 8 disponível em:
<<https://www.band.uol.com.br/m/conteudo.asp?id=100000842144&Programa=Cidades>>. Acesso em: 18 ago. 2019.

3.2.3 Descrição da área de estudo

3.2.3.1 Recursos hídricos

O principal corpo hídrico da região é o Rio Juqueri, componente do Sistema Cantareira de Abastecimento de Água que atende a Região Metropolitana de São Paulo – RMSP, pertencente à Bacia Hidrográfica do Alto Tietê – BAT. A Bacia Hidrográfica do Alto Tietê é dividida em seis sub-bacias de drenagem com uma disponibilidade hídrica superficial de 115,36 m³/s: Billings – Tamanduateí, Cabeceiras, Cotia – Guarapiranga, Penha – Pinheiros, Pinheiros – Pirapora e Juqueri – Cantareira, essa última com os Reservatórios Paiva Castro e Águas Claras, com uma área de 848,71 km² e uma disponibilidade hídrica *per capita* de 974,52 m³/hab/ano. Os municípios que constituem a BAT são: Arujá, Barueri, Biritiba Mirim, Caieiras, Cajamar, Carapicuíba, Cotia, Diadema, Embu, Embu-Guaçu, Ferraz de Vasconcelos, Francisco Morato, Franco da Rocha, Guarulhos, Itapeverica da Serra, Itapevi, Itaquaquecetuba, Jandira, Jquitiba, Mairiporã, Mauá, Mogi das Cruzes, Osasco, Pirapora do Bom Jesus, Poá, Ribeirão Pires, Rio Grande da Serra, Salesópolis, Santana do Parnaíba, Santo André, São Bernardo do Campo, São Caetano do Sul, São Lourenço da Serra, São Paulo, Suzano, Taboão da Serra (FABHAT, 2019).

O Rio Juqueri está muito assoreado e recebe grande parte do esgoto não tratado dos municípios de Caieiras, Francisco Morato e Franco da Rocha durante anos (PREFEITURA MUNICIPAL DE CAIEIRAS, 2017). Há vários cursos d'água ao redor da Linha Rubi, com destaque a esse rio. Em Francisco Morato, os rios que cruzam a cidade são os ribeirões Tapera Grande e Euzébio Matoso (CÂMARA MUNICIPAL DE FRANCISCO MORATO, 2020). A linha férrea percorre o fundo de vale dos rios Tapera Grande e Juqueri, recebendo a descarga de vários córregos afluentes, cujas bacias muito ramificadas encontram-se tomadas por assentamentos urbanos, por vezes, precários. As interferências com recursos hídricos são frequentes, ou seja, há uma alta incidência de pontos de inundações ao longo da linha. (AMORIM *et al.*, 2019).

É importante destacar que a Represa Paiva Castro pode agravar inundações e alagamentos na região ao abrir suas comportas. Em 11 de março de 2016, a Companhia de Saneamento Básico do Estado de São Paulo (Sabesp) abriu parcialmente as comportas da represa devido ao risco de transbordamento, depois de um forte temporal atingir a Grande São Paulo, liberando a água do Rio Juqueri (10 metros cúbicos por segundo), que corta os municípios de Mairiporã, Franco da Rocha e Caieiras (Figura 9) (LEITE; PINHO, 2016).

Como o volume de água que entrou no reservatório continuou aumentando por causa da chuva, a descarga de água chegou a 50 metros cúbicos por segundo. Franco da Rocha foi uma das cidades que sofreu mais prejuízos com a chuva: 210 moradias foram interditadas, 10 famílias e 18 moradores de rua foram para abrigos, e uma pessoa morreu afogada (LEITE; PINHO, 2016).

Figura 9 - Sabesp abriu comportas da represa em 11 de março de 2016.

Modificado de: LEITE; PINHO, 2016.

3.2.3.2 Cobertura e uso do solo

O município de Caieiras está inserido no bioma da Mata Atlântica e tem 16,85% de seu território com cobertura vegetal composta por Cerrado, Capoeira e Mata, enquanto 20,64% da região é formada pelas áreas de silvicultura (eucaliptos) (PREFEITURA MUNICIPAL DE CAIEIRAS, 2017). Em contrapartida, o município

de Franco da Rocha é constituído de mata (27,09%), campo (22,09%), capoeira (8,66%), vegetação de várzea (0,05%), além de áreas de reflorestamento (18,68%) (FEHIDRO/EMPLASA, 2006). Quanto a Francisco Morato, o município é constituído de capoeira (16,70%) e áreas de reflorestamento (1,34%) (SECRETARIA DO MEIO AMBIENTE DO ESTADO DE SÃO PAULO, 2020).

A cobertura do solo da área de estudo (Figura 10) contempla classes como Floresta Plantada, Formação Florestal, Infraestrutura Urbana, Mosaico de Agricultura e Pastagem, Outras Lavouras Temporárias (exceto soja e cana), Outras Áreas não Vegetadas (exceto mineração), Pastagem, além de Corpos Hídricos (PROJETO MAPBIOMAS, 2020). A via férrea está localizada sobre áreas de Infraestrutura Urbana e de Mosaicos de Agricultura e Pastagem, o que pode ser relacionado com as inundações que ocorrem frequentemente na região, pois o uso e a cobertura do solo influenciam diretamente o regime hidrológico de uma bacia hidrográfica; a urbanização e o desmatamento (por exemplo, a substituição de áreas florestais por áreas agrícolas) contribuem para a redução da infiltração de água no solo e, portanto, para o aumento do escoamento (RODRIGUES *et al.*, 2019).

É uma área com ocupação urbana dispersa, porém com zonas densamente povoadas, principalmente nas regiões centrais de Francisco Morato e Franco da Rocha. Detectou-se nessa área uma alta incidência de interferências da ferrovia com Áreas de Preservação Permanente (APPs) e fragmentos de cobertura vegetal (AMORIM *et al.*, 2019). As APPs são “espaços territoriais legalmente protegidos, ambientalmente frágeis e vulneráveis, podendo ser públicas ou privadas, urbanas ou rurais, cobertas ou não por vegetação nativa” (MMA, 2020).

Figura 10 - Mapa de Cobertura e Uso do Solo do Entorno da Linha 7 – Rubi.

Fonte: Elaboração própria, 2020.

3.2.3.3 Relevô

A área de estudo se caracteriza pelo relevo predominantemente acidentado, com presença de formas onduladas de morros e morrotes. As características geomorfológicas dessa área da ferrovia sofreram significativas intervenções nas feições naturais do terreno durante a implantação do traçado da ferrovia; portanto, é possível verificar uma sequência de cortes de terreno e taludes artificiais, e, conseqüentemente, as mais graves incidências de eventos geodinâmicos da Linha Rubi, resultando num *continuum* de áreas suscetíveis à escorregamentos, que frequentemente ocasionam danos ao sistema ferroviário, podendo provocar até paralisações dos serviços (AMORIM *et al.*, 2019).

3.2.3.4 Transporte

Os núcleos urbanos de Caieiras, Franco da Rocha e Francisco Morato encontram-se estruturados ao longo da infraestrutura ferroviária – a Linha Rubi da CPTM – que conecta essas cidades a São Paulo e seu sistema de metrô (PREFEITURA MUNICIPAL DE CAIEIRAS, 2017). Em 2010, as estações da Caieiras e Franco da Rocha da CPTM foram tombadas pelo Conselho de Defesa do Patrimônio Histórico, Arqueológico, Artístico e Turístico – Condephaat (GOVERNO DO ESTADO DE SÃO PAULO, 2010).

E uma nova Estação Francisco Morato foi inaugurada em 31 de agosto de 2020, na qual a CPTM construiu um “sistema antienchente” (constituído por diques e bombas com a finalidade de drenar a água por debaixo da terra), um sistema de contenção de água para não invadir a estação, além de outro sistema que transforma a água da chuva em um “turbilhão” que retira a água do local, descendo pelas colunas da estação, para não acumular água e evitar vazamentos. A CPTM fez essas obras com o intuito de reduzir os riscos de interrupção na circulação de trens e, conseqüentemente, aumentar a segurança dos passageiros em períodos de chuva (CPTM, 2020). O trem é o meio de deslocamento mais importante na região, uma vez que a maioria dos moradores e trabalhadores da área o utilizam diariamente.

A Companhia Paulista de Trens Metropolitanos oferece serviços de transporte público para favorecer a mobilidade urbana no estado de São Paulo (CPTM, 2018a). A CPTM foi criada em 28 de maio de 1992, pela autorização da Lei nº 7.861, para substituir a Companhia Brasileira de Trens Urbanos (CBTU) e a Ferrovia Paulista S/A (FEPASA), com o objetivo de melhorar e garantir a continuidade dos sistemas de trens da Região Metropolitana de São Paulo - RMSP (CPTM, 2018b). A CPTM possui 94 estações, com 273 km de extensão das linhas operacionais e transporta em média 2,8

milhões de passageiros por dia útil. A Companhia Paulista de Trens Metropolitanos dispõe das seguintes linhas: a linha 7 - Rubi, linha 8 - Diamante, linha 9 - Esmeralda, linha 10 - Turquesa, linha 11- Coral, linha 12 - Safira e a linha 13 - Jade (Figura 9 a seguir) (CPTM, 2018c).

A Linha 7 – Rubi, a via férrea a ser considerada nessa pesquisa, apresenta 60,1 km de extensão (AMORIM *et al.*, 2019) e consiste nas seguintes estações: Luz, Palmeiras – Barra Funda, Água Branca, Lapa, Piqueri, Pirituba, Vila Clarice, Jaraguá, Perus, Caieiras, Franco da Rocha, Baltazar Fidélis, Francisco Morato, Botujuru, Campo Limpo Paulista, Várzea Paulista e Jundiaí (CPTM, 2018c). A Linha 7 transporta cerca de 450 mil usuários por dia e se localiza em uma região caracterizada pelas deficiências em acessibilidade por modos de transporte rodoviários; desse modo, a ferrovia desempenha um papel fundamental na conexão desses municípios com a porção central da RMSP e de Jundiaí, polos que concentram os empregos da população local (AMORIM *et al.*, 2019). É importante ressaltar que a preocupação desse estudo não é apenas os escorregamentos e as inundações que afetam a via férrea, mas também como esses eventos de desastres naturais interferem nas viagens dos passageiros da CPTM e no dia a dia dos moradores e trabalhadores da região.

3.3 Levantamento de fontes/dados oficiais na área de estudo

Dados úteis foram considerados para este estudo, segundo a bibliografia consultada, como: censo demográfico de 2010, relevo, geologia, hidrografia, cobertura e uso do solo, modelo digital de superfície, divisão territorial, educação, saúde, imagens de satélite, ortofotos, mapas de suscetibilidade e de risco de escorregamentos e inundações, ocorrências de escorregamentos e inundações, sistema viário, rede urbana etc. Procurou-se levantar informações que fossem pertinentes à análise de áreas suscetíveis a escorregamentos e inundações, e também características físicas, ambientais, sociais e econômicas da área de estudo. Esse banco de dados não só visa complementar as referências de um sistema de alerta de desastres, como também apoiar os tomadores de decisão.

Buscou-se preferencialmente informações gratuitas, disponíveis para a população, porém também se teve acesso a dados do IPT, da Defesa Civil dos municípios de Caieiras e Francisco Morato e da CPTM. Os dados foram organizados no formato “.xlsx” e o levantamento foi estruturado em 10 categorias:

1. A primeira coluna é denominada “Categoria”, que se refere a informação geral do dado; por exemplo: “Relevo do município de Caieiras”, “Hidrografia do Brasil”, “Transporte do Estado de São Paulo” etc.;
2. A segunda coluna se refere a “Classe” do dado, especifica a informação; como “Curvas de nível”, “Trechos de drenagem”, “Estações da CPTM”, entre outros;
3. A terceira coluna identifica o “Nome do Arquivo”, por exemplo: “CurvaDeNivel”, “hid_trecho_drenagem_1” etc;
4. A coluna seguinte mostra o “Formato Primário” do dado: se é .csv (*Comma-separated-values*), .dbf (*dBase*), .ecw (*Enhanced Compression Wavelet*), .kml (*Keyhole Markup Language*), .gdb (*Geodatabase*), .gpkg (*GeoPackage*), .lyr, *MapServer*, .ovr, .shp (*Shapefile*), .tiff (*Tagged Image File Format*), WMS (*Web Map Service*), .xlsx (*Excel*), .jpeg (*Joint Photographics Experts Group*), ou uma página de site, entre outros formatos;
5. A quinta coluna mostra a “Feição” da informação: se é *raster*, vetor (ponto, linha e polígono) e quando não se aplica (quando não possui um formato);
6. A próxima coluna informa sobre a “Escala” do dado; por exemplo, 1:10.000, 1:250.000 etc;
7. A sétima coluna é nomeada “Ano” e nos informa qual é o ano de publicação do dado;
8. A coluna subsequente, “Fonte”, indica de onde as informações foram extraídas. As seguintes fontes foram consultadas: Agência Nacional de Águas – ANA; Agência Nacional de Energia Elétrica – ANEEL; Agência Nacional de Transportes Terrestres – ANTT; Agência Espacial Europeia – ESA; Companhia de Pesquisa de Recursos Minerais - CPRM; Centro de Estudos da Metrópole – CEM; Centro Nacional de Monitoramento e Alertas de Desastres Naturais - Cemaden; Companhia Paulista de Trens Metropolitanos – CPTM; Coordenadoria Estadual de Proteção e Defesa Civil do Estado de São Paulo (Defesa Civil de Caieiras e de Francisco Morato); Companhia Ambiental do Estado de São Paulo – Cetesb; Departamento de Estradas e Rodagem do Estado de São Paulo – DER; Departamento Nacional de Infraestrutura de Transportes – DNIT; Departamento de Águas e Energia Elétrica – DAEE; Empresa Brasileira de Pesquisa Agropecuária – Embrapa; Empresa Paulista de Planejamento Metropolitano S.A – Emplasa; Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP; Fundação para a Conservação e a Produção Florestal do Estado de São Paulo –

FF; Fundação Sistema Estadual de Análise de Dados - SEADE; Fundação SOS Mata Atlântica; *Google Earth*; *Google Earth Engine* - GEE; Instituto Brasileiro de Geografia e Estatística – IBGE; Instituto Florestal – IF; Instituto Geográfico e Cartográfico de São Paulo – IGC–SP; Instituto Geológico do Estado de São Paulo – IG-SP; Instituto Nacional de Pesquisas Espaciais – INPE; Instituto de Pesquisas Tecnológicas do Estado de São Paulo – IPT; Instituto Chico Mendes de Conservação da Biodiversidade - ICMBio; Ministério da Saúde; Mapbiomas; Ministério dos Transportes; Ministério do Meio Ambiente – MMA; Fundação Sistema Estadual de Análise de Dados – SEADE; Secretaria da Saúde do Estado de São Paulo; Secretaria de Estrutura e Meio Ambiente do Estado de São Paulo; *United States Geological Survey* – USGS; Secretaria da Educação do Estado de São Paulo; Secretaria de Saneamento e Recursos Hídricos do Estado de São Paulo – SSRH – SP; Sistema Ambiental Paulista – DataGeo; Universidade de São Paulo – USP.

9. Na nona coluna aparece o link de onde a informação foi extraída;
10. E a última coluna mostra se o dado é gratuito, pago ou privado.

3.4 Ocorrência de eventos/Geocodificação automática

Para analisar a espacialização de eventos na região foram utilizados os dados de localização das ocorrências de desastres naturais da Defesa Civil de Caieiras e Francisco Morato, em formato “.xlsx”; excepcionalmente usou-se os dados de ocorrências de escorregamentos e inundações, que ocorreram entre 2005 e 2013. Também se empregou as ocorrências de escorregamentos e inundações da Linha 7, da CPTM (2019), em formato *shapefile*. Posteriormente, esses pontos foram comparados com os mapas de suscetibilidade de escorregamentos e inundações, de Caieiras (2013), de Franco da Rocha (2017) e de Francisco Morato (2013), na escala 1:25.000, do IPT/CPRM, em formato “.shp”.

Os endereços da Defesa Civil foram geocodificados automaticamente pela Autora com a utilização dos *softwares Microsoft Excel, R Studio, API do Google Maps* e *ArcGIS* 10.2. A planilha Excel de endereços das ocorrências foi convertida no formato “.csv” que foi inserido no *R Studio*, onde se operou a função “*geocode*” associada a “*API key*” do *API do Google Maps*; após o processo de geocodificação no R, o “.csv” foi transformado em “.xls” e geocodificado no *ArcGIS* 10.2, usando a ferramenta “*Add XY Data*”.

Para verificar quais classes de suscetibilidade de escorregamentos e inundações coincidiram com a área de estudo foi utilizada a função “*Select by Attributes*” do *ArcGIS*. E para analisar a distribuição das ocorrências de escorregamentos e inundações em cada classe de suscetibilidade, utilizou-se a opção “*Select by Location*” do *software* de SIG.

3.5 Relação de ocorrências de desastres naturais com cobertura e uso do solo

Nessa etapa analisou-se quais tipos de cobertura do solo coincidiram com as ocorrências de escorregamentos e inundações. O objetivo dessa análise era identificar a relação do uso e cobertura do solo com o número de ocorrências de desastres naturais na área de estudo.

Utilizou-se os dados geocodificados do número de ocorrências de desastres naturais da Defesa Civil de Caieiras e Francisco Morato (2013) e das ocorrências de escorregamentos e inundações da Linha 7 da CPTM (2019), além dos dados de uso e cobertura do solo, de ambos os municípios, do Projeto Mapbiomas (2020), do ano 2019, Coleção 5, em formato *.tif*. Primeiramente, as informações foram recortadas para a área de estudo, usando a ferramenta “*Clip*” do *ArcGIS 10.2*.

Após, o arquivo *raster* do uso do solo foi vetorizado através da ferramenta “*Raster to Polygon*” do *ArcMap* para facilitar a análise com o arquivo vetorial das classes de suscetibilidade. Posteriormente, foi realizada a interseção das classes de cobertura do solo com os pontos de escorregamentos e inundações através das ferramentas “*Select by Attributes*” e “*Select by Location*” do *software* de SIG.

Também se adicionou uma nova coluna chamada “Área” na tabela de atributos do arquivo (*shapefile*) de uso do solo da área de estudo para calcular a área de cada classe de cobertura do solo, utilizando a opção “*Calculate Geometry*” e escolhendo hectares como unidade de medida. Para este estudo também era interessante verificar se as inundações poderiam estar associadas à ocupação urbana em APPs de hidrografia.

Para fazer as APPs dos principais cursos hídricos da área de estudo (no caso, Rio Juqueri e Ribeirão Tapera Grande) foi utilizada a ferramenta “*Buffer*” do *ArcMap* e aplicou-se um *buffer* de 30 m para cada rio, pois de acordo com o inciso I do artigo 4º da Lei nº 12.651 de 2012, “as faixas marginais de qualquer curso d’água natural perene e intermitente, excluídos os efêmeros, desde a borda da calha do leito regular, em largura

mínima de 30 (trinta) metros, para os cursos d'água de menos de 10 (dez) metros de largura” (BRASIL, 2012).

3.6 Relação das ocorrências de inundações e escorregamentos com o Índice Paulista de Vulnerabilidade Social (IPVS – SEADE)

Esta pesquisa realizou uma análise do IPVS nos setores censitários na área de estudo para verificar a vulnerabilidade social da população que mora entorno da via férrea e, portanto, pode potencialmente, utilizar o trem como meio de transporte. Utilizou-se o *shapefile* dos setores censitários do Estado de São Paulo, do IBGE (2010), e o *.csv* do Índice Paulista de Vulnerabilidade Social do SEADE (2010).

Para gerar o resultado, usou-se o *software ArcGIS* e sua ferramenta “*Join*” para unir as informações do *.csv* ao arquivo *.shp*. Posteriormente, esse mapa foi associado com os pontos de inundações e de escorregamentos, através da função “*Intersect*” para verificar se o maior número de ocorrências de desastres naturais está relacionado às áreas de alta vulnerabilidade social.

3.7 Indicador de Lixo Acumulado

Para esta pesquisa era relevante conhecer quais setores censitários da área de estudo possuíam lixo acumulado nos logradouros e averiguar se estes poderiam estar relacionados a ocorrência de inundações e, conseqüentemente, a doenças transmissíveis por água contaminada; portanto, analisou-se os locais em que os domicílios possuíam lixo acumulado em seu entorno, com intuito de identificar os lugares com maior interferência para a ocorrência de desastres naturais na ferrovia. Este estudo elaborou seus índices ao nível dos setores censitários porque os setores são comumente utilizados para coletar e analisar dados para políticas e planejamento no governo e na saúde pública (FLANAGAN *et al.*, 2011).

Assim, usou-se----- os dados do “Arquivo Entorno 01” do Censo do IBGE (2010). As variáveis utilizadas foram escolhidas de acordo com a literatura consultada, onde os critérios são explicados no item 2.5.3.1 desta Dissertação. As variáveis consideradas são:

- a. V001: número de domicílios particulares permanentes;

- b. V056: número de domicílios particulares permanentes próprios – existe lixo acumulado nos logradouros;
- c. V058: número de domicílios particulares permanentes alugados – existe lixo acumulado nos logradouros;
- d. V060: número de domicílios particulares permanentes cedidos – existe lixo acumulado nos logradouros.

Baseado na lógica do cálculo do “Indicador Sem Lixo Acumulado” de Fagundes (2017), o Indicador de Lixo Acumulado foi calculado da seguinte forma (Equação 1):

$$\text{Ind. Lixo} = (V056 + V058 + V060) / V001 \quad (1)$$

Posteriormente, os resultados desse indicador foram adequados para a proporção de área dos setores censitários que estão dentro do *buffer* de 500 m (área de estudo). Nem todos os setores censitários estão totalmente inclusos dentro do *buffer*; dessa forma, se calculou a proporção da área de cada setor dentro do *buffer* em relação ao tamanho total da área do setor censitário. As áreas dos setores censitários foram calculadas no *QGIS* 3.10: criou-se uma coluna “Área” na tabela de atributos, através da função “Novo Campo”, e a área foi calculada pela opção “Calculadora de Campo” em unidades de medida “hectares (ha)”. Depois, o valor da proporção de área do setor censitário foi multiplicado pelo valor do Indicador de Lixo Acumulado do setor calculado anteriormente.

Por fim, os resultados do Indicador de Lixo Acumulado foram normalizados entre 0 e 1. A normalização dos dados significa “uma conversão de escala para que todos os dados estivessem ponderados pelo número total de casos no universo, gerando um valor relativo, portanto, comparativo” (FAGUNDES, 2017, p.176). Para a construção dos índices, a normalização dos dados seguiu critérios semelhantes aos adotados pelo Programa das Nações Unidas para o Desenvolvimento (PNUD) para o cálculo do Índice de Desenvolvimento Humano (IDH): “o valor de cada índice é igual ao quociente entre a diferença entre o valor observado e o mínimo possível; e a diferença entre os limites máximos e mínimos possíveis” (MORATO, 2008, p. 149).

No caso do Indicador de Lixo Acumulado, o valor 0 representa os setores censitários com menor concentração de lixo acumulado nos logradouros, enquanto o valor 1 refere-se aos setores com maior concentração de lixo acumulado nas ruas. Os cálculos foram realizados, primeiramente, no *Microsoft Excel* e, posteriormente, os dados das

variáveis no *Excel* foram espacializados, unidos ao *shapefile* dos setores censitários da região, no *software QGIS* 3.10, através da ferramenta “Uniões”, e os resultados foram validados através da repetição dos cálculos na “Calculadora de Campo” da tabela de atributos no *QGIS*. Enfim, os dados foram classificados em 5 classes (de menor concentração de lixo para maior concentração), usando o método “Intervalos Iguais” no *QGIS*. O método Intervalo Igual divide um intervalo de valores de atributo em subintervalos de tamanhos iguais, permitindo que o usuário especifique o número de intervalos, enquanto as quebras de classes são determinadas automaticamente pelo *software* (ARCGIS PRO, 2020).

3.8 Indicador de Esgoto à Céu Aberto

Para este estudo era interessante verificar quais setores censitários da área apresentavam esgoto à céu aberto e confirmar se estes estariam associados a uma provável inundação insalubre, além da capacidade de deixar o terreno mais úmido, favorecendo escorregamentos, principalmente em taludes; portanto, averiguou-se os setores em que os domicílios apresentavam esgoto à céu aberto em seu entorno, com objetivo de identificar os locais com maior interferência para a ocorrência de escorregamentos e inundações na ferrovia. Desse modo, utilizou-se os dados do “Arquivo Entorno 01” do Censo 2010.

Escolheu-se as variáveis utilizadas de acordo com a literatura consultada, onde os critérios são explicados no item 2.5.3.2 desta Dissertação. As variáveis consideradas são:

- a. V001: número de domicílios particulares permanentes;
- b. V050: número de domicílios particulares permanentes próprios – existe esgoto a céu aberto;
- c. V052: número de domicílios particulares permanentes alugados – existe esgoto a céu aberto;
- d. V054: número de domicílios particulares permanentes cedidos – existe esgoto a céu aberto.

Com base na lógica do cálculo do “Indicador Sem Esgoto à Céu Aberto” de Fagundes (2017), o Indicador de Esgoto à Céu Aberto foi calculado do seguinte modo (Equação 2):

$$\text{Ind. Esgoto} = (V050 + V052 + V054) / V001 \quad (2)$$

Após, os resultados desse indicador foram adequados para a proporção de área dos setores censitários; as áreas dos setores censitários foram calculadas no *QGIS*; e a seguir, o valor da proporção de área do setor censitário foi multiplicado pelo valor do Indicador de Esgoto à Céu Aberto do setor calculado.

Depois, os resultados do Indicador de Esgoto à Céu Aberto foram normalizados entre 0 e 1. Para o Indicador de Esgoto à Céu Aberto, o valor 0 representa os setores censitários com menor concentração de esgoto aberto no entorno dos domicílios, enquanto o valor 1 refere-se aos setores com maior concentração de esgoto à céu aberto. Enfim, os dados foram classificados em 5 classes (de menor concentração de esgoto aberto para maior concentração), usando o método “Intervalos Iguais” no *QGIS*.

3.9 Indicador de Abastecimento por Água Não-Encanada

Para esta pesquisa era relevante conhecer quais setores censitários não tinham acesso à um abastecimento de água da rede geral, pois esses locais podem estar relacionados à criação de poços e fossas, que têm potencial para favorecer escorregamentos, principalmente em taludes, e potenciais inundações insalubres. Para essa finalidade, foram utilizados os dados do “Arquivo Domicílio” do Censo Demográfico do IBGE (2010). No Censo 2010 não há dados sobre os domicílios que não possuem abastecimento de água da rede geral, mas ele fornece informações sobre os domicílios particulares que têm abastecimento de água da rede geral.

As variáveis utilizadas foram escolhidas de acordo com a literatura consultada, onde os critérios são explicados no item 2.5.3.3 desta Dissertação. As variáveis consideradas são:

- a. V002: número de domicílios particulares permanentes;
- b. V012: número de domicílios particulares permanentes com abastecimento de água da rede geral.

Baseado no cálculo do “Indicador de Água Encanada” de Fagundes (2017), primeiramente calculou-se um Indicador de Abastecimento de Água da seguinte forma (Equação 3):

$$\text{Ind. Abastecimento de Água} = V012 / V002 \quad (3)$$

Porém, para criar um Indicador de Água Não-Encanada, o Indicador de Abastecimento de Água teve que ser classificado do menor ao maior valor porque, ao contrário das outras variáveis, um valor mais alto indica mais domicílios com abastecimento de água da rede geral e, portanto, menor interferência externa à ferrovia. Então, o cálculo utilizado para essa inversão de valores foi (Equação 4):

$$\text{Ind. Não Abastecimento de Água} = 1 - (V012 / V002) \quad (4)$$

Posteriormente, os resultados do Indicador de Água Não-Encanada foram adequados para a proporção de área dos setores censitários. Por fim, os resultados do Indicador de Água Não-Encanada foram normalizados entre 0 e 1, onde o valor 0 representa os setores censitários com menor concentração de domicílios com abastecimento de água não encanada, enquanto o valor 1 refere-se aos setores com maior concentração de moradias com abastecimento de água não encanada. Os dados foram classificados em 5 classes usando o método “Intervalos Iguais”.

3.10 Indicador de Baixa Renda

Para esta Dissertação, interessava saber quais setores censitários da área de estudo apresentavam uma menor renda salarial e verificar se seus moradores estariam mais expostos à ocorrência de desastres naturais, e, concomitantemente, se a população desses locais também poderia contribuir para a ocorrência de escorregamentos e inundações entorno da ferrovia; portanto, considerou-se os setores em que os domicílios particulares permanentes recebiam, no máximo, até um salário-mínimo. Desse modo, foram utilizados os dados do “Arquivo de Renda dos Domicílios” do Censo Demográfico do IBGE (2010).

Escolheu-se as variáveis de acordo com a literatura consultada, onde os critérios são explicados no item 2.5.3.4 desta Dissertação. As variáveis consideradas são:

- a. V001a: número de domicílios particulares permanentes;
- b. V001b: número de domicílios particulares improvisados;
- c. V005: número de domicílios particulares com rendimento nominal mensal domiciliar *per capita* de até 1/8 salário-mínimo;
- d. V006: número de domicílios particulares com rendimento nominal mensal domiciliar *per capita* de mais de 1/8 a 1/4 salário-mínimo;
- e. V007: número de domicílios particulares com rendimento nominal mensal domiciliar *per capita* de mais de 1/4 a 1/2 salário-mínimo;

- f. V008: número de domicílios particulares com rendimento nominal mensal domiciliar *per capita* de mais de 1/2 a 1 salário-mínimo;
- g. V014: número de domicílios particulares sem rendimento nominal mensal domiciliar *per capita*.

Com base na lógica do Indicador Rendimento *Per Capita* Maior 1 S.M. calculado por Fagundes (2017); o Indicador de Baixa Renda foi calculado da seguinte forma (Equação 5):

$$\text{Ind. Renda} = (V005 + V006 + V007 + V008 + V014) / (V001a + V001b) \quad (5)$$

Normalizou-se os resultados do Indicador de Baixa Renda entre 0 e 1, onde o valor 0 representa os setores censitários com uma menor concentração de domicílios com renda baixa, enquanto o valor 1 refere-se aos setores com maior concentração de moradias que têm uma baixa renda. Enfim, os dados foram classificados em 5 classes, usando o método “Intervalos Iguais”.

3.11 Indicador de Idosos

Para este estudo também era interessante conhecer a proporção de idosos em cada setor censitário região, pois quanto mais idosos, maior seria a dificuldade em uma possível evacuação por causa de um evento de desastre natural. Deste modo, verificou-se o número de pessoas residentes, em domicílios particulares e domicílios coletivos, a partir de 65 anos de idade. Utilizou-se os dados do “Arquivo Idade” do Censo Demográfico do IBGE (2010).

As variáveis utilizadas foram escolhidas de acordo com a literatura consultada, onde os critérios são explicados no item 2.5.3.5 desta Dissertação. As variáveis consideradas são:

- a. V001: número de pessoas residentes em domicílios particulares e domicílios coletivos;
- b. V099: número de pessoas com 65 anos de idade;
- c. V100: número de pessoas com 66 anos de idade;
- d. V101: número de pessoas com 67 anos de idade;
- [...]
- e. V134: número de pessoas com 100 anos ou mais de idade.

O Indicador de Idosos foi calculado da seguinte maneira (Equação 6):

$$\text{Ind. Idosos} = (V099 + V100 + V101 [...] + V134) / V001 \quad (6)$$

Os resultados do Indicador de Idosos foram normalizados entre 0 e 1, onde o valor 0 representa os setores censitários com uma menor concentração de idosos, enquanto o valor 1 refere-se aos setores com maior concentração de pessoas idosas. Os dados foram classificados em 5 classes, usando o método “Intervalo Igual”.

3.12 Indicador de Crianças

Para esta pesquisa, era relevante averiguar a taxa de crianças em cada setor censitário da área de estudo, pois quanto mais crianças, maior seria a dificuldade de elas serem retiradas de um possível evento de desastre natural e de pedir ajuda. Assim, verificou-se o número de pessoas residentes, em domicílios particulares e domicílios coletivos, menores de 5 anos de idade. Considerou-se os dados do “Arquivo Idade” do Censo Demográfico do IBGE (2010).

As variáveis utilizadas foram escolhidas de acordo com a literatura consultada, onde os critérios são explicados no item 2.5.3.6 desta Dissertação. As variáveis consideradas são:

- a. V001: número de pessoas residentes em domicílios particulares e domicílios coletivos;
- b. V022: número de pessoas com menos de 1 ano de idade;
- c. V035: número de pessoas de 1 ano de idade;
- d. V036: número de pessoas com 2 anos de idade;
- e. V037: número de pessoas com 3 anos de idade;
- f. V038: número de pessoas com 4 anos de idade.

O Indicador de Crianças foi calculado do seguinte modo (Equação 7):

$$\text{Ind. Crianças} = (V022 + V035 + V036 + V037 + V038) / V001 \quad (7)$$

Os resultados do Indicador de Crianças foram normalizados entre 0 e 1, onde o valor 0 representa os setores censitários com uma menor concentração de crianças, enquanto o valor 1 refere-se aos setores com maior concentração de crianças. Enfim, os dados foram classificados em 5 classes (de menor concentração de crianças para maior concentração), usando o método “Intervalos Iguais”.

3.13 Indicador de Analfabetismo

Para esta pesquisa era interessante conhecer a proporção de pessoas não-alfabetizadas em cada setor censitário da área de estudo, pois quanto mais não-alfabetizados, maior seria a dificuldade deles em se ter informação suficiente para interpretar avisos ou cartilhas de prevenção de riscos de desastres, preparação inadequada para fugir, além da restrição de mobilidade, caso ocorresse algum evento de desastre natural. Desse modo, verificou-se o número de pessoas não-alfabetizadas em relação ao número total de moradores. Considerou-se os dados do “Arquivo Alfabetização” e do “Arquivo Idade” do Censo Demográfico do IBGE (2010). No Censo 2010 não há dados sobre os moradores que não possuem alfabetização, mas ele fornece informações sobre os moradores que são alfabetizados.

As variáveis utilizadas foram escolhidas de acordo com a literatura consultada, onde os critérios são explicados no item 2.5.3.7 desta Dissertação. As variáveis consideradas são:

- a. V001: número de pessoas alfabetizadas com 5 ou mais anos de idade;
- b. V062: número de pessoas alfabetizadas com 65 anos de idade;
- c. V063: número de pessoas alfabetizadas com 66 anos de idade;
- [...]
- d. V077: número de pessoas alfabetizadas com 80 anos ou mais de idade;
- e. V039: número de pessoas com 5 anos de idade;
- f. V040: número de pessoas com 6 anos de idade;
- [...]
- g. V098: número de pessoas com 64 anos de idade.

Primeiro, calculou-se um Indicador de Alfabetização, porém nesse indicador não poderia estar incluído os idosos alfabetizados, pois todos idosos já foram considerados no “Indicador de Idosos”. Assim, subtraiu-se o número de idosos alfabetizados (pessoas alfabetizadas a partir de 65 anos de idade) do número total de pessoas alfabetizadas. Posteriormente, o número de pessoas alfabetizadas (com exceção dos moradores até 4 anos e a partir dos 65 anos) foi dividido pelo número de pessoas entre 5 anos e 64 anos de idade (Equação 8):

$$\text{Ind. Alfabetização} = [V001 - (V062 + V063 \dots + V077)] / (V039 + V040 \dots + V098) \quad (8)$$

Porém, para criar um Indicador de Analfabetismo, o Indicador de Alfabetização teve que ser classificado do menor ao maior valor porque, ao contrário das outras variáveis, um valor mais alto indica mais pessoas alfabetizadas e, portanto, menor vulnerabilidade social. Então, o cálculo utilizado para essa inversão de valores foi (Equação 9):

$$\text{Ind. Analfabetismo} = 1 - \{[V001 - (V062 + V063 \dots + V077)] / (V039 + V040\dots + V098)\} \quad (9)$$

Os resultados do Indicador de Não-Alfabetizados foram normalizados entre 0 e 1, onde o valor 0 representa os setores censitários com uma menor concentração de pessoas não-alfabetizadas, enquanto o valor 1 refere-se aos setores com maior concentração de pessoas que não possuem alfabetização. Os dados foram classificados em 5 classes, usando o método “Intervalos Iguais”.

3.14 Indicador dos Residentes Sem Moradia Própria

Para esta pesquisa era relevante averiguar a proporção de pessoas que não possuíam moradia própria na área de estudo, em cada setor censitário, pois quanto mais pessoas sem moradia própria, maior seria a carência financeira dessas pessoas, o que aumentaria a vulnerabilidade desses moradores a ocorrências de desastres naturais. Utilizou-se os dados do “Arquivo Domicílio, características gerais” do Censo Demográfico do IBGE (2010).

As variáveis utilizadas foram escolhidas de acordo com a literatura consultada, onde os critérios são explicados no item 2.5.3.8 desta Dissertação. As variáveis consideradas são:

- a. V002: número de domicílios particulares permanentes;
- b. V007: número de domicílios particulares permanentes próprios em aquisição (refere-se aos domicílios que eram de propriedade, total ou parcial, de um ou mais moradores e ainda não estava integralmente pago (IBGE, 2011));
- c. V008: número de domicílios particulares permanentes alugados (refere-se aos domicílios que eram alugados e o aluguel era pago por um ou mais moradores (IBGE, 2011));
- d. V009: número de domicílios particulares permanentes cedidos por empregador (refere-se aos domicílios que eram cedidos por empregador de qualquer um dos

moradores, ainda que mediante uma taxa de ocupação ou conservação (IBGE, 2011));

- e. V010: número de domicílios particulares permanentes cedidos de outra forma (refere-se aos domicílios cedidos gratuitamente por pessoa que não era moradora ou por instituição que não era empregadora de algum dos moradores, ainda que mediante uma taxa de ocupação ou de conservação (IBGE, 2011));
- f. V011: número de domicílios particulares permanentes em outra condição de ocupação (quando o domicílio era ocupado de forma diferente das anteriormente relacionadas (IBGE, 2011)).

O Indicador de Residentes Sem Moradia Própria foi calculado da seguinte maneira (Equação 10):

$$\text{Ind. Moradia} = (V007 + V008 + V009 + V010 + V011) / V002 \quad (10)$$

Os resultados do Indicador de Residentes Sem Moradia Própria foram normalizados entre 0 e 1, onde o valor 0 representa os setores censitários com uma menor concentração de domicílios que não são próprios, enquanto o valor 1 refere-se aos setores com maior concentração de domicílios que não são próprios. Os dados foram classificados em 5 classes (de menor concentração de domicílios não-próprios para maior concentração), usando o método “Intervalos Iguais”.

3.15 Índice de Interferência Socioeconômica à Ferrovia

O Índice de Interferência Socioeconômica Externa à Ferrovia (IISEF) foi criado com a intenção de conhecer quais locais da área de estudo possuem fatores socioeconômicos externos que podem contribuir para a ocorrência de escorregamentos e inundações. As variáveis consideradas são:

- a. Indicador de Lixo Acumulado: os setores mais próximos do valor 1 teriam maior concentração de lixo acumulado nos logradouros e, portanto, representariam uma interferência maior para a ocorrência de desastres naturais na região;
- b. Indicador de Esgoto à Céu Aberto: os setores mais próximos do valor 1 possuiriam maior concentração de esgoto à céu aberto no entorno e, portanto, apresentariam uma interferência maior para a ocorrência de desastres naturais na área de estudo;
- c. Indicador de Abastecimento por Água Não-Encanada: os setores mais próximos do valor 1 teriam maior concentração de domicílios com abastecimento de água

sem ser da rede geral de distribuição (poços, nascentes, cisternas etc.) e, portanto, representariam uma interferência maior para a ocorrência de desastres naturais na região;

- d. Indicador de Baixa Renda: os setores mais próximos do valor 1 possuiriam maior concentração de domicílios com baixa renda e, portanto, estariam associados a uma interferência maior para a ocorrência de desastres naturais na área de estudo.

Converteu-se ps arquivos vetoriais de cada indicador em arquivos matriciais através da ferramenta “*Polygon to Raster*” do *ArcGIS* 10.3. Posteriormente, os dados *raster* foram transferidos para o *Google Earth Engine* através dos “*Assets > GeoTiff*” e depois utilizou-se o algoritmo “*ee.Image*”. Três métodos foram utilizados e comparados neste estudo:

3.15.1 Média Aritmética Simples

No *ArcGIS*, utilizando a ferramenta “*Raster Calculator*”, o Índice de Interferência Socioeconômica Externa foi calculado pela média aritmética simples dos indicadores de lixo acumulado, esgoto à céu aberto, não-abastecimento de água e de baixa renda (Equação 11); sendo que todas as variáveis possuem, supostamente, a mesma importância para composição do IISEF:

$$\text{IISEF} = (\text{Ind. Lixo Acumulado} + \text{Ind. Esgoto Aberto} + \text{Ind. Não-Abastecimento Água} + \text{Ind. Baixa Renda}) / 4 \quad (11)$$

Por fim, os dados foram classificados em 3 classes (Interferência Baixa, Interferência Média e Interferência Alta), utilizando o método de classificação “*Geometrical Interval*” do *ArcGIS*. O Intervalo Geométrico cria quebras de classe baseando-se em intervalos de classe que possuem uma série geométrica; o algoritmo cria intervalos geométricos, minimizando a soma dos quadrados do número de elementos em cada classe, garantindo que cada intervalo de classe tenha aproximadamente o mesmo número de valores em cada classe e que a mudança entre os intervalos seja razoavelmente consistente. Esse método é um meio-termo entre o Intervalo Igual, Quebras Naturais (*Jenks*) e Quantis; a sua maior vantagem é que ela funciona bem em dados que não são distribuídos normalmente (ARCGIS PRO, 2020).

No GEE não foi possível realizar um método de classificação da simbologia do dado (como no *software* de SIG). Portanto, considerou-se os intervalos de valores fornecidos pelo *software* de SIG e elaborou-se um fatiamento dos valores do histograma,

com as funções “*ee.Image*”, “*.mask*” e “*.toByte*”, usando os métodos para corte “*.lte*” e “*.gte*”.

3.15.2 *Outliers*

Segundo Carey *et al.* (1997), os *outliers*, que estão presentes em um conjunto de dados, são como um subconjunto de observações que parecem ser inconsistentes com as demais observações que seguem uma distribuição hipotética. Os *outliers* são exibidos graficamente - junto com o centro dos dados, a amplitude dos dados e a assimetria ou simetria dos dados – em *boxplots*. Eles se localizam abaixo do limite inferior do *boxplot* ou acima do seu limite superior (SIM; GAN; CHANG, 2012).

Para identificar os *outliers* de cada índice foram elaborados *boxplots* no *software* R Studio, utilizando a função “*boxplot* (dados)”. No *software* de SIG, considerou-se somente os setores censitários *outliers* dos indicadores de baixa renda, de lixo acumulado, de esgoto aberto e de não-abastecimento de água encanada. Os *outliers* foram considerados, dependendo dos valores, como “Interferência Alta” ou “Interferência Média”. Classificou-se os setores censitários que não eram *outliers* como “Interferência Baixa”. O método de classificação utilizado também era o “*Geometrical Interval*”, que definiu a quebra das classes.

3.15.3 Análise de Componente Principal de Objetos Espaciais

A Análise de Componentes Principais (ACP) é um método linear, na qual a transformação entre os dados originais e a nova representação dimensional reduzida é uma projeção linear. A sua principal finalidade é a redução da dimensionalidade, mas também pode ser usada para explorar relações entre variáveis. Assim como a ACP, a ACP de Objetos Espaciais (*Spatial Objects PCA*), que é empregada nesta Dissertação, é comumente utilizada para redução de dimensionalidade ou como um método de pré-processamento de dados, como acontece com qualquer aplicativo não espacial (Demšar *et al.*, 2013).

Exemplos podem ser encontrados em muitas áreas das geociências e nas ciências sociais. Uma aplicação comum nas ciências sociais é para desenvolver índices sociais compostos para o desenvolvimento, saúde, ou qualidade de vida por meio das primeiras CPs usando variáveis de entrada sociodemográficas (Demšar *et al.*, 2013), como é o caso nesta Dissertação.

A ACP foi realizada no *R Studio*, utilizando os pacotes *plyr*, *magrittr*, *readr*, *dplyr*, *tidyr*, *tibble*, *ggplot2*, *dplyr*, *tidyverse*, *ggstatsplot*, *RColorBrewer*, *scales*, *GGally*, *pander*, *rpart*, *rpart.plot*, *tidyverse*, *VennDiagram*, *DiagrammeR*, *factoextra* e *ggpubr*. A ACP foi baseada na matriz de correlação amostral entre os quatro indicadores (Renda, Lixo, Esgoto e Abastecimento) e a função usada era “*res.pca.cor*”. Por fim, os resultados das CPs de cada setor censitário foram classificados em 3 classes (Interferência Baixa, Interferência Média e Interferência Alta), usando o método de classificação “*Geometrical Interval*”.

3.16 Índice de Vulnerabilidade Social à Ocorrências de Desastres Naturais

O Índice de Vulnerabilidade Social à Ocorrências de Desastres Naturais (IVSDN) foi elaborado com o objetivo de identificar quais setores censitários demonstravam uma maior concentração de pessoas vulneráveis aos eventos de escorregamentos e inundações, para utilizá-lo no monitoramento de desastres e examinar seu valor potencial. As variáveis consideradas são:

- a. Indicador de Idosos: os setores mais próximos do valor 1 teriam maior concentração de idosos (com 65 anos ou mais) e, portanto, representariam uma vulnerabilidade maior em relação à ocorrência de desastres naturais na região;
- b. Indicador de Crianças: os setores mais próximos do valor 1 apresentariam maior concentração de crianças (até 4 anos de idade) e, assim, estariam associados a uma vulnerabilidade maior em relação à ocorrência de desastres naturais na área de estudo;
- c. Indicador de Analfabetismo: os setores mais próximos do valor 1 teriam maior concentração de analfabetos e, portanto, representariam uma vulnerabilidade maior em relação à ocorrência de desastres naturais na região;
- d. Indicador de Sem Moradia Própria: os setores mais próximos do valor 1 apresentariam maior concentração de domicílios que não são próprios e, assim, estariam associados a uma vulnerabilidade maior em relação à ocorrência de desastres naturais na área de estudo;
- e. Indicador de Baixa Renda: os setores mais próximos do valor 1 teriam maior concentração de domicílios com baixa renda e, portanto, representariam uma vulnerabilidade maior em relação à ocorrência de desastres naturais na região.

Os arquivos vetoriais de cada índice também foram convertidos em arquivos *raster*, para, posteriormente, serem inseridos no GEE. Utilizou-se e comparou-se três métodos nesta pesquisa:

3.16.1 Média Aritmética Simples

No *software* de SIG, utilizando a ferramenta “*Raster Calculator*”, o Índice de Vulnerabilidade Social à Ocorrências de Desastres Naturais foi calculado pela média aritmética simples (Equação 12):

$$IVSDN = (\text{Ind. Idosos} + \text{Ind. Crianças} + \text{Ind. Analfabetismo} + \text{Ind. Sem Moradia Própria} + \text{Ind. Baixa Renda}) / 5 \quad (12)$$

Ao final, os dados foram classificados em 3 classes (Vulnerabilidade Baixa, Vulnerabilidade Média e Vulnerabilidade Alta), usando o método de classificação “*Geometrical Interval*”.

3.16.2 Outliers

Para identificar os *outliers* de cada índice, foram elaborados *boxplots* no *software* R Studio com a função “*boxplot*”. No *software* de geoprocessamento, considerou-se somente os setores censitários *outliers* dos indicadores de baixa renda, de idosos, de crianças, de analfabetos e de domicílios não-próprios. Os *outliers* foram considerados, dependendo dos valores, como “Vulnerabilidade Alta” ou “Vulnerabilidade Média”. Os setores censitários que não eram *outliers* foram classificados como “Vulnerabilidade Baixa”. O método de classificação utilizado também foi o “*Geometrical Interval*”.

3.16.3 Análise de Componente Principal de Objetos Espaciais

A Análise de Componente Principal foi realizada no *R Studio*, utilizando os pacotes já apresentados anteriormente. A ACP baseou-se na matriz de correlação amostral entre os cinco indicadores (Idosos, Crianças, Analfabetismo, Moradia e Renda) e a função usada era “*res.pca.cor*”. Por fim, os resultados das CPs de cada setor censitário foram classificados em 3 classes (Vulnerabilidade Baixa, Vulnerabilidade Média e Vulnerabilidade Alta), utilizando o método de classificação “Intervalos Geométricos”.

3.17 Inserção e visualização dos dados de um ambiente SIG para um ambiente *Big Data*

3.17.1 GeoMesa: implementação e passagem dos dados

Para a instalação do GeoMesa é necessário instalar um sistema Linux e para isso deve-se escolher uma distribuição - ou distro (uma versão desse sistema operacional pronta e disponível para o usuário). Existem diversas distribuições para o Linux como Mint, Debian, Fedora, Slackware, entre outras; para este estudo foi escolhido a distro Ubuntu por ser mais “amigável” para usuários iniciantes. Inicialmente a imagem do Ubuntu 20.04 (UBUNTU, 2018) deve ser baixada e depois deve-se criar um *pen drive* com essa imagem. Após, é preciso acessar a BIOS (Sistema Básico de Entrada e Saída) do computador para habilitar o *boot* (refere-se ao processo de inicialização do computador) através do USB. Assim, é possível realizar a instalação do Ubuntu.

Para trabalhar com o Ubuntu deve-se instalar uma linguagem de programação. Uma das mais utilizadas é a linguagem Python (PYTHON, 2020). Através do terminal do computador, instala-se o Python 2.7. Também é fundamental instalar o Java: outra linguagem de programação e uma plataforma computacional (JAVA, 2020). Para isso, deve-se instalar o OpenJDK (uma implementação *open source* do Java) no terminal do Linux.

Posteriormente, deve-se instalar um banco de dados. Um dos mais conhecidos é o *Apache Cassandra* (APACHE CASSANDRA, 2016), um sistema NoSQL (dados não-estruturados), que possui alta escalabilidade e disponibilidade, além de uma grande tolerância a falhas. Para instalar o Cassandra é necessário escolher um método de instalação; o mais simples para a maioria dos usuários é o “tarball binário”. O *download* da versão 3.11.6 deve ser realizada no site oficial, o arquivo deve ser descompactado e criar-se-á uma variável de ambiente para finalizar a instalação (GEOMESA, 2020b).

Após a instalação do Cassandra é necessário integrá-lo ao GeoMesa. Um espaço chave deve ser criado no Cassandra e definir-se uma variável de ambiente. Os objetos do GeoMesa Cassandra devem ser baixados da versão binária mais recente (3.0.0) do *GitHub*¹. O GeoMesa Cassandra inclui um *plug-in GeoServer* para adicionar repositórios de dados Cassandra no *GeoServer* (GEOMESA, 2020b). O *GeoServer* é um servidor de mapas funcional, que possibilita a visualização e edição de dados geoespaciais, sendo um servidor de *Web Map Service* (WMS), *Web Feature Service* (WFS) e de *Web Coverage*

¹ “*GitHub* é um site de hospedagem de código colaborativo desenvolvido com base do sistema de controle de versão *git*” (Kalliamvakou *et al.*, 2014).

Service (WCS). Os arquivos do *plug-in* serão instalados por *script* através do terminal do Ubuntu, assim como os arquivos JARs do Cassandra.

Para a inserção dos dados no GeoMesa-Cassandra, inicialmente, deve-se subir o Cassandra, através do comando “*cassandra*” e verificar se obteve conexão através do comando “*cqlsh*”, e o *Geoserver*, indo até o diretório através do comando “*cd geoserver-2.17.1/bin*” e inicializando-o através do comando “*sh startup.sh*”, no terminal do Ubuntu. Em seguida, acessa-se o site do *Geoserver* (*localhost:8080/geoserver/web*) para verificar se a conexão foi estabelecida. Logo após, é possível fazer a ingestão dos dados através do GeoMesa.

No terminal do Ubuntu, deve-se escrever “*geomesa-cassandra*” para aparecer todos os comandos disponíveis; no caso, utilizaremos o comando “*ingest*” para a passagem dos dados. Para fazer a inserção de um *shapefile*, no terminal do Ubuntu (com o Cassandra e o Geoserver ativados), deve-se escrever o script “*geomesa-cassandra ingest diretório do arquivo/nome do arquivo.shp -k nomedokeyspace -c nomedocatalog -P 127.0.0.1:9042*”, onde o *keyspace* e o *catalog* se referem ao banco de dados utilizado. Para se fazer a ingestão de um *.csv*, deve-se adequá-lo a uma tabela de atributos através de um conversor, pelo comando: “*geomesa-cassandra ingest diretório do arquivo/nome do arquivo.csv -k nomedokeyspace -c nomedocatalog -P 127.0.0.1:9042 -s nomedosfts -C nomedoconversor*”, onde -s e -C correspondem, respectivamente, à tabela de atributos e conversor criados no arquivo “*application.conf*” do GeoMesa. O GeoMesa, por padrão, aceita os seguintes tipos de arquivos: *Delimited text* (*.csv*, *.tsv*), *JSON* (*JavaScript Object Notation*), *XML* (*Extensible Markup Language*), *Avro* (Apache Hadoop), *Parquet* (Apache Hadoop), *Shapefile*, *Data Fixed-Width Text* e *JDBC* (Java) (GEOMESA, 2020b).

3.17.2 Google Earth Engine: inserção dos dados

O *Google Earth Engine* – GEE, a princípio, aceita arquivos no formato *tif.*, *tfrecord.* + *json.*, *shapefile* e *csv*. Para fazer a transferência de um *shapefile* para o GEE, deve-se ir ao editor de código, em “*Assets*” e selecionar a opção “*Shapefiles*”. Ao selecionar um arquivo *shp.*, deve-se selecionar todos os arquivos *dbf.*, *shx.* e *prj.* relacionados; caso um arquivo *prj.* não seja fornecido, o GEE assumirá como padrão as coordenadas WGS84 (GEE GUIDE, 2020a).

Para fazer a passagem de um *.csv* deve-se também clicar em “*Assets*”, na opção “*CSV file*” e selecionar o arquivo desejado. O *.csv* deve conter uma linha para cada recurso e o mesmo número de colunas que forem as propriedades ou variáveis do conjunto de recursos. Se os recursos forem geoespaciais, eles devem ter uma geolocalização definida por uma *string* de geometria (GeoJSON, WKT) ou propriedades de posição X e Y; o GEE detectará uma coluna de geometria primária e, caso não seja especificado de outra forma, assumirá que a projeção de dados é WGS84 (GEE GUIDE, 2020a).

Para importar um *.tif*, é necessário clicar em “*Assets*” e depois em “*GeoTIFF ou TFRecord*”. É possível fazer *upload* de arquivos de imagem GeoTIFF de até 10 GB para a pasta de usuário (GEE GUIDE, 2020b). A Figura 11 a seguir mostra os métodos do GeoMesa e do GEE apresentados de forma resumida.

Figura 11 - Método da passagem de dados de um ambiente SIG para o GeoMesa e GEE.

4 RESULTADOS E DISCUSSÕES

A seguir serão apresentados os resultados obtidos e suas discussões.

4.1 Análise bibliométrica

Na etapa da Identificação (Anexo B), encontrou-se 36.727 registros com a pesquisa por palavra-chave no banco de dados, além de 23 registros adicionais de livros, documentos e sites oficiais (totalizando 36.750 publicações). Posteriormente, 121 artigos duplicados foram removidos, resultando em 36.629 registros (CORREIA; CORSI; QUINTANILHA, 2020).

Na etapa da Triagem (Anexo C), um filtro foi aplicado para retornar apenas os estudos publicados entre 2008 e 2019, e outro filtro para considerar os idiomas português, inglês e espanhol. Esses filtros reduziram o total de resultados para 23.959 (além dos 23 registros adicionais). Na etapa de Elegibilidade (Anexo D), os 20 primeiros artigos mais relevantes foram utilizados como critério para determinar estudos nos quais o resumo, os objetivos e a conclusão eram mais semelhantes ao tema da pesquisa. Dessa forma, 23.801 registros foram excluídos por não estarem relacionados ao tema deste estudo, produzindo 158 resultados finais (CORREIA; CORSI; QUINTANILHA, 2020).

A etapa final compreende os estudos utilizados na pesquisa. Após análise qualitativa e quantitativa, 90 referências foram incluídas nessa pesquisa (67 registros na base de dados e 23 registros adicionais). Na análise qualitativa, observou-se que existem muitos artigos relacionados à identificação de desastres naturais, principalmente escorregamentos e inundações, mas ainda há um número limitado de estudos relacionados à aplicação de técnicas de *Big Data* na análise de desastres naturais, particularmente em referência a ferrovias. Ainda falta conhecimento sobre as informações geoespaciais nos ambientes de *Big Data*, particularmente sobre desastres naturais e transporte ferroviário (CORREIA; CORSI; QUINTANILHA, 2020).

Em relação aos temas desastres naturais, *Big Data* e aprendizado de máquina, tiveram um aumento de pesquisas desde 2008, com um pico em 2019 (135 citações). Quando comparado a desastres naturais (ND), sensoriamento remoto (RS) e SIG (GIS), é notável que houve um aumento no número de artigos nessa área (148 em 2008 e 255 em 2019). Com relação ao tema de desastres naturais e ferrovias, alguns estudos aparecem nos anos posteriores, com mais publicações em 2014 (22). Portanto, há um crescimento

no período entre 2008 e 2019, com um pico em 2018 (Gráfico 3) (CORREIA; CORSI; QUINTANILHA, 2020).

Gráfico 3 - Número de publicações encontradas entre 2008 e 2019.

Elaborado por: Correia; Corsi; Quintanilha, 2020.

Também se pesquisou na base de dados Scopus, quais países tiveram publicações sobre o tema de estudo nos últimos anos. Nesse caso, todos os idiomas foram considerados, desde a primeira publicação em 2014, resultando em um total de 23 registros (Tabela 2) (CORREIA; CORSI; QUINTANILHA, 2020).

Tabela 2 - Número de registros sobre o tema de estudo por país.

Países	Nº de Estudos
EUA	3
Indonésia	3
Taiwan	2
Europa	9
Ásia	5
Indefinido	1

Elaborado por: Correia; Corsi; Quintanilha, 2020.

Essa análise bibliométrica foi publicada, em 01 de abril de 2020, em forma de artigo científico, na revista *Smart Cities*, edição especial *Geographic Knowledge Discovery and Big Data Analytics in Smart Cities*, com o título “*Big Data for natural disasters in an urban railroad neighborhood: A Systematic Review*” (Anexo E).

Uma revisão sistemática similar sobre *Big Data*, escorregamentos e técnicas de *machine learning*, utilizando o *software* Parsifal, também foi submetida, no dia 22 de

fevereiro de 2020, para o XIII Simpósio Internacional de Deslizamentos em Cartagena, na Colômbia, cujo título é “*A Systematic Review: Big Data Analytics and Landslides*” (Anexo F) e foi aprovada em 07 de julho de 2020. Este estudo teve como objetivo investigar os principais *features* utilizados nas técnicas de aprendizado de máquina para predição de escorregamentos e identificou que a técnica de aprendizado de máquina mais utilizada (nos estudos selecionados) foi a máquina de vetores de suporte (*support vector machine*). Os autores também apontaram que estudos sobre *Big Data*, *machine learning* e predição de escorregamentos ainda são escassos, principalmente no Brasil; porém, notou-se um crescimento sobre o tema nos últimos anos (AZEVEDO *et al.*, 2020)

4.2 Levantamento de fontes/dados oficiais na área de estudo

Esta Dissertação selecionou 165 informações relacionadas à área de estudo pertinentes à finalidade dessa pesquisa (Anexo G). Esses dados se distinguem quanto aos tipos de formatos de arquivos entre .csv (1,2%), .ecw (0,6%), .lyr (0,6%), .ovr (7,9%), .shp (67,3%), páginas de sites (2,4%), .tif (6,7%), WMS (9,7%) e .xlsx (3,6%) (Gráfico 4); em relação ao formato primário dessas fontes, 72,7% são dados vetoriais (sendo 50% em forma de polígonos, 30% como linhas e 20% como pontos), 22,4% são dados do tipo *raster* e 4,8% das informações não apresentava feição (Gráfico 5). Quanto à disponibilidade das fontes, 96% dos dados são gratuitos e 4% são informações privadas (como Defesas Civas Municipais de Caieiras e de Francisco Morato, IPT e CPTM).

Gráfico 4 - Formatos primários dos dados levantados.

Fonte: elaboração própria, 2020.

Gráfico 5 - Tipos de formatos dos dados selecionados.

Fonte: elaboração própria, 2020.

4.3 Ocorrência de eventos/Geocodificação automática

4.3.1 Escorregamentos

A área de estudo possui 1.771 hectares; identificou-se que a classe baixa de suscetibilidade de escorregamentos está presente em 77% (1371,6 ha) da área de estudo, a classe média está em 14% (248,92 ha) da área e a classe alta, em 9% (149,65 ha) da área de estudo. Dos 379 pontos de escorregamentos obtidos da Defesa Civil de Caieiras e de Francisco Morato e da CPTM, 124 intersectavam a área de estudo. Acreditava-se que a maioria destes pontos cruzariam com as áreas de alta suscetibilidade de escorregamento, por apresentarem um maior potencial de ocorrência deste tipo de desastre natural; mas, somente 2 eventos de escorregamento (2%) intersectavam as zonas de alta suscetibilidade, enquanto 117 pontos (94%) estão nas áreas de baixa suscetibilidade e 5 (4%) sobrepõem as áreas de média suscetibilidade de escorregamento (Figura 12). Os escorregamentos podem ter ocorrido, em maior parte, em áreas de baixa suscetibilidade devido ao fato da maioria da área de estudo ser composta por locais de baixa suscetibilidade à escorregamentos, mas também podem estar relacionados à fatores como cobertura e uso do solo, e aspectos socioeconômicos, o que será verificado nos próximos resultados.

Em relação à ferrovia, percebeu-se que ela está em sua maior parte sobre zonas de baixa suscetibilidade à escorregamentos, porém apresenta áreas de média e alta suscetibilidade em seu entorno, o que explica as ocorrências de escorregamentos que

afetam a Linha Rubi. É importante ressaltar que alguns pontos estão sobrepostos, o que significa que houve mais de uma ocorrência de escorregamento no mesmo local e em datas diferentes.

Figura 12 - Mapa de suscetibilidade e de ocorrências de escorregamentos na área de estudo.

Fonte: elaboração própria, 2020.

4.3.2 Inundações

Dos 1.771 hectares do *buffer* de estudo, a classe baixa de suscetibilidade de inundações ocorre em 6% (110 ha) da área de estudo, a classe média está presente em 7% (132,85 ha) da área e a classe alta, em 9% (158,35 ha) da área de estudo. Dos 61 pontos de inundações fornecidos pela Defesa Civil de Caieiras e da CPTM, 19 coincidem com a área de estudo. Assim como os escorregamentos, pressupunha-se que a maioria destes pontos cruzariam com as áreas de alta suscetibilidade de inundação, mas, somente 2 ocorrências de inundações (10%) intersectaram as zonas de alta suscetibilidade, enquanto 11 pontos (58%) estão nas áreas de baixa suscetibilidade e 6 (32%) sobrepõe as áreas de média suscetibilidade de inundação (Figura 13).

As inundações que ocorreram em áreas de baixa suscetibilidade à inundação podem estar relacionadas à fatores como cobertura e uso do solo, e aspectos socioeconômicos, o que será apresentado a seguir. Quanto à ferrovia, notou-se que ela está em sua maior parte sobre áreas de alta suscetibilidade a inundações, o que pode explicar as ocorrências de inundações que atingem a Linha 7 da CPTM. Vale destacar que, assim como os escorregamentos, existe a coincidência de alguns pontos, o que significa que houve mais de uma ocorrência de inundação no mesmo lugar e em períodos distintos.

Em relação aos dados de suscetibilidade e aos pontos de ocorrências de escorregamentos e inundações, é importante ressaltar que há uma diferença de escala entre os dados; portanto o resultado pode não ter 100% de precisão, mas é possível se ter uma ideia da distribuição das ocorrências de escorregamentos e de inundações nas classes de suscetibilidade a esses tipos de desastres naturais. Este estudo teve acesso aos dados de ocorrências de desastres naturais da Defesa Civil de Caieiras e de Francisco Morato, mas não obteve até o momento os dados de ocorrências do município de Franco da Rocha; portanto, esse resultado é uma visão parcial da distribuição dos eventos de escorregamentos e de inundações na área de estudo.

Figura 13 - Mapa de suscetibilidade e de ocorrências de inundações na área de estudo.

Fonte: elaboração própria, 2020.

4.4 Relação de ocorrências de desastres naturais com cobertura e uso do solo

Verificou-se que 66% da área de estudo é constituída por área urbana, 17% da área é composta por mosaicos de agricultura e pastagem, 14% é formada por florestas, 1% é representa por florestas plantadas, 0,7% por pastagem, 0,4% por lavouras temporárias, 0,2% por hidrografia e 0,06% por áreas não vegetadas.

4.4.1 Escorregamentos

Na análise espacial foi detectado que 93% dos escorregamentos ocorreram em área urbana e 7% dos escorregamentos, em mosaicos de agricultura e pastagem (Quadro 2). Esse resultado pode estar relacionado ao fato de áreas de intensa urbanização, com planejamento de uso e ocupação do solo inadequados, além de locais onde foram instaladas obras de engenharia de grande porte, nesse caso a ferrovia, serem ambientes favoráveis a ocorrências de escorregamentos (MINISTÉRIO DAS CIDADES/IPT, 2007).

Identificou-se 2 pontos de escorregamentos, na área de estudo, que estão próximos aos taludes de cortes nos maciços ou a ocupação irregular do local, que, associados a chuvas intensas, podem desencadear movimentos de massa. Também se verificou que 22% das ocorrências de escorregamentos estão associadas a elementos da paisagem como as estações da CPTM (Estações Caieiras, Franco da Rocha, Baltazar Fidélis e Francisco Morato) e a Rodovia SP-023. Além do risco desses locais serem afetados por escorregamentos, os seus passageiros podem ser afetados negativamente por estes eventos, como por exemplo, atrasos nas viagens e até acidentes. Também se notou que os escorregamentos ocorrem em uma faixa de 59 m a 2.500 m de distância das estações de trem.

Quadro 2 - Proporção de uso do solo e de ocorrências na área de estudo (*buffer* de 500 m).

Uso do Solo	Área	Escorregamentos	Inundações
Infraestrutura Urbana	1167,78 ha	116	18
Mosaico de Agricultura e Pastagem	297,66 ha	8	1
Formação Florestal	256,24 ha	-	-
Floresta Plantada	26,59 ha	-	-
Pastagem	11 ha	-	-
Outras Lavouras Temporárias	7,06 ha	-	-
Hidrografia	3,49 ha	-	-
Outras Áreas Não Vegetadas	1,01 ha	-	-
Total	1771 ha	124	19

Fonte: elaboração própria, 2020.

4.4.2 Inundações

Esse estudo identificou que 95% das inundações ocorreram em área urbana enquanto 5% delas coincidem em mosaicos de agricultura e pastagem (Figura 14). Essa análise é validada pelo fato que a urbanização impermeabiliza os solos (como o asfaltamento), provocando o aumento do volume e da velocidade da água escoada superficialmente, e a redução do tempo de resposta da bacia hidrográfica, tornando essa bacia mais sensível a precipitações mais curtas e intensas (MINISTÉRIO DAS CIDADES/IPT, 2007). Quanto às ocorrências nas áreas de pastagens, as inundações podem estar relacionadas a compactação do solo devido à passagem do gado e de máquinas, provocando uma menor infiltração da água de chuvas e, consequentemente, maior escoamento superficial, que pode ser intensificado pelo aumento da declividade e da pouca presença de vegetação na superfície do solo (MIGUEL; VIEIRA; GREGO, 2009).

15% das inundações na área de estudo intersectaram Áreas de Preservação Permanente (APPs) de hidrografia (referentes aos Rios Juqueri e Ribeirão Tapera Grande), além da própria ferrovia. Dentre as diversas funções das APPs estão a proteção dos corpos hídricos a fim de evitar enchentes, assoreamento dos rios e poluição das águas, e a manutenção da permeabilidade do solo e do regime hídrico, prevenindo enxurradas e inundações (MMA, 2020).

26% das inundações no entorno da ferrovia estão próximas a trechos de drenagem que cruzam a linha férrea, o que indica a probabilidade de haver uma tubulação mal dimensionada nesses locais, causando um refluxo de água, proporcionado por uma possível obra não adequada na ferrovia ou isso acontece devido ao aumento da densidade demográfica e, conseqüentemente, o grau de impermeabilização do solo (UFRRJ, 2020). 16% das ocorrências de inundações estão associadas a elementos da paisagem como as estações da CPTM e a Rodovia SP-023. Também se verificou que as inundações ocorrem em uma faixa de 55 m a 2.915 m de distância das estações de trem, ou seja, se houver uma paralisação da viagem de trem por causa de uma inundação e o passageiro for obrigado a retornar à estação, esse passageiro pode ter que andar sobre a linha férrea por até 2,9 km até a estação ferroviária mais próxima.

Figura 14 - Relação de cobertura e uso do solo com ocorrências de escorregamentos e inundações na área de estudo.

Fonte: elaboração própria, 2020.

4.5 Relação das ocorrências de inundações e escorregamentos com o Índice Paulista de Vulnerabilidade Social (IPVS do SEADE)

A área de estudo intersecta as classes de vulnerabilidade social do IPVS (SEADE, 2010): muito baixa (24%), baixa (18%), média (8%), alta (28%), muito alta (2%) e não classificado (20%) (Figura 15 a seguir). Os setores “não classificados” podem não ter sido categorizados porque, provavelmente, não possuem no mínimo 50 domicílios particulares permanentes (um dos requisitos para se elaborar o IPVS de um local), o que faz sentido já que se observou que esses setores estão sobrepostos em sua maior parte por florestas.

Esta pesquisa identificou que 53% dos escorregamentos coincidem com setores de vulnerabilidade social baixa e muito baixa, assim como 85% das inundações ocorreram em áreas de vulnerabilidade social muito baixa (Quadro 3). Essas ocorrências em áreas de baixa vulnerabilidade social estão, na verdade, relacionadas ao uso do solo: todos esses pontos coincidem com a área urbana, o que explica que os eventos de escorregamentos e inundações, podem estar relacionados ao crescimento desordenado das cidades, incluindo ocupação em áreas impróprias (TOMINAGA, 2009).

Quadro 3 - Relação IPVS com escorregamentos e inundações na área de estudo.

Classe IPVS	Área	Escorregamentos	Inundações
Vuln. Alta (urbanos)	485,19 ha	24	1
Vuln. Muito Baixa	426,70 ha	35	16
Não Classificado	369,88 ha	11	1
Vuln. Baixa	313,37 ha	31	-
Vuln. Média	148,53 ha	18	1
Vuln. Muito Alta (aglomerados subnormais)	26,89 ha	5	-

Fonte: elaboração própria, 2020.

Detectou-se também que 23% dos escorregamentos da área de estudo ocorreram em locais de alta e muito alta vulnerabilidade social. Apesar de todos os cidadãos que vivem em áreas de risco serem considerados como vulneráveis, os impactos sociais da exposição ao risco são desproporcionais sobre as pessoas ou grupos mais vulneráveis da sociedade, como crianças, idosos, pobres e pessoas com deficiência. Estes grupos são os menos preparados para uma situação de emergência, pois eles têm menos recursos financeiros para se preparar a uma situação de perigo; tendem a viver nos locais de maior risco, em habitações precárias; e pouco sabem sobre as conexões sociais e políticas

necessárias para usufruir dos recursos que poderiam os auxiliarem na sua recuperação em uma situação de desastre (NRC, 2006; DUNNING, 2009; CAMARINHA, 2016).

Lembrando que há uma diferença de escala entre os dados; portanto o resultado não pode ter 100% de precisão, mas é possível se ter uma ideia da distribuição das ocorrências de escorregamentos e de inundações nas classes do IPVS; além deste estudo não obter até o momento os dados de ocorrências do município de Franco da Rocha; portanto, esse resultado é uma visão parcial da distribuição dos eventos de escorregamentos e de inundações na área de estudo.

Figura 15 - Mapa do IPVS da área de estudo.

Fonte: elaboração própria, 2020.

4.6 Indicador de Lixo Acumulado

O Indicador de Lixo Acumulado (Figura 16) é representado em 5 classes, onde o valor 0 (a cor verde mais clara) refere-se aos setores censitários com uma menor concentração de domicílios com lixo em seu entorno, enquanto o valor 1 (a cor vermelha) equivale aos setores com uma maior concentração de domicílios com lixo em seu entorno. O Indicador de Lixo Acumulado mostra que, em média, apenas 2% dos setores censitários da área de estudo apresentam lixo acumulado nos logradouros.

Em relação às ocorrências de desastres naturais, todos os pontos de inundações intersectam com classes de baixa concentração de lixo acumulado, enquanto 99% dos eventos de escorregamento também coincidem com valores baixos do Indicador de Lixo Acumulado. Somente 1% dos escorregamentos ocorreram na classe mais alta do Indicador de Lixo Acumulado, que coincidem com aglomerados urbanos subnormais.

No Brasil, os aglomerados subnormais são conhecidos por diferentes nomes, como favelas, invasões, grotas, baixadas, comunidades, vilas, ressacas, loteamentos irregulares, mocambos e palafitas etc. (IBGE, 2020). De acordo com Cynamon e Monteiro (1985), o lixo constitui-se num problema social, econômico, sanitário e ambiental de muitas favelas. Além da precariedade dos serviços sanitários, as companhias de limpeza urbana têm dificuldades em recolher o lixo na área devido à falta de traçado urbano, topografia irregular da região e dimensões incompatíveis com os equipamentos convencionais.

Isso significa que 1% dos escorregamentos da área de estudo podem estar relacionados a ausência ou precariedade da coleta de lixo. E as classes altas do Indicador de Lixo Acumulado mostram setores censitários que podem oferecer uma maior probabilidade de se ter escorregamentos de depósitos de lixo, pois o acúmulo de lixo em bolsões pode causar a desestabilização dos mesmos; e também indica a maior possibilidade de ter inundações, principalmente em épocas de chuva, devido ao entupimento de canais e de redes de esgoto, além da obstrução do leito de rios provocada pelo lançamento de lixo pelas populações de favelas situadas a suas margens (CYNAMON; MONTEIRO, 1985).

4.7 Indicador de Esgoto à Céu Aberto

O Indicador de Esgoto à Céu Aberto (Figura 16) é caracterizado em 5 classes, onde o valor 0 (a cor verde mais clara) refere-se aos setores censitários com uma menor concentração de domicílios com esgoto à céu aberto em seu entorno, enquanto o valor 1 (a cor vermelha) equivale aos setores com uma maior concentração de domicílios com esgoto à céu aberto em seu entorno. O Indicador de Esgoto à Céu Aberto demonstra que, em média, 13% dos setores censitários da área de estudo possuem esgoto à céu aberto no entorno dos domicílios.

Quanto às ocorrências de desastres naturais, todos os pontos de inundações coincidem com classes de baixa concentração de esgoto à céu aberto. Já os escorregamentos, 88% intersectam as classes mais baixas do Indicador de Esgoto à Céu Aberto (0 – 0,40), 7% correspondem a classe média do indicador (0,41 – 0,60) e 5% sobrepõem as classes mais altas (0,61 – 1).

Os 5% de escorregamentos nas classes altas são compatíveis às áreas urbanas, tendo uma ocorrência em um aglomerado subnormal. O esgoto a céu aberto pode estar sendo lançado nos taludes de corte ou de aterro desse modo influenciam diretamente no grau de umidade da área, além da associação com a ocupação desordenada e do modo de ocupação da encosta.

E as classes altas do Indicador de Esgoto à Céu Aberto apontam os setores censitários que têm uma maior chance de contribuir para a ocorrência de escorregamentos, pois a infiltração de água servida (despejo irregular do esgoto, fossas negras etc.) é um dos fatores mais comuns encontrados na maioria das áreas de risco a escorregamentos no Brasil (CEMADEN, 2017); e também demonstram uma maior probabilidade de se ter inundações insalubres porque nas áreas urbanas, principalmente no período chuvoso, as inundações são o principal fator de risco para a ocorrência de doenças transmissíveis pela água, como a leptospirose (GUIMARÃES *et al.*, 2014).

Figura 16 - Indicadores de Lixo Acumulado nos Logradouros e de Esgoto à Céu Aberto na área de estudo.

Fonte: elaboração própria, 2020.

4.8 Indicador de Abastecimento por Água Não-Encanada

O Indicador de Abastecimento por Água Não-Encanada (Figura 17) é representado em 5 classes, onde o valor 0 (a cor verde mais clara) refere-se aos setores censitários com uma menor concentração de domicílios com abastecimento de água não-encanada, enquanto o valor 1 (a cor vermelha) equivale aos setores com uma maior concentração de domicílios com abastecimento de água não-encanada (ou seja, mais moradias com abastecimento de água por poços, nascentes, água de chuva em cisterna etc.). O Indicador de Abastecimento por Água Não-Encanada mostra que, em média, 6% dos setores censitários da área de estudo têm abastecimento de água sem ser da rede geral de distribuição.

Em relação às ocorrências de desastres naturais, 95% das inundações ocorrem nas classes mais baixas do Indicador de Abastecimento por Água Não-Encanada e, portanto, 5% encontram-se na classe mais alta do índice. Quanto aos escorregamentos, 93% localizam-se nas classes mais baixas do indicador e 7% na classe intermediária (0,41 – 0,60). As classes altas do Indicador de Abastecimento por Água Não-Encanada mostram setores censitários que podem estar relacionados a uma maior probabilidade de se ter escorregamentos, uma vez que a distribuição de mangueiras nas encostas pode contribuir na taxa de umidade delas.

4.9 Indicador de Baixa Renda

O Indicador de Baixa Renda (Figura 17) é caracterizado em 5 classes, onde o valor 0 (a cor verde mais clara) refere-se aos setores censitários com uma menor concentração de domicílios com renda baixa, enquanto o valor 1 (a cor vermelha) equivale aos setores com uma maior concentração de domicílios com renda baixa. O Indicador de Baixa Renda mostra que, em média, 53% dos domicílios recebem no máximo 1 salário-mínimo.

Quanto às ocorrências de desastres naturais, 95% dos pontos de inundações coincidem com as classes de mais baixas concentração Indicador de Baixa Renda (0 – 0,20 e 0,21 – 0,40), enquanto 5% sobrepõem a classe mais alta do indicador (0,81 – 1). Já os escorregamentos, 67% intersectam as classes mais baixas do Indicador de Baixa Renda (0 – 0,40), 27% correspondem a classe intermediária do indicador (0,41 – 0,60) e 6% coincidem com as classes mais altas (0,61 – 1).

As classes de 0,41 a 0,60 e 0,61 a 0,80 do Indicador de Baixa Renda correspondem a aglomerados subnormais, o que faz sentido já que os moradores dessas áreas possuem uma menor renda. De acordo com o IBGE (2019), em 2010, considerando o rendimento nominal domiciliar *per capita*, 31,6% da população brasileira residente em aglomerados subnormais estava na faixa de até meio salário-mínimo.

As classes altas do Indicador de Baixa Renda apontam os setores censitários que têm uma maior probabilidade de contribuir para a ocorrência de escorregamentos e inundações através da ocupação de áreas com uma suscetibilidade maior devido ao empobrecimento geral da população, implicando em falta de recursos para aquisição de terrenos em áreas mais aptas a ocupação urbana. Ao mesmo tempo, as classes mais altas do Indicador de Baixa Renda apresentam os setores censitários com maior vulnerabilidade à ocorrência de desastres naturais. A dificuldade do acesso à terra e a moradia em áreas urbanas, mais a baixa atuação do poder público, levou a um quadro de ocupação irregular dessas áreas, principalmente nos grandes centros urbanos, e isso mostra que a degradação ambiental está vinculada a pobreza da população e a não aquisição de terras, tornando a população de baixa renda a mais vulnerável a desastres naturais (MAFFRA; MAZZOLA, 2007).

Figura 17 - Indicadores de Não-Abastecimento por Água Encanada e de Baixa Renda na área de estudo.

Fonte: elaboração própria, 2020.

4.10 Indicador de Idosos

O Indicador de Idosos (Figura 18) é caracterizado em 5 classes, onde o valor 0 (a cor verde mais clara) refere-se aos setores censitários com uma menor concentração de moradores idosos, enquanto o valor 1 (a cor vermelha) equivale aos setores com uma maior concentração de moradores idosos. O Indicador de Idosos mostra que, em média, 5% dos moradores são idosos a partir de 65 anos de idade.

Em relação às ocorrências de desastres naturais, 58% das inundações ocorrem nas classes mais baixas do Indicador de Idosos (0 – 0,20 e 0,21 – 0,40) e 42% encontram-se na classe intermediária do indicador (0,41 – 0,60). Quanto aos escorregamentos, 92% localizam-se nas classes mais baixas do indicador (de 0 a 0,40) e 8% na classe intermediária (0,41 – 0,60).

As classes mais altas do Indicador de Idosos mostram os setores censitários que têm uma população mais vulnerável às ocorrências de desastres naturais, nesse caso, os idosos. Os idosos que vivem nos setores com Indicador de Idosos mais próximo de 1, principalmente quando estes estão associados aos pontos de escorregamentos e inundações, por apresentarem um declínio da capacidade funcional, terão variáveis como percepção de risco, estado de alerta, atenção e mobilidade reduzidas, aumentando a vulnerabilidade do indivíduo e a probabilidade de ele ser vitimado por um desastre natural (BODSTEIN; LIMA; BARROS, 2014).

4.11 Indicador de Crianças

O Indicador de Crianças no mapa (Figura 18) é caracterizado em 5 classes, onde o valor 0 (a cor verde mais clara) refere-se aos setores censitários com uma menor concentração de moradores crianças, enquanto o valor 1 (a cor vermelha) equivale aos setores com uma maior concentração de moradores crianças. O Indicador de Crianças mostra que, em média, 7% dos moradores são crianças menores que 5 anos.

Em relação às ocorrências de desastres naturais, 95% das inundações ocorrem nas classes mais baixas do Indicador de Crianças (0 – 0,20 e 0,21 – 0,40) e 5% encontram-se na classe mais alta do indicador (0,81 – 1). Quanto aos escorregamentos, 74% localizam-se nas classes mais baixas do indicador (de 0 a 0,40), 25% na classe intermediária (0,41 – 0,60) e 1% nas classes mais altas (0,61 – 1).

As classes mais altas do Indicador de Crianças mostram os setores censitários que têm uma população mais vulnerável às ocorrências de desastres naturais, nesse caso, as crianças. As crianças que vivem nos setores com Indicador de Crianças mais próximo de 1, principalmente quando estes estão associados aos pontos de escorregamentos e inundações, têm sua vida, seus direitos e suas necessidades ameaçados: a perda de meios de subsistência pode acarretar à pobreza extrema, à evasão escolar e à desnutrição (MUDAVANHU, 2014); sendo assim, as crianças fazem parte das populações vulneráveis a desastres, o que se confirma pela quantidade de mortes, nas consequências no pós desastre e na dificuldade em obter ajuda (OLIVEIRA, 2018).

Figura 18 - Indicadores de Idosos e de Crianças na área de estudo.

Fonte: elaboração própria, 2020.

4.12 Indicador de Analfabetismo

O Indicador de Analfabetismo (Figura 19) é caracterizado em 5 classes, onde o valor 0 (a cor verde mais clara) refere-se aos setores censitários com uma menor concentração de moradores analfabetos, enquanto o valor 1 (a cor vermelha) equivale aos setores com uma maior concentração de moradores analfabetos. O Indicador de Analfabetismo mostra que, em média, 9% dos moradores não são alfabetizados.

Em relação às ocorrências de desastres naturais, 95% das inundações ocorrem nas classes mais baixas do Indicador de Analfabetismo (0 – 0,20 e 0,21 – 0,40) e 5% encontram-se na classe mais alta do indicador (0,81 – 1). Quanto aos escorregamentos, 96% localizam-se nas classes mais baixas do indicador (de 0 a 0,40), e 4% na classe intermediária (0,41 - 0,60).

As classes mais altas do Indicador de Analfabetismo mostram os setores censitários que têm uma população mais vulnerável às ocorrências de desastres naturais, nesse caso, os analfabetos. As pessoas não-alfabetizadas que vivem nos setores com Indicador de Analfabetismo mais próximo de 1, principalmente quando estes estão associados aos pontos de escorregamentos e inundações terão capacidade restrita de compreender informações de alerta e acesso a informações de recuperação (CUTTER *et al.*, 2003).

4.13 Indicador de Residentes Sem Moradia Própria

O Indicador de Residentes Sem Moradia Própria (Figura 19) é caracterizado em 5 classes, onde o valor 0 (a cor verde mais clara) refere-se aos setores censitários com uma menor concentração de domicílios que não são próprios, enquanto o valor 1 (a cor vermelha) equivale aos setores com uma maior concentração de domicílios não-próprios. O Indicador de Residentes Sem Moradia Própria mostra que, em média, 24% dos domicílios não são próprios.

Em relação às ocorrências de desastres naturais, todas as inundações e escorregamentos coincidem com as classes mais baixas do Indicador de Residentes de Sem Moradia Própria (do 0 a 0,20 e do 0,21 a 0,40). Porém, as classes mais altas do Indicador de Residentes de Sem Moradia Própria mostram os setores censitários que têm uma população mais vulnerável às ocorrências de desastres naturais, nesse caso, as pessoas que não têm casa própria.

As pessoas que não possuem casa própria (como moradores de casas alugadas) que vivem nos setores com Indicador de Residentes de Sem Moradia Própria mais próximo de 1, principalmente quando estes estiverem associados a eventos de escorregamentos e inundações, se encontram em situação transitória ou não têm recursos financeiros para adquirir casa própria. Desse modo, muitas vezes falta acesso às informações sobre ajuda financeira durante a recuperação pós-desastre; nos casos mais graves, esses moradores não terão renda suficiente para pagar hospedagens temporárias (CUTTER *et al.*, 2003; AVILA, 2015).

Figura 19 - Indicadores de Analfabetismo e de Residentes Sem Moradia Própria na área de estudo.

Fonte: elaboração própria, 2020.

4.14 Índice de Interferência Socioeconômica Externa à Ferrovia (IISEF)

4.14.1 Média Aritmética Simples

O Índice de Interferência Socioeconômica Externa à Ferrovia calculado através do método da média aritmética simples mostra que a maior parte da área de estudo (57%) é constituída pelas classes de média interferência externa, seguido pelas classes de alta interferência (35%) e de baixa interferência (8%). Em relação aos pontos de inundações, 84% sobrepõem a classe média do IISEF, enquanto 16% correspondem à classe alta. Quanto aos eventos de escorregamentos, 61% ocorrem na classe média do índice de interferência, 33% intersectam a classe alta do IISEF e 6% coincidem com a classe baixa.

Identificou-se também que todos os aglomerados urbanos subnormais da área de estudo coincidem com a classe alta do IISEF, o que pode ser explicado pela maior concentração de domicílios de baixa renda, esgoto a céu aberto, abastecimento por água não encanada e lixo acumulado nesses locais. Os setores censitários com um alto IISEF apontam os locais que possuem maior relação com a ocorrência de escorregamentos e inundações na área de estudo; nota-se que o trecho ferroviário em Francisco Morato sobrepõe com muitas classes de alta interferência socioeconômica, o que faz sentido já que esse trecho é frequentemente destacado na mídia pelos transtornos causados aos passageiros da CPTM devido à ocorrência de desastres naturais (Figura 22).

4.14.2 Outliers

Para identificar os *outliers*, primeiro observou-se os *boxplots* de cada índice (Figura 20). No Indicador de Baixa Renda, os *outliers* são os setores censitários com o valor de indicador igual a 1; o Indicador de Lixo Acumulado mostra que quase todos os setores da área de estudo têm lixo coletado, portanto, os *outliers* indicam os locais onde têm lixo acumulado no entorno; no Indicador de Esgoto à Céu Aberto, os *outliers* referem-se aos setores censitários com valor de indicador a partir de 0,2 a 1; e o Indicador de Abastecimento por Água Não-Encanada mostra que a grande maioria dos setores possuem abastecimento de água via rede geral de distribuição, ou seja, os *outliers* indicam os lugares onde não há abastecimento de água encanada.

Figura 20 - *Boxplots* dos indicadores que compõem o IISEF.

Elaborado por: Michelle Trintinalia, 2020.

O Índice de Interferência Socioeconômica Externa calculado através do método de *outliers* mostra que a maioria da área de estudo (42%) é composta pelas classes de alta interferência externa, seguido pelas classes de baixa interferência (27%) e de média interferência (31%). Em relação aos pontos de inundações, 63% sobrepõem a classe baixa do IISEF, 21% correspondem à classe média e 16%, a classe alta de interferência. Quanto aos eventos de escorregamentos, 42% ocorrem na classe alta do índice de interferência, 37% intersectam a classe baixa do IISEF e 27% coincidem com a classe média.

Identificou-se também que todos os aglomerados urbanos subnormais da área de estudo coincidem com a classe alta do IISEF, o que pode ser explicado pela maior concentração de domicílios de baixa renda, esgoto a céu aberto, abastecimento por água não-encanada e lixo acumulado nesses locais. Os setores censitários com maior IISEF indicam os lugares que possuem maior relação com a ocorrência de escorregamentos e inundações na área de estudo; e percebe-se que a via férrea entre as estações Baltazar Fidélis e Francisco Morato tem muitos setores com classificados com alta interferência externa ao seu redor, confirmando as notícias de ocorrências de desastres naturais que afetam esse trecho (Figura 22).

4.14.3 Análise de Componente Principal de Objetos Espaciais

O percentual das variâncias nos mostrou que a primeira componente principal - CP1 explica 46% da variabilidade dos dados, seguida pela segunda componente principal - CP2 (26%), pela terceira componente principal - CP3 (18%) e pela quarta componente principal - CP4 (10%). A matriz de correlação dos dados nos permitiu observar que as variáveis que têm maior relação entre si são: Indicador de Esgoto à Céu Aberto e Indicador de Lixo Acumulado (0,45), Indicador de Esgoto à Céu Aberto e Indicador de Baixa Renda (0,37), e Indicador de Lixo Acumulado e Indicador de Baixa Renda (0,30).

O Quadro 4 mostra os resultados das quatro componentes principais. Dentre todas as componentes, a CP1 foi escolhida para a elaboração do ISEF, pois, além de ela representar a maior variabilidade dos dados, os seus resultados são mais interessantes para este estudo do que as outras componentes. A CP1 indica que os setores censitários que têm renda baixa, também possuem esgoto à céu aberto; o que faz sentido já que as áreas com moradias de baixa renda podem apresentar precariedade dos serviços sanitários, como coleta de esgotos (CYNAMON; MONTEIRO, 1985).

Quadro 4 - Matriz das componentes principais das variáveis do ISEF.

	CP1	CP2	CP3	CP4
Ind_Baixa_Renda	0,795	-0,044	0,604	0,040
Ind_Lixo_Acumulado	0,292	0,160	-0,314	-0,889
Ind_Esgoto_Aberto	0,504	0,292	-0,673	0,456
Ind_Sem_Abast_Agua	-0,169	0,942	0,290	0,012

Elaborado por: Michelle Trintinalia, 2020.

O Índice de Interferência Socioeconômica calculado através do método da ACP mostra que a maior parte da área de estudo (49%) é constituída pelas classes de média interferência externa, seguido pelas classes de baixa interferência (36%) e de alta interferência (15%). Em relação aos pontos de inundações, 68% sobrepõem a classe média do ISEF, enquanto 32% correspondem à classe alta. Quanto aos eventos de escorregamentos, 58% ocorrem na classe média do índice de interferência, 27% intersectam a classe baixa do ISEF e 6% coincidem com a classe alta.

Identificou-se também que 90% dos aglomerados urbanos subnormais da área de estudo coincidem com a classe alta do ISEF, enquanto os 10% sobrepõem a classe média do índice; isso pode ser explicado pela maior concentração de domicílios de baixa renda,

esgoto a céu aberto, abastecimento por água não encanada e lixo acumulado nesses locais. Os setores censitários com um alto ISEF apontam os locais que possuem maior relação com a ocorrência de escorregamentos e inundações na área de estudo; nota-se que o trecho ferroviário entre as estações Francisco Morato e Baltazar Fidélis possui uma maior concentração de setores com interferência externa alta no entorno (Figura 22).

4.14.4 Comparação entre os métodos

O ISEF calculado pelos *outliers* possui a maior área da classe de alta de interferência socioeconômica à ferrovia e também a maior concentração de escorregamentos na classe alta do índice; porém, o maior número de ocorrências de inundações está presente na classe média do índice de interferência externa calculado pela média aritmética. Vale ressaltar que há uma diferença de escala entre os dados; portanto o resultado não pode ter 100% de precisão, mas é possível se ter uma ideia da distribuição das ocorrências de escorregamentos e de inundações em cada classe do ISEF. A classe alta do ISEF calculado pela média e pelos *outliers* coincidem com todos os setores censitários que possuem vulnerabilidade social muito alta (aglomerados urbanos), enquanto a classe alta do ISEF dos *outliers* corresponde com uma maior proporção de setores com vulnerabilidade social alta (urbanos).

Ambos apresentam uma maior concentração de setores censitários com alto ISEF entre as estações Baltazar Fidélis e Francisco Morato, enquanto os ISEF calculados pelos *outliers* e pela ACP mostram uma maior concentração de setores com baixo ISEF entre as estações Caieiras e Franco da Rocha. Como a ACP está invertida, porque se trata da 1ª CP, no ISEF da ACP os setores em amarelo (classe média) ficariam iguais ao verde (classe baixa) e o vermelho, igual aos locais em amarelo. No ISEF de *outliers*, à exceção de área inferior à esquerda que é amarela no de médias, já está contido no ISEF de médias (o ISEF de médias corresponde ao ISEF de *outliers* mais outras áreas), porém o índice de *outliers* registra os piores casos, ou seja, uma possível intervenção começaria nesses setores censitários, mesmo na área não indicada no índice de médias, mas que tem valor médio no índice da ACP. Em relação à ACP, ela priorizou as áreas com esgoto à céu aberto e as áreas de baixa renda que, dependendo do parâmetro, pode ajudar a estabelecer um critério para intervenção, considerando esses 2 indicadores como prioritários.

Em suma, os locais da classe alta do ISEF estão associados ao acúmulo de lixo nas ruas, ao não tratamento de esgoto, ao não-abastecimento de água via geral de distribuição e às moradias de baixa renda, que, segundo a literatura, são fatores que podem

contribuir para a ocorrência de escorregamentos e inundações, principalmente em ocupações de áreas de maior suscetibilidade; e que também podem interferir na via férrea (Figura 22). Ou seja, além dos passageiros da CPTM estarem expostos a transtornos durante a viagem, como atrasos e até paralisação do trem (que também podem ser causados por um escorregamento que afeta a linha de eletricidade da ferrovia), eles podem ter que se arriscar ao enfrentar uma inundação insalubre, por causa do lixo e esgoto aberto, sendo capaz de até adquirirem doenças, como leptospirose.

Todos os métodos utilizados são válidos e proporcionaram resultados de IIEF semelhantes, mas caso seja necessário escolher apenas um deles, recomenda-se, por precaução, escolher o método que apresentou o pior cenário; assim não haverá o risco de uma área com alta interferência socioeconômica ser considerada como uma área de baixa interferência. O método que apresentou o pior cenário é o IIEF calculado com os *outliers*, pois ele apresenta uma maior proporção de área de classes altas de interferência externa, uma maior concentração de escorregamentos na classe alta do índice, e também uma maior sobreposição da classe alta do índice de interferência com áreas de alta e muito alta vulnerabilidade social. Os três métodos do IIEF também foram inseridos no GEE e apresentaram visualização semelhante ao do *software* de SIG (Figura 21).

Figura 21 - Exemplo do IIEF, método dos *outliers*, no GEE.

Fonte: elaboração própria, 2021.

Figura 22 - Comparação do IISEF para cada método.

Fonte: elaboração própria, 2021.

4.15 Índice de Vulnerabilidade Social à Ocorrência de Desastres Naturais (IVSDN)

4.15.1 Média Aritmética Simples

O Índice de Vulnerabilidade Social à Ocorrência de Desastres Naturais calculado através do método da média aritmética simples mostra que a maior parte da área de estudo (43%) é constituída pelas classes de média vulnerabilidade social, seguido pelas classes de alta vulnerabilidade (40%) e de baixa vulnerabilidade (17%). Em relação aos pontos de inundações, 74% sobrepõem a classe média do IVSDN, 16% correspondem a classe baixa e 10%, a classe alta. Quanto aos eventos de escorregamentos, 46% ocorrem na classe alta do índice de vulnerabilidade social, 40% intersectam a classe média do IVSDN e 14% coincidem com a classe baixa.

Identificou-se também que a classe alta do IVSDN coincide com todos os aglomerados urbanos subnormais da área de estudo. Os setores censitários com um alto IVSDN apontam os locais que possuem maior relação com a ocorrência de escorregamentos e inundações na área de estudo; nota-se que o trecho ferroviário entre as estações Franco da Rocha e Francisco Morato sobrepõe com muitas classes de alta vulnerabilidade social, o que sentido é coerente já que esse trecho é frequentemente destacado na mídia pelos transtornos causados aos passageiros da CPTM devido à ocorrência de escorregamentos e inundações (Figura 25).

4.15.2 *Outliers*

Para identificar os *outliers*, primeiro observou-se os *boxplots* de cada índice (Figura 23). No Indicador de Idosos, os *outliers* são os setores censitários com os valores de índices entre 0,4 e 1; o Indicador mostra que os *outliers* são os setores com valor igual a 1; no Indicador de Sem Moradia Própria, os *outliers* referem-se aos setores censitários com valor de índice a partir de 0,5 a 1; o Indicador de Analfabetismo mostra que a grande maioria dos setores possuem moradores alfabetizados, ou seja, os *outliers* indicam os lugares onde não há pessoas alfabetizadas; e o Indicador de Baixa Renda não possui *outliers*, portanto ele não foi incluído neste método.

Figura 23 - *Boxplots* dos indicadores que compõem o IVSDN.

Elaborado por: Michelle Trintinalia, 2020.

O Índice de Vulnerabilidade Social calculado através do método de *outliers* mostra que a maioria da área de estudo (80%) é composta pelas classes de baixa vulnerabilidade social, seguido pelas classes de alta vulnerabilidade (11%) e de média vulnerabilidade (9%). Em relação aos pontos de inundações, 53% sobrepõem a classe baixa do IVSDN, 42% correspondem à classe média e 5%, a classe alta de interferência. Quanto aos eventos de escorregamentos, 85% ocorrem na classe baixa do índice de vulnerabilidade, 12% intersectam a classe média do IVSDN e 3% coincidem com a classe alta.

Identificou-se também que 30% dos aglomerados urbanos subnormais da área de estudo coincidem com a classe alta do IVSDN. Os setores censitários com maior IVSDN indicam os lugares que têm uma população mais vulnerável a escorregamentos e inundações na área de estudo; percebe-se que os setores de alto IVSDN estão próximos da Estação Caieiras e da Estação Francisco Morato (Figura 25).

4.15.3 Análise de Componente Principal de Objetos Espaciais

O percentual das variâncias nos mostrou que a primeira componente principal - CP1 explica 65% da variabilidade dos dados, seguida pela segunda componente principal - CP2 (15%), pela terceira componente principal - CP3 (11%), pela quarta componente

– CP4 (6%) e pela quinta componente – CP5 (3%). A matriz de correlação dos dados nos permitiu observar que as variáveis que têm maior relação entre si são: Indicador de Crianças com Indicador de Baixa Renda (0,92), Indicador de Crianças e Indicador de Residentes Sem Moradia Própria (0,55), e Indicador de Baixa Renda e Indicador de Residentes Sem Moradia Própria (0,51).

O Quadro 5 mostra os resultados das cinco componentes principais. Dentre todas as componentes, a CP3 foi escolhida para a elaboração do IVSDN, pois os seus resultados são mais interessantes para este estudo do que as outras componentes. A CP3 indica que os setores censitários que apresentam um maior número de analfabetos, também têm uma maior concentração de domicílios não-próprios; o que faz sentido já que os analfabetos têm menos chances de conseguirem empregos melhores, portanto recebem salários menores, o que dificulta a aquisição de uma casa própria.

Quadro 5 - Matriz das componentes principais das variáveis do IVSDN.

	CP1	CP2	CP3	CP4	CP5
Ind. Idosos	-0,062	0,503	-0,411	-0,742	0,151
Ind. Crianças	-0,532	-0,041	0,019	-0,162	-0,830
Ind. Baixa Renda	-0,783	-0,225	0,203	-0,091	0,536
Ind. Sem Moradia Própria	-0,313	0,433	-0,555	0,636	0,042
Ind. Analfabetismo	0,044	-0,712	-0,694	-0,100	0,011

Elaborado por: Michelle Trintinalia, 2020.

O Índice de Vulnerabilidade calculado através do método da ACP mostra que a maior parte da área de estudo (67%) é constituída pelas classes de média vulnerabilidade social, seguido pelas classes de alta vulnerabilidade (19%) e de baixa interferência (14%). Em relação aos pontos de inundações, 78% sobrepõem a classe média do IVS, enquanto 11% correspondem à classe alta e 11%, à classe baixa. Quanto aos eventos de escorregamentos, 77% ocorrem na classe média do índice de vulnerabilidade, 19% intersectam a classe baixa do IVSDN e 4% coincidem com a classe alta.

Identificou-se também que todos aglomerados urbanos subnormais da área de estudo coincidem com a classe alta do IVSDN; isso pode estar relacionado a maior concentração de analfabetismo, de domicílios não-próprios e de baixa renda nesses locais. Os setores censitários com um alto IVSDN apontam os locais que possuem maior vulnerabilidade aos eventos de escorregamentos e inundações na área de estudo; nota-se

que o trecho ferroviário entre as estações Francisco Morato e Baltazar Fidélis possui uma maior concentração de setores com vulnerabilidade social alta no entorno (Figura 25).

4.15.4 Comparação entre os métodos

O IVSDN calculado pela média aritmética simples possui a maior área da classe de alta de vulnerabilidade social; a maior concentração de inundações ocorre na classe média do IVSDN do método da ACP e a maior proporção de escorregamentos está presente na classe baixa do IVSDN calculado pelos *outliers*. É importante lembrar que esta pesquisa não obteve até o momento os dados de ocorrências do município de Franco da Rocha; portanto, esse resultado é uma visão parcial da distribuição dos eventos de escorregamentos e de inundações na área de estudo, mas é possível se ter uma noção de proporção de escorregamentos e inundações em cada classe do IVSDN.

A classe alta do IVSDN calculado pela média aritmética e pela componente principal coincidem com todos os setores censitários que possuem o IPVS (do SEADE) muito alto (aglomerados urbanos). O IVSDN dos métodos da média e da ACP apresentam uma maior concentração de setores censitários com alto IVSDN entre as estações Baltazar Fidélis e Francisco Morato, enquanto o IVSDN calculado pelos *outliers* mostram uma maior concentração de setores com baixo IVSDN por quase toda a linha férrea. O IVSDN dos *outliers* está diferente dos outros métodos provavelmente porque há poucos *outliers*, sendo que todos os dados não-*outliers* foram considerados como baixa vulnerabilidade; além disso, esse método não considerou o Indicador de Baixa Renda, pois este não apresentou *outliers*.

A ACP por sua vez, priorizou as áreas com crianças e as áreas de baixa renda que dependendo do parâmetro, pode ajudar a estabelecer um critério para intervenção, considerando esses 2 indicadores como prioritários.

Todos os métodos utilizados são válidos, mas somente os métodos de média aritmética e ACP proporcionaram resultados de IVSDN semelhantes. Caso seja necessário escolher apenas um deles, recomenda-se, por precaução, escolher o método que apresentou o pior cenário; assim não haverá o risco de uma área com alta vulnerabilidade social ser considerada como uma área de baixa vulnerabilidade. O método que apresentou o pior cenário é o IVSDN calculado pelo método da média aritmética simples, pois ele apresenta uma maior proporção de área de classes alta de

vulnerabilidade social e uma grande sobreposição da classe alta do IVSDN com aglomerados subnormais (Figura 25).

Em suma, os locais da classe alta do IVSDN estão associados à maior concentração de idosos, de crianças, de analfabetos, de moradias não-próprias e de domicílios de baixa renda, que, segundo a literatura, são grupos mais vulneráveis à ocorrência de escorregamentos e inundações, pois possuem maior dificuldade de evacuação e acessibilidade a informações durante um evento de desastre e maior dificuldade de recuperação financeira no pós-desastre. O mapeamento desses dados revela padrões geográficos de população potencialmente vulneráveis aos desastres e podem ser usados na mitigação, preparação, resposta e recuperação (MORROW, 1999). Os três métodos do IVSDN também foram inseridos no GEE e apresentaram visualização semelhante ao do *software* de SIG (Figura 24).

Figura 24 - Exemplo do IVSDN, método da média aritmética, no GEE.

Fonte: elaboração própria, 2021.

Figura 25 - Comparação do IVSDN para cada método.

Fonte: elaboração própria, 2021.

4.16 Inserção e visualização dos dados de um ambiente SIG para um ambiente *Big Data*

4.16.1 GeoMesa: visualização dos dados e dificuldades encontradas

Em relação às dificuldades encontradas na passagem dos dados de um ambiente SIG para um ambiente *Big Data*, no GeoMesa, há a passagem de um dado no formato *shapefile*, onde as colunas não podem ter o nome “ID” na tabela de atributos, sendo assim, as colunas IDs dos *shapefiles* tiveram que ser renomeadas; no *.csv* deve-se configurar um conversor e uma nova tabela de atributos de acordo com o GeoMesa antes de subi-lo no GeoServer; e para os arquivos WMS e GeoTiff é necessário um maior conhecimento em programação para se fazer a inserção dos dados. Outro problema encontrado relaciona-se com o ambiente *cluster*, que é um equipamento caro para se adquirir, não é tão fácil de encontrar e necessita de conhecimento técnico para usá-lo. Portanto, a inserção dos dados do GeoMesa foi realizada em *single node*, apresentando uma menor capacidade de armazenamento e processamento de dados comparado a um *cluster* (CORREIA *et al.* 2020).

Todos os dados nos formatos *.shp* e *.csv* foram visualizados no *Geoserver* (Figura 26 a seguir). O GeoMesa apresentou uma grande capacidade de armazenamento dos dados (foi possível subir para o *Geoserver* dados representados em todo território brasileiro) e com uma ótima velocidade de processamento. 110 *layers* (67% dos dados totais) foram transferidos para esse tipo de ambiente *Big Data* geoespacial. Não se percebeu nenhuma modificação nos dados espaciais depois da passagem para o GeoMesa (como distorções, perda de georreferenciamento, problemas no sistema de projeção etc).

Figura 26 - Visualização de parte dos dados em SIG no GeoMesa/GeoServer.

Elaborado por: Lucas Sandre, 2020.

4.16.2 GEE: visualização dos dados e dificuldades encontradas

No GEE, para obter a visualização dos dados inseridos, é necessário utilizar a função *Map.addLayer* através do comando “*Map.addLayer (nome da variável, {parâmetros de visualização}, ‘nome do layer’)*” (GEE, 2020). Em relação às dificuldades encontradas no GEE, por enquanto não foi possível fazer a inserção de um WMS na plataforma; quanto as informações tipo .csv, foi possível transferir o dado, porém é um processo extremamente trabalhoso fazer com que o GEE reconheça a geometria da feição (ponto, linha e polígono) e a posição geográfica do dado (latitude e longitude); portanto, é mais rápido e simples converter o .csv em .shp para subi-lo na ferramenta. Outro problema encontrado refere-se a impressão do console do editor de código limitada a 5.000 elementos, ou seja, os dados que apresentavam mais de 5.000 *layers* (por exemplo, eixos de logradouro da RMSP) não foram representados em sua totalidade; nesses casos, foi necessário filtrar apenas aqueles elementos que intersectavam na área de estudo (CORREIA *et al.* 2020).

Somente foi possível realizar a passagem dos dados .shp e .tif para o GEE, representando 76% dos dados totais (125 camadas) (Figura 27 a seguir). Vale ressaltar que dados como imagens Landsat, Sentinel e os dados de cobertura do solo do Mapbiomas não foram exatamente importados; estes foram carregados diretamente do GEE, pois se encontram disponíveis na plataforma e o GEE permite um upload de Geotiff até 10 GB. Assim como o GeoMesa, não se notou nenhuma alteração nos dados espaciais após a passagem para o GEE.

Figura 27 - Visualização de alguns dados em SIG no GEE.

Fonte: elaboração própria, 2020.

Estes resultados foram apresentados no relatório “Do SIG ao *Big Data*: Estruturação dos Dados Espaciais do Entorno das Estações Ferroviárias Caieiras e Francisco Morato, SP” no 34º ANPET e publicado nos anais do evento em 14 de dezembro de 2020 (Anexo H).

5 CONCLUSÕES E RECOMENDAÇÕES

Nesta Dissertação percebeu-se que a Linha 7 – Rubi da CPTM está em sua maior parte sobre zonas de baixa suscetibilidade à escorregamentos, mas com áreas de alta suscetibilidade no seu entorno, enquanto a maioria da via férrea está sobre áreas de alta suscetibilidade a inundações, o que explica a ocorrência destes tipos de desastres naturais na região, que frequentemente afetam a ferrovia. Esta pesquisa também identificou que a todos os escorregamentos da área de estudo ocorreram em área urbana e de pastagem, que pode estar relacionado à um planejamento inadequado de uso e ocupação do solo. Notou-se também que todas as inundações ocorreram em área urbana e de pastagem, o que pode estar associado à impermeabilização dos solos.

Este trabalho criou dois índices: um Índice de Interferência Social Externa à Ferrovia (IISFE), para avaliar quais fatores socioeconômicos podem contribuir para a ocorrência de desastres naturais no entorno da via férrea, e um Índice de Vulnerabilidade Social à Ocorrência de Desastres Naturais (IVSDN), para se obter um perfil da população mais afetada pelos eventos de escorregamentos e inundações. Os índices foram calculados e comparados a partir de três métodos: média aritmética simples, *outliers* e análise de componente principal (ACP).

O método *outliers* proporcionou um melhor resultado do IISFE, mostrando as regiões com maior interferência externa à escorregamentos e inundações ao redor da Linha 7 da CPTM, que indicam lixo acumulado nos logradouros, presença de esgoto à céu aberto, abastecimento de água sem ser da rede de distribuição geral e domicílios de baixa renda. Esses fatores, que associados a ocupação de áreas de risco, urbanização acelerada e a chuvas intensas, podem colaborar para a ocorrência de escorregamentos e inundações.

Em contrapartida, o método da média aritmética gerou um melhor resultado do IVSDN, indicando os locais com maior concentração de cidadãos mais vulneráveis aos eventos de desastres naturais no entorno da Linha Rubi, como idosos, crianças,

analfabetos, pessoas sem moradia própria e domicílios de baixa renda. Essas variáveis estão relacionadas com a dificuldade desses grupos evacuarem, de ter acesso à informação durante um evento de desastre e de se recuperarem financeiramente em um pós-desastre.

Em suma, a população que já é carente de recursos financeiros, de saneamento básico, é ainda mais afetada a cada evento que interrompe a linha férrea. Os passageiros da CPTM ficam sem transporte, terão que andar pela via férrea, se expondo mais ainda a escorregamentos, a inundações insalubres etc. O ISEF e o IVSDN podem auxiliar os órgãos municipais e estaduais (como a Defesa Civil e o IPT) preocupados com todas as fases do ciclo de desastre natural.

Os índices foram elaborados em ambiente SIG e depois inseridos em ambiente *Big Data* Geoespacial. Uma desvantagem desses índices é que se utilizou os dados do Censo 2010 do IBGE, que, apesar de ser a informação mais recente desse tipo, já faz mais de 10 anos que foram elaborados. Portanto, esses índices podem não condizer com a realidade atual, porém eles nos permitem obter uma estimativa da situação socioeconômica da região. Outro fator negativo era a indisponibilidade de informações na escala dos setores censitários, como o número de moradores que vivem em unidades residenciais superlotadas, o número de pessoas vulneráveis que precisam de assistência social e de assistência previdenciária, número de residentes que possuem algum tipo de deficiência física, portanto teriam uma maior vulnerabilidade aos eventos de desastres etc. Houve a indisponibilidade de alguns dados utilizados, nem todos eram públicos, como as ocorrências de escorregamentos e de inundações da Defesa Civil Municipal de Caieiras e de Franco da Rocha e da CPTM.

Esta Dissertação demonstrou como uma abordagem mais holística da avaliação da vulnerabilidade social, que utiliza dados socioeconômicos e espaciais, e pode apontar quais comunidades são mais vulneráveis aos eventos de desastres naturais. Conhecendo a localização dos lugares com maior interferência externa e das comunidades socialmente vulneráveis, os gestores podem direcionar e apoiar de modo mais eficaz os esforços comunitários para mitigar e se preparar para as ocorrências de escorregamentos e inundações. As fatalidades e os impactos na comunidade da área de estudo podem ser reduzidos se as autoridades se prepararem para a evacuação daqueles que estão em risco iminente.

A maioria das análises foram feitas em ambiente SIG, mas o propósito era a implementação de um ambiente *Big Data* estruturado que esbarrou em problemas como uso do *software* livre, que requer várias configurações específicas, como um *cluster*, que na maioria das vezes é pago e não é barato, e como conhecimento técnico nessa área; ainda é difícil encontrar profissionais com esse conhecimento e, por enquanto, é uma área bem complexa para programadores iniciantes, além de demandar tempo para se estruturar um banco de dados nesse tipo de ambiente. Até o momento, apesar da dedicação e da qualificação da equipe técnica, o IPT ainda não conseguiu realizar a estruturação de dados em *Big Data* a tempo da defesa desta Dissertação. Esta pesquisa enfrentou dificuldades na compatibilidade entre os dados: dados de diversos formatos, de diferentes anos, de escalas distintas, de diferentes sistemas de referência, de legendas distintas e problemas de qualidade na tabela de atributos dos dados.

Uma forma de contornar esse problema era inserir os dados analisados no SIG para as ferramentas de *Big Data* Geoespacial: GEE e GeoMesa. O GeoMesa apresentou uma maior capacidade para comportar uma maior quantidade de dados, mas o GEE aceitou uma maior quantidade de formatos diferentes de dados. Porém, no GeoMesa o processamento de dados dessa pesquisa poderia ter sido mais eficiente e rápido se tivéssemos tido acesso a um *cluster*.

Esta pesquisa ressalta que utilizou, em sua maioria, *softwares* gratuitos (QGIS, R Studio, GEE e GeoMesa) para mostrar que qualquer pessoa (independentemente de sua situação financeira), as secretarias, prefeituras e outras de instituições, podem realizar esse tipo de análise. Sugere-se buscar por outras ferramentas de *Big Data* e aproveitar mais as vantagens que o *Big Data* pode proporcionar ao estudo, procurando dar preferência a *softwares* gratuitos e livres.

Além disso, esta pesquisa constatou que faltam referências sobre o assunto, principalmente no Brasil. Novos estudos sobre esse tema seriam de grande utilidade para a mobilidade urbana, pois principalmente na Região Metropolitana de São Paulo há ocorrências de escorregamentos e inundações nas ferrovias, que interrompem a circulação dos trens.

Recomenda-se o desenvolvimento de novas pesquisas em outros municípios (principalmente entorno de feições lineares, como ferrovias, rodovias, hidrovias, linhas de transmissão de energia etc.), em outros anos, utilizando dados de relevo, modelos

hidrológicos, usando outras variáveis para calcular os índices, elaborando outros métodos e outras ferramentas de *Big Data* podem ser utilizadas. Após a finalização deste estudo, pretende-se inserir os resultados dessa pesquisa na Infraestrutura de Dados Espaciais Ambientais do Estado de São Paulo (DataGeo) ou no Repositório de dados científicos da USP.

6 REFERÊNCIAS

AJI, A.; WANG, F.; VO, H.; LEE, R.; LIU, Q.; ZHANG, X.; SALTZ, J. Hadoop-GIS: A High Performance Spatial Data Warehousing System over MapReduce. **Proceedings VLDB Endowment**, v. 6, n. 11, p. 1009-1020, 2013.

ALLAM, Z.; DHUNNY, A. On big data, artificial intelligence and smart cities. **Cities**, v. 89, p. 80-91, 2019.

ALCÁNTARA-AYALA, I. Geomorphology, natural hazards, vulnerability and prevention of natural disasters in developing countries. **Geomorphology**, v. 47. p. 107 – 124, 2002.

AL-SAI, Z. A.; ABDULLAH, R.; HUSIN, M. H. Big Data Impacts and Challenges: A Review. **IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology**, p. 150-155, 2019.

AL-SAI, Z. A.; ABUALIGAH, L. M. Big data and E-government: A review. **8th International Conference on Information Technology**, p. 580-587, 2017.

AMORIM, D., FUSCO, M., LAPA, D., SOUZA, C.B. Gestão de Processos de Inundação, Enchente e Movimentação de Solo na Ferrovia. 25ª Semana de Tecnologia Metroferroviária, **AEAMESP**, São Paulo, 2019.

ANEJIONU, O. C. D.; THAKURIAH, P. V.; MCHUGH, A.; SUN, Y.; MCARTHUR, D.; MASON, P.; WALPOLE, R. Spatial urban data system: A cloud-enabled big data infrastructure for social and economic urban analytics. **Future Generation Computer Systems**, v. 98, p. 456-473, 2019.

APACHE ACCUMULO. **Apache Accumulo® is a sorted, distributed key/value store that provides robust, scalable data storage and retrieval**, 2020. Disponível em: <<https://accumulo.apache.org/>>. Acesso em: 12 mar. 2020.

APACHE CASSANDRA. **Installing Cassandra**, 2016. Disponível em: <https://cassandra.apache.org/doc/latest/getting_started/installing.html#prerequisites>. Acesso em: 01 ago. 2020.

APACHE HADOOP. **Apache Hadoop**, 2020. Disponível em: <<http://hadoop.apache.org/>>. Acesso em: 08 mar. 2020.

ARCGIS PRO. **Data classification methods**, 2020. Disponível em: <[https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/data-classification-methods.htm#:~:text=Natural%20breaks%20\(Jenks\),-With%20natural%20breaks&text=Class%20breaks%20are%20created%20in,differences%20in%20the%20data%20values](https://pro.arcgis.com/en/pro-app/latest/help/mapping/layer-properties/data-classification-methods.htm#:~:text=Natural%20breaks%20(Jenks),-With%20natural%20breaks&text=Class%20breaks%20are%20created%20in,differences%20in%20the%20data%20values)>. Acesso em: 30 dez. 2020.

ASSEMBLEIA LEGISLATIVA DO ESTADO DE SÃO PAULO. Decreto N° 64.592, de 14 de Novembro de 2019, 2019. Disponível em: <<https://www.al.sp.gov.br/repositorio/legislacao/decreto/2019/decreto-64592-14.11.2019.html>>. Acesso em: 16 dez. 2020.

AVILA, L. O. **Vulnerabilidade das áreas sob ameaça de desastres naturais na cidade de Santa Maria/RS**. 2015. Tese (Doutorado em Geografia) – Instituto de Geociências, Universidade de Federal do Rio Grande do Sul, Porto Alegre.

AZEVEDO, C. S.; CORREIA, T. P.; CORSI, A. C.; SPINOLA, M. M.; QUINTANILHA, J. A. A Systematic Review: Big Data Analytics and Landslides. **SCG-XIII International Symposium On Landslides**, Cartagena, p. 1-8, 2020.

BACHIEGA, J.; REIS, M. A. S.; HOLANDA, M.; ARAUJO, A. P. F. An Architecture for Cost Optimization in the Processing of Big Geospatial Data in Public Cloud Providers. **2018 IEEE International Congress on Big Data (Big Data Congress)**, San Francisco, CA, p. 190-197, 2018.

BADEA, A. C.; BADEA, G. The advantages of creating compound GIS functions for automated workflow. **Geology & Mining Ecology Management**, v. 1, p. 943-949, 2013.

BAIG, F.; VO, H.; KURC, T.; SALTZ, J.; WANG, F. SparkGIS: Resource Aware Efficient In-Memory Spatial Query Processing. **Proc ACM SIGSPATIAL Int Conf Adv Inf.** 2017, p; 1-26, 2017.

BARRECA, A., CURTO, R., ROLANDO, D. Assessing social and territorial vulnerability on real estate submarkets. **Buildings**, v. 7, n. 4, 2017.

BEDNARIK, M.; MAGULOVÁ, B.; MATYS, M.; MARSCHALKO, M. Landslide susceptibility assessment of the Kraľovany - Liptovský Mikuláš railway case study. **Physics and Chemistry of the Earth**, v. 35, p. 162-171, 2010.

BODSTEIN, A.; LIMA, V. V. A.; BARROS, A. M. A. A vulnerabilidade do idoso em situações de desastres: necessidade de uma política de resiliência eficaz. **Ambiente & Sociedade**, v. 17, n. 2, 2014.

BRAGHIROLI, G. **Mapeamento 3D de Áreas de Risco a Inundação a Partir de Dados Laser Escâner**. 2017. Dissertação (Mestrado em Planejamento Territorial e Desenvolvimento Socioambiental) – Centro de Ciências Humanas e da Educação, Universidade do Estado de Santa Catarina, Florianópolis.

BRASIL. Senado Federal. **Atividade Legislativa: Constituição Federal de 1988 – Art. 7º**. Disponível em: <https://www.senado.leg.br/atividade/const/con1988/con1988_02.07.2020/art_7_.asp>.

BRASIL. Casa Civil da Presidência da República. **Lei nº 12.651, de 25 de maio de 2012**. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2012/lei/112651.htm>.

BRASIL. Ministério da Integração Nacional, Secretaria Nacional de Proteção e Defesa Civil, Centro Nacional de Gerenciamento de Riscos e Desastres. **Anuário brasileiro de desastres naturais**: 2013. Brasília: CENAD, 2014. Disponível em: <http://www.mi.gov.br/c/document_library/get_file?uuid=fee4007a-ab0b-403e-bb1a-8aa00385630b&groupId=10157>.

CÂMARA MUNICIPAL DE FRANCISCO MORATO. Aspectos gerais, 2020. Disponível em: <<http://www.camarafranciscomorato.sp.gov.br/site/index/aspectosgerais>>. Acesso em: 01 nov. 2020.

CAMARINHA, P. I. M. **Vulnerabilidade aos Desastres Naturais decorrentes de Deslizamentos de Terra em Cenários de Mudanças Climáticas na Porção Paulista da Serra do Mar**. 2016. Tese (Doutorado em Ciência do Sistema Terrestre) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos.

CAREY, V. J.; WALTERS, E. E.; WAGER, C. G.; ROSNER, B. A. Resistant and Test-Based Outlier Rejection: Effect on Gaussian One- and Two-Sample Inference, **Technometrics**, v. 39, p. 320–330, 1997.

CASTRO, J. P. C. **Analisando Sistemas Analíticos Espaciais Baseados em Hadoop e Spark: Uma Perspectiva de Usuário**. 2019. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.

CEMADEN - Centro Nacional de Monitoramento e Alertas de Desastres Naturais. **Missão**, 2016. Disponível em: <<http://www.cemaden.gov.br/missao-do-cemaden/>>. Acesso em: 15 dez. 2020.

CENAD - CENTRO NACIONAL DE GERENCIAMENTO DE RISCOS E DESASTRES. **Centro Nacional de Gerenciamento de Riscos e Desastres – CENAD**, 2019. Disponível em: <<https://antigo.mdr.gov.br/protecao-e-defesa-civil/centro-nacional-de-gerenciamento-de-riscos-e-desastres-cenad>>. Acesso em: 15 dez. 2020.

CHAN, F.K.S.; GRIFFITHS, J.A.; HIGGITT, D.; XU, S.; ZHU, F.; TANG, Y.-T.; XU, Y.; THORNE, C.R. “Sponge City” in China—A breakthrough of planning and flood risk management in the urban context. **Land Use Policy**, v. 76, p. 772-778, 2018.

CHANG, F.; DEAN, J.; GHEMAWAT, S.; HSIEH, W. C.; WALLACH, D. A.; BURROWS, M.; CHANDRA, T.; FIKES, A.; GRUBER, R. E. Bigtable: A Distributed Storage System for Structured Data. **ACM Transactions on Computer Systems**, v. 26, n. 2, 2008.

CHEETHAM, M.; CHIROUZE, F.; BREDIER, L. RISK VIP: Evaluation of Flood Risk on the French Railway Network Using an Innovative GIS Approach. **E3S Web of Conferences**, v. 7, n. 10004, 2016.

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

CIAN, F.; MARCONCINI, M.; CECCATO, P. Normalized Difference Flood Index for rapid flood mapping: Taking advantage of EO big data. **Remote Sensing of Environment**, v. 209, p. 712-730, 2018.

CLARIVATE ANALYTICS. **Opções de classificação**, 2019. Disponível em: <https://images.webofknowledge.com/WOKRS514B4/help/pt_BR/WOK/hs_sort_options.html>. Acesso em: 29 dez. 2019.

CNM – CONFEDERAÇÃO NACIONAL DE MUNICÍPIOS. **Decretações de anormalidades causadas por desastres nos Municípios Brasileiros**. Brasília, 2018. Disponível em: <<https://www.cnm.org.br/cms/biblioteca/documentos/Decretacoes-de-anormalidades-causadas-por-desastres-nos-Municipios-Brasileiros-10-10-2018-v2.pdf>>.

CORRÊA, C. V. S. A Geotecnologia no Monitoramento, Alerta e Prevenção de Desastres - Os Desastres Naturais e Tecnológicos no Cenário Brasileiro e Estudos de Caso do Emprego de Geotecnologias como Subsídio a sua Mitigação. In: MAGNONI JUNIOR, L. *et al.* **Redução do risco de desastres e a resiliência no meio rural e urbano**. São Paulo: CPS, 2020, 865 p.

CORREIA, T. P.; CORSI, A. C.; QUINTANILHA, J. A. Big Data for Natural Disasters in an Urban Railroad Neighborhood: A Systematic Review. **Smart Cities**, v 3, n. 2, p. 202–211, 2020.

CORREIA, T. P.; QUINTANILHA, J. A.; CORSI, A. C.; SANDRE, L. H. Do SIG ao *Big Data*: Estruturação dos Dados Espaciais do Entorno das Estações Ferroviárias Caieiras e Francisco Morato, SP. **34º Congresso de Pesquisa e Ensino em Transporte da ANPET**, p. 294-298, 2020.

CPRM - COMPANHIA DE PESQUISA DE RECURSOS MINERAIS/SERVIÇO GEOLÓGICO DO BRASIL. **Setorização de Risco Geológico**, 2020. Disponível em: <<http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-de-Desastres/Setorizacao-de-Risco-Geologico-5389.html>>. Acesso em: 15 dez. 2020.

CPTM - COMPANHIA PAULISTA DE TRENS METROPOLITANOS. **A CPTM – Código de Conduta e Integridade**, 2018a. Disponível em: <<http://www.cptm.sp.gov.br/a-companhia/Pages/a-companhia.aspx>>. Acesso em: 29 mai. 2018.

CPTM - COMPANHIA PAULISTA DE TRENS METROPOLITANOS. **CPTM entrega à população a nova Estação Francisco Morato**, 2020. Disponível em: <[http://www.cptm.sp.gov.br/noticias/Pages/CPTM-entrega-%C3%A0-](http://www.cptm.sp.gov.br/noticias/Pages/CPTM-entrega-%C3%A0)

popula%C3%A7%C3%A3o-a-nova-Esta%C3%A7%C3%A3o-Francisco-Morato.aspx>. Acesso em: 03 nov. 2020.

CPTM - COMPANHIA PAULISTA DE TRENS METROPOLITANOS. **Mapa da Rede**, 2018c. Disponível em: <<http://www.cptm.sp.gov.br/publicidade/Pages/Mapa-da-Rede.aspx>>. Acesso em: 29 mai. 2018.

CPTM - COMPANHIA PAULISTA DE TRENS METROPOLITANOS. **Nossa História**, 2018b. Disponível em: <<http://www.cptm.sp.gov.br/a-companhia/Pages/Nossa-Historia.aspx>>. Acesso em: 29 mai. 2018.

CRED/USAID - CENTRE FOR RESEARCH ON THE EPIDEMIOLOGY OF DISASTERS/ UNITED STATES AGENCY FOR INTERNATIONAL DEVELOPMENT. **Annual Disaster Statistical Review 2017**, 2018. Disponível em: <<https://www.cred.be/annual-disaster-statistical-review-2017>>. Acesso em: 04 mar. 2019.

CUTTER, S. L. GI Science, Disaster, and Emergency Management. **Transactions in GIS**, v. 7, n. 4, p. 439–445, 2003.

CUTTER, S. L.; BORUFF, B. J.; SHIRLEY, W. L. Social Vulnerability to Environmental Hazards. **Social Science Quarterly**, v. 84, n. 2, p. 242-261, 2003.

CYNAMON, S. E.; MONTEIRO, T. C. N. Solução para remoção de lixo nas favelas: um projeto de estudo. **Cad. Saúde Pública**, v.1, n. 1, 1985.

DAVIS, C.; CÂMARA, G. Arquitetura de Sistemas de Informação Geográfica. In:_____. **Introdução à Ciência da Geoinformação**. INPE, São José dos Campos, 2011, 345 p.

DE FREITAS, C. M.; XIMENES, E. F. Enchentes e saúde pública – uma questão na literatura científica recente das causas, consequências e respostas para prevenção e mitigação. **Rev. Ciência & Saúde Coletiva**, v. 17, n. 6, p. 1601-1615, 2012.

DEMŠAR, U.; HARRIS, P.; BRUNSDON, C.; FOTHERINGHAM, A. S.; MCLOONE, S. Principal Component Analysis on Spatial Data: An Overview. **Annals of the Association of American Geographers**, v. 103, n. 1, p. 106-128, 2013.

DE OLIVEIRA, G. G.; FLORES, T.; BRESOLIN JUNIOR, N. A.; HAETINGER, C.; ECKHARDT, R. R.; QUEVEDO, R. P. Análise da suscetibilidade a inundações e

enxurradas na Bacia Hidrográfica do Rio Forqueta, RS, Brasil. **Geociências**, v. 37, n. 2, p. 437 – 453, 2018.

DO AMARAL, R.; RIBEIRO, R. R. **Inundações e enchentes**. In: TOMINAGA, L. K.; SANTORO, J.; AMARAL, R. (Org.). Desastres Naturais: conhecer para prevenir. São Paulo: Instituto Geológico, 2009. p. 39-52.

DOMENEGHETTI, A.; SCHUMANN, G. J. P.; TARPANELLI, A. Preface: Remote Sensing for Flood Mapping and Monitoring of Flood Dynamics. **Remote Sens.**, v. 11, n. 8, p. 2 – 4, 2019.

DOS SANTOS, D. N.; DE OLIVEIRA, M. A. M.; FORMIGONI, A.; MAIELLARO, J. R.; SAAD, A. R.; FERNANDES, M. E. River Network and the Risk of Flooding for the Railway and Subway Systems in the City of São Paulo, Brazil. **Jourof Geo. Res. and Eng.**, v. 7, p. 328-334, 2016.

DUNNING, M.C. **Social vulnerability analysis methods for corps planning**, 2009. Disponível em: <<https://www.iwr.usace.army.mil/Portals/70/docs/iwrreports/2011-R-07.pdf>>.

ECHENDU, A. J. The impact of flooding on Nigeria's sustainable development goals (SDGs). **Ecosystem Health and Sustainability**, v. 6, n. 1, p. 1-13, 2020.

EIRD - ESTRATÉGIA INTERNACIONAL PARA A REDUÇÃO DE DESASTRES. **Marco de Ação de Hyogo 2005-2015: Aumento da resiliência das nações e das comunidades frente aos desastres**. Disponível em: <<http://urbanismo.mppr.mp.br/arquivos/File/MarcodeAcaodeHyogoCidadesResilientes20052015.pdf>>. Acesso em: 25 dez. 2020.

ELDAWY, A.; ALARABI, L.; MOKBEL, M. F. Spatial partitioning techniques in spatial hadoop. **Proc. Int. Confmon Very Large Data Bases**, v. 8, n. 12, p. 1602-1605, 2015.

ELDAWY, A.; MOKBEL, M. F.; JONATHAN, C. (2016) Hadoopviz: A mapreduce framework for extensible visualization of big spatial data. **Proceedings of the IEEE International Conference on Data Engineering**, p 601-612, 2016.

ENGÉLINUS, J.; BADARD, T. Elcano: A Geospatial Big Data Processing System based on SparkSQL. **Proceedings of the 4th International Conference on Geographical Information Systems Theory, Applications and Management**, v. 1, p. 119-128, 2018.

ESRI. **ArcIndia News**, 2015. Disponível em: <<https://www.esri.in/~media/esri-india/files/pdfs/news/arcindianews/Vol5/mapping.pdf>>. Acesso em: 08 mar. 2020.

ESRI. **GIS Tools for Hadoop**, 2013. Disponível em: <<http://esri.github.io/gis-tools-for-hadoop/>>. Acesso em: 09 mar. 2020.

ESTRATÉGIA ODS. **O que são os ODS?**, 2020. Disponível em: <<https://estrategiaods.org.br/conheca-os-ods/>>. Acesso em: 11 dez. 2020.

FABHAT – Fundação Agência da Bacia Hidrográfica do Alto Tietê. **Plano de Bacia Hidrográfica do Alto Tietê**. São Paulo: FEHIDRO, 2019. Disponível em: <https://drive.google.com/drive/folders/19vsyJ_IwJxHiLssvSIFL-_ZavCXGjTC6>.

FAGUNDES, C. A. **Sorocaba, Cidade do Futuro e Futuro da Cidade: Avaliação das Políticas Públicas Urbanas e da Qualidade de Vida**. 2017. Tese (Doutorado em Geografia na Área de Análise Ambiental e Dinâmica Territorial) - Instituto de Geociências, Universidade Estadual de Campinas, Campinas.

FEHIDRO/EMPLASA - Fundo Estadual de Recursos Hídricos/ Empresa Paulista de Planejamento Metropolitano S.A. **Atlas de Uso e Ocupação do Solo do Município de Franco da Rocha**. São Paulo: Secretaria de Economia e Planejamento do Estado de São Paulo, 2006. Disponível em: <https://emplasa.sp.gov.br/Cms_Data/Sites/EmplasaDev/Files/Documentos/Cartografia/Atlas/RMSP/Atlas_FcodaRocha.pdf>.

FERNANDES, N. F.; AMARAL, C. P. Movimentos de massa: uma abordagem geológica – geomorfológica. In: GUERRA, A. J. T.; CUNHA, S. B. (orgs.). **Geomorfologia e meio ambiente**. Rio de Janeiro: Bertrand Brasil, 1996. p. 123–194.

FERNANDES, L. S.; GRIFFITH, J. J.; DA FONSECA, D. M.; DIAS, L. E.; RUIZ, H. A. Uso de geomantas no controle da erosão brasuperficial hídrica em um talude em corte de estrada. **Rev. Bras. Ciênc. Solo**, v. 33, n. 1, p. 199-206, 2009.

FLANAGAN, B. E.; GREGORY, E. W.; HALLISEY, E. J.; HEITGERD, J. L.; LEWIS, B. A Social Vulnerability Index for Disaster Management. **Journal of homeland security and emergency management**, v. 8, n.1, p. 1-22, 2011.

FORNASARI FILHO, N.; FARIA, M. S. C. **Relevo e Dinâmica Superficial**. In: OLIVEIRA, A. M. S.; MONTICELI, J. J. (Org.). **Geologia de Engenharia e Ambiental**,

Vol. 2. São Paulo: Associação Brasileira de Geologia de Engenharia e Ambiental, 2018, p. 204-228.

FUZARO, P. A.; CARNIELLO, M. F. Estudo Comparativo do Índice de Desenvolvimento Humano Municipal e Índice Firjan de Desenvolvimento Municipal na Região Metropolitana do Vale do Paraíba-SP. **Desenvolvimento em Questão**, v. 18, n. 52, p. 75-100, 2020.

G1. **Chuva alaga estações da CPTM na Grande SP e interrompe circulação**, 2016. Disponível em: <<http://g1.globo.com/sao-paulo/noticia/2016/03/chuva-alaga-estacoes-da-cptm-na-grande-sp-e-interrompe-circulacao.html>>. Acesso em: 19 ago. 2019.

G1. **Temporal provoca alagamentos na Grande SP; caminhoneiros ficam ilhados em Caieiras**, 2018. Disponível em: <<https://g1.globo.com/sp/sao-paulo/noticia/2018/12/01/temporal-provoca-alagamentos-em-franco-da-rocha-caieiras-e-francisco-morato.ghtml>>. Acesso em: 18 ago. 2019.

GEOMESA. **Bigtable Data Store**, 2021. Disponível em: <<https://www.geomesa.org/documentation/2.0.2/user/bigtable/index.html>>. Acesso em: 11 mar. 2021.

GEOMESA. **Store, index, query, and transform spatio-temporal data at scale in HBase, Accumulo, Cassandra, Kafka and Spark**, 2020a. Disponível em: <<https://www.geomesa.org/>>. Acesso em: 12 mar. 2020.

GEOMESA. **Installing GeoMesa Cassandra**, 2020b. Acesso em: 01 ago. 2020. Disponível em: <<https://www.geomesa.org/documentation/stable/user/cassandra/install.html>>.

GEOSPARK. **Introduction**, 2020. Disponível em: <<http://geospark.datasyslab.org/>>. Acesso em: 11 mar. 2020.

GHOFRANI, F.; HE, Q.; GOVERDE, R. M. P.; LIU, X. Recent applications of big data analytics in railway transportation systems: A survey. **Transportation Research Part C: Emerging Technologies**, v. 90, p. 226-246, 2018.

GOOGLE EARTH ENGINE GUIDE - GEE. **Importing Table Data**, 2020a. Disponível em: <<https://developers.google.com/earth-engine/importing>>. Acesso em: 05 ago. 2020.

GOOGLE EARTH ENGINE GUIDE - GEE. **Importing Raster Data**, 2020b. Disponível em: <https://developers.google.com/earth-engine/image_upload>. Acesso em: 05 ago. 2020.

GORELICK, N.; HANCHER, M.; DIXON, M.; ILYUSHCHENKO, S.; THAU, D.; MOORE, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. **Remote Sensing of Environment**, v. 202, n. 1, p. 18-27, 2017.

GOVERNO DO ESTADO DE SÃO PAULO. **Oito estações da CPTM são tombadas pelo Patrimônio Histórico**, 2010. Disponível em: <<http://www.saopaulo.sp.gov.br/spnoticias/ultimas-noticias/oito-estacoes-da-cptm-sao-tombadas-pelo-patrimonio-historico/>>. Acesso em: 22 mai. 2019.

GUIMARÃES, R. M.; CRUZ, O. G.; PARREIRA, V. G.; MAZOTO, M. L.; VIEIRA, J. D.; ASMUS, C. I. R. F. A. Análise temporal da relação entre leptospirose e ocorrência de inundações por chuvas no município do Rio de Janeiro, Brasil, 2007-2012. **Ciênc. saúde coletiva**, v. 19, n. 9, 2014.

HONG, L.; OUYANG, M.; PEETA, S.; YAN, Y. Vulnerability assessment and mitigation for the Chinese railway system under floods. **Reliability Engineering and System Safety**, v. 137, p. 58–68. 2015.

HORITA, F. E. A. **An approach for improving decision-making with heterogeneous geospatial big data: an application using spatial decision support systems and volunteered geographic information to disaster management**. 2017. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos.

HORITA, F. E. A.; ALBUQUERQUE, J. P.; MARCHEZINI, V.; MENDIONDO, E. M. Bridging the gap between decision-making and emerging big data sources: An application of a model-based framework to disaster management in Brazil. **Decision Support Systems**, v. 97, p. 12-22, 2017.

HUGHES, J. N.; ANNEX, A.; EICHELBERGER, C. N.; FOX, A.; HULBERT, A.; RONQUEST, M. GeoMesa: a distributed architecture for spatio-temporal fusion. **Proceedings Volume 9473, Geospatial Informatics, Fusion, and Motion Video Analytics V**, 2015.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Aglomerados Subnormais**, 2020. Disponível em: <<https://www.ibge.gov.br/geociencias/organizacao-do-territorio/tipologias-do-territorio/15788-aglomerados-subnormais.html?=&t=o-que-e>>. Acesso em: 30 dez. 2020.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. Base de informações do Censo Demográfico 2010: Resultados do Universo por setor censitário. Rio de Janeiro, 2011. Disponível em: <https://www.ipea.gov.br/redeipea/images/pdfs/base_de_informacoess_por_setor_censitario_universo_censo_2010.pdf>.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Censo 2010 mostra as características territoriais dos aglomerados subnormais e suas diferenças das demais áreas das cidades**, 2019. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/14504-asi-censo-2010-mostra-as-caracteristicas-territoriais-dos-aglomerados-subnormais-e-suas-diferencas-das-demais-areas-das-cidades>>. Acesso em: 31 dez. 2020.

IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Franco da Rocha**, 2018. Disponível em: <<https://cidades.ibge.gov.br/brasil/sp/franco-da-rocha/panorama>>. Acesso em: 16 dez. 2018.

INTRIERI, E.; BARDI, F.; FANTI, R.; GIGLI, G.; FIDOLINI, F.; CASAGLI, N.; COSTANZO, S.; RAFFO, A.; DI MASSA, G.; CAPPARELLI, G.; VERSACE, P. Big data managing in a landslide earlywarning system: experience from a ground-based interferometric radar application. **Natural Hazards and Earth System Sciences**, v. 17, p. 1713-1723, 2017.

IPEA - INSTITUTO DE PESQUISA ECONÔMICA APLICADA. **11. Cidades e Comunidades Sustentáveis – Objetivos de Desenvolvimento Sustentável**, 2019. Disponível em: <<https://www.ipea.gov.br/ods/ods11.html>>. Acesso em: 11 dez. 2020.

IPT - INSTITUTO DE PESQUISAS TECNOLÓGICAS DO ESTADO DE SÃO PAULO. **Mapeamento de Áreas de Risco de Municípios do Estado de São Paulo – Município de Caieiras (SP)**. São Paulo, 2005. (Relatório Técnico nº 82 912-205).

IPT - INSTITUTO DE PESQUISAS TECNOLÓGICAS DO ESTADO DE SÃO PAULO. **Seção de Investigações, Riscos e Desastres Naturais – Sirden**, 2020. Disponível em: <https://www.ipt.br/centros_tecnologicos/CTGeo/laboratorios_e_sessoes/48-secao_de_investigacoes_riscos_e_desastres_naturais___sirden.htm>. Acesso em: 15 dez. 2020.

JAHN, G. F. **Uma Proposta de Arquitetura para Tratamento de Dados Não Estruturados no Âmbito dos Institutos Federais de Educação**. 2017. Dissertação (Mestrado em Ciência da Computação) – Centro de Informática, Universidade Federal de Pernambuco, Recife.

JAISWAL, P.; VAN WESTEN, C. J.; JETTEN, V. Quantitative assessment of landslide hazard along transportation lines using historical records. **Landslides**, v. 8, p. 279-291, 2011.

JAVA. **O que é a Tecnologia Java e porque preciso dela?** Disponível em: <https://www.java.com/pt_BR/download/faq/whatis_java.xml>. Acesso em: 01 ago. 2020.

NASCIMENTO JÚNIOR, L.; SANT'ANNA NETO, J. L. A. Índice de Vulnerabilidade Social à Desastres Naturais para Áreas Urbanas. **Revista de Geociências do Nordeste**, v. 6, n. 2, p. 65-72, 2020.

KALLIAMVAKOU, E.; GOUSIOS, G.; BLINCOE, K.; SINGER, L.; GERMAN, D.; DAMIAN, D. The Promises and Perils of Mining GitHub. **MSR 2014: Proceedings of the 11th Working Conference on Mining Software Repositories**, p. 92-101, 2014. Disponível em: <<https://dl.acm.org/doi/pdf/10.1145/2597073.2597074>>.

KATZMAN, R. **Vulnerabilidad, activos y exclusion social em Argentina y Uruguay**. Santiago de Chile, OITFord, 1999.

KELLERMANN, P.; SCHÖBEL, A.; KUNDELA, G.; THIEKEN, H. Estimating flood damage to railway infrastructure – the case study of the March River flood in 2006 at the Austrian Northern Railway. **Nat. Hazards Earth Syst. Sci.**, v. 15, p. 2485-2496, 2015.

KHAN, Z.; ANJUM, A.; KIANI, S. L. Cloud Based Big Data Analytics for Smart Future Cities. Proceedings of the 2013 IEEE/ACM 6th International Conference on Utility and Cloud Computing. **IEEE Computer Society**, p. 381–386, 2013.

KOBIYAMA, M.; MENDONÇA, M.; MORENO, D.A.; MARCELINO, I.P.V.O; MARCELINO, E.V.; GONÇALVES, E.F.; BRAZETTI, L.L.P.; GOERL, R.F.; MOLLERI, G.S.F.; RUDORFF, F.M. **Prevenção de Desastres Naturais: Conceitos Básicos**. Curitiba: Ed. Organic Trading. 109 p., 2006. Disponível em: <<http://www.labhidro.ufsc.br/publicacoes.html>>.

LEITE, I.; PINHO, M./G1 SÃO PAULO. **Saiba como e quando comportas de represas em SP são abertas por cheia, 2016**. Disponível em: <<http://g1.globo.com/sao-paulo/noticia/2016/03/saiba-como-e-quando-comportas-de-represas-em-sp-sao-abertas-por-cheia.html>>. Acesso em: 01 nov. 2020.

LEONARDI, I. **Esri lança ferramenta GIS para analisar Big Data**, 2013. Disponível em: <<https://mundogeo.com/2013/05/10/esri-lanca-ferramenta-gis-para-analisar-big-data/>>. Acesso em: 09 mar. 2020.

LI, X.; ZHAO, J. Research on the Construction of Emergency Mapping Information System of Natural Disasters Based on Mapping Geographic Information. **Rev. de la Facultad de Ingeniería U.C.V**, v. 32, n. 13, p. 374-379, 2017.

LI, Z.; SHI, W.; LU, P.; YAN, L.; WANG, Q.; MIAO, Z. Landslide mapping from aerial photographs using change detection-based Markov random field. **Remote Sensing of Environment**, v.187, p. 76-90, 2016.

LI, Z.; YANG, C.; JIN, B.; YU, M.; LIU, K.; SUN, M.; ZHAN, M. Enabling big geoscience data analytics with a cloud-based, MapReduce-enabled and service-oriented workflow framework. **Plos One**, v. 10, 2015.

LOPES, E. S. S. **Modelagem espacial dinâmica em sistema de informação geográfica: uma aplicação ao estudo de movimentos de massa em uma região da Serra do Mar Paulista**. 2006. Tese (Doutorado em Geociências e Meio Ambiente) – Instituto de Geociências e Ciências Exatas, Universidade Estadual Paulista “Júlio de Mesquita Filho”- UNESP, Rio Claro, 2006.

LUKIC, T.; GAVRILOV, M. B.; MARKOVIC, S. B.; KOMAC, B; ZORN, M.; MLADAN, D.; DORDEVIC, J.; MILANOVIC, M.; VASILJEVIC, D. A.; VUJICIC, M. D.; KUZMANOVIC, B.; PRENTOVIC, R. Classification of Natural Disasters between the Legislation and Application: Experience of the Republic of Serbia. **Acta geographica Slovenica**, v. 53, n. 1, p. 149-164, 2013.

MAFFRA, C. Q. T.; MAZZOLA, M. **As razões dos desastres em território brasileiro**. In: DOS SANTOS, R. F. (Org.). *Vulnerabilidade Ambiental: Desastres naturais ou fenômenos induzidos?* Brasília: Ministério do Meio Ambiente – MMA, 2007. p. 9-12.

MALIK, P. Governing Big Data: Principles and practices. **IBM of Research and Development**, v. 57, n. 3, p. 1-13, 2013.

MANFRE, L. A.; ALBUQUERQUE, N. G.; QUINTANILHA, J. A. Mapeamento do risco de deslizamento associado ao Complexo Nuclear Almirante Álvaro Alberto, Rio de Janeiro, Brasil. **Bol. Ciênc. Geod.**, v. 24, n. 1, p. 125-141, 2018.

MANIFOLD. **Home**, 2020. Disponível em: <<http://www.manifold.net/>>. Acesso em: 14 mar. 2020.

MARCELINO, E. V.; FORMAGGIO, A. R.; MAEDA, E. E. Landslide inventory using image fusion techniques in Brazil. **International Journal of Applied Earth Observation and Geoinformation**, v. 11, n. 3, p. 181-191, 2009.

MARTINES, M. R.; CAVAGIS, A. D. M.; KAWAKUBO, F. S.; MORATO, R. G.; FERREIRA, R. V; TOPPA, R. H. Spatial segregation in floodplain: An approach to correlate physical and human dimensions for urban planning. **Cities**, v. 97, 2020.

MARTÍNEZ-ÁLVARES, F.; MORALES-ESTEBAN, A. Big data and natural disasters: New approaches for spatial and temporal massive data analysis. **Computers & Geosciences**, v. 129, p. 38-39, 2019.

MAVRIDIS, I.; KARATZA, H. Performance evaluation of cloud-based log file analysis with Apache Hadoop and Apache Spark. **Journal of Systems and Software**, v. 125, p. 133-151, 2017.

MIGUEL, F. R. M.; VIEIRA, S. R.; GREGO, C. R. Variabilidade espacial da infiltração de água em solo sob pastagem em função da intensidade de pisoteio. **Pesq. Agropec. Bras.**, v. 44 n. 11, 2009.

MINISTÉRIO DAS CIDADES; INSTITUTO DE PESQUISAS TECNOLÓGICAS DO ESTADO DE SÃO PAULO – IPT. **Mapeamento de riscos em encostas e margem de rios**. In: CARVALHO, C. S.; MACEDO, E. S.; OGURA, A. T. (Org.). Brasília: Ministério das Cidades; Instituto de Pesquisas Tecnológicas – IPT, 176 p., 2007.

MMA – MINISTÉRIO DO MEIO AMBIENTE. **Áreas de Preservação Permanente**, 2020. Disponível em: <<https://www.mma.gov.br/cidades-sustentaveis/areas-verdes-urbanas/%C3%A1reas-de-prote%C3%A7%C3%A3o-permanente.html>>.

MIRANDA, L. H.; MORATO, R. G.; KAWAKUBO, F. S. Mapeamento da Qualidade de Vida Urbana em Pouso Alegre, Sul de Minas Gerais. **Revista do Departamento de Geografia – USP**, v. 24, p.24-36, 2012.

MORATO, R. G. Análise da Qualidade de Vida Urbana no Município de Osasco/SP utilizando Geoprocessamento. **Geografia e Pesquisa**, v. 2, n. 1, p. 139-155, 2008.

MORETTIN, P. A., SINGER, J. M. **Introdução à Ciência de Dados - Fundamentos e Aplicações**. Versão Preliminar. Departamento de Estatística - Universidade de São Paulo, 2019.

MORROW, B.H. Identifying and Mapping Community Vulnerability. **Disasters**, v. 23, n. 1, p. 1–18, 1999.

MUDAVANHU, C. The impacto of flood disasters on child education in Muzarabani District, Zimbabwe. **Journal of Disaster Risk Studies**, v. 6, n. 1, p. 1-8, 2014.

MURPHY, A. T.; GOULDBY, B.; COLE, S. J.; MOORE, R. J.; KENDALL, H. Real-time flood inundation forecasting and mapping for key railway infrastructure: a UK case study. **E3S Web of Conferences**, v. 7, n. 18020, p. 01-08, 2016.

NAHAS, M. I. P. **Bases teóricas, metodologia de elaboração e aplicabilidade de indicadores intraurbanos na gestão municipal da qualidade de vida urbana em grandes cidades: o caso de Belo Horizonte**. 2002. Tese (Doutorado em Ecologia e Recursos Naturais) - Centro de Ciências Biológicas e da Saúde, Universidade Federal de São Carlos. Disponível em: <<https://repositorio.ufscar.br/handle/ufscar/1607>>

NG, G. H.; WICKERT, A. D.; SOMERS, L. D.; SABERI, L.; CRONKITE-RATCLIFF, C.; NISWONGER, R. G.; MCKENZIE, J. M. GSFLOW-GRASS v1.0.0: GIS-enabled hydrologic modeling of coupled groundwater-surface-water systems. **Geoscientific Model Development**, v. 11, p. 4755-4777, 2018.

NRC - NATIONAL RESEARCH COUNCIL FACING HAZARDS AND DISASTERS. **Understanding Human Dimensions**. Committee on Disaster Reduction in the Social

Sciences: future challenges and opportunities. Washington D.C.: National Academy Press, 2006.

AL NUAIMI, E., AL NEYADI, H., MOHAMED, N.; AL-JAROODI, J. Applications of big data to smart cities. **J. Internet Serv. Appl.**, v. 6, n. 25, 2015.

OLIVEIRA, D. R. **Crianças em situações de riscos e desastres: Atenção Psicossocial, Saúde Mental e Direitos Humanos.** 2018. Tese (Doutorado em Psicologia – Psicologia Escolar e do Desenvolvimento Humano) – Instituto de Psicologia, Universidade de São Paulo, São Paulo.

OPOLOT, E. Application of Remote Sensing and Geographical Information Systems in Flood Management: A Review. **Research Journal of Applied Sciences, Engineering and Technology**, v. 6, n. 10, p. 1884-1894, 2013.

OTA, N. Problems Related to and Approaches for Developing Countermeasures against Natural Disasters. **Quarterly Report of Railway Technical Research Institute**, v. 59, n. 1, p. 6-9, 2018.

OUMA, Y. O.; TATEISHI, R. Urban Flood Vulnerability and Risk Mapping Using Integrated Multi-Parametric AHP and GIS: Methodological Overview and Case Study Assessment. **Water**, v. 6, p. 1.515 – 1.545, 2014.

OUSSOUS, A.; BENJELLOUN F. Z., A.; LAHCEN, A. A.; BELFKIH, S. Big Data technologies: A survey. **Journal of King Saud University - Computer and Information Sciences**, v. 30, n. 4, p. 431-448, 2018.

OZAKI, N.; WATANABE, T.; FUKUSAWA, N. Algorithm based on gustand flood forecasts for managing train operation control and passenger evacuation. **QR of RTRI**, v. 59, n. 3, 2018.

POMANTE, L.; DI FELICE, P. WSN and GIS integration for a cost-effective real-time monitoring of landslides on railway stations and lines. **IEEE 29th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)**, p. 396-400, 2018.

PREFEITURA MUNICIPAL DE CAIEIRAS. **História**, 2018. Disponível em: <<http://www.educacao.caieiras.sp.gov.br/caieiras/?id=historia>>. Acesso em: 11 abr. 2018.

PREFEITURA MUNICIPAL DE CAIEIRAS. **Relatório de Acompanhamento 2 – Diagnóstico municipal**. Objeto: Plano de Mobilidade de Caieiras (SP) – PlanMob Caieiras, 167 p., 2017. Disponível em: <https://caieiras.sp.gov.br/secretarias/downloads/caieiras_diagnostico.pdf>. Acesso em: 22 mai. 2019.

PREFEITURA MUNICIPAL DE FRANCO DA ROCHA. **Decreto 2595/2017**, 2017. Disponível em: <<http://francodarocha.sp.gov.br/franco/servico/lei/7800>>. Acesso em: 16 dez. 2018.

PROJETO MAPBIOMAS – **Coleção 4.1 da Série Anual de Mapas de Cobertura e Uso de Solo do Brasil**. Disponível em: <<https://plataforma.mapbiomas.org/map#coverage>>. Acesso em: 06 jan. 2020.

PYTHON. **Home**, 2020. Disponível em: <<https://www.python.org/>>. Acesso em: 01 ago. 2020

QINGQUAN, L.; DEREN, L. Big data GIS. **Geomatics and Information Science of Wuhan University**, v. 39, p. 641-644, 2014.

RAGINI, J. R.; ANAND, P. M. R.; BHASKAR, V. Big data analytics for disaster response and recovery through sentimento analysis. **International Journal of Information Management**, v. 42, p. 13-24, 2018.

RAMOS, A. M.; ALVES, L.; DINIZ, A. F.; MARENGO, J. A. Anuário Climático Do Brasil – 2019. Ano 2, n. 2, 2020.

RASCH, R. J. Assessing urban vulnerability to flood hazard in Brazilian municipalities. **Environment and urbanization**, v. 28, n. 1, p. 145-168, 2016.

RASCH, R. Income inequality and urban vulnerability to flood hazard in Brazil. **Social Science Quarterly**, v. 98, n. 1, p. 299-325, 2017.

RESTREPO-ESTRADA, C.; DE ANDRADE, S. C.; ABE, N.; FAVA, M. C.; MENDIONDO, E. M.; DE ALBUQUERQUE, J. P. Geo-social media as a proxy for hydrometeorological data for streamflow estimation and to improve flood monitoring. **Computers and Geosciences**, v. 111, p. 148-158, 2018.

RIBEIRO, H.; VARGAS, H. C. Qualidade Ambiental Urbana: Ensaio de uma definição. In: RIBEIRO, H.; VARGAS, H. C. (Orgs.) *Novos Instrumentos de Gestão Ambiental Urbana*. São Paulo, Edusp, p.13-19, 2001.

ROBAINA, L. E. S. Espaço urbano: relação com os acidentes e desastres naturais no Brasil. **Rev. Ciência e Natura – UFSM**, v. 30, n. 2, p. 93-105, 2008.

RODRIGUES, A. L. M.; REIS, G. B.; DOS SANTOS, M. T.; DA SILVA, D. D.; DOS SANTOS, V. J.; CASTRO, J. S.; CALIJURI, M. L. Influence of land use and cover's change on the hydrological regime at a Brazilian southeast urbanized watershed. **Environmental Earth Sciences**, p. 1 – 13, 2019.

ROY J.; TSCHAKERT, P.; WAISMAN, H. Sustainable Development, Poverty Eradication and Reducing Inequalities. In *Global Warming of 1.5 C. IPCC: Intergovernmental Panel on Climate Change*, 2018. Disponível em: <https://www.ipcc.ch/site/assets/uploads/sites/2/2019/05/SR15_Chapter5_Low_Res.pdf>.

RUSSOM, P. Managing Big Data. **TDWI Research**, 2013.

SAGIROGLU, S.; SINANC, D. Big data: A review. **International Conference on Collaboration Technologies and Systems**, p. 42-47, 2013.

SANYAL, J.; LU, X. X. Remote sensing and GIS based flood vulnerability assessment of human settlements: A case study of Gangetic West Bengal, India. **Hydrol. Process.**, v. 19, p. 3699-3716, 2005.

SMARZARO, R.; LIMA, T. F. M.; DAVIS JR., C. A. Quality of Urban Life Index From Location-Based Social Networks Data: A Case Study in Belo Horizonte, Brazil. In:_____. **Volunteered Geographic Information and the Future of Geospatial Data**. BrazilPublisher, IGI GlobalEditors, 2017, 357 p.

SEADE - FUNDAÇÃO SISTEMA ESTADUAL DE ANÁLISE DE DADOS DO ESTADO DE SÃO PAULO. **Índice Paulista de Vulnerabilidade Social: IPVS** Versão 2010, p. 1-18, 2013. Disponível em: <<http://ipvs.seade.gov.br/view/pdf/ipvs/metodologia.pdf>>.

SECRETARIA DO MEIO AMBIENTE DO ESTADO DE SÃO PAULO. **Mapa florestal dos municípios do estado de São Paulo**: Francisco Morato. Disponível em: <<http://s.ambiente.sp.gov.br/sifesp/franciscomorato.pdf>>. Acesso em: 02 nov. 2020.

SIM, C. H.; GAN, F. F.; CHANG, T. C. Outlier Labeling With Boxplot Procedures, **Journal of the American Statistical Association**, v. 100, n. 470, p. 642-652, 2005.

SIMÕES, S. J. C.; GOMES, L.; MENDES, R. M.; MENDES, T. S. G. SIG e Modelos de Escorregamentos: Avaliando Métodos para Reduzir as Incertezas de Dados de Solos e Precipitação. **Revista Brasileira de Cartografia**, n. 68/9, p. 1737-1746, 2016.

SINGH, H.; BAWA S. Spatial data analysis with ArcGIS and MapReduce. **International Conference on Computing, Communication and Automation (ICCCA)**, p. 45-49, 2016.

SIVARAJAH, M.; KAMAL, M.; IRANI, Z.; WEERAKKODY, V. Critical analysis of Big Data challenges and analytical methods. **Journal of Business Research**, v. 70, p. 263-286, 2017.

SOUZA, N. C.; GIANOTTI, M. A.; LAROCCA, A. P. C.; QUINTANILHA, J. A.; ALMEIDA FILHO, G. S. Análise da Associação entre Características Topográficas e Tipos de Ocorrência de Processos Erosivos ao Longo de Linha Férrea - Estudo de Caso: Malha Paulista – All (SP). **Bol. Ciênc. Geod.**, v. 20, n. 4, p. 947-969, 2014.

SUNG, C. H.; LIAW, S. C. A GIS-based approach for assessing social vulnerability to flood and debris flow hazards. **International Journal of Disaster Risk Reduction**, v. 46, p. 1-10, 2020.

TAN, Q.; BAI, M.; ZHOU, P.; HU, J.; QIN, X. Geological hazard risk assessment of line landslide based on remotely sensed data and GIS. **Measurement**, v. 169, p. 1-10, 2021.

TANG, M.; YU, Y.; MALLUHI, Q. M.; OUZZANI, M.; AREF, W. G. LocationSpark: A Distributed In-Memory Data Management System for Big Spatial Data. **Proceedings of the VLDB Endowment**, v. 9, n. 13, 2016.

TAPSELL, S. M.; PENNING-ROWSELL, E. C.; TUNSTALL, S. M.; WILSON, T. L. Vulnerability to flooding: health and social dimensions. **Philosophical transactions of the royal society of London. Series A: Mathematical, Physical and Engineering Sciences**, v. 360, n. 1796, p. 1511-1525, 2002.

TECHAMERICA FOUNDATION'S FEDERAL BIG DATA COMMISSION.
Demystifying big data: A practical guide to transforming the business of Government, 2012. Disponível em:
 <https://bigdatawg.nist.gov/_uploadfiles/M0068_v1_3903747095.pdf>.

TECUCI, G. Artificial intelligence. **Wiley Interdisciplinary Reviews: Computational Statistics**, v. 4, p. 168-180, 2012.

THADURI, A.; GALAR, D.; KUMAR, U. Railway assets: A potential domain for big data analytics. **Procedia Computer Science**, v. 53, p.457-467, 2015.

TOBIN, G. A.; MONTZ, B. E. Natural hazards and disasters: when potential becomes reality. In: _____. **Natural Hazards: explanation and integration**. Nova York: The Guilford Press, 1997, 388 p.

TOMINAGA, L. K. **Desastres Naturais: por que ocorrem?** In: TOMINAGA, L. K.; SANTORO, J.; AMARAL, R. (Org.). **Desastres Naturais: conhecer para prevenir**. São Paulo: Instituto Geológico, 2009. p. 13-23.

UBUNTU. **Ubuntu 20.04.1 LTS (Focal Fossa)**, 2018. Disponível em:
 <<https://releases.ubuntu.com/20.04/>>. Acesso em: 01 ago. 2020.

UFFRJ - UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO. **Mapa Mental dos Problemas das Enchentes Urbanas**, 2020. Disponível em:
 <<http://www.uffrj.br/institutos/it/de/acidentes/mma10.htm>>. Acesso em: 02 jan. 2021.

UNDRR - United Nations Office for Disaster Risk Reduction. **About UNDRR**, 2020. Disponível em: <<https://www.undrr.org/about-undrr>>. Acesso em: 15 dez. 2020.

UNISDR - INTERNATIONAL STRATEGY FOR DISASTER REDUCTION. **Revealing Risk, Redefining Development**, 2011. Disponível em:
 <<https://www.preventionweb.net/english/hyogo/gar/2011/en/home/intro.html>>. Acesso em: 04 mar. 2019.

VARANDA, E.; MAHLER, C. F.; OLIVEIRA, L. C. D. Análise de risco de escorregamento com uso de SIG. **Revista Luso-Brasileira de Geotecnia**, n. 119, p. 55-68, 2010.

VEDOVELLO R.; MACEDO, E. S. **Deslizamentos de Encostas**. In: DOS SANTOS, R. F. (Org.). *Vulnerabilidade Ambiental: Desastres naturais ou fenômenos induzidos?* Brasília: Ministério do Meio Ambiente – MMA, 2007. p. 75-93.

VENKATRAMAN, S.; VENKATRAMAN, R. Big data security challenges and strategies. **AIMS Mathematics**, v. 4, n. 3, p. 860-879, 2019.

WANG, S.; ZHONG, Y.; WANG, E. An integrated GIS platform architecture for spatiotemporal big data. **Future Generation Computer Systems**, v. 94, p. 160-172, 2019.

WISNER, B.; BLAIKIE, P.; CANNON, T.; DAVIS, I. **At Risk: Natural Hazards, People's Vulnerability and Disasters**. Second edition, 496 p., 2004.

WHITE, T. **Hadoop: The Definitive Guide**. Yahoo Press, 2010.

YANG, C.; HUANG, Q.; LI, Z.; LIU, K.; HU, F. Big Data and cloud computing: innovation opportunities and challenges. **International Journal of Digital Earth**, v. 10, n. 1, p. 13-53, 2017a.

YANG, C.; YU, M.; HU, F.; JIANG, Y.; LI, Y. Utilizing Cloud Computing to address big geospatial data challenges. **Computers, Environment and Urban Systems**, v. 61, p. 120-128, 2017b.

YOU S.; ZHANG, J.; GRUENWALD, L. Large-scale spatial join query processing in Cloud, **2015 31st IEEE International Conference on Data Engineering Workshops**, p. 34-41, 2015.

YU, M.; YANG, C.; LI, Y. Big Data in Natural Disaster Management: A Review. **Geosciences**, v. 8, n. 165, p. 1-26, 2018.

YU, J.; WU, J.; SARWAT, M. GeoSpark: A Cluster Computing Framework for Processing Large-Scale Spatial Data. **Proceeding of the ACM International Conference on Advances in Geographic Information Systems ACM SIGSPATIAL GIS 2015**, p. 1-4, 2015.

YU, J.; ZHANG, Z.; SARWAT, M. Spatial data management in apache spark: the GeoSpark perspective and beyond. **GeoInformatica an International Journal on Advances of Computer Science for Geographic Information Systems**, v. 17, n. 4. 2018.

ZAHARIA, M.; CHOWDHURY, M.; DAS, T.; DAVE, A.; MA, J.; MCCAULY, M.; FRANKLIN, M. J.; SHENKER, S.; STOICA, I. Resilient Distributed Datasets: A Fault-Tolerant Abstraction for In-Memory Cluster Computing. **NSDI**, p. 15-28, 2012.

ZAINAL, N. Z. B.; HUSSIN, H.; NAZRI, M. N. M. Big data initiatives by governments – Issues and challenges: A review. **Proc. – 6th International Conference on Information and Communication Technology for The Muslim World ICT4M 2016**, p. 304-309, 2017.

ZHANG, J.; XU, L.; ZHANG, Y.; LIU, G.; ZHAO, L.; WANG, Y. An On-Demand Scalable Model for Geographic Information System (GIS) Data Processing in a Cloud GIS. **ISPRS International Journal of Geo-Information**, v. 8, n. 9, p. 1-13, 2019.

ZHANG, T.; XIE, S.; FAN, J.; HUANG, B.; WANG, Q.; YUAN, W.; ZHAO, H.; CHEN, J. P.; LI, H.; LIU, G.; TONG, L.; SOUSA, J. J. Detection of Active Landslides in Southwest China using Sentinel-1 and ALOS-2 Data. **Procedia Computer Science**, v. 181, p. 1138-1145, 2021.

ZHANG, W.; WANG, W.; WU, F. The Application of Multi-variable Optimum Regression Analysis to Remote Sensing Imageries in Monitoring Landslide Disaster. **Energy Procedia**, v. 16, part A, p. 190-196, 2012.

ZHENG, Y.; CAPRA, L.; WOLFSON, O. E.; YANG, H. Urban Computing: Concepts, Methodologies, and Applications. **ACM Transactions on Intelligent Systems and Technology**, v. 5, n. 38, p. 1-55, 2014.

ZHOU, C., YIN, K., CAO, Y., AHMED, B., LI, Y., CATANI, F., POURGHASEMI, H.R. Landslide susceptibility modeling applying machine learning methods: a case study from Longju in the Three Gorges Reservoir area, China. **Computers & Geosciences**, v. 112, p. 23-37, 2018.

ANEXOS

Em Caieiras, na Grande SP, pessoas ficam 'ilhadas' em estação de trem

Chuva interrompeu circulação de trens entre Caieiras e Francisco Morato. Pessoas tentam caminhos alternativos para ir ao trabalho ou para casa.

Modificado de: Mirella Nascimento/G1, 2009.

Chuva alaga estações da CPTM na Grande SP e interrompe circulação

Passageiros ficam ilhados e dormem nas estações da Linha 7-Rubi. Situação não estava normalizada na manhã desta sexta-feira (11).

Modificado de: G1, 2016.

Diário da CPTM atualizou o próprio status.
21 de janeiro de 2017 · 🌐

#L7 - 22:40 - Por motivo de deslizamento de terra próximo à via na região da estação de PERUS/Alagamento na região de Caieiras, a circulação de trens está interditada entre as estações PERUS e FRANCO DA ROCHA, o trecho está sendo atendido por ônibus da operação PAESE. Pedimos a sua compreensão. Atualização.

Modificado de: Diário da CPTM/Facebook, 2017.

Temporal provoca alagamentos na Grande SP; caminhoneiros ficam ilhados em Caieiras

Não há informações sobre deslizamentos e feridos. Cidade de SP também teve alagamentos e Marginal Tietê ficou bloqueada.

Modificado de: G1, 2018.

Pelo 3º dia consecutivo, deslizamento de terra afeta circulação de Linha da CPTM

Veículos circulam com velocidade reduzida e com maior tempo de parada entre as estações Perus e Francisco Morato desde sexta-feira (20).

Modificado de: G1, 2017.

Diário da CPTM
5 de fevereiro de 2015 · 🌐

Devido a alagamento, os trens da Linha 7-Rubi estão circulando com velocidade reduzida e maior tempo de parada entre as estações CAIEIRAS e BALTAZAR FIDELIS.

Modificado de: Diário da CPTM/Facebook, 2015.

Anexo B – Primeira seleção (Identificação). Elaborado por: Thaís Correia, 2019.

Estratégia de Busca							N° de Registros Obtidos na Base de Dados				
Termo Principal	Operador	Termo Secundário	Operador	Termo Secundário	Operador	Termo Secundário	Identificação				
							Scopus	Web of Science	Scielo	Google Scholar	Total de Resultados
Natural Disaster*	AND	Big Data					431	4	0	9	36.727
		Big Data	AND	Spatial Analysis			32	1	0	0	
		Big Data		Spatial Analysis	AND	GIS	6	0	0	0	
		Machine Learning					317	1	1	3	
		GIS					1.694	12	0	72	
		Remote Sensing					2.027	27	1	122	
		Railway OR Railroad					347	6	0	15	
Landslide* OR Flood*	AND	Big Data					1.474	19	0	50	
		Big Data	AND	Spatial Analysis			121	0	0	1	
		Big Data		Spatial Analysis	AND	GIS	23	0	0	0	
		Machine Learning					1.170	115	0	232	
		GIS					9.121	1.000	2	4.040	
		Remote Sensing					9.665	528	3	1.861	
		Railway OR Railroad					1.859	65	0	250	
N° Resultados							28.287	1.778	7	6.655	

Anexo C – Segunda seleção (Triagem). Elaborado por: Thaís Correia, 2019.

Estratégia de Busca							N° de Registros Obtidos na Base de Dados				
Termo Principal	Operador	Termo Secundário	Operador	Termo Secundário	Operador	Termo Secundário	Triagem				
							Scopus	Web of Science	Scielo	Google Scholar	Total de Resultados
Natural Disaster*	AND	Big Data					339	4	0	9	23.959
		Big Data	AND	Spatial Analysis			29	1	0	0	
		Big Data		Spatial Analysis	AND	GIS	4	0	0	0	
		Machine Learning					254	1	1	3	
		GIS					1.172	10	0	38	
		Remote Sensing					1.384	18	1	47	
		Railway OR Railroad					163	3	0	8	
Landslide* OR Flood*	AND	Big Data					1.036	18	0	41	
		Big Data	AND	Spatial Analysis			94	0	0	1	
		Big Data		Spatial Analysis	AND	GIS	20	0	0	0	
		Machine Learning					909	94	0	199	
		GIS					6.095	709	2	2.467	
		Remote Sensing					6.315	375	3	1.127	
		Railway OR Railroad					795	25	0	145	
N° Resultados							18.609	1.258	7	4.085	

Anexo D – Terceira seleção (Eligibilidade). Elaborado por: Thaís Correia, 2019.

Estratégia de Busca							Nº de Registros Obtidos na Base de Dados				
Termo Principal	Operador	Termo Secundário	Operador	Termo Secundário	Operador	Termo Secundário	Eligibilidade				
							Scopus	Web of Science	Scielo	Google Scholar	Total de Resultados
Natural Disaster*	AND	Big Data					10	2	0	3	158
		Big Data	AND	Spatial Analysis			4	1	0	5	
		Big Data		Spatial Analysis	AND	GIS	1	0	0	2	
		Machine Learning					3	0	1	2	
		GIS					2	1	0	2	
		Remote Sensing					4	4	1	3	
		Railway OR Railroad					3	0	0	0	
Landslide* OR Flood*	AND	Big Data					4	4	0	3	
		Big Data	AND	Spatial Analysis			3	0	0	5	
		Big Data		Spatial Analysis	AND	GIS	3	0	0	2	
		Machine Learning					8	10	0	9	
		GIS					5	9	2	7	
		Remote Sensing					5	4	3	6	
		Railway OR Railroad					5	2	0	5	
Nº Resultados							60	37	7	54

Review

Big Data for Natural Disasters in an Urban Railroad Neighborhood: A Systematic Review

Thais P. Corneia ^{1,*} , Alessandra C. Corsi ² and José A. Quintanilha ³

¹ Polytechnic (Engineering) School, University of São Paulo-PTR-EPUSP, São Paulo 05508-070, Brazil

² Investigations, Risks and Natural Disasters Section, Center for Geoenvironmental Technologies, Institute for Technological Research-Sirden-CTGeo-IPT, São Paulo 05508-901, Brazil; accorsi@ipt.br

³ Institute of Energy and Environment, University of São Paulo-IEEUSP, São Paulo 05508-010, Brazil; jaquinta@usp.br

* Correspondence: thais.pasos.corneia@usp.br

Received: 31 January 2020; Accepted: 26 March 2020; Published: 1 April 2020

Abstract: Landslides and floods are among the most common disasters in Brazil and are responsible for losses on social, environmental, and economic scales, even resulting in deaths. Floods can negatively affect the structure and operations of a railway network, causing travel delays, train service cancellations, and major fines for the railway. The objective of this article is to conduct a bibliographic review of what is available in publications on natural disasters, particularly landslides and floods, big data techniques, and railroads, at international and national levels. A bibliometric analysis was carried out according to the “PRISMA Flow Diagram” guidelines. The analysis in this study was conducted through searches of the following reference databases: Scopus, Web of Science, Scielo, and Google Scholar. After the keyword search was completed, the absence of available data and references relating to Brazil was verified. This justified the development of this and other related papers, and the efforts necessary to turn these data into useful information for the managers of cities and environmental institutions. The aim of this study is to fill the gap in the research, focusing on Brazil, related to big data, smart cities, and natural disasters (particularly, landslides and floods), and to propose other papers that can be developed in this subject area.

Keywords: landslide; flood; railway; PRISMA; Brazil

1. Introduction

Landslides and floods are among the most common disasters in Brazil, and they are responsible for social, environmental, and economic damage, often resulting in loss of life [1,2]. It is necessary to define the time, location, and severity of the impacts of natural disasters in order to develop mitigation techniques, appropriate responses, and firm recovery plans [3].

Railway infrastructure is crucial for transportation and contributes to economic and social welfare [4]. However, transportation is subject to frequent natural disasters, such as floods, earthquakes, and landslides [5].

Climate change has provoked an increase in rainfall, generating concern that tornados and floods have also increased [5]. While expanding civil-engineering infrastructure projects, the capacity of resistance of these structures, in relation to external forces such as natural disasters, should be verified [6].

Rivers may also overflow, depending on the duration of the rainfall, which can flow to the roads and railroads [6]. Floods can negatively affect the structure and operation of a railroad network, including travel delays, service cancellations, and major fines to the railroad company because of the delay [7,8].

Anexo F – Artigo “A Systematic Review: Big Data Analytics and Landslides”, submetido no XII Simpósio Internacional de Deslizamentos, em Cartagena, Colômbia; aceito em 07 de julho de 2020.

A Systematic Review: Big Data Analytics and Landslides

Caio da Silva Azevedo¹, Thaís Passos Correia², Alessandra Cristina Corsi³, Mauro de Mesquita Spinola⁴ and José Alberto Quintanilha⁵

1 Postgraduate Program in Production Engineering at the Polytechnic School of the University of São Paulo – PRO - EPUSP

2 Postgraduate Program in Transportation Engineering at the Polytechnic School of the University of São Paulo – PTR - EPUSP

3 Institute for Technological Research - IPT

4 Production Engineering Department at the Polytechnic School of the University of São Paulo – PRO – EPUSP

5 Institute of Energy and Environment of the University of São Paulo - IEEUSP

caio.aze@usp.br

Abstract

The landslides are among the natural disasters that most affect the population in Brazil and around the world, that they're responsible for human, economic, social and environmental losses. High-resolution aerial and satellite imagery, that contribute to identify this type of event, are valuable sources of information, even if they are not the only ones. Machine learning (ML) techniques are often required to identify relevant patterns, as high performance computing and data visualization are widely applied successfully to natural disaster-related data, specifically for landslides. The greater the volume of data and its processing as machine learning occurs, the better the accuracy of the results obtained. Among other reasons, Big Data (BD) has stood out for its ability to integrate and generate a large volume of data from different sources and in different formats. The use of BD concepts and techniques has become ordinary in several areas of knowledge, in particular for the storage and analysis of data related to natural accidents and, thus, helping to improve decision making as well as prevention of disasters, for example. With the emergence of the BD came the challenge of aligning decision making processes so that there is no shortage of information or useless data for decision makers. In the case of Brazil and considering, especially, applications to landslide data, there is a lack of articles dealing with the subject. The aim of this study is to investigate the main features used on machine learning techniques, to map the literature and point out new routes and opportunities in the field. In the systematic review, the eligibility criteria adopted are studies published in the last 5 years (from January 2015 to February 2020), in English, at national and international levels, including publications in scientific events and relevant journals, through databases such as Scopus and Web of Science. The search was performed in a structured form with the terms "big data", "machine learn*", "landslid*", "forecast*" and "predict*" and the like through the boolean operator "AND" or "OR" in order to restrict the theme to the aspects to be discussed.*

Anexo G – Levantamento de dados/fontes oficiais referentes à área de estudo.

Categoria	Classe	Nome Arquivo	Formato Primário	Feição	Escala	Ano	Fonte	Site	Link	Possui VMS?	VMS	Gratuito, Pago ou Privado?
Ocorrências Esborregamento	Ocorrências Esborregamento	Ocorrências Caieiras	shp.	Não se Aplica	Não se Aplica	2013	Defesa Civil Caieiras	Não se Aplica		Não se Aplica		Privado
Ocorrências Inundações Caieiras	Ocorrências Inundações	Ocorrências Caieiras	shp.	Não se Aplica	Não se Aplica	2013	Defesa Civil Caieiras	Não se Aplica		Não se Aplica		Privado
Relevo Caieiras	Curva de Nível	CurvaDeNivel	shp.	Linha	1:25.000	2002	Emplasa	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Relevo Caieiras	Pontos Cotados	PontoCotado	shp.	Ponto	1:25.000	2002	Emplasa	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Hidrografia Caieiras	Cursos d'água	CursoDagua	shp.	Linha	1:25.000	2000, 2002, 2007	IBGE, Emplasa, Plano da	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Hidrografia Caieiras	Massas d'água	MassaDagua	shp.	Polígono	1:25.000	2002, 2007	Emplasa, Plano da Bacia	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Limites Caieiras	Área Urbanizada Edificada	AreaUrbanizadaEdificada	shp.	Polígono	1:25.000	2010/2011	Emplasa	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Viário Caieiras	Ferrovia	Ferrovia	shp.	Linha	1:25.000	2002	Emplasa	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Viário Caieiras	Linha de Transmissão	LinhaTransmissao	shp.	Linha	1:25.000	2002, 2010/2011	Emplasa	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Viário Caieiras	Rodovia	Rodovia	shp.	Linha	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Ortofotos Caieiras	Ortofoto	Caieiras	ecw.	Raster	1:10.000	2010/2011	Emplasa	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Modelo Digital de Superfície	MDE	mde	tif.	Raster	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Declividade Caieiras	Fatiameto	fatiameto	ovr.	Raster	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito

Categoria	Classe	Nome Arquivo	Formato Primário	Feição	Escala	Ano	Fonte	Site	Link	Possui VMS?	VMS	Gratuito, Pago ou Privado?
Declividade Caieiras	Declividade	declividade	ovr.	Raster	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Relevo Sombreado Caieiras	Relevo Sombreado	sombra	ovr.	Raster	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Relevo Caieiras	Tipos de Relevo	Relevo	shp.	Polígono	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Feições Caieiras	Feições	feicoes	shp.	Polígono	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Geologia Caieiras	Geologia	Geologia	shp.	Polígono	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Geologia Caieiras	Lineamentos	Lineamentos	shp.	Linha	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Pedologia Caieiras	Pedologia	Pedologia	shp.	Polígono	1:500.000	1999	Oliveira et al.	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Pluviosidade Caieiras	Isoietas	Isoietas_PMA_1977_2006	shp.	Linha	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Mapa Suscetibilidade Movimento de Massa Caieiras	Susctibilidade Movimento de Massa	ibildidadeMovimentoDeMassa	shp.	Polígono	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Mapa Suscetibilidade Movimento de Massa Caieiras	Susctibilidade Movimento de Massa	oetibilidadeMovimentoDeMa	shp.	Polígono	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito
Mapa Suscetibilidade Movimento de Massa Caieiras	Susctibilidade Movimento de Massa	SusctibilidadeBacias	shp.	Polígono	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-DefesaCivil/Prevencao-DefesaCivilCaieiras.aspx	Não		Gratuito

Anexo G – Levantamento de dados/fontes oficiais referentes à área de estudo.

Categoria	Classe	Nome Arquivo	Formato Primário	Feição	Escala	Ano	Fonte	Site	Link	Possui VMS?	VMS	Gratuito, Pago ou Privado?
Mapa Suscetibilidade Inundação Cateiras	Susceptibilidade Inundação	Susceptibilidadelnundacao	shp.	Polígono	1:25.000	2013	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Relevo Franco da Rocha	Relevo	Relevo	shp.	Polígono	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Feições Franco da Rocha	Feições	Feicoes	shp.	Polígono	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Geologia Franco da Rocha	Geologia	Geologia	shp.	Polígono	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Lineamentos Franco da Rocha	Lineamentos	Lineamentos	shp.	Linha	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Pedologia Franco da Rocha	Pedologia	Pedologia	shp.	Polígono	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Isoietas Franco da Rocha	Isoietas	Isoietas_FMA_1977_2006	shp.	Linha	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Localidades Franco da Rocha	Localidades	localidades	shp.	Ponto	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Mapa Suscetibilidade Movimentos de Massa Franco da Rocha	Susceptibilidade de Movimentos de Massa	ceptibilidadeMovimentoDeMa	shp.	Polígono	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Mapa Suscetibilidade Inundação Franco da Rocha	Susceptibilidade de Inundação	Susceptibilidadelnundacao	shp.	Polígono	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Áreas de Risco de Inundação de Franco da Rocha	Áreas de Risco de Inundação	Inundação	shp.	Polígono	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Categoria	Classe	Nome Arquivo	Formato Primário	Feição	Escala	Ano	Fonte	Site	Link	Possui VMS?	VMS	Gratuito, Pago ou Privado?
Modelo Digital de Terreno de Franco da Rocha	MDT	mdt	tif.	Raster	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Declividade Franco da Rocha	Declividade	declividade	ovr.	Raster	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Fatiamento Franco da Rocha	Fatiamento	fatiamento	ovr.	Raster	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Relevo Sombreado Franco da Rocha	Relevo Sombreado	sombra	ovr.	Raster	1:25.000	2017	IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-	Não		Gratuito
Mapa Cobertura do Solo	Cobertura do Solo		tif.	Raster	1:250.000	2018	Mapbiomas (Coleção 4.0)	Mapbiomas	http://www.mapbiomas.org/fma	Não		Gratuito
Mapa de Mudanças de Cobertura e Uso do Solo	Cobertura do Solo		tif.	Raster	1:250.000	2018	Mapbiomas (Coleção 4.0)	Mapbiomas	http://www.mapbiomas.org/fma	Não		Gratuito
Mapa Desmatamento Brasil	Desmatamento		tif.	Raster	1:250.000	2018	Mapbiomas (Coleção 4.0)	Mapbiomas	http://www.mapbiomas.org/fma	Não		Gratuito
Energia Brasil	Linha de Transmissão	linha_de_transmissao	shp.	Linha	Não Informada	2017	Operadora Nacional do	Mapbiomas	http://www.mapbiomas.org/fma	Não		Gratuito
Energia Brasil	Subestação	subestacao	shp.	Ponto	Não Informada	2017	Operadora Nacional do	Mapbiomas	http://www.mapbiomas.org/fma	Não		Gratuito
Perfil Municípios Brasil	Perfil Municípios	Base_MUNIC_2017_tifs	tifs	Não se Aplica	Não se Aplica	2017	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
PIB Municípios Brasil	PIB	Municípios - base de dados	tifs	Não se Aplica	Não se Aplica	2016	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
Estrutura Econômica Brasil	Extração Mineral	eco_ext_mineral	shp.	Ponto e Polígono	1:250.000	2017	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
Limites Brasil	Limites Municípios	lim_municipio_a	shp.	Polígono	1:250.000	2017	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
Limites Brasil	Limite País	lim_pais_a	shp.	Polígono	1:250.000	2017	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
Limites Brasil	Limites de Unidade de Proteção	lim_unidade_protecao_integral	shp.	Polígono	1:250.000	2017	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
Localidades Brasil	Área Edificada	loc_area_edificada_a	shp.	Polígono	1:250.000	2017	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
Localidades Brasil	Capital	loc_capital_p	shp.	Ponto	1:250.000	2017	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
Localidades Brasil	Cidade	loc_cidade_p	shp.	Ponto	1:250.000	2017	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
Transporte Brasil	Travessia	tra_travessia	shp.	Linha e Ponto	1:250.000	2017	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
Transporte Brasil	Trecho Ferroviário	tra_trecho_ferroviario_l	shp.	Linha	1:250.000	2017	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
Transporte Brasil	Trecho Rodoviário	tra_trecho_rodoviario_l	shp.	Linha	1:250.000	2017	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
Folhas Topográficas Vetoriais SP, MG e RJ	Capital Estadual	CAPITAL_ESTADUAL	shp.	Ponto	1:1.000	2003	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito
Folhas Topográficas Vetoriais SP, MG e RJ	Cidade	CIDADE	shp.	Ponto	1:1.000	2003	IBGE	IBGE	http://www.ibge.gov.br/download	Não		Gratuito

Anexo G – Levantamento de dados/fontes oficiais referentes à área de estudo.

Categoria	Classe	Nome Arquivo	Formato Primário	Feição	Escala	Ano	Fonte	Site	Link	Possui VMS?	VMS	Gratuito, Pago ou Privado?
Folhas Topográficas Vetoriais SP, MG e RJ	Ferrovia	FERROVIA	shp.	Linha	1:10.000	2003	IBGE	IBGE	sibge.gov.br/downloads	Não		Gratuito
Folhas Topográficas Vetoriais SP, MG e RJ	Estradas	OUTRAS_ESTRADAS	shp.	Linha	1:10.000	2003	IBGE	IBGE	sibge.gov.br/downloads	Não		Gratuito
Folhas Topográficas Vetoriais SP, MG e RJ	Rio	RIO_PERMANENTE	shp.	Linha	1:10.000	2003	IBGE	IBGE	sibge.gov.br/downloads	Não		Gratuito
Folhas Topográficas Vetoriais SP, MG e RJ	Rodovia	RODOVIA_PAVIMENTADA	shp.	Linha	1:10.000	2003	IBGE	IBGE	sibge.gov.br/downloads	Não		Gratuito
Folhas Topográficas Vetoriais SP, MG e RJ	Limite Estadual	E_POLITICO_ADMINISTRATIVO	shp.	Polígono	1:10.000	2003	IBGE	IBGE	sibge.gov.br/downloads	Não		Gratuito
Folhas Topográficas Vetoriais SP, MG e RJ	Área Edificada	AREA_EDIFICADA	shp.	Polígono	1:10.000	2003	IBGE	IBGE	sibge.gov.br/downloads	Não		Gratuito
Unidades de Conservação e Áreas Verdes RMSP	Parques Públicos e Ucs Simil	AvPar18_RMSP_CEM	shp.	Polígono	1:10.000	2018	Fundação SEADE, Secretaria do Verde do Estado de São Paulo, Google Earth, Instituto Florestal, Fundação Florestal, Prefeituras dos municípios da Região	CEM	http://centrodametropole.fhch.usp.br/pt-br/download-de-dados?%5B0%5D=facets_temas%3Ameio%20ambiente	Não		Gratuito
Unidades de Conservação e Áreas Verdes RMSP	Áreas Verdes pela Legislação	AvEsp18_RMSP_CEM	shp.	Polígono	1:10.000	2018	Fundação SEADE, Secretaria do Verde do Estado de São Paulo, Google Earth, Instituto Florestal, Fundação Florestal, Prefeituras dos municípios da Região	CEM	http://centrodametropole.fhch.usp.br/pt-br/download-de-dados?%5B0%5D=facets_temas%3Ameio%20ambiente	Não		Gratuito
Sistema Viário	Logradouros	LOG2018_CEM_RMSP	shp.	Linha	1:10.000	2018	Secretaria de Transporte do Município de São Paulo, Base de CEP's do correio, Google Earth,	CEM	http://centrodametropole.fhch.usp.br/pt-br/download-de-dados?%5B0%5D=facets_temas%3Ameio%20ambiente	Não		Gratuito
Educação RMSP	Escolas	Esc2013_RMSP_CEM	shp.	Ponto	1:10.000	2016	Secretaria de Educação do Estado de São Paulo, Cadastro Nacional de	CEM	http://centrodametropole.fhch.usp.br/pt-br/download-de-dados?%5B0%5D=facets_temas%3Ameio%20ambiente	Não		Gratuito
Saúde RMSP	Estabelecimentos de Saúde	SAU2016_CEM_RMSP	shp.	Ponto	1:10.000	2016	Estabelecimentos de	CEM	http://centrodametropole.fhch.usp.br/pt-br/download-de-dados?%5B0%5D=facets_temas%3Ameio%20ambiente	Não		Gratuito

Categoria	Classe	Nome Arquivo	Formato Primário	Feição	Escala	Ano	Fonte	Site	Link	Possui VMS?	VMS	Gratuito, Pago ou Privado?
Saúde RMSP	Estabelecimentos de Saúde	SAU2016_CEM_RMSP	shp.	Ponto	1:10.000	2016	Cadastro Nacional de Estabelecimentos de Saúde - CNES - do	CEM	http://centrodametropole.fhch.usp.br/pt-br/download-de-dados?%5B0%5D=facets_temas%3Ameio%20ambiente	Não		Gratuito
Hidrografia RMSP	Trechos de Drenagem	HidroL17_RMSP_CEM	shp.	Linha	1:10.000	2016	Secretaria do Medio Ambiente do Estado de São Paulo, DAEE, Sabesp, Prefeituras dos	CEM	http://centrodametropole.fhch.usp.br/pt-br/download-de-dados?%5B0%5D=facets_temas%3Ameio%20ambiente	Não		Gratuito
Divisão Territorial Brasil	Regiões metropolitanas em 2	RmeBr12_CEM	shp.	Polígono	Não Informada	2013	IBGE	CEM	http://centrodametropole.fhch.usp.br/pt-br/download-de-dados?%5B0%5D=facets_temas%3Ameio%20ambiente	Não		Gratuito
Divisão Territorial Brasil	Regiões Metropolitanas divic	MunRm12_CEM	shp.	Polígono	Não Informada	2013	IBGE	CEM	http://centrodametropole.fhch.usp.br/pt-br/download-de-dados?%5B0%5D=facets_temas%3Ameio%20ambiente	Não		Gratuito
Divisão Territorial Regiões Metropolitanas	Setores CEP's	CepRM07	shp.	Polígono	1:10.000	2000	Secretaria de Transporte do Município de São	CEM	http://centrodametropole.fhch.usp.br/pt-br/download-de-dados?%5B0%5D=facets_temas%3Ameio%20ambiente	Não		Gratuito
Divisão Territorial RMSP	Subprefeituras	SubSA007	shp.	Polígono	Não Informada	2003	IBGE	CEM	http://centrodametropole.fhch.usp.br/pt-br/download-de-dados?%5B0%5D=facets_temas%3Ameio%20ambiente	Não		Gratuito
Imageamento Ortoimagem	Imagens RapidEye		wms	Raster	Não Informada	2013	MMA	DataGeo	ambiente.sp.gov.br/app/?	Sim	http://datageo.am	Gratuito
Imageamento Ortoimagem	Ortofotos do Estado de São Paulo		wms	Raster	1:25.000	2010	Emplasa	DataGeo	sp.gov.br/images/inform	Sim	http://datageo.am	Gratuito
Transporte SP	Malha Rodoviária		wms	Polígono	Não Informada	2012	DER	DataGeo	ambiente.sp.gov.br/app/?	Sim	http://datageo.am	Gratuito
Relevo SP	Pedologia	DOLOGICO_ROSSI_2017_P	shp.	Polígono	1:250.000	2017	IF	DataGeo	https://www.infraestruturameioambiente.sp.gov.br/	Sim	http://datageo.am	Gratuito
Relevo SP	Recursos Minerais		wms	Ponto	1:750.000	2005	CPRM	DataGeo	ambiente.sp.gov.br/app/?	Sim	http://datageo.am	Gratuito
Relevo SP	Declividade em Graus		wms	Raster	1:50.000	2015	SMA	DataGeo	ambiente.sp.gov.br/app/?	Sim	http://datageo.am	Gratuito
Relevo SP	Declividade em Porcentagem		wms	Raster	1:50.000	2013	SMA	DataGeo	ambiente.sp.gov.br/app/?	Sim	http://datageo.am	Gratuito
Relevo SP	Modelo Digital de Elevação	MDE_ESP_v2	tif.	Raster	1:50.000	2013	SMA	DataGeo	sp.gov.br/cpl/2013/10/f	Sim	http://datageo.am	Gratuito
Relevo SP	Relevo Sombreado		wms	Raster	1:50.000	2013	SMA	DataGeo	ambiente.sp.gov.br/app/?	Sim	http://datageo.am	Gratuito
Condição de Vida SP	Escolaridade	EscolaridadeMunicipio2014F	shp.	Polígono	Não Informada	2014	SEADE	DataGeo	ambiente.sp.gov.br/app/?	Sim	http://datageo.am	Gratuito

Anexo G – Levantamento de dados/fontes oficiais referentes à área de estudo.

Categoria	Classe	Nome Arquivo	Formato Primário	Feição	Escala	Ano	Fonte	Site	Link	Possui VMS?	VMS	Gratuito, Pago ou Privado?
Condição de Vida SP	Longevidade	LongevidadeMunicípio2014F	shp.	Polígono	Não Informada	2014	SEADE	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Condição de Vida SP	Riqueza	seRiquezaMunicípio2014Pol	shp.	Polígono	Não Informada	2014	SEADE	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Condição de Vida SP	IPRS - Índice Paulista de Res	IPRS2014Polygon	shp.	Polígono	Não Informada	2014	SEADE	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Demografia SP	Densidade Demográfica	isidadeDemografica2018Poly	shp.	Polígono	1:50.000	2018	SEADE	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Fiscalização SP	Autorização Ambiental		wms	Ponto	Não Informada	2015	SMA	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Fiscalização SP	Auto de Infração Ambiental - AIA		wms	Ponto	Não se Aplica	2019	Coordenadoria de Fiscalização e	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Mapeamento de Áreas de Ri	Áreas de Risco de Escorrega	VWM_AREA_RISCO_ESCORREGAMENTO_IG_2014	shp.	Polígono	1:25.000	2014	IG-SP	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Gerenciamento Preventivo de Riscos Naturais SP	Acidentes relacionados a escorregamentos (período	VWM_DESASTRE_ESCORREGAMENTO_IG_2000_	shp.	Polígono	Não Informada	2013	IG-SP	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Gerenciamento Preventivo de Riscos Naturais SP	Acidentes relacionados a inu	TRE_INUNDACAO_IG_2000_	shp.	Polígono	Não Informada	2013	IG-SP	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Gerenciamento Preventivo de Riscos Naturais SP	Óbitos por acidentes e desastres devido a eventos	ASTRE_OBITOS_IG_2000_2	shp.	Polígono	Não Informada	2013	IG-SP	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Gerenciamento Preventivo de Riscos Naturais SP	Pessoas afetadas por acidentes e desastres	VWM_DESASTRE_PESSOAS_AFETADAS_IG_200	shp.	Polígono	Não Informada	2013	IG-SP	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Mananciais SP	Limite da Bacia Alto Juquery	IMA_J_LIMITE_BACIA_2015_	shp.	Polígono	1:10.000	2015	IPT	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Unidades de Conservação SP	Ues Estaduais de Proteção Integral		wms	Polígono	Não Informada	2018	IF e FF	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Limites Administrativos SP	Limite Estadual	LimiteEstadualPolygon	shp.	Polígono	1:50.000	2010	SMA	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Limites Administrativos SP	Regiões Metropolitanas	ES_METROPOLITANAS_20	shp.	Polígono	1:50.000	2016	IGC-SP	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Limites Administrativos SP	Sedes Municipais	SedesMunicipaisPoint	shp.	Ponto	1:50.000	2010	IGC-SP	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Sistema Viário SP	Logradouros		wms	Linha	Não Informada	2013	SMA	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Rede Urbana SP	Rede Urbana Emplasa 2011	RedeUrbanaEmplasa2011Polyg	shp.	Polígono	1:50.000	2011	Emplasa	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito

Categoria	Classe	Nome Arquivo	Formato Primário	Feição	Escala	Ano	Fonte	Site	Link	Possui VMS?	VMS	Gratuito, Pago ou Privado?
Rede Urbana SP	Rede Urbana Emplasa 2011	RedeUrbanaEmplasa2011Polyg	shp.	Polígono	1:50.000	2011	Emplasa	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Censo 2010	Setores Censitários IBGE - 2	33SEE250GC_SIR	shp.	Polígono	1:250.000	2010	IBGE	DataGeo	mbiente.sp.gov.br/app/?v=1	Sim	http://dados.agedo.am	Gratuito
Energia Brasil	Linhas de transmissão	abilidades e zee_linhas_transm	wms e shp	Linha	1:2.000.000	2012	SIGEL	MMA	http://mapas.mma.gov.br/geonetwork/srv/br/metadata.show?id=1333	Não		
Limites Administrativos Bras	Limites Municipais com Dado	municipalocio	shp.	Polígono		2001	IBGE	IBGE	https://www.ibge.gov.br/	Não		
Vulnerabilidade a Inundações	Trechos de cursos d'água com vulnerabilidade a	SUM_vulnerabilidade_brasil	lgr.	Linha		2019	ANA	ANA	http://portal.snirh.gov.br	Não		Gratuito
Eventos Hidrológicos Críticos	Eventos críticos de cheia no	ntos_criticos_cheias_inunda	shp.	Ponto	1:1.000.000	2015	ANA, Ministério da Integra	ANA	https://metadados.ana.gov.br	Não		Gratuito
Eventos Hidrológicos Críticos	Vulnerabilidade a Inundações asil	vulnerabilidade_inundaco	shp.	Linha	1:1.000.000	2014	ANA	ANA	https://portal.snirh.gov.br	Não		Gratuito
Precipitação Brasil	Precipitação média anual (197	SNIRH_Isoietas	shp.	Linha	1:1.000.000	2016	ANA, CPRM	ANA	https://www.cprm.gov.br	Não		Gratuito
Educação e Cultura SP	Escolas		wms.	Ponto	1:250.000	2013	Secretaria da Educação di	SIM - Emplasa	https://sim.emplasa.sp.gov.br	Sim	https://dados.agedo.am	Gratuito
Transporte SP	Linhas ferroviárias		wms.	Polígono		2013	Emplasa, Promaps	SIM - Emplasa	https://sim.emplasa.sp.gov.br	Sim	https://dados.agedo.am	Gratuito
Transporte SP	Estações CPTM		wms.	Ponto		2012	STM	SIM - Emplasa	https://sim.emplasa.sp.gov.br	Sim	https://dados.agedo.am	Gratuito
Transporte SP	Rodoanel		wms.	Polígono		2016	Emplasa	SIM - Emplasa	https://sim.emplasa.sp.gov.br	Sim	https://dados.agedo.am	Gratuito
Dados Pluviométricos Brasi	Estações pluviométricas		xls.			2013 - 2020	Cemaden	Cemaden	http://www.cemaden.gov.br	Não		Gratuito
Imagem NDVI	Landsat 8 NDVI Composite		Site	Raster		2013 - 2019	Google	GEE	https://code.earthengine.net	Não		Gratuito
Imagem Landsat 8	Landsat 8 Reflectância de Superfície		Site	Raster		2013 - 2019	USGS	GEE	https://code.earthengine.net	Não		Gratuito
Imagem Vegetação	Mudança de Cobertura Florestal		Site	Raster		2000 - 2018	Hansen/UMD/Google/US	GEE	http://earthenginepartners.net	Não		Gratuito
Imagem Queimadas	MODIS Área Queimada Mensal 500 m		Site	Raster		2000 - 2019	NASA LP DAAC at the US	GEE	https://code.earthengine.net	Não		Gratuito
Imagem CBERS	Imagem 2012/2018	RS_4_AVF1_20181220_151_12	tif.	Raster		2018	INPE	INPE	http://www.dgi.inpe.br/CJ	Não		Gratuito
Imagem Landsat 8	Imagem 2012/2018	LD82190762018354CUB00	tif.	Raster		2018	INPE	INPE	http://www.dgi.inpe.br/CJ	Não		Gratuito
Imagem Sentinel 1	Imagem 15/02/2019	S1A_IW_GRDH_1SDV_20190215T080210_20190215T080210	tif.	Raster		2019	ESA	ESA	https://scihub.copernicus.eu	Não		Gratuito
Imagem Sentinel 1	Imagem 17/03/2019	S1A_IW_GRDH_1SDV_20190317T083051_20190317T083051	tif.	Raster		2019	ESA	ESA	https://scihub.copernicus.eu	Não		Gratuito
Hidrografia Francisco Morat	Cursos d'água	CursoDagua	shp.	Linha	1:25.000	2000, 2002, 2007	IBGE, Emplasa, Plano da I	CPRM	http://www.cprm.gov.br/publique/Gestao	Não		Gratuito
Hidrografia Franco da Rocha	Cursos d'água	CursoDagua	shp.	Linha	1:25.000	2000, 2002, 2007	IBGE, Emplasa, Plano da I	CPRM	http://www.cprm.gov.br/publique/Gestao	Não		Gratuito

Anexo G – Levantamento de dados/fontes oficiais referentes à área de estudo.

Categoria	Classe	Nome Arquivo	Formato Primário	Feição	Escala	Ano	Fonte	Site	Link	Possui VMS?	VMS	Gratuito, Pago ou Privado?
Hidrografia Franco da Rocha	Massas d'água	MassaDagua	shp.	Polígono	1:25.000	2002, 2007	Emplasa, Plano da Bacia	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Condição de Vida SP	IPVS - Índice Paulista de Vulnerabilidade	BaseIPVS2010	csv.	Não se Aplica	Não se Aplica	2010	SEADE			Não		Gratuito
Altimetria Franco da Rocha	Curva de Nível	CurvaDeNivel	shp.	Linha	1:25.000	2017	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-aw/index.php?seil-oca=100	Não		Gratuito
Altimetria Franco da Rocha	Ponto Cotado	PontoCotado	shp.	Ponto	1:25.000	2017	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-aw/index.php?seil-oca=100	Não		Gratuito
Limites Franco da Rocha	Localidades	Localidades	shp.	Ponto	1:25.000	2017	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-aw/index.php?seil-oca=100	Não		Gratuito
Limites Franco da Rocha	Área Urbanizada Edificada	AreaUrbanizadaEdificada	shp.	Polígono	1:25.000	2017	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-Territorial/Prevencao-aw/index.php?seil-oca=100	Não		Gratuito
Altimetria Francisco Morato	Curva de Nível	CurvaDeNivel	shp.	Linha	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Altimetria Francisco Morato	Ponto Cotado	PontoCotado	shp.	Ponto	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Hidrografia Francisco Morato	Cursos d'água	CursoDagua	shp.	Linha	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Limites Francisco Morato	Localidades	Localidades	shp.	Ponto	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Limites Francisco Morato	Área Urbanizada Edificada	AreaUrbanizadaEdificada	shp.	Polígono	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Viário Francisco Morato	Ferrovias	Ferrovias	shp.	Linha	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Viário Francisco Morato	Linhas de transmissão	LinhaTransmissao	shp.	Linha	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Viário Francisco Morato	Rodovias	Rodovia	shp.	Linha	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Modelo Digital de Superfície Francisco Morato	Hipsometria	Hipsometria	shp.	Raster	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Categoria	Classe	Nome Arquivo	Formato Primário	Feição	Escala	Ano	Fonte	Site	Link	Possui VMS?	VMS	Gratuito, Pago ou Privado?
Modelo Digital de Superfície Francisco Morato	Declividade	declividade	ovr.	Raster	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Modelo Digital de Superfície Francisco Morato	Fatiamento	fatiamento	ovr.	Raster	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Modelo Digital de Superfície Francisco Morato	Relevo Sombreado	sombra	ovr.	Raster	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Ortofoto Francisco Morato	Ortofoto	FranciscoMorato	tif.	Raster	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Modelo Digital de Superfície Francisco Morato	Padrões de Relevo	Relevo	shp.	Polígono	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Feições Francisco Morato	Feições	Feicoes	shp.	Polígono	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Geologia Francisco Morato	Geologia	Geologia	shp.	Polígono	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Geologia Francisco Morato	Lineamentos	Lineamentos	shp.	Linha	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Pedologia Francisco Morato	Pedologia	Pedologia	shp.	Polígono	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Pluviometria Francisco Morato	Isoietas	etas Anuais Médias 1977 a 2010	shp.	Linha	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Mapeamento de Suscetibilidade à Inundação	Susceptibilidade de Inundação	SusceptibilidadeInundacao	shp.	Polígono	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Mapeamento de Suscetibilidade à Inundação	Susceptibilidade de Movimento de Massa	SusceptibilidadeMovimentoDeMassa	shp.	Polígono	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Cobertura do Solo Brasil	Cobertura do Solo	SP_atlas_2016	shp.	Polígono	1:50.000	2016	Fundação SOS Mata Atlântica	Fundação SOS Mata Atlântica	http://mapas.sosma.org.br/	Não		Gratuito
Modelo Digital de Superfície Caieiras	Hipsometria	Hipsometria	shp.	Raster	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Modelo Digital de Superfície Caieiras	Declividade	declividade	ovr.	Raster	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Modelo Digital de Superfície Caieiras	Fatiamento	fatiamento	ovr.	Raster	1:25.000	2013	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito
Modelo Digital de Terreno de Franco da Rocha	Hipsometria	Hipsometria	shp.	Raster	1:25.000	2017	CPRM/IIPT	CPRM	http://www.cprm.gov.br/publique/Gestao-aw/index.php?seil-oca=100	Não		Gratuito

Anexo G – Levantamento de dados/fontes oficiais referentes à área de estudo.

Categoria	Classe	Nome Arquivo	Formato Primário	Feição	Escala	Ano	Fonte	Site	Link	Possui VMS?	VMS	Gratuito, Pago ou Privado?
Modelo Digital de Terreno de Franco da Rocha	Fatiamento	fatiamento	ovr.	Raster	1:25.000	2017	CPRM/IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-	Não		Gratuito
Viário Franco da Rocha	Viário	Viario	shp.	Linha	1:25.000	2017	CPRM/IPT	CPRM	http://www.cprm.gov.br/publique/Gestao-	Não		Gratuito
Processos Morfodinâmicos	Escorregamentos	PROCESSOS_MORFO	shp.	Ponto	1:150.000	2019	CPTM	CPTM		Não		Privado
Inundações na CPTM	Inundações	PT_INUNDACOES.shp	shp.	Ponto	1:150.000	2019	CPTM	CPTM		Não		Privado
Estações CPTM	Estações	ESTACOES	shp.	Ponto		2019	CPTM			Não		Privado
Linha CPTM	Ferrovias	LINHAS_CPTM	shp.	Linha		2019	CPTM			Não		Privado
Escorregamentos Francisco	Escorregamentos	Escorregamentos Francisco Morato	xls.			2013	Defesa Civil Francisco Morato			Não		Privado
Censo 2010	Base de informações por setor censitário		xls. ou osv.			2010	Censo 2010/IBGE	IBGE	ftp://ftp.ibge.gov.br/Censos/Censo_Demografico/2010/	Não		Gratuito

Anexo H – Relatório “Do SIG ao *Big Data*: Estruturação dos Dados Espaciais do Entorno das Estações Ferroviárias Caieiras e Francisco Morato, SP” apresentado no 34º ANPET e publicado nos anais do evento em 14 de dezembro de 2020.

34º Congresso de Pesquisa e Ensino em Transporte da ANPET
100% Digital, 16 a 21 de novembro de 2020

DO SIG AO *BIG DATA*: ESTRUTURAÇÃO DOS DADOS ESPACIAIS DO ENTORNO DAS ESTAÇÕES FERROVIÁRIAS CAIEIRAS E FRANCISCO MORATO, SP

Thais P. Correia

Departamento de Engenharia de Transportes – Escola Politécnica
Universidade de São Paulo (PTR - EPUSP)

José A. Quintanilha

Instituto de Energia e Ambiente
Universidade de São Paulo (IEE - USP)

Alessandra C. Corsi

Seção de Investigações, Riscos e Desastres Naturais do Centro de Tecnologias Geomambientais
Instituto de Pesquisas Tecnológicas do Estado de São Paulo (Sirden – CTGEO – IPT)

Lucas H. Sandre

Instituto de Geociências
Universidade Estadual de Campinas (IG – Unicamp)

RESUMO

As inundações e os escorregamentos estão entre os desastres naturais mais comuns no Brasil, que são responsáveis por prejuízos de ordem social, ambiental e econômica, podendo resultar em perdas de vidas. As geotecnologias tornaram-se uma opção para auxiliar no mapeamento desses eventos e, consequentemente, ajudar na previsão desses eventos, na avaliação de vulnerabilidades e na quantificação de perdas. As técnicas de *Big Data* também vêm se destacando por sua capacidade de integração de uma grande quantidade de bancos de dados, de diversas fontes e de formatos distintos; e colaborando para a melhoria da tomada de decisão. O estudo terá como foco o entorno das estações Caieiras e Francisco Morato da Linha 7 - Rubi da CPTM (Companhia Paulista de Trens Metropolitanos), na qual esses eventos são frequentes e podem afetar a circulação de trens, além de acarretarem perdas econômicas e danos sociais aos moradores e trabalhadores da região. Essa pesquisa identificará as necessidades e dificuldades de se realizar a transferência de dados (estruturados e não-estruturados) em um Sistema de Informações Geográficas - SIG para um ambiente *Big Data*, contribuindo para uma melhor compreensão dos desastres naturais, principalmente escorregamentos e inundações, e como eles afetam o sistema ferroviário.

1. PROPOSTA DA PESQUISA

A proposta deste estudo é identificar as necessidades e as dificuldades de realizar a passagem de um dado em ambiente SIG para um ambiente *Big Data*, analisando dados estruturados e não-estruturados de diferentes bancos de dados espaciais, contribuindo para uma melhor compreensão dos desastres naturais e como eles afetam o sistema ferroviário. Pretende-se também obter um diagnóstico, em termos de disponibilidade e viabilidade de uso dos dados existentes da região de estudo, sobre variáveis de interesse para a tomada de decisão com foco nos desastres naturais, e mostrar o contexto atual das referências bibliográficas associados ao problema tratado.

2. REVISÃO BIBLIOGRÁFICA

2.1. Desastres naturais

De acordo com o Centro Nacional de Gerenciamento de Riscos e Desastres – CENAD (Brasil, 2014, p. 15): “Desastre é o resultado de eventos adversos, naturais ou provocados pelo homem, sobre um cenário vulnerável, causando grave perturbação ao funcionamento de uma comunidade ou sociedade, envolvendo extensivas perdas e danos humanos, materiais, econômicos ou ambientais, que excede a sua capacidade de lidar com o problema usando meios próprios”. Segundo Tominaga (2009), no Brasil, os principais desastres naturais derivam da dinâmica externa da Terra, ou seja, aqueles que são orientados pelo clima e por processos atmosféricos. Os eventos mais comuns são inundações e enchentes,

anpet