

MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA
INSTITUTO MILITAR DE ENGENHARIA
CURSO DE DOUTORADO EM ENGENHARIA DE DEFESA

LIVIA DOS SANTOS ABDALLA

**MODELAGEM BASEADA EM DADOS PARA PREVISÃO DA
EMERGÊNCIA DE ZOONOSES: UM ESTUDO DE CASO DA FEBRE
AMARELA SILVESTRE NO BRASIL**

**RIO DE JANEIRO
AGOSTO 2019**

INSTITUTO MILITAR DE ENGENHARIA

LIVIA DOS SANTOS ABDALLA

**MODELAGEM BASEADA EM DADOS PARA PREVISÃO DA
EMERGÊNCIA DE ZOOSE: UM ESTUDO DE CASO DA FEBRE
AMARELA SILVESTRE NO BRASIL**

Tese de Doutorado apresentada ao Curso de Engenharia de Defesa do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Doutora em Ciências em Engenharia de Defesa.

Orientadores: Prof. Leonardo Castro de Oliveira – D.E.
Prof. Eduardo Krempser – D.Sc.

Rio de Janeiro
2019

c2019

INSTITUTO MILITAR DE ENGENHARIA

Praça General Tibúrcio, 80 – Praia Vermelha

Rio de Janeiro - RJ CEP: 22290-270

Este exemplar é de propriedade do Instituto Militar de Engenharia, que poderá incluí-lo em base de dados, armazenar em computador, microfilmar ou adotar qualquer forma de arquivamento.

É permitida a menção, reprodução parcial ou integral e a transmissão entre bibliotecas deste trabalho, sem modificação de seu texto, em qualquer meio que esteja ou venha a ser fixado, para pesquisa acadêmica, comentários e citações, desde que sem finalidade comercial e que seja feita a referência bibliográfica completa.

Os conceitos expressos neste trabalho são de responsabilidade do(s) autor(es) e do(s) orientador(es).

Abdalla, Livia dos Santos

MODELAGEM BASEADA EM DADOS PARA PREVISÃO DA EMERGÊNCIA
DE ZOONOSES: UM ESTUDO DE CASO DA FEBRE AMARELA SILVESTRE
NO BRASIL / Livia dos Santos Abdalla, orientada por Leonardo Castro de Oliveira e
Eduardo Krempser – Rio de Janeiro: Instituto Militar de Engenharia, 2019.

522 p.: il.

Tese (Doutorado) – Instituto Militar de Engenharia, Rio de Janeiro, 2019.

1. Curso de Engenharia de Defesa – Tese. Instituto Militar de Engenharia.

INSTITUTO MILITAR DE ENGENHARIA

LIVIA DOS SANTOS ABDALLA

**MODELAGEM BASEADA EM DADOS PARA PREVISÃO DA
EMERGÊNCIA DE ZOONOSES: UM ESTUDO DE CASO DA FEBRE
AMARELA SILVESTRE NO BRASIL**

Tese de Doutorado apresentada ao Curso de Doutorado em Engenharia de Defesa do Instituto Militar de Engenharia, como requisito parcial para a obtenção do título de Doutora em Ciências em Engenharia de Defesa.

Orientadores: Prof. Leonardo Castro de Oliveira - D.E.

Prof. Eduardo Krempser – D.Sc.

Aprovada em 28 de agosto de 2019 pela seguinte Banca Examinadora:

Prof. Leonardo Castro de Oliveira - D.E./IME - Presidente

Eduardo Krempser - D.Sc./Fiocruz

Adriano Pinter dos Santos – D.Sc./SUCEN-SP

Prof. Fabio Ribeiro Cerqueira - D.Sc./UFF

Douglas Adriano Augusto - D.Sc./Fiocruz

Marcia Chame – D.Sc./Fiocruz

Prof. Ivanildo Barbosa - D.Sc./IME

AGRADECIMENTOS

Aos meus pais, Sonia Abdalla e Wagner Abdalla, pela sólida base familiar, valores, estímulo, incentivo e paciência. À minha irmã Aline Abdalla pelo cuidado, amizade, alegria e permanente entusiasmo! À vocês, meu eterno amor, agradecimento e respeito. Sem vocês, nada teria sido possível.

Ao meu marido Guilherme Giminiani pela cumplicidade, companheirismo, respeito, incentivo, compreensão e carinho sem fim. Obrigada por me escutar e me ajudar a pensar todas as vezes que minha ansiedade e nervosismo se sobrepuseram sobre minha capacidade de raciocínio. Você é, definitivamente, a melhor pessoa que eu poderia ter ao meu lado.

Aos meus estimados orientadores, Prof. Dr. Leonardo C. Oliveira e Prof. Dr. Eduardo Krempser. Muito obrigada pela dedicação, paciência, atenção, instigações e questionamentos brilhantes! Vocês fizeram toda a diferença para o meu aprimoramento acadêmico, intelectual e pessoal. Muito obrigada! Não há maneiras para retribuir, mas sim reconhecer!

À Marcia Chame, minha coordenadora na Fiocruz, pelos ensinamentos e impulsionamentos (científicos, profissionais e pessoais), pela confiança e constante encorajamento. Te agradeço pela extensa, valiosa e memorável contribuição à minha formação técnica, científica e profissional.

Aos amigos da Fiocruz: Douglas Augusto e Marina Galvão Bueno, obrigada pelos conselhos técnico-científicos sobre diferentes aspectos da tese (modelos, métodos, primatas, mosquitos). Ao Luiz Carlos pelas conversas reflexivas e acolhedoras. E às amigas: Rita Braune e Raquel M. Marques, pelo carinho, empatia, desabafos, conselhos e inumeráveis risadas (vocês me ajudaram muito a recuperar o fôlego durante essa árdua caminhada).

Aos amigos da Geo-UFF Paula Pitassi e Victor Duba, pela leal e duradoura amizade. Vocês que sempre se fizeram presentes na minha vida, novamente me ajudaram a conduzir com maior leveza e contentamento essa dura jornada do doutorado. Obrigada pela construção da nossa linda parceria e trajetória.

Por fim, agradeço ao Instituto Militar de Engenharia – IME e à Fundação Oswaldo Cruz – Fiocruz, alicerces do meu conhecimento técnico-científico e espaços propulsores ao meu aperfeiçoamento profissional.

“O conhecimento serve para encantar pessoas,
não para humilhá-las”

MARIO SERGIO CORTELLA

SUMÁRIO

RESUMO.....	15
1. INTRODUÇÃO.....	16
1.1 Contextualização.....	16
1.2 Hipótese.....	20
1.3 Objetivo.....	20
1.4 Justificativa.....	20
1.5 Estrutura da tese.....	22
2. ESTADO DA ARTE, CONCEITOS E FUNDAMENTOS TEÓRICOS.....	25
2.1. Modelagem de emergência de zoonoses: Estado do conhecimento.....	25
2.2 Modelagem explicativa x preditiva.....	34
2.3 Nomenclaturas distintas na literatura técnico-científica.....	37
2.4 Modelagem Baseada em Dados - MBD.....	40
2.5 Aprendizagem de Máquina - AM.....	43
2.6 Métodos de classificação.....	47
2.7 Febre amarela.....	54
2.7.1 Aspectos gerais.....	54
2.7.2 Breve histórico da Febre amarela no Brasil.....	56
2.7.3 Sistema de vigilância da Febre Amarela Silvestre (FAS) – 1999 a 2018.....	60
3. METODOLOGIA PARA PREVISÃO DA EMERGÊNCIA DA FEBRE AMARELA SILVESTRE NO BRASIL POR MODELAGEM BASEADA EM DADOS.....	66
3.1 Aquisição das bases de dados.....	69
3.2 Produção de municípios pseudonegativos.....	75
3.3 Pré-processamento e estruturação.....	82
3.4 Descrição estatística das características socioambientais dos municípios.....	84
3.5 Base de dados com as estatísticas socioambientais e as ocorrências de febre amarela por município.....	87
3.6 Modelagem baseada em dados.....	93
3.7 Validação dos modelos.....	99
4. RESULTADOS E DISCUSSÃO.....	104

4.1 Modelos experimentais.....	104
4.1.1 Modelos recorte 1999 a 2017.....	106
4.1.2 Modelos recorte 1999 a 2003.....	111
4.1.3 Modelos recorte 2007 a 2009.....	114
4.1.4 Modelos recorte 2015 a 2017.....	118
4.2 Principais aspectos sobre os dados/variáveis temáticas indicadas nos modelos experimentais.....	121
4.2.1 Perfil demográfico.....	126
4.2.2 Primatas Não Humanos.....	132
4.2.3 Biomas.....	145
4.2.4 Unidades de conservação.....	160
4.2.5 Tipos de uso e cobertura da terra.....	163
4.2.6 Temperatura e precipitação.....	176
4.2.7 Mineração.....	206
4.2.8 Ferrovias.....	214
4.2.9 Solos.....	217
4.2.10 Relevô.....	221
4.3 Modelos finais.....	223
4.3.1 Modelos finais – Método J48.....	224
4.3.1.1 Recorte 01 (1999 a 2017).....	224
4.3.1.2 Recorte 02 (1999 a 2003).....	235
4.3.1.3 Recorte 03 (2007 a 2009).....	240
4.3.1.4 Recorte 04 (2015 a 2017).....	246
4.3.2 Modelos finais – Método <i>Random Forest</i>	252
4.3.2.1 Recorte 01 (1999 a 2017).....	252
4.3.2.2 Recorte 02 (1999 a 2003).....	254
4.3.2.3 Recorte 03 (2007 a 2009).....	256
4.3.2.4 Recorte 04 (2015 a 2017).....	258
5. CONCLUSÕES.....	260
5.1 Recomendações da pesquisa.....	265

REFERÊNCIAS BIBLIOGRÁFICAS.....	266
---------------------------------	-----

LISTA DE ILUSTRAÇÕES

Figura 2.1: Aprendizagem por modelagem baseada em dados (Adaptado de SOLOMATINE e OSTFELD, 2008).....	42
Figura 2.2: Configuração lógica da modelagem baseada em dados (Fonte: ABU-MOSTAFA et al, 2012).....	44
Figura 2.3: Função/estrutura de neurônios (Adaptado de ABU-MOSTAFA, 2019).....	48
Figura 2.4: Estrutura rede neuronal (Adaptado de ABU-MOSTAFA, 2019).....	49
Figura 2.5: Classificação de um objeto pelo método KNN.....	50
Figura 2.6: Hiperplanos de separação de amostras - método SVM.....	51
Figura 2.7: Classificação de amostras pelo método de Árvores de decisão.....	52
Figura 2.8: Seleção de amostras e classificação por Programação Genética.....	53
Figura 2.9: Países com risco de transmissão do vírus da febre amarela.....	54
Figura 2.10: Ciclo de transmissão da febre amarela (Fonte).....	55
Figura 2.11: Provável rota da disseminação da febre amarela no Brasil (1932-1942) Fonte: Corrigir autoria.....	59
Figura 2.12: Distribuição dos casos humanos e epizootias confirmadas para febre amarela no período de 1999 a 2018.....	61
Figura 2.13: Casos humanos e epizootias confirmadas para febre amarela no período de 1999 a 2018 (Fonte: Sistema de vigilância da Febre Amarela Silvestre - GT-Arboviroses).....	62
Figura 2.14: Distribuição dos casos humanos e epizootias confirmadas para febre amarela no período de 1999 a 2003.....	63
Figura 2.15: Distribuição dos casos humanos e epizootias confirmadas para febre amarela no período de 2007 a 2009.....	64
Figura 2.16: Distribuição dos casos humanos e epizootias confirmadas para febre amarela no período de 2015 a 2018.....	65
Figura 3.1: Fluxograma da modelagem de emergência favorabilidade de emergência da febre amarela silvestre.....	68
Figura 3.2: Municípios positivos e neutros.....	69

Figura 3.3: Distância euclidiana no espaço de atributos entre municípios.....	76
Figura 3.4: Distância euclidiana no espaço geográfico entre municípios.....	76
Figura 3.5: Representação de uma árvore de decisão binária (Fonte: GAMA, 2004).....	95
Figura 3.6: Geração da Curva ROC de um classificador (MERSCHMANN, 2018).....	103
Figura 4.1: Contagem população - 2015.....	127
Figura 4.2: Densidade população - 2015.....	128
Figura 4.3: Ciclos de transmissão da febre amarela na África (Fonte: WHO, 2016).....	131
Figura 4.4: Total de animais com diagnóstico confirmado para Febre Amarela por gênero (Fonte: Sistema de vigilância da Febre Amarela Silvestre entre 1999 e 2018 - GT-Arboviroses).....	133
Figura 4.5: Área de distribuição <i>Callithrix jacchus</i> e <i>Callithrix penicillata</i>	134
Figura 4.6: Área de distribuição <i>Alouatta caraya</i> , <i>Alouatta guariba</i> , <i>Alouatta macconnelli</i> e <i>Alouatta belzebul</i>	137
Figura 4.7: Área de distribuição <i>Callicebus personatus</i> , <i>Callicebus moloch</i> e <i>Callicebus nigrifrons</i>	140
Figura 4.8: Área de distribuição <i>Aotus trivirgatus</i>	142
Figura 4.9: Biomas Amazônia, Caatinga e Mata Atlântica.....	146
Figura 4.10: Área da classe Floresta no bioma Amazônia entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018).....	151
Figura 4.11: Área da classe Floresta no bioma Caatinga entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018).....	152
Figura 4.12: Área da classe Agropecuária no bioma Amazônia entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018).....	153
Figura 4.13: Área da classe Agropecuária no bioma Caatinga entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018).....	153
Figura 4.14: Área da classe Floresta no bioma Mata Atlântica entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018).....	154
Figura 4.15: Área da classe Floresta Plantada no bioma Mata Atlântica entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018).....	155
Figura 4.16: Área da classe Agropecuária no bioma Mata Atlântica entre 1985 e 2017 (Fonte:	

MAPBIOMAS, 2018).....	157
Figura 4.17: Área da classe Floresta no bioma Cerrado entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018).....	158
Figura 4.18: Área da classe Agropecuária no bioma Cerrado entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018).....	158
Figura 4.19: Limite de Unidades de Conservação.....	160
Figura 4.20: Classe Formações Florestais.....	164
Figura 4.21: Classe Agricultura ou Pastagem.....	167
Figura 4.22: Projeção de regiões a serem convertidas de florestas a áreas agrícolas (Fonte: ESTRADA <i>et al</i> , 2017).....	169
Figura 4.23: Classes Formações Savânicas e Vegetação Campestre.....	170
Figura 4.24: Classes Corpos d'água e Áreas úmidas naturais não florestais.....	173
Figura 4.25: Classe Praias e Dunas.....	175
Figura 4.26: Precipitação do trimestre mais frio – Bio19 (<i>Worldclim</i>).....	178
Figura 4.27: Precipitação total mensal – Prec6 (<i>Worldclim</i>).....	179
Figura 4.28: Precipitação total mensal – Prec7 (<i>Worldclim</i>).....	180
Figura 4.29: Precipitação total mensal – Prec8 (<i>Worldclim</i>).....	181
Figura 4.30: Temperatura média mensal de Agosto – Tmean8 (<i>Worldclim</i>).....	182
Figura 4.31: Temperatura máxima mensal de Agosto – Tmax8 (<i>Worldclim</i>).....	183
Figura 4.32: Precipitação total mensal – Prec10 (<i>Worldclim</i>).....	185
Figura 4.33: Precipitação total mensal – Prec11 (<i>Worldclim</i>).....	186
Figura 4.34: Temperatura média do trimestre mais úmido – bio8 (<i>Worldclim</i>).....	188
Figura 4.35: Precipitação do trimestre mais chuvoso – Bio16 (<i>Worldclim</i>).....	189
Figura 4.36: Precipitação do trimestre mais quente – Bio18 (<i>Worldclim</i>).....	190
Figura 4.37: Precipitação do trimestre mais chuvoso – Bio13 (<i>Worldclim</i>).....	191
Figura 4.38: Precipitação total mensal – Prec12 (<i>Worldclim</i>).....	192
Figura 4.39: Precipitação total mensal – Prec2 (<i>Worldclim</i>).....	193
Figura 4.40: Sazonalidade da temperatura – bio4 (<i>Worldclim</i>).....	194
Figura 4.41: Sazonalidade da precipitação – Bio15 (<i>Worldclim</i>).....	196
Figura 4.42: Precipitação anual – Bio12 (<i>Worldclim</i>).....	197

Figura 4.43: Isotermalidade – bio3 (<i>Worldclim</i>).....	198
Figura 4.44: Amplitude térmica anual – bio7 (<i>Worldclim</i>).....	199
Figura 4.45: Variação térmica diária – bio2 (<i>Worldclim</i>).....	200
Figura 4.46: Climas Zonais e suas Subunidades Regionais.....	202
Figura 4.47: Tipos de exploração mineral.....	207
Figura 4.48: Sistema ferroviário brasileiro.....	215
Figura 4.49: Tipos de solos.....	219
Figura 4.50: Tipos de relevo.....	222
Figura 4.51: Árvore de decisão (método J48) para o recorte 1999 a 2017.....	226
Figura 4.52: Árvore de decisão (método J48) com atributos não normalizados para o recorte 1999 a 2017.....	230
Figura 4.53: Modelo de favorabilidade de emergência de febre amarela (método J48) para o recorte 1999 a 2017.....	234
Figura 4.54: Árvore de decisão (método J48) para o recorte 1999 a 2003.....	235
Figura 4.55: Árvore de decisão (método J48) com atributos não normalizados para o recorte 1999 a 2003.....	238
Figura 4.56: Modelo de favorabilidade de emergência de febre amarela (método J48) para o recorte 1999 a 2003.....	239
Figura 4.57: Árvore de decisão (método J48) para o recorte 2007 a 2009.....	240
Figura 4.58: Árvore de decisão (método J48) com atributos não normalizados para o recorte 2007 a 2009.....	242
Figura 4.59: Modelo de favorabilidade de emergência de febre amarela (método J48) para o recorte 2007 a 2009.....	245
Figura 4.60: Árvore de decisão (método J48) para o recorte 2015 a 2017.....	246
Figura 4.61: Árvore de decisão (método J48) com atributos não normalizados para o recorte 2015 a 2018.....	248
Figura 4.62: Modelo de favorabilidade de emergência de febre amarela (método J48) para o recorte 2015 a 2017.....	251
Figura 4.63: Modelo de favorabilidade de emergência de febre amarela (método <i>Random Forest</i>) para o recorte 1999 a 2017.....	253

Figura 4.64: Modelo de favorabilidade de emergência de febre amarela (método <i>Random Forest</i>) para o recorte 1999 a 2003.....	255
Figura 4.65: Modelo de favorabilidade de emergência de febre amarela (método <i>Random Forest</i>) para o recorte 2007 a 2009.....	257
Figura 4.66: Modelo de favorabilidade de emergência de febre amarela (método <i>Random Forest</i>) para o recorte 2015 a 2017.....	259

LISTA DE TABELAS

Tabela 3.1: Camadas temáticas que compõem a base de dados.....	71
Tabela 3.2: Valor da mediana das menores distâncias calculadas entre municípios positivos e neutros no espaço de atributos.....	81
Tabela 3.3: Estatísticas calculadas das feições/variáveis no escopo dos municípios.....	86
Tabela 3.4: Quantitativo de atributos, dados de treino, teste e validação para cada recorte temporal.....	89
Tabela 3.5: Quantitativo de atributos, dados de treino, teste e validação para cada recorte temporal e seus respectivos sub-recortes temáticos.....	92
Tabela 3.6: Interpretação dos Valores da Estatística Kappa (Fonte: LANDIS e KOCH, 1977)	100
Tabela 3.7: Estrutura de uma Matriz de Confusão.....	102
Tabela 4.1: Métricas de validação dos modelos experimentais (J48 e <i>Random Forest</i>) do recorte temporal 1999 a 2017.....	107
Tabela 4.2: Métricas de validação da acurácia preditiva utilizando positivos de 2018 para o recorte temporal 1999 a 2017.....	109
Tabela 4.3: Métricas de validação dos modelos experimentais (J48 e <i>Random Forest</i>) do recorte temporal 1999 a 2003.....	112
Tabela 4.4: Métricas de validação da acurácia preditiva utilizando positivos de 2007 a 2009 para o recorte temporal 1999 a 2003.....	113
Tabela 4.5: Métricas de validação dos modelos experimentais (J48 e <i>Random Forest</i>) do recorte temporal 2007 a 2009.....	115
Tabela 4.6: Métricas de validação da acurácia preditiva utilizando positivos de 2015 a 2018	

para o recorte temporal 2007 a 2009.....	116
Tabela 4.7: Métricas de validação dos modelos experimentais (J48 e <i>Random Forest</i>) do recorte temporal 2015 a 2017.....	119
Tabela 4.8: Métricas de validação da acurácia preditiva utilizando positivos de 2018 para o recorte temporal 2015 a 2017.....	120
Tabela 4.9: Frequência de indicação dos dados ou variáveis temáticas nas árvores de decisão	123
Tabela 4.10: Métricas de acurácia com modelos 1.4 com dados normalizados e não normalizados.....	229
Tabela 4.11: Métricas de acurácia com modelos 2.0 com dados normalizados e não normalizados.....	237
Tabela 4.12: Métricas de acurácia com modelos 3.4 com dados normalizados e não normalizados.....	241
Tabela 4.13: Métricas de acurácia com modelos 4.5 com dados normalizados e não normalizados.....	247

RESUMO

A modelagem de favorabilidade de emergência de zoonoses proposta nesta tese pretende transcender os estudos epidemiológicos de transmissão de doenças infecto parasitárias. A intenção é identificar fatores/relações ambientais, ecológicos, climáticos e sociais que favorecem a ocorrência de determinada doença, formulando hipóteses não indicadas em estudos anteriores. Considerando a complexidade da dinâmica de ocorrência de infecções zoonóticas, especialmente da febre amarela silvestre, foram utilizadas 728 camadas de dados, correspondentes a diferentes temas, tais como: uso do solo; temperatura; precipitação; densidade e contagem da população; altimetria; entre outros. Entre os principais desafios, encontra-se a adoção como unidade espacial de agregação dos dados temáticos os limites municipais, visto que os dados de doenças e agravos de notificação do Brasil não possuem georreferenciamento. A solução proposta nesta tese, para a ausência do georreferenciamento, foi caracterizar estatisticamente os 5.571 municípios brasileiros com base nas diversas camadas de dados temáticos que representam as feições naturais e artificiais que compõem as paisagens desses municípios. Assim foi construído o perfil estatístico descritivo das características socioambientais dos municípios brasileiros, o que representa outra importante e original contribuição deste trabalho. Essa etapa produziu uma base de dados com aproximadamente 23 mil atributos (valores normalizados e não normalizados), contendo medidas de tendência central e de dispersão para cada tipo de classes ou variáveis. A partir do cômputo das estatísticas, foi organizada uma base de dados íntegra e estruturada, composta pelas estatísticas descritivas das características socioambientais dos municípios e dados de ocorrência de febre amarela silvestre. A partir dessa base de dados foram realizados 4 recortes temporais, em razão dos 3 surtos de manifestação da doença, e 7 sub-recortes de atributos para cada um desses recortes. Sobre todos os recortes e sub-recortes foram aplicados dois métodos de aprendizagem de máquina (J48 e *Random Forest*), ambos baseados em árvores de decisão, com o objetivo de construir modelos que permitam tanto a interpretação das relações existentes entre as camadas socioambientais e a ocorrência da febre amarela, como a obtenção de modelos que indiquem com maior acurácia a favorabilidade de novas regiões para emergência da doença. Os 40 modelos produzidos foram avaliados, quanto a sua acurácia, por diferentes métricas de validação (*Precision*, *Recall*, *F-Measure*, Curva ROC), e assim foram selecionados os 8 modelos conclusivos da tese que indicam a favorabilidade dos municípios para a ocorrência da febre amarela silvestre. A partir da metodologia construída foi possível ainda a verificação dos atributos evidenciados nas 28 árvores de decisão dos modelos produzidos pelo método J48. A partir da verificação da frequência de indicação foram descritos cada um dos atributos evidenciados (agrupados em 10 grupos temáticos) e discutidas suas prováveis relações com a emergência da febre amarela. Os modelos de previsão de favorabilidade de ocorrência da febre amarela silvestre produzidos por esta pesquisa poderão servir de instrumento para direcionar as políticas de vigilância epidemiológica e principalmente de priorização de áreas para vacinação.

Palavras chave: estatísticas descritivas; caracterização socioambiental; aprendizagem de máquina; árvores de decisão; interpretabilidade; acurácia preditiva; favorabilidade de ocorrência.

1. INTRODUÇÃO

1.1 Contextualização

Muitos agentes infecciosos, parasitos ou patógenos que acometem humanos tem a sua origem em animais silvestres, como por exemplo: gripes; malária; febre amarela; tuberculose; brucelose; triquinose; leishmanioses; toxoplasmose; leptospirose; febres hemorrágicas virais; febre de Lassa; hantavírus; raiva; peste bubônica; doença de Chagas; vírus de imunodeficiência adquirida, Oropouche, Marburg, Ebola, Nipah, Oeste do Nilo, Coronavírus, dentre outras. Diversas podem ser as formas de transmissão: a) por via oral, com a ingestão, por exemplo, de carne crua e sangue de bovinos silvestres, javalis, esquilos, roedores e macacos; b) a ingestão de formas embrionárias de agentes etiológicos dispostas na água, no ambiente ou alimentos; c) por intermédio de vetores; d) pela transmissão direta (WOODALL, 2009). Destaca-se que entre as doenças infecciosas que acometem os humanos, 71,8% são causadas por patógenos com origem na vida silvestre (JONES *et al*, 2008).

A Febre Amarela desperta especial interesse, pois trata-se de zoonose infecciosa com grande relevância epidemiológica, principalmente nas regiões tropicais do Planeta. Em linhas gerais, seu principal ciclo de transmissão ocorre entre mosquitos e primatas, sendo os humanos no Brasil infectados por circunstâncias acidentais (BRASIL/MS/SVS, 2017). A partir do histórico de ocorrência da doença no Brasil (desde os anos de 1600), verifica-se como a manifestação da doença é complexa e mutável, principalmente na região extra-Amazônica (que conta com maior número de indivíduos não vacinados) (COSTA, 2011). Os surtos recentes (1999 a 2003; 2007 a 2009; 2015 a 2018) evidenciaram heterogêneos graus de intensidade e propagação em sua distribuição espaço-temporal. Assim sendo, torna-se urgente e necessária a compreensão sobre suas especificidades, a investigação sobre as lacunas do conhecimento ainda existentes, bem como desenvolver pesquisas que indiquem antecipadamente novas regiões que oferecem condições socioambientais para emergência da doença.

A emergência de doenças oriundas de animais silvestres está fortemente associada a mudanças de comportamento e às alterações ambientais, incluindo mudanças climáticas, impactos naturais e antropogênicos (ESTRADA-PEÑA *et al*, 2014). As relações que

conectam a perda da biodiversidade e a saúde humana não são diretas, estas tratam de questões complexas, indiretas e dispersas no espaço e no tempo (OSTFELD *et al.*, 2008). Atualmente os impactos antrópicos são os principais causadores da perda da biodiversidade, pois promovem a fragmentação de matas e o isolamento dos ecossistemas naturais, resultando na perda de *habitats* e na simplificação/redução da diversidade ecológica dos ambientes. Em decorrência desses impactos, muitos microrganismos causadores de infecções rompem as barreiras biológicas de seus ciclos naturais e passam a acometer novas espécies e indivíduos, inclusive a espécie humana (LOVEJOY, 1993).

Estudos recentes mostraram o efeito de diluição da biodiversidade na modulação, dispersão e dinâmica de transmissão de patógenos, especialmente nas doenças transmitidas por artrópodes (STEPHENS *et al.*, 2016; POULIN e FORBERS, 2012; XAVIER *et al.*, 2012; KEESING *et al.*, 2006). Desta forma, a biodiversidade cumpre papel importante na proteção da saúde humana, pois a riqueza de espécies no ecossistema reduz a possibilidade de transmissão de parasitos de um hospedeiro para outros. Da mesma forma, a presença de predadores exerce controle sobre a população de pequenos animais, como por exemplo, roedores, potenciais transmissores de doenças (CHAME *et al.*, 2014). Segundo os mesmos autores (p. 1-2):

“Considerando a baixa capacidade de reverter às mudanças climáticas e os impactos ambientais, além da forma de produção e consumo de recursos naturais por parte da sociedade atual, parece clara a necessidade de monitoramento de detecção de surtos de doenças, despertando medidas de prevenção, previsão e controle das doenças transmitidas de animais silvestres para humanos.”

Considerando a grande diversidade de parasitos (bactérias, vírus, fungos, helmintos e protozoários), hospedeiros e vetores, bem como as suas respostas diante das diferentes influências antrópicas sobre seu ciclo biológico, pesquisas elaboram modelos que permitem compreender a ecologia da dinâmica de doenças parasitárias e a influência das transformações ambientais nesse processo. Segundo Chame e Labarthe (2013), os modelos da ecologia de transmissão de doenças parasitárias nem sempre são adequados às premissas dos modelos estatísticos convencionais, visto que esses devem ser capazes de agregar crescente número de parâmetros e analisá-los sempre de modo integrado.

Nesse sentido, compreende-se que existe a necessidade de inovação metodológica para a modelagem de emergência de zoonoses. Inovação metodológica essa que agregue conhecimento além da modelagem epidemiológica de transmissão de doenças, na qual já se conhecem os fatores de risco e as equações que descrevem o fenômeno de interesse (ROTHMAN *et al.*, 2011; MEDRONHO *et al.*, 2008; YANG, 2001; FORATTINI, 1999; GRENFELL e DOBSON, 1995), e que seja capaz de explorar as múltiplas relações e fatores da dinâmica de transmissão de doenças para a construção de modelos mais adequados à complexidade deste fenômeno (CHAME *et al.*, 2019). Segundo Barcellos *et al.* (2008) para que as bases de dados de saúde sejam utilizadas como meio de análise, estas devem ser georreferenciadas, integradas a dados ambientais e socioeconômicos, e submetidas a procedimentos de avaliação de sua distribuição espacial.

A complexidade da dinâmica de transmissão de doenças (COURA, 2013; REY, 2008) direciona as pesquisas epidemiológicas para a análise de grandes bancos de dados e cria demanda por novas formas de análise de dados complexos e desestruturados (CHIAVEGATTO FILHO, 2015). Ainda, segundo o mesmo autor, o crescimento da quantidade e complexidade dos dados tem também gerado alguns desafios em relação à escolha da metodologia. As técnicas estatísticas tradicionais de análise de dados apresentam algumas limitações para análise de grandes bancos de dados, principalmente em relação aos dados com muitas dimensões e/ou no caso da presença de correlações espúrias. Por isso, na área de análise de dados tem aumentado o interesse especificamente em um conjunto de técnicas denominado Modelagem Baseada em Dados e métodos de Aprendizagem de Máquina.

A Aprendizagem de Máquina é uma área da ciência de computação e uma subárea da inteligência artificial estabelecida em 1959 por Arthur Samuel, que tem sido utilizada para os mais diversos fins (LODOLI *et al.*, 2016). Uma propriedade importante de alguns métodos de aprendizagem de máquina, como árvores de decisão, algoritmos de extração de regras e algoritmos genéticos, é que o modelo é, ele próprio, a representação explícita do conhecimento extraído dos dados (ABU-MOSTAFA *et al.*, 2012; BISHOP, 2006; KOZA, 1992). Mais especificamente, é a identificação - passível de interpretação humana - das relações existentes entre os dados de entrada e saída.

Segundo Lemos *et al.* (2009), a Modelagem Baseada em Dados (MBD) vem sendo extensivamente utilizada na construção de modelos capazes de reproduzir comportamentos típicos de sistemas complexos, por meio de conjuntos de amostras dos mesmos. Nesse sentido, a modelagem, utilizando-se dos algoritmos de aprendizagem de máquina, se caracteriza como potencial arcabouço metodológico para investigação da emergência de zoonoses, no qual o fenômeno será descrito pelos dados e o modelo será construído a partir desses dados.

Além da necessidade de inovação metodológica para a modelagem de emergência de zoonoses, existe outra problemática notável nas pesquisas que utilizam dados de doenças e agravos. A série histórica dos dados de ocorrência de doenças e agravos de notificação compulsória, fornecida pelo Ministério da Saúde (MS), não possui georreferenciamento nos dados de ocorrência, os dados são agrupados por municípios (BARCELLOS *et al.*, 2008; SKABA *et al.*, 2004; BARCELLOS e RAMALHO, 2002; CÂMARA e MONTEIRO, 2001). A ausência de georreferenciamento desses dados é fator que limita as análises e modelagens espaciais pretendidas nos estudos ecológicos/epidemiológicos, porque a produção de modelos espacialmente generalistas e imprecisos (onde aproximadamente todo o País é indicado como área potencial para ocorrência de determinada doença) não é relevante para a distribuição de recursos e ações de vigilância epidemiológica e ambiental nas esferas estaduais e municipais.

Neste contexto, essa pesquisa buscou por alternativas para a utilização eficiente desses dados, de modo que os dados de doenças e agravos, sem georreferenciamento, não representem um impeditivo para a evolução das pesquisas com abordagem espacial. A solução encontrada foi adoção do limite municipal para extrair informações estatísticas sobre as feições naturais e artificiais que compõem as paisagens desses municípios. Assim, foi construído o perfil estatístico descritivo das características socioambientais dos municípios brasileiros, o que representa outra contribuição original deste trabalho.

O conhecimento das características ambientais, sociais e econômicas do local de ocorrência de determinado evento epidemiológico é essencial para compreendê-lo. A simples identificação de presença ou ausência das espécies de interesse (hospedeiros, vetores e agentes etiológicos) e da ocorrência de doenças em humanos não é suficiente para compreender o padrão de distribuição espacial de eventos epidemiológicos complexos e

dinâmicos. Para a compreensão dos padrões espaciais de dispersão ou limitação da febre amarela silvestre foi considerado um amplo e complexo espectro de elementos sociais e ambientais que compõem o espaço geográfico.

Em síntese, a modelagem desenvolvida nesta pesquisa é inovadora em sua abordagem metodológica baseada em dados e traz a descrição estatística das características socioambientais dos municípios brasileiros como alternativa original à ausência de georreferenciamento dos dados epidemiológicos brasileiros. Serve ainda como base para estudos ainda mais aprofundados, ou ainda, futura aplicação conjunta com outras metodologias e áreas de estudo. Por tudo isso, considera-se que a modelagem desenvolvida nesta pesquisa terá a capacidade e eficiência para investigação do complexo fenômeno de ocorrência de zoonoses emergentes e reemergentes, utilizando a febre amarela silvestre como estudo de caso.

1.2 Hipótese

Esta tese tem como hipótese que a modelagem computacional baseada em dados é capaz de tratar a diversidade e complexidade da dinâmica de distribuição e emergência de zoonoses, além de poder indicar caminhos para o entendimento do fenômeno, pois pressupõe que a descrição estatística das características socioambientais dos municípios brasileiros contribuirá para a construção de modelos preditivos de favorabilidade de emergência da febre amarela.

1.3 Objetivo

Investigar e propor metodologia baseada em métodos de aprendizagem de máquina para a modelagem de emergência de zoonoses, tomando-se como estudo de caso a febre amarela silvestre no Brasil. Para tanto, a modelagem proposta deve utilizar a base de dados produzida por esta pesquisa, contendo a descrição estatística das feições naturais e artificiais que compõem as paisagens dos 5.571 municípios brasileiros.

1.4 Justificativa

É importante destacar que esta tese propõe uma abordagem metodológica diferenciada da abordagem estatística utilizada nos estudos epidemiológicos de transmissão de doenças. A

modelagem proposta transcende os estudos epidemiológicos de transmissão de doenças, pois não trata somente de identificar as espécies de vetores e hospedeiras de parasitos e sua distribuição geográfica, ou inferir causalidade baseada em construções teóricas. O que se pretende é produzir cenários preditivos, bem como identificar a associação de fatores socioambientais que favorecem (ou podem favorecer) a ocorrência de determinada doença, formulando novos paradigmas e hipóteses, além dos estudos epidemiológicos de transmissão de doenças.

Esta pesquisa surge também com a proposta de preencher lacunas de conhecimento sobre os limites geográficos das zoonoses. A ausência do georreferenciamento dos dados brasileiros de doenças e agravos influencia diretamente na construção de modelos espacialmente distribuídos. E, portanto, foi formulada a estratégia da descrição estatística das características socioambientais dos municípios para minimização desse problema.

A diferença fundamental entre a modelagem proposta e a modelagem epidemiológica é que os estudos de transmissão de doenças aplicam modelos estatísticos aos dados para testar hipóteses causais sobre construções teóricas, ou seja, tem como objetivo verificar associações entre variáveis conceituais para a inferência de causalidade. Os estudos de transmissão de doenças utilizam-se ainda da modelagem matemática clássica, na qual já se conhecem as equações que descrevem (aproximadamente) o fenômeno de interesse, como por exemplo, as interações entre parasitos e hospedeiros.

Na abordagem estatística compartilha-se a premissa básica de utilizar um conjunto de observações para descobrir o processo subjacente. Neste caso, o processo é uma distribuição de probabilidade e as observações são amostras dessa distribuição. Na estatística é dada ênfase a situações em que a maioria das questões pode ser respondida com provas matemáticas rigorosas. Como resultado, as estatísticas se concentram em modelos muitas vezes idealizados e singulares. Assim, apesar de os modelos estatísticos serem eficientes em muitos estudos epidemiológicos de transmissão de doenças, eles podem não se adequar ao estudo de eventos complexos, os quais não se baseiam em relações/associações diretas entre variáveis; ou nos eventos nos quais são completamente desconhecidos os caminhos de associações entre os fatores observáveis.

A modelagem baseada em dados utiliza-se de pressupostos menos restritivos e trata-se de

modelos mais generalistas do que nas estatísticas tradicionais. Nesse cenário, as equações não são viáveis ou conhecidas, pois o fenômeno é tipicamente muito complexo. Nesta abordagem, o fenômeno é implicitamente descrito pelos dados, e o modelo computacional é construído a partir desses dados. Assim, tem-se como principal vantagem a representação mais apropriada possível sobre o fenômeno em estudo, ou pelo menos sobre o conjunto de dados representativos deste.

Deve-se destacar que a pesquisa proposta é relevante aos interesses do Exército Brasileiro (EB), pela grande circulação e exposição do seu efetivo de militares entre as milhares de espécies animais, vegetais e microorganismos que compõem as mais diversas regiões geográficas brasileiras e também pelo controle de circulação de doenças em zonas de fronteiras. A inserção dos militares em ambientes naturais, habitando áreas de riscos ou áreas degradadas, favorece a interação com agentes etiológicos, vetores e reservatórios, que por consequência, aumenta a exposição a doenças e também o risco de dispersão. Nesse sentido, estudos de monitoramento, modelagem e previsão de zoonoses fornecem subsídios ao EB para a garantia da soberania nacional e para o acompanhamento e promoção da saúde no território brasileiro, tanto em razão do seu efetivo de militares, quanto da população brasileira.

A partir da metodologia proposta deverão ser enfrentados desafios de múltiplas dimensões como: a coleta e estruturação de dados; a exploração desses dados; a extração de conhecimento sobre a dinâmica de emergência de zoonoses; a construção de modelos de previsão, entre outras. Tratam-se, portanto, de diferentes atividades, com distintos níveis de complexidade, envolvendo o conhecimento de diferentes áreas: Biologia; Ecologia; Epidemiologia; Geografia; Cartografia; Estatística e Computação. Dessa forma, caracteriza pesquisa interdisciplinar e que objetiva a modelagem de um fenômeno dinâmico e complexo de grande relevância para o Exército Brasileiro e para as políticas de saúde pública do Brasil de prevenção e controle de doenças que acometem (ou podem acometer) a sociedade brasileira.

1.5 Estrutura da tese

Esta tese compõe-se de cinco capítulos principais: Introdução; Estado da arte, conceitos e fundamentos teóricos; Metodologia para previsão da emergência da febre amarela

silvestre no Brasil por modelagem baseada em dados; Resultados e discussão; Conclusões. Além destes principais capítulos, também fazem parte deste documento a lista de Referências Bibliográficas e os Apêndices A (Dicionário de dados), B (Lista de distâncias no espaço de atributos entre positivos e neutros), C (Bases de dados) e D (Modelos experimentais).

No capítulo Introdução são inicialmente apresentados e contextualizados os principais temas de relevância para a pesquisa: emergência de doenças infecciosas; relevância epidemiológica da febre amarela silvestre no Brasil; complexidade da dinâmica de ocorrência de doenças; modelagem baseada em dados; aprendizagem de máquina; ausência de georreferenciamento de dados de doenças e agravos; além da breve indicação de estratégias metodológicas elaboradas e desenvolvidas para a construção da pesquisa. Além da contextualização dos principais temas, também foram descritos o Objetivo principal, a Hipótese e as Justificativas para o desenvolvimento da pesquisa.

No segundo capítulo, “Estado da arte, conceitos e fundamentos teóricos”, inicialmente é apresentada a visão global sobre quinze publicações científicas recentes relacionadas ao tema de modelagem de emergência de zoonoses, e ainda são pontuadas e discutidas as principais diferenciações sobre as abordagens metodológicas empregadas por esta Tese e por essas publicações. Em seguida estão formalizados a apresentação e a descrição de conceitos e fundamentos teóricos que embasam a pesquisa, sob o aspecto da modelagem computacional empregada (Modelagem preditiva, Modelagem baseada em dados, Aprendizagem de máquina, Métodos de classificação), de acordo com autores especialistas no tema. Seguindo a mesma lógica são abordados os principais aspectos sobre o estudo de caso da Tese, ou seja, a emergência da Febre Amarela Silvestre. Assim, são apresentados os aspectos gerais sobre a doença, um breve histórico da doença no Brasil, bem como a síntese sobre as ocorrências de febre amarela entre 1999 e 2018 fornecidos pelo sistema de vigilância do Ministério da Saúde do Brasil.

No terceiro capítulo, “Metodologia para previsão da emergência da febre amarela silvestre no Brasil por modelagem baseada em dados”, são descritas todas as etapas metodológicas e as fases de desenvolvimento da pesquisa. A partir da apresentação de fluxograma contendo as etapas metodológicas são descritas as bases de dados utilizadas, o pré-processamento e estruturação dos dados adquiridos, a rotina automatizada para cálculo

das estatísticas socioambientais dos municípios, a produção de 28 bases de dados contendo diferentes recortes temporais e temáticos em função dos dados, os métodos de classificação empregados na construção dos modelos (J48 e *Random Forest*), e as estatísticas de validação utilizadas para avaliação e seleção dos modelos finais. Também são descritas as plataformas computacionais utilizadas para o cumprimento dessas etapas, bem como as estratégias de resolução de problemas pontuais (fora do escopo/propósito dessa pesquisa), que são inerentes às particularidades ou especificidades oriundas das fontes dos dados adquiridas e/ou do formato de armazenamento do dado adotado (ausência de georreferenciamento dos dados de doenças, inexistência de municípios comprovadamente negativos e reamostragem de *rasters* de uso e cobertura).

O quarto capítulo, “Resultados e discussão”, apresenta todas as métricas de acurácia (interna e preditiva) de todos os quarenta modelos experimentais produzidos (oriundos das 28 bases de dados contendo diferentes recortes e da aplicação dos dois métodos de classificação). A partir da verificação dessas métricas são indicados oito modelos finais, os quais devem combinar os maiores valores de acurácia interna e preditiva. Entre os modelos experimentais, 28 desses, produzidos pelo método J48, possuem árvore de decisão interpretável. Dessa forma, com o intuito de verificar os atributos evidenciados por essas árvores de decisão, são listados os atributos em ordem de frequência de indicação, segundo o recorte temporal do surto e também o número total de indicações nos 28 modelos. A partir dessa listagem de frequência dos atributos, são discutidos cada um dos atributos evidenciados (agrupados nos subcapítulos de acordo ao grupo temático ao qual pertencem) e suas prováveis relações com emergência da febre amarela.

Ainda nesse capítulo são apresentados os modelos finais produzidos pelo método J48, contendo, para cada recorte temporal dos surtos, as suas árvores de decisão correspondentes e a geoespacialização da classificação preditiva dos municípios, bem como a geoespacialização dos modelos produzidos pelo método *Random Forest*, já que as árvores de decisão produzidas por este método não seriam viáveis de serem diretamente interpretadas.

No quinto e último capítulo, o de Conclusões, são pontuadas e destacadas as principais contribuições metodológicas adotadas por essa Tese, assim como a indicação de sugestões de idéias/soluções e recomendações para o desenvolvimento de trabalhos futuros.

2. ESTADO DA ARTE, CONCEITOS E FUNDAMENTOS TEÓRICOS

2.1. Modelagem de emergência de zoonoses: Estado do conhecimento

A abordagem metodológica desta pesquisa foi idealizada a partir da identificação de lacunas metodológicas e de resultados nas pesquisas em modelagem de emergência de zoonoses. Foram pesquisados artigos científicos recentes desta área, com enfoque epidemiológico e geográfico. Neste capítulo são discutidos alguns dos artigos encontrados e são indicados os tipos de dados utilizados, métodos e resultados obtidos.

Já foi indicado que pesquisas epidemiológicas de transmissão de doenças aplicam métodos estatísticos aos dados para testar hipóteses causais sobre construções teóricas, nas quais já se conhecem os fatores de risco e as equações que descrevem o fenômeno de interesse (ROTHMAN *et al.*, 2011; MEDRONHO *et al.*, 2008; YANG, 2001; FORATTINI, 1999; GRENFELL e DOBSON, 1995). Porém duas publicações recentes demonstram a utilização dos métodos de aprendizagem de máquina para a identificação de padrões e correlações em eventos de ocorrência de doenças infecciosas.

Allen *et al* (2017) utilizam modelos de regressão linear para analisar correlações entre eventos de ocorrência global de doenças infecciosas emergentes (*emerging infectious disease* - *EID*) com dados demográficos, ambientais e biológicos. Como resultado, apresentam que o risco de *EID* zoonótico é elevado em regiões tropicais florestadas, que experimentam mudanças no uso da terra e possuem grande diversidade de espécies de mamíferos. Destaca-se que segundo os mesmos autores:

“Até o momento, nenhum estudo quantificou o risco ou a incidência da febre amarela em zonas de risco global em alta resolução geoespacial, ou avaliou o potencial de receptividade à febre amarela em áreas fora das atuais zonas de risco”.

Hamrick *et al* (2017) buscaram identificar padrões espaciais e a relação entre fatores geográficos e ambientais com a distribuição de casos humanos de febre amarela nas Américas. Os casos foram agregados por município e foram utilizados oito fatores geoambientais para configurar os aspectos naturais dos municípios. Foi aplicado o método de regressão logística multivariável para identificar correlações e associações entre os dados. Como resultado, a presença da febre amarela foi significativamente associada à altitude,

precipitação, temperatura e diversidade de primatas não humanos.

Estes artigos possuem relações com esta tese, porém propõe-se aqui contribuições metodológicas distintas, com destaque para: a experimentação de dois diferentes métodos de aprendizagem de máquina, com diferentes níveis de complexidades de análise; a escala municipal de análise dos dados (com maior resolução espacial do que as pesquisas publicadas anteriormente a esta pesquisa), com a caracterização dos municípios baseada em ampla descrição estatística com aproximadamente 700 camadas socioambientais, além da abordagem temporal como importante mecanismo de controle e correspondência entre o período de ocorrência e o contexto das condições naturais e antrópicas da paisagem que coexistiram no tempo e espaço para emergência da febre amarela, permitindo assim a compreensão dos ciclos epidemiológicos e as particularidades dos surtos. Todas estas diferenças, portanto, caracterizam a contribuição original dessa tese.

Também é pertinente destacar que dentre os trabalhos restritos à saúde humana, alguns estudos similarmente estão evoluindo na utilização de métodos de aprendizagem de máquina. Apesar desse tipo de modelagem ser de distinta complexidade e utilizar apenas variáveis referentes ao indivíduo, é interessante mostrar como os métodos de aprendizagem de máquina vêm sendo incorporados nos estudos de saúde humana. Portanto, são apresentados quatro artigos científicos recentes de pesquisas em saúde humana que utilizam técnicas de aprendizagem de máquina.

Cerqueira *et al.* (2014) apresentaram um simulador (NICeSim) de código aberto e baseado em métodos de aprendizagem de máquina (redes neurais e máquina de vetor suporte), que auxilia profissionais de saúde a identificar a probabilidade de morte de recém-nascidos internados em unidades neonatal. O sistema auxilia a pesquisa médica na identificação e compreensão de correlações entre variáveis-chave com caso clínico em análise. Os resultados obtidos com o estudo retornaram acurácia de 86,7% e Curva ROC de 0.84. O sistema demonstrou ser uma ferramenta flexível (permite a inclusão e exclusão de variáveis sobre o paciente) e bastante conveniente aos médicos para obtenção do prognóstico.

Rose (2013) utilizou a combinação de diferentes algoritmos para prever risco de morte em 5 anos, entre 2.066 indivíduos com 54 anos ou mais, residentes em Snoma – Califórnia. Utilizou como variáveis sexo, idade, saúde “autoavaliada”, atividade física,

tabagismo, histórico cardiovascular e presença de doenças crônicas. Combinou a predição de 12 algoritmos (*superlearner*) para prever em um único modelo o risco de morte. A combinação dos algoritmos promoveu a redução do erro quadrático médio para 0,09, e produziu ganho de 44% em relação ao método que obteve pior desempenho, neste caso as redes neurais.

Kessler *et al* (2014) utilizaram 5 métodos de predição e combinaram esses métodos para prever a presença de Transtorno de Estresse Pós-Traumático (TEPT) entre 47.466 pacientes expostos a experiências traumáticas, que responderam um inquérito sobre saúde mental, conduzido pela Organização Mundial da Saúde (OMS) em 24 países. Utilizaram como variáveis sexo, idade, escolaridade, estado civil, 29 tipos de experiências traumáticas, presença anterior ao mesmo tipo de experiência traumática e de distúrbios mentais. Os resultados indicaram que entre os 95,6% dos casos de TEPT, 10% das observações foram classificadas pelo algoritmo como as de maior risco. Os algoritmos utilizados ainda ordenaram as observações de acordo com o maior e menor risco.

Caruana *et al* (2015) apresentaram dois estudos de caso em que modelos aditivos generalizados com inclusão de interação (GA²Ms) foram aplicados para prever probabilidade de morte em pacientes com pneumonia e probabilidade de reinternação em 30 dias. Os GA²Ms apresentaram acurácia superior ao método de redes neurais, que possui maior complexidade. Os modelos produzidos permitiram a clara interpretação sobre a contribuição de cada variável.

Além dos trabalhos epidemiológicos, de carácter rigorosamente estatístico e dos trabalhos que também utilizam métodos de aprendizagem de máquina, muitos artigos e pesquisas recentes na área epidemiológica contam com forte abordagem ecológica para a modelagem de distribuição de zoonoses. Estas pesquisas direcionam-se para a utilização dos métodos e técnicas da “Modelagem de Distribuição Potencial (MDP) de espécies” ou “Modelagem de Nicho Ecológico” (MNE) para prever a distribuição de doenças zoonóticas (PETERSON *et al*, 2011). Essa metodologia é próxima da metodologia proposta nesta tese, porém, existem características da MDP que comprometem a modelagem de distribuição potencial de doenças. Nesse sentido, são apresentados nove artigos científicos de pesquisas em doenças zoonóticas que utilizam MDP. Ao término da apresentação dos artigos, são evidenciadas as restrições da

MDP e são indicados os principais diferenciais metodológicos desta tese para a modelagem de distribuição de zoonoses.

Oliveira (2013) estimou a potencialidade de distribuição da síndrome cardiopulmonar por hantavírus (SCPH) nos biomas cerrado e mata atlântica. Estimou os nichos ecológicos dos roedores *Necromys lasiurus* e *Oligoryzomys nigripes* a partir do algoritmo Maxent (IWASHITA, 2007), utilizando variáveis climáticas da base Worldclim (FICK e HIJMANS, 2017) e índice de vegetação baseada no sensor AVHRR (ROBEL e GRAUMANN, 2014), todas as bases com resolução espacial de 25 km². De modo independente e complementar ao modelo de nicho ecológico dos roedores, se fez análise multicritério dos indicadores epidemiológicos, socioeconômicos e demográficos associados à distribuição dos casos humanos no Brasil para estimar a vulnerabilidade de municípios quanto a ocorrência da SCPH. O autor concluiu que ambos os métodos empregados contribuíram, de forma complementar, para o entendimento epidemiológico da SCPH.

Escobar *et al* (2015) analisaram relatórios de raiva transmitidos por morcegos (*Tadarida brasiliensis*) no Chile, de 2002 a 2012, para estabelecer associações entre ocorrências de raiva e fatores ambientais para a produção de um modelo de nicho ecológico. Utilizam a plataforma Maxent, versão 3.3.3. Foram utilizadas 726 ocorrências da variante do vírus da raiva AgV4 nesta espécie de morcego, e relacionaram-se as ocorrências com dados do índice de vegetação de diferença normalizada (NDVI) de imagens MODIS (resolução 500x500 m) (TUCKER *et al*, 2005). Como resultado obtiveram modelos de previsão com indicação de áreas com circulação de raiva em morcegos para antecipação da ocorrência da raiva humana. Os autores destacaram que os modelos possuem restrições referentes a seleção reducionista de variáveis ambientais e ao viés de amostragem e, portanto, devem ser utilizados com cautela.

Peterson e Samy (2016) utilizaram técnicas de modelagem de nicho ecológico para prever a ocorrência de filovírus, filamento de vírus RNA do gênero que causa febres hemorrágicas severas em humanos e primatas, e que inclui os vírus Ebola e Marburg. O conjunto de dados ambientais desse estudo foi composto pelo Índice de Vegetação de Diferença Normalizada – NDVI – do satélite AVHRR, com resolução espacial de 1 km², correspondentes ao período de fevereiro de 1995 a janeiro de 1996, com dados de uso e cobertura da terra do mesmo período. Foram testados algoritmos disponíveis nas plataformas

openModeller (MUÑOZ *et al*, 2011) e Maxent, utilizando técnicas de *bootstrapping* para avaliação das incertezas e *jackknife* (EFRON, 1979) para identificação de variáveis representativas. O modelo produzido indicou áreas emergentes nas quais não existem registros de ocorrência do vírus, como o sul da Etiópia e sudoeste da África Oriental. Os resultados indicaram que os dados de ocorrência foram insuficientes para permitir o desenvolvimento rigoroso do modelo.

Samy *et al* (2016a) utilizaram a modelagem de nicho ecológico para prever a ocorrência do Zika vírus (ZIKV) em humanos. Para tanto, também utilizaram o algoritmo de entropia máxima implementada no software Maxent versão 3.3. Foram produzidos quatro modelos, com resolução espacial de 25 km², baseados nas ocorrências de Zika em humanos e em diferentes combinações de variáveis climáticas, socioeconômicas, cobertura terrestre e registros de ocorrência de mosquitos. Os modelos foram calibrados para todo o México, América do Sul e Central, e posteriormente estendidos a todo o globo terrestre. Os resultados publicados foram pioneiros para o conhecimento sobre a potencial dispersão do Zika vírus em diferentes regiões, como a Austrália, partes da Nova Zelândia, bacia amazônica e o norte da América do Sul, Florida, sul do Texas e Louisiana nos EUA. O risco na Europa e no norte da Índia apareceu impulsionado pela acessibilidade e a fatores socioeconômicos.

Escobar *et al* (2016) consideraram casos de *Chikungunya* (CHIKV) em diversos países, e integraram informações de viagens aéreas, distâncias geográficas e conectividade entre cidades e adequação climática para espécies de vetores (*Aedes albopictus* e *Ae. aegypti*) para compreender e antecipar a disseminação de CHIKV nas Américas. Para a modelagem de nicho ecológico dos vetores foi utilizado o método ENM, disponível no software NicheA (QIAO *et al*, 2016). Foi assumido que a transmissão seria limitada apenas por condições climáticas, com maior capacidade reprodutiva em cerca de 25 °C e 200 mm de precipitação. A modelagem de distribuição dos vetores foi baseada nos padrões climáticos das últimas décadas (1950-2000), representados pelo conjunto de dados climáticos WorldClim. Os dados climáticos utilizados correspondiam a resolução espacial de 4 km². Foram especificamente utilizados os dados de temperatura média anual, faixa de temperatura diurna média, isoterminia, temperatura sazonal, temperatura máxima do mês mais quente, temperatura mínima do mês mais frio, temperatura anual, temperaturas médias dos períodos mais quentes e mais frios, a

precipitação anual, a precipitação dos meses mais úmidos e mais secos, a estagnação da precipitação e a precipitação do trimestre mais úmido, a temperatura média dos trimestres mais úmidos e mais secos, a precipitação dos trimestres mais quentes e mais frios e a precipitação do trimestre mais seco. Foi aplicada análise de componentes principais (PCA) (HONGYU *et al*, 2016) sobre essas variáveis climáticas para reduzir o número e correlação entre elas. Os três primeiros componentes explicaram 84,9% da variância geral nas variáveis. Como resultado apresentaram modelos de distribuição global do vírus, mas destacaram precauções sobre a confiabilidade dos modelos. Destacaram que os países com monitoramento contínuo de casos contribuíram para modelos com maior assertividade, enquanto que os países que possuíam relatórios epidemiológicos interrompidos e atrasados produziram previsões precárias.

Samy *et al* (2016b) empregaram 348 registros de diferentes espécies de vetor envolvidas no ciclo de transmissão da leishmaniose cutânea: patógeno (*Leishmania major*); vetor (*Phlebotomus papatasi*); e potenciais mamíferos reservatórios (*Psammomys obesus*, *Meriones libycus*, *Meriones shawi*, *Gerbillus gerbillus*). Esses registros foram reunidos a partir da coleta dos próprios autores e do banco de dados PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>). As variáveis ambientais utilizadas foram NDVI produzido a partir de imagens do satélite NOAA-AHRR, com resolução espacial de 1 km², para os anos de 1982 a 2000; 35 camadas climáticas globais da base CliMond (KRITICOS *et al*, 2012); modelo digital de elevação obtido a partir de imagens da missão espacial SRTM da NASA, com resolução espacial de 1 km². Todas as variáveis foram reamostradas para a resolução espacial de 20 km². Para a modelagem da distribuição de cada uma das seis espécies de vetores foi utilizado o algoritmo Maxent. A distribuição do nicho da doença foi estimada com base na sobreposição espacial e ecológica dos nichos estimados das espécies de patógenos, vetores e hospedeiros. Foram realizados testes de validação do modelo final com registros de casos humanos de leishmaniose independentes dos dados de entrada da modelagem. Os modelos indicaram que os nichos das espécies *L. major* e *P. papatasi* referem-se a regiões de baixa elevação e temperaturas máximas de 25 a 37 °C. As outras espécies apresentaram respostas semelhantes às condições ambientais, no entanto, eles tendem a ser mais amplamente distribuídos. Ao fim, os autores discutem que as atividades agrícolas e as mudanças antropogênicas serão fatores-

chave para afetar as distribuições de vetores e reservatórios da leishmaniose cutânea em toda a Líbia. Os efeitos desses dois fatores deveriam ser refletidos entre algumas das variáveis ambientais incluídas no estudo, e a ausência dessas variáveis marca uma limitação ao estudo.

Outros artigos ainda mais recentes utilizaram métodos e bases de dados notavelmente semelhantes às pesquisas descritas anteriormente para a modelagem de distribuição de outras espécies transmissoras de doenças, como por exemplo: Altamiranda-Saavedra *et al* (2017), que modelaram o potencial de distribuição de espécies de mosquitos transmissores de malária na Colômbia; Samy *et al* (2017), que simularam a distribuição do vírus da febre do Vale do Rift na África e Península Arábica; e Peterson e Talukdar (2017), que produziram mapas preditivos do risco de transmissão da Doença da Floresta de Kyasanur no Sul da Índia.

Conforme destacado na apresentação sintética dos artigos, os trabalhos de Modelagem de Distribuição Potencial de espécies baseiam-se na relação entre os pontos de ocorrência de espécies e um conjunto de dados topográficos, climáticos e de precipitação estimada, em escala mundial, disponibilizada por plataformas de base de dados como WorldClim, Climond e USGS Earth Explorer. As camadas climáticas e topográficas possuem restrições de resolução espacial e temporal. Como por exemplo, a base Worldclim, na qual as menores unidades de mapeamento (*pixels*) dos dados correspondem a áreas de 1 km² para todas as suas camadas, e restrição temporal, pois suas bases climáticas e de temperatura compreendem valores estimados para o período de 1950 a 2000 referentes às Normais Climatológicas para o Brasil. Dependendo do recorte temporal do estudo, os dados climáticos pretéritos podem tornar os modelos obsoletos e incompatíveis com o clima e pluviometria correspondente ao período. Dependendo também da espécie de vetor ou hospedeiro que se está trabalhando, essa representação generalista da informação pode mascarar a capacidade real de circulação da espécie no espaço. Como exemplo, a espécie de mosquito *Aedes aegypti* (transmissor do vírus da dengue, ZIKV e CHIKV), possui raio de vôo variando de 40 a 240 m, em regiões de aglomerados urbanos, que são os locais preferenciais dessa espécie. Escobar *et al* (2016) estimaram o nicho dos vetores *Ae. aegypti* e *Ae. albopictus*, utilizando dados recentes de ocorrência da CHIKV em humanos. Na modelagem foram utilizadas variáveis climáticas da base Worldclim com resolução espacial de 1 km² e temporal do período de 1950 a 2000, resoluções estas incompatíveis com a capacidade de circulação do mosquito e com o período

do estudo.

Outra restrição identificada corresponde às plataformas computacionais nas quais a MDP está implementada (Open Modeler, Maxent), pois essas não estão preparadas para trabalhar com múltiplas resoluções espaciais. Isso torna a metodologia de MDP refém de uma mesma resolução espacial, geralmente dos dados climáticos de baixa resolução espacial, existindo, portanto, a necessidade de reamostragem de toda e qualquer nova variável a ser incorporada na modelagem para o mesmo tamanho de célula dos dados climáticos (500 m², 1 km², 20 km²), mesmo quando a pesquisa dispõe de variáveis que podem, por exemplo, possuir maior resolução espacial (30 m², 90 m²), a exemplo de variáveis de uso e cobertura do solo. Nesse aspecto, e conforme discutido em Peterson *et al* (2011), a Modelagem de Distribuição Potencial de espécies precisa ser aprimorada no que se refere ao desenvolvimento de implementações computacionais capazes de associar pontos de ocorrência de espécies com variáveis em diferentes resoluções espaciais.

Nestes trabalhos, além das restrições espaciais e temporais inerentes aos dados de entrada, que são repassadas aos modelos, percebe-se a utilização facultativa de variáveis antrópicas, as quais são influentes na paisagem e que, fundamentalmente, deveriam ser incorporadas na modelagem, pois estas variáveis são determinantes para a ocorrência ou não das espécies. Como exemplo, dados sobre desmatamento, degradação florestal, fragmentação de *habitats* naturais, avanço agropecuário, pressão antrópica/imobiliária, entre outras. As influências antrópicas representam papel determinante no padrão de distribuição de espécies, tornando-se essencial entender como as atividades humanas afetam a vida silvestre nestes locais. Mesmo em áreas protegidas, que ainda são consideradas importantes refúgios para muitas espécies, também sofrem com a ação humana.

Muitas publicações de Modelagem de Distribuição Potencial utilizam dados de ocorrência de doenças oriundos de coletas de campo, coleções biológicas e dos sistemas governamentais de vigilância em saúde. Nas coleções biológicas podem ser encontradas informações de importância ambiental e epidemiológica sobre bactérias, fungos, protozoários, helmintos, artrópodes, moluscos, amostras histopatológicas humanas e animais. Diversas pesquisas científicas se utilizam dessas bases de dados e, especialmente para as pesquisas ecológicas/epidemiológicas, o acesso aos registros de ocorrência de animais vertebrados

(hospedeiros de zoonoses), invertebrados (transmissores de zoonoses) e de agentes etiológicos (de relevância epidemiológica) são essenciais.

No Brasil, muitos dados de coleções biológicas mantidos pelas instituições de pesquisa apresentam imprecisões e/ou ausência de posicionamento geográfico da coleta da espécie ou organismo. A referência espacial desses dados são muitas vezes associadas ao nome de localidades, bairros, municípios ou estados. O processo de espacialização dessas informações é, portanto, associada às unidades espaciais de referência, condicionando a escala de análise para regiões ou para áreas de abrangência. Os sistemas governamentais de vigilância em saúde, mantidos pelo Ministério da Saúde (MS) do Brasil, também não possuem georreferenciamento nos dados de ocorrência, e os dados são agrupados por municípios. A ausência do georreferenciamento dos dados biológicos/epidemiológicos é um fator que limita as análises e modelagens espaciais pretendidas nos estudos ecológicos/epidemiológicos.

Nas pesquisas de MDP, a ausência do georreferenciamento dos dados de ocorrência representa um grave problema. Isso porque existe a necessidade de utilizar a ocorrência da espécie ou da doença vinculada a um ponto, com coordenadas geográficas. Quando os dados utilizados não possuem informações de latitude e longitude, as ocorrências são vinculadas às coordenadas do centróide do município, o que representa uma aposta grosseira sobre a localização da espécie. O centróide do município provavelmente não corresponde ao local de vida da espécie, e portanto não deveria ser utilizado como referência aos algoritmos para o aprendizado sobre as características ambientais preferenciais da espécie.

No contexto da ausência de georreferenciamento dos dados de ocorrência das espécies e das doenças, da supressão da resolução espacial das camadas e da utilização facultativa de camadas referentes à influência antrópica, esta pesquisa buscou por alternativas de tratamento dessas limitações para o aprimoramento e evolução das pesquisas de modelagem de distribuição de doenças. Assim esta tese traz contribuições sobre dois principais aspectos: 1) modelagem baseada em um conjunto diverso e expressivo de variáveis temáticas, em diferentes resoluções espaciais e temporais, que são representativas do vasto e complexo espaço geográfico brasileiro; 2) perfil estatístico descritivo das características socioambientais dos municípios brasileiros, para viabilizar a utilização eficiente dos dados de ocorrência das espécies, oriundas de coleções biológicas, e dos dados epidemiológicos do Ministério da

Saúde do Brasil, os quais não possuem georreferenciamento.

2.2 Modelagem explicativa x preditiva

O termo modelo refere-se a uma abstração da realidade, que se aproxima do comportamento que se observa em um sistema, mas que normalmente é mais simples do que o próprio sistema real. Diferentes modelos servem a diferentes propósitos, logo eles não são excludentes, mas sim complementares. Não há modelo perfeito que represente a realidade, nem existe um modelo único; entretanto, pode existir aquele que melhor descreve determinado fenômeno.

Segundo Shmueli (2010) os modelos podem ser caracterizados sob duas abordagens distintas: explicativa ou preditiva. A modelagem explicativa trata da aplicação de modelos estatísticos aos dados para testar hipóteses causais sobre construções teóricas. Já a modelagem preditiva é o processo de aplicação de um modelo estatístico ou algoritmo de aprendizagem de máquina com a finalidade de prever observações novas ou futuras.

Feelders (2002) descreveu esta distinção entre os modelos explicativos e preditivos no campo da economia. A disparidade entre os modelos surge porque o objetivo na modelagem explicativa é combinar a função estimada (f) e a função alvo (F) o mais próximo possível da inferência estatística para se aplicar hipóteses teóricas. Os dados X (amostras) e Y (classes) são utilizados para estimar f , que por sua vez é usado para testar as hipóteses causais.

As principais etapas do processo de modelagem possuem distinções fundamentais de acordo com a abordagem explicativa e preditiva (SHMUELI, 2010):

- desenho do estudo e coleta de dados: a modelagem preditiva é prospectiva, isto é, f é construída para prever novas observações; já a modelagem explicativa é retrospectiva, de modo que f é usado para testar um conjunto de hipóteses já existentes. Na modelagem explicativa, o objetivo é estimar f baseado na teoria, com precisão adequada e utilizá-lo para inferência, sendo o poder estatístico a principal consideração. Na modelagem preditiva, o próprio f é determinado a partir dos dados, exigindo assim uma amostra maior para alcançar um menor viés (possibilidade de erro sistemático) e variação (possibilidade de dispersão do valor alvo);

- manipulação de valores faltantes: a maioria dos conjuntos de dados reais possuem valores ausentes, o que torna necessário identificar o tipo e a extensão dos valores faltantes, e a partir disso, identificar a melhor maneira de solucionar essas ausências. Na modelagem preditiva, a solução depende fortemente se os valores faltantes estão nas amostras (X) de treinamento e / ou nas classes a serem previstas. Na modelagem explicativa, a imputação de dados é predominantemente baseada no conhecimento empírico e teórico;
- particionamento de dados: na modelagem explicativa a partição de dados é menos usual devido à redução do poder estatístico. Quando usado, geralmente é feito para o propósito retrospectivo de avaliar a robustez de f . Na modelagem preditiva a partição dos dados deve evitar um número pequeno de dados para teste/avaliação, pois isso implica em uma estimativa enviesada sobre a acurácia do modelo;
- análise exploratória de dados: a análise exploratória de dados é um passo inicial fundamental em modelagem explicativa e preditiva. Consiste em resumir os dados numericamente e graficamente, reduzindo sua dimensão e "preparando" os dados para modelagem efetiva. Embora o mesmo conjunto de técnicas exploratórias dos dados possa ser usado em ambos os casos, elas são usadas de forma diferente. Na modelagem explicativa, a exploração é canalizada nas relações causais teoricamente especificadas, enquanto na modelagem preditiva a análise é usada de forma menos restrita, com propósito de identificar e expressar relacionamentos que talvez sejam desconhecidos ou, menos formalmente formulados. Métodos como análise de componentes principais (PCA) ou outros métodos de compressão de dados são geralmente realizados na abordagem preditiva;
- escolha de variáveis: na modelagem explicativa a escolha da variável é baseada na estrutura causal teórica e na própria operacionalização. Na modelagem preditiva, o foco é na associação e não na causalidade, os critérios para escolher os preditores são a qualidade da associação entre os preditores e a resposta, a qualidade dos dados e a

disponibilidade dos preditores no momento da predição;

- escolha dos métodos: a modelagem explicativa requer modelos estatísticos interpretativos de f que se conectam ao modelo teórico subjacente F . Na modelagem preditiva, o objetivo principal é gerar previsões precisas de novas observações e f é muitas vezes desconhecida. Entre os métodos utilizados inclui-se principalmente os métodos de aprendizagem de máquina;
- validação, avaliação de modelo e seleção de modelo: na modelagem explicativa, a validação consiste em duas partes: a validação do modelo valida que f representa adequadamente F , e o ajuste do modelo valida que f se ajusta aos dados. Na modelagem preditiva a validação é focada na generalização, que é a capacidade de f para prever novos dados. As medidas do poder preditivo são geralmente métricas fora da amostra ou suas aproximações na amostra (viés e variância), que dependem do tipo de previsão requerida. O poder preditivo é avaliado usando métricas calculadas a partir de um conjunto de retenção ou usando validação cruzada. Na modelagem explicativa, o foco é minimizar o viés para obter a representação mais precisa da teoria subjacente; em contraste, a modelagem preditiva procura minimizar a combinação de viés e variância de estimativa, sacrificando ocasionalmente a precisão teórica para uma melhor precisão empírica.

O processo de modelagem possui fortes distinções em função da abordagem explicativa ou preditiva. As diferenças conceituais e práticas levam invariavelmente a diferenças entre um modelo explicativo e um preditivo, mesmo que eles possam usar o mesmo conjunto de dados de treinamento. Assim, a determinação a priori do principal objetivo do estudo como explicativo ou preditivo é essencial para a construção de modelos adequados.

Importante destacar que nesta tese aplicou-se a modelagem preditiva, dado que aspectos fundamentais dessa abordagem serão adotados como referência, como por exemplo: a identificação de f (função estimada) será a partir da associação entre X (amostras) e Y (classes); o modelo será construído para prever novas observações; e ocasionalmente, a precisão teórica poderá ser sacrificada para uma melhor precisão empírica.

2.3 Nomenclaturas distintas na literatura técnico-científica

No levantamento bibliográfico desta pesquisa foi possível identificar a grande diversidade de terminologias, conceitos e definições, e notável ausência de um consenso na área sobre qual conjunto de termos utilizar. As pesquisas de modelagem evoluíram de maneira independente em vários campos científicos e técnicos. Esses campos desenvolveram diferentes ênfases e jargões próprios, principalmente porque a utilização das técnicas/métodos/algoritmos iniciou historicamente em diferentes momentos e domínios. Como resultado, são encontrados diversos termos na literatura com muitas nomenclaturas criadas por diferentes autores: Inteligência Artificial, Ciência de dados, *KDD (Knowledge Discovery in Databases)*, *DDM (Data-driven modelling)*, Mineração de dados, Aprendizagem de máquina, *Big Data*. Neste subcapítulo são identificados os principais termos da área, seus significados, suas definições (segundo diferentes autores) e o contexto em que são utilizados. Por fim, serão definidas e justificadas as terminologias adotadas para este trabalho.

Inteligência Artificial (IA) trata-se da macro área que abrange muitos dos conceitos abordados nesta seção. Esta refere-se ao estudo abrangente de como a inteligência humana pode ser incorporada nos computadores (SOLOMATINE *et al.*, 2008). Busca entender e construir entidades inteligentes, para entender a inteligência humana, restringindo-se a métodos inspirados na biologia (RUSSEL e NORVIG, 2009). McCarthy (2007) define a área como a ciência e engenharia de máquinas inteligentes, especialmente programas inteligentes de computador (sistemas especialistas ou *expert systems*). Atualmente, a IA envolve basicamente os sistemas especialistas, diversos métodos de busca – como busca em profundidade e busca em largura —, alguns sistemas baseados em agentes e sistemas de raciocínio ou inferência baseados em lógica.

Já a Ciência de Dados trata-se da abordagem holística e interdisciplinar, que considera a integração e a combinação de dados provenientes de diferentes fontes (LOUKIDES, 2011). Novo paradigma da ciência que unifica teoria, experimentação e simulação, considerando o uso intensivo da computação na exploração de dados (HEY *et al.*, 2009). Finzer (2013) relata que o termo ciência de dados foi mencionado pela primeira vez em 2001, em um texto de autoria de William S. Cleveland intitulado “Ciência de dados: um plano estratégico para a expansão das áreas técnicas no campo da Estatística”, o qual tinha por objetivo aliar a

estatística à programação e à computação. Conaway (2013) acrescenta que a ciência de dados constitui-se como a congruência entre domínios computacionais, habilidades de matemática e estatística, e especialidade na área de aplicação de dados.

A sigla *KDD* corresponde ao amplo processo de descoberta do conhecimento em base de dados, em inglês *Knowledge Discovery in Databases* – KDD. Na primeira conferência internacional sobre KDD, realizada na cidade de Montreal no Canadá, em 1995, foi proposto que a terminologia se referisse a todo o processo de extração de conhecimentos a partir de dados: Aquisição de dados; Pré-processamento dos dados; Mineração de dados; Validação dos resultados (CASTRO e FERRARI, 2016).

Modelagem baseada em dados (MBD), em inglês *Data-driven modelling* - DDM, trata-se do processo de análise dos dados sobre um sistema. Em particular, que encontra conexões entre as variáveis de estado do sistema (entrada, variáveis internas e de saída) sem conhecimento explícito do comportamento físico do mesmo (SOLOMATINE *et al.*, 2008; SOLOMATINE e OSTFELD, 2008). Os principais objetivos da MBD são: extrair e reconhecer padrões nos dados; interpretar ou explicar as observações; testar a validade de hipóteses; procurar o espaço de hipóteses (KUTZ, 2013).

Similarmente, mineração de dados corresponde ao processo sistemático, interativo e iterativo, de preparação e extração de conhecimento a partir de grandes bases de dados (HAN e KAMBER, 2006). O termo mineração de dados foi elaborado como alusão ao processo/atividade econômica de extração de minerais valiosos, de modo que a base de dados (mina) usando algoritmos (ferramentas) adequados para obter conhecimento (minerais preciosos) (CASTRO e FERRARI, 2016). A mineração de dados é um campo prático que se concentra em encontrar padrões, correlações ou anomalias em grandes bancos de dados relacionais. Como exemplo, é possível analisar os registros médicos dos pacientes e tentar detectar uma relação de causa-efeito entre um determinado medicamento e efeitos em longo prazo. Pode-se também analisar os padrões de gastos de cartão de crédito e tentar detectar possíveis fraudes. Como os bancos de dados utilizados na mineração geralmente são grandes, os problemas computacionais são muitas vezes críticos na mineração de dados (ABU-MOSTAFA, 2012). Tem como foco grandes bases de dados e está associada a aplicativos em serviços bancários, financeiros e de recursos de clientes. A mineração de dados é vista como

etapa do amplo processo de KDD. Os métodos utilizados são principalmente de estatística e aprendizagem de máquina (SOLOMATINE *et al.*, 2008).

Já a aprendizagem de máquina trata-se de área de pesquisa, pertencente à macro área da Inteligência Artificial, que objetiva o desenvolvimento de técnicas, algoritmos ou programas computacionais capazes de automaticamente melhorar seu desempenho e aprender a resolver problemas por meio de experiências passadas, também chamadas de exemplos ou simplesmente dados de entrada. A aprendizagem de máquina tem como foco extrair informação a partir de dados de maneira automatizada. (CASTRO e FERRARI, 2016; ABU-MOSTAFA, 2012; ROGERS e GIROLAMI, 2011; ALPAYDIN, 2009; MARSLAND, 2009; MITCHELL, 1997).

Por sua vez, *Big Data* é a terminologia amplamente utilizada na área de vendas, *marketing* empresarial. Trata-se do crescente acúmulo de dados disponíveis no sistema global de redes de computadores interligados. Cinco aspectos (“Vs”) são fundamentais na definição do termo: volume (quantidade de dados); variedade (heterogeneidade e diversidade de formatos e tipos de dados estruturados e não-estruturados); velocidade (produção de dados em fluxo ininterrupto e em tempo real); viabilidade (identificação de relacionamentos entre variáveis e padrões latentes); valor (aplicação e tradução desses relacionamentos e padrões em situações reais e práticas) (BIEHN, 2013; LANEY, 2001). Ward e Barker (2013) concluem que existem três fatores críticos comuns carregados no conceito *Big Data*: o tamanho, ou seja, o grande volume dos conjuntos de dados; a complexidade, que corresponde à estrutura, comportamento e permutações dos conjuntos de dados; e, por fim, as tecnologias, ferramentas e técnicas que são utilizadas para processar um conjunto de dados bastante grande e complexo.

Conforme foi apresentado, a literatura está permeada por diferentes nomenclaturas e entendimentos, muitas vezes sutis, sobre os métodos, processos e etapas da área. Todo esse universo está em constante desenvolvimento e aplicação por diferentes pesquisadores, grupos de pesquisa, empresas, consultores, empregando os mais variados recursos teóricos, práticos, computacionais ou fontes de inspiração, desde a estatística até fenômenos apenas observados na natureza, causando confusão e dificuldade de entendimento.

Esta tese de doutorado concentrou sua abordagem nos termos Modelagem Baseada em

Dados e Aprendizagem de Máquina, por considerar que esses termos são mais coerentes e representativos sobre o processo de construção de modelos científicos. A escolha para esses termos deve-se principalmente ao fato de não exigirem um modelo teórico inicial pré-concebido, sendo suficiente apenas o conhecimento prévio do fenômeno. Dessa maneira, a construção do modelo será feita com base em conexões entre as variáveis de estado do sistema (variáveis de entrada, internas e de saída), tornando possível a extração do conhecimento sobre o fenômeno, por meio da interpretabilidade dos modelos, o que se faz indispensável para a compreensão da emergência da febre amarela, sobre a qual existem poucas convicções sobre os fatores impulsionadores do fenômeno.

2.4 Modelagem Baseada em Dados - MBD

Modelagem Baseada em Dados corresponde ao amplo processo de extração de conhecimento a partir de dados, no qual o modelo é construído baseando-se em informações extraídas de um conjunto de dados e o conhecimento de especialistas é utilizado como informação complementar (ISMAIL-ZADEH *et al.*, 2016; KUTZ, 2013; ABU-MOSTAFA *et al.*, 2012; SOLOMATINE *et al.*, 2008; SOLOMATINE e OSTFELD, 2008; BISHOP, 2006; BALDI e BRUNAK, 2001; BABUSKA, 1998; MITCHELL, 1997; ANGELOV e GUTHKE, 1997; JANG *et al.*, 1997; KASABOV, 1996; CHIU, 1994; YAGER e FILEV, 1994; JANG, 1993; WANG e MENDEL, 1992).

A MBD é utilizada em situações nas quais não existem soluções analíticas, porém os dados possibilitam construir soluções empíricas (ABU-MOSTAFA *et al.*, 2012). Existem dois principais objetivos na MBD: a construção de modelos com alta acurácia¹, reproduzindo resultados semelhantes ao sistema real e com o menor erro possível; o desenvolvimento de modelos com estruturas transparentes, de forma que seja possível interpretar os resultados obtidos e/ou inserir conhecimento a priori na sua estrutura (LEMOS *et al.*, 2009; ESPÍNDOLA, 2004).

A MBD é baseada na análise dos dados que caracterizam o sistema em estudo. Um

¹Acurácia incorpora tanto tendência (erros sistemáticos) quanto precisão (erros aleatórios). Para avaliar a acurácia devem existir valores de referência para a estimativa da tendência. (MONICO *et al.*, 2009).

modelo pode então ser definido com base em conexões entre as variáveis do sistema (entrada, interna e saída) com apenas um número limitado de suposições sobre o comportamento físico e real do sistema (SOLOMATINE e OSTFELD, 2008). A modelagem baseada em dados também é utilizada para construir modelos de modelos em vez de modelos de sistemas naturais. Tais modelos são frequentemente referidos como emulação ou metamodelo, que é um modelo de menor custo computacional, baseado em um número limitado de elementos resultantes de experimentos computacionais, que pode ser usado no lugar do modelo de simulação (KREMPSE, 2014; FONSECA, 2009; BARTON e MECKESHEIMER, 2006; KHU *et al.*, 2008; SOLOMATINE e TORRES, 1996).

O termo “modelagem” enfatiza o fato de que o processo de MBD está próximo dos objetivos tradicionais de modelagem e segue as etapas de modelagem tradicionalmente aceitas, o que não inclui apenas a análise ou extração de dados. O processo de construção segue os princípios gerais adotados na modelagem: estudar o problema; coletar dados; selecionar a estrutura do modelo; construir o modelo; testar o modelo; e (possivelmente) iterar. Nesse sentido, a MBD geralmente é composta por cinco etapas. Essas cinco etapas são correlacionadas e interdependentes de tal forma que a abordagem ideal para extrair informações relevantes em banco de dados consiste em considerar as inter-relações entre cada uma dessas etapas e sua influência no resultado final. As cinco etapas do processo de modelagem são descritas a seguir (CASTRO e FERRARI, 2016):

1. compreensão do fenômeno: estudo substancial do problema em análise;
2. organização da base de dados: composição de uma base de dados composta por valores quantitativos ou qualitativos referentes ao conjunto de informações referentes ao fenômeno estudado;
3. pré-processamento dos dados: inclui tarefas de preparação dos dados como limpeza, integração, seleção ou redução e normalização da base de dados, para análise eficiente e eficaz por parte dos algoritmos;
4. aprendizagem de máquina: utilização de técnicas (análise descritiva, detecção de anomalias, agrupamento, associação e predição) e de algoritmos capazes de extrair conhecimento de um conjunto de dados pré-processados;
5. avaliação ou validação do conhecimento: identificação do conhecimento

verdadeiramente útil e não trivial; analisar a capacidade de generalização do método.

Algumas considerações complementares são fundamentais para as pesquisas de modelagem baseada em dados, tais como: verificar que as características da base de dados influenciam todos os resultados; necessidade de análises preliminares dos dados (com a utilização de métodos estatísticos, técnicas de análises descritivas, Análise de Componentes Principais – PCA, entre outros); buscar pela parcimônia (a escolha entre os modelos deve considerar a complexidade do modelo resultante).

A MBD utiliza os métodos de aprendizagem de máquina para a construção de modelos que incorporam as dependências desconhecidas entre os dados de entradas e as saídas de um sistema (real) modelado. O modelo é utilizado para prever (ou deduzir efetivamente) as saídas do sistema baseada em seus valores de entrada (figura 2.1).

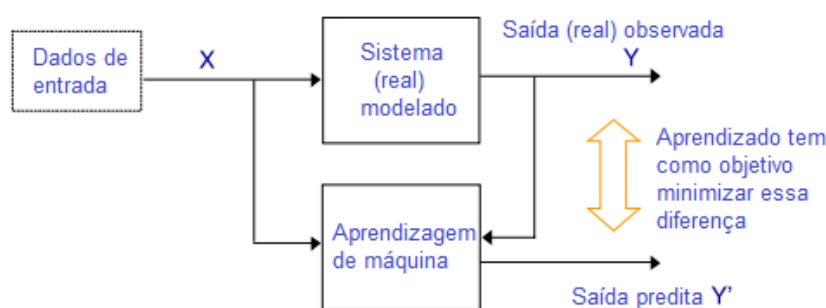


Figura 2.1: Aprendizagem por modelagem baseada em dados
(Adaptado de SOLOMATINE e OSTFELD, 2008)

Na modelagem baseada em dados, os parâmetros e a estrutura do modelo estão sujeitos às tarefas de otimização (KREMPSER e BARBOSA, 2009; FONSECA, 2009; RAO, 1996). Trata-se do processo de minimização ou maximização de uma dada função (função objetivo) para o entendimento de porções de seu domínio e obtenção de soluções ótimas localmente (KREMPSER, 2009; AUGUSTO, 2009). O conjunto de todas as possíveis soluções é chamado de espaço de busca. Um problema de otimização é caracterizado quando é procurada a melhor solução em um problema que admite inúmeras soluções.

Nesse contexto, se destaca o princípio da navalha de Occam (ABU-MOSTAFA *et al*,

2012; MITCHELL, 1997), no qual os modelos parcimoniosos são valorizados. Um exemplo de modelo parcimonioso são os modelos de regressão linear ou de rede neuronal com um pequeno número de nós ocultos. Esses modelos podem ser construídos pelo uso deliberado da chamada regularização, na qual a função objetivo — que representa o desempenho geral do modelo —, inclui não apenas o erro de validação do modelo, mas também deprecia a função objetivo de acordo com a complexidade do modelo (ex: aumento do número de nós ocultos em uma rede neuronal).

O processo de MBD é complexo e composto por um conjunto de etapas que empregam diferentes técnicas e métodos de classificação ou regressão. Para cada propósito devem ser testadas as técnicas que melhor contribuem para a construção de determinado modelo. Não existem regras fechadas sobre qual conjunto de técnicas utilizar, os ganhos e as perdas com determinada técnica ou método devem ser avaliados em cada conjuntura com base em testes e medidas de erro.

2.5 Aprendizagem de Máquina - AM

A AM é uma área da Inteligência Artificial cujo objetivo é o desenvolvimento de técnicas computacionais para o aprendizado, bem como a construção de sistemas capazes de adquirir conhecimento de forma automática, sem a necessidade de intervenção humana contínua (CHIAVEGATTO FILHO, 2015).

Tem como propósito principal encontrar a configuração lógica do modelo (figura 2.2), a qual corresponde, fundamentalmente, a estimar a função alvo, ou seja, ao treinamento dos algoritmos de aprendizagem - com um conjunto amostral de dados - para a produção e seleção de hipótese(s) (ABU-MOSTAFA, 2012).

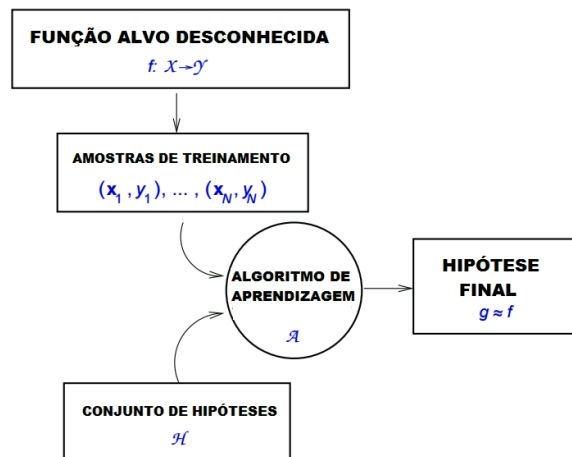


Figura 2.2: Configuração lógica da modelagem baseada em dados (Fonte: ABU-MOSTAFA et al, 2012)

As técnicas de aprendizagem de máquina têm se mostrado como opções eficientes para reconhecimento de padrões em conjuntos de dados (LODOLI *et al.*, 2016). O campo de reconhecimento de padrões está focado na descoberta automática de regularidades em dados por meio de algoritmos, valendo-se dessas regularidades para, por exemplo, realizar ações como classificar os dados em diferentes categorias (BISHOP, 2006).

O resultado da execução do algoritmo de AM pode ser expresso como uma função que recebe um novo registro x como entrada e que gera um vetor de saída y , codificado da mesma maneira que os vetores de destino. A forma precisa da função $f(x)$ é determinada durante a fase de treinamento, também conhecida como fase de aprendizagem, com base nos dados de treinamento (SOLOMATINE e OSTFELD, 2008).

Uma vez que o modelo é treinado, pode então determinar a identidade de novos registros, que são escolhidas para compor um conjunto de testes. A capacidade de categorizar corretamente novos exemplos que diferem daqueles usados para treinamento é conhecida como generalização (DOMINGOS, 2015; DIGNUM e POLI, 2007). Em aplicações práticas, a variabilidade dos vetores de entrada será tal que os dados de treinamento podem incluir apenas uma pequena fração de todos os vetores de entrada possíveis e, portanto, a generalização é um objetivo central no reconhecimento de padrões (ABU-MOSTAFA *et al.*, 2012; BISHOP, 2006).

Para a maioria das aplicações práticas, as variáveis de entrada originais são tipicamente pré-processadas para transformá-las em algum novo espaço de variáveis, no qual o problema de reconhecimento de padrões será mais fácil de resolver. O pré-processamento também pode ser realizado para acelerar o processo de computação. Por exemplo, se o objetivo é a detecção de rosto em tempo real em um fluxo de vídeo de alta resolução, o computador deve lidar com grandes números de *pixels* por segundo, e apresentá-los diretamente em um complexo algoritmo de reconhecimento de padrões, que pode ser computacionalmente inviável. Em vez disso, o objetivo é encontrar recursos úteis que são rápidos para computar e, no entanto, também preservam informações discriminatórias úteis que permitem, por exemplo, que rostos sejam distinguidos de não-faces. Esses recursos são então usados como entradas para o algoritmo de reconhecimento de padrões.

Existe uma divisão importante nas aplicações dos métodos de AM, podendo esse ser supervisionado ou não supervisionado. As aplicações em que os dados de treinamento compreendem exemplos dos vetores de entrada com seus vetores-alvo correspondentes são conhecidos como problemas de aprendizagem supervisionada. Exemplo clássico é de reconhecimento de dígitos escritos a mão, em que o objetivo é atribuir cada vetor de entrada a um número finito de categorias discretas. Se a saída desejada consistir em uma ou mais variáveis contínuas, a tarefa é chamada de regressão.

Em outros problemas de reconhecimento de padrões, os dados de treinamento consistem em um conjunto de vetores de entrada sem valores-alvo correspondentes. O objetivo em tais problemas de aprendizagem não supervisionados pode ser descobrir grupos de exemplos similares nos dados, que formam um agrupamento (*clustering*), ou para determinar a distribuição de dados dentro do espaço de entrada, conhecido como estimativa de densidade, ou para projetar os dados de um alto espaço dimensional até duas ou três dimensões para fins de visualização.

Existem ainda os métodos de aprendizagem por reforço (SUTTON e BARTO, 2012), que tratam de situações onde um agente aprende por tentativa e erro ao atuar sobre um ambiente dinâmico. Desta maneira, não é necessária uma entidade externa que forneça exemplos ou um modelo a respeito da tarefa a ser executada: a única fonte de aprendizado é a própria experiência do agente, cujo objetivo formal é adquirir uma política de ações que maximize

seu desempenho geral.

A AM corresponde a um conjunto de etapas/tarefas minuciosas e iterativas do processo de MBD, sua categorização varia entre autores e áreas de aplicação como exemplo destacam-se (CAMILO e SILVA, 2009; LAROSE, 2005):

1. “descrição: é a tarefa utilizada para descrever os padrões e tendências revelados pelos dados. A descrição geralmente oferece uma possível interpretação para os resultados obtidos. A tarefa de descrição é muito utilizada em conjunto com as técnicas de análise exploratória de dados para comprovar a influência de certas variáveis no resultado obtido”.
2. “classificação: uma das tarefas mais comuns, visa identificar a qual classe um determinado registro pertence. Nesta tarefa, o modelo analisa o conjunto de registros fornecidos, com cada registro já contendo a indicação à qual classe pertence, a fim de “aprender” como classificar um novo registro (treinamento supervisionado). Por exemplo, categoriza-se cada registro de um conjunto de dados contendo as informações sobre os colaboradores de uma empresa: Perfil Técnico, Perfil Negocial e Perfil Gerencial. O modelo analisa os registros e então é capaz de dizer em qual categoria um novo colaborador se encaixa”.
3. “estimação ou regressão: a estimação é similar à classificação, porém é usada quando o registro é identificado por um valor numérico e não um categórico. Assim, pode-se estimar o valor de uma determinada variável analisando-se os valores das demais, este refere-se ao processo denominado treinamento supervisionado, onde os rótulos de classes são conhecidos a priori; e o modelo pode ser utilizado para classificar ou estimar”;
4. “agrupamento: a tarefa de agrupamento (*clustering*) visa identificar e agrupar os registros similares entre si. Esta tarefa difere da classificação pois não necessita que os registros sejam previamente categorizados (treinamento não-supervisionado). Além disso, ela não tem a pretensão de classificar, estimar ou prever o valor de uma variável, ela identifica a similaridade dos registros e os agrupa”;
5. “associação: a tarefa de associação consiste em identificar quais atributos estão relacionados. Corresponde à descoberta de regras de associação que apresentam valores de atributos que ocorrem concomitantemente em uma base de dados”;
6. “detecção de anomalias: trata-se da identificação de objetos que não seguem o comportamento ou não possuem característica comum dos dados ou de um modelo que os represente”.

Essa pesquisa utiliza a abordagem supervisionada, pois são conhecidas as classes alvo. E cumpre as tarefas de classificação e estimação/predição, pois deseja-se indicar se as condições socioambientais de determinado município o aproxima das características dos municípios positivos ou negativos para febre amarela; e predição, pois deseja-se indicar antecipadamente quais municípios são favoráveis a manifestação da doença, para assim direcionar as políticas de vigilância em saúde e de priorização de campanhas de vacinação contra a febre amarela.

2.6 Métodos de classificação

Existem diferentes algoritmos/métodos de classificação baseado em técnicas de aprendizagem de máquina (Redes Neurais, *KNN – k-nearest neighbors*, *SVM – support vector machine*, Árvore de Decisão, Programação genética, entre outros). Dependendo da aplicação da modelagem, existem algoritmos mais ou menos apropriados para a abordagem pretendida. A escolha para descrição de tais métodos a seguir, reside no fato de terem sido avaliados e listados como algumas das melhores e mais populares técnicas de aprendizagem de máquina atualmente disponíveis (ALMEIDA e YAMAKAMI, 2012; ALMEIDA *et al.*, 2011; WU *et al.*, 2007). Restringindo-se este trabalho à aplicação somente de árvores de decisão pelo foco na interpretabilidade dos resultados, como será mais profundamente discutido nas seções posteriores.

- **Redes Neurais:** As redes neurais são modelos computacionais com inspiração biológica, que se baseiam na forma como funciona o cérebro humano (figura 2.3). Os modelos de rede neurais são desenvolvidos treinando a rede para representar as relações e os processos inerentes aos dados. São modelos de regressão essencialmente não lineares, que executam um mapeamento de entrada e saída usando um conjunto de nós de processamento simples, neurônios interligados (MITCHELL, 1997; HALL *et al.*, 2009; TAN *et al.*, 2006). Cada neurônio recebe entradas de forma externa ou de outros neurônios e passa através de uma função de ativação ou transferência, como uma curva logística ou sigmoide. Os dados entram na rede através das unidades de entrada dispostas naquilo que é chamado de camada de entrada. Esses dados são então alimentados para frente através de camadas sucessivas, incluindo as camadas ocultas intermediárias ou internas (figura 2.4). As entradas podem ser qualquer combinação de variáveis importantes para a previsão da saída; portanto, é importante ter o conhecimento mínimo do sistema que está sendo modelado (WANG *et al.*, 2010). A camada oculta é o componente essencial que permite que a rede neuronal aprenda as relações nos dados, como mostrado originalmente em Werbos (1994) e Rumelhart *et al.* (1986), que popularizaram também o algoritmo de *backpropagation* para treinar

uma rede neuronal *feedforward*. Esta configuração também é referida como um *perceptron* multicamada (MLP) e representa uma das redes neurais mais utilizadas. O algoritmo de *backpropagation* é uma variação do algoritmo de otimização de redução do gradiente de erro, o qual minimiza o erro entre os valores de saída previstos e reais. As conexões ponderadas entre os neurônios são ajustadas após cada ciclo de treinamento até atingir o nível de saturação do erro, ou seja, quando a taxa de erro volta a aumentar. Uma vez que as redes são treinadas, elas podem ser colocadas em operação quando os novos dados de entrada são fornecidos a uma rede previamente treinada para produzir as saídas do modelo desejado. Para contribuir na capacidade de generalização do modelo, é importante garantir que o conjunto de dados de treinamento contenha uma amostra representativa sobre o conjunto dos dados (SOLOMATINE *et al.*, 2008; HONG *et al.*, 2003).

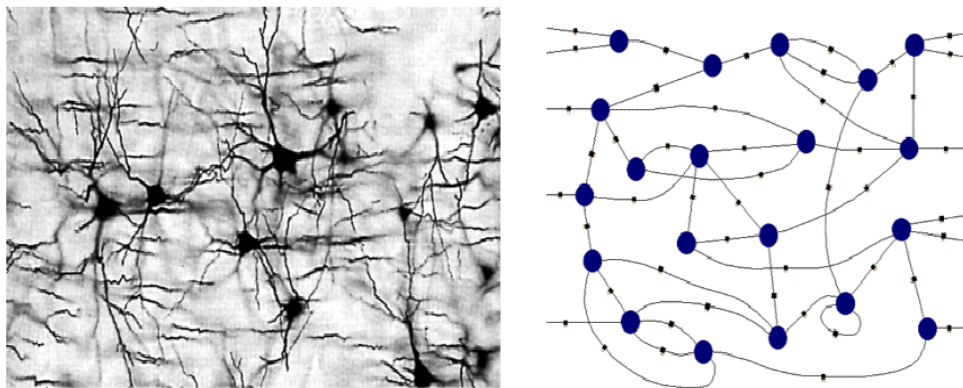


Figura 2.3: Função/estrutura de neurônios (Adaptado de ABU-MOSTAFA, 2019)

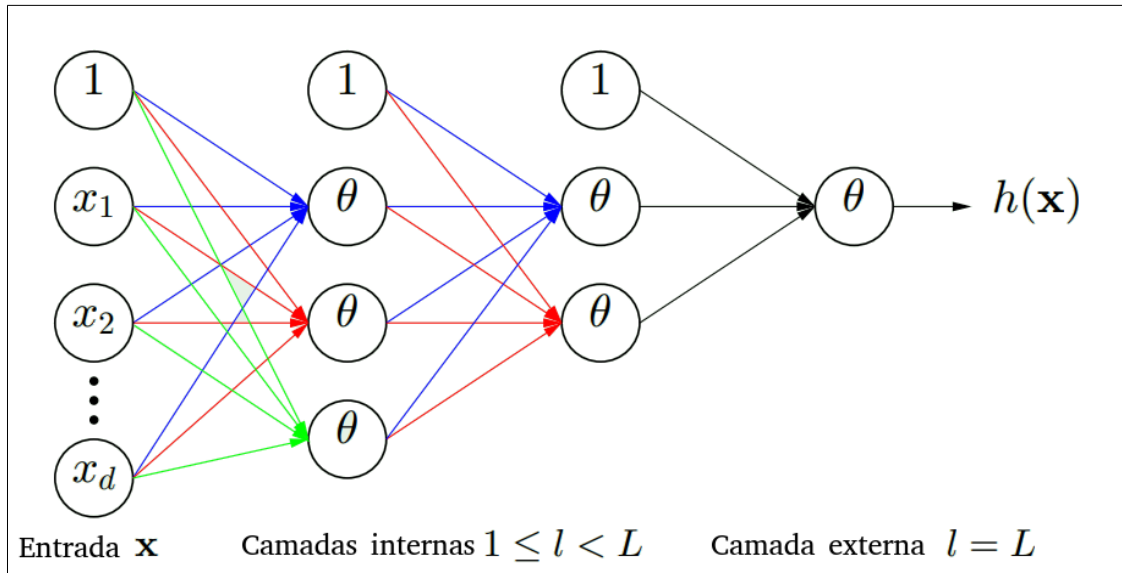


Figura 2.4: Estrutura rede neuronal (Adaptado de ABU-MOSTAFA, 2019)

- **KNN:** O método dos k -vizinhos mais próximos (*k-nearest neighbors*) é um dos classificadores não paramétricos, baseados em distâncias, mais simples e conhecidos da literatura (LI *et al.*, 2008; LIU e SUN, 2011). O KNN opera da seguinte maneira: dado um objeto X cuja classe se deseja inferir, encontra-se os k objetos X_i , $i = 1, \dots, k$ da base que estejam mais próximos a X , e depois se classifica o objeto X como pertencente à classe da maioria dos k vizinhos (figura 2.5). Quando todos os objetos da base são apresentados numericamente, a medida mais comum de ser empregada é a distância Euclidiana, mas essa escolha deve considerar o contexto da aplicação e as características da base. Apesar da simplicidade, o método KNN apresenta bons resultados em diversos cenários e, geralmente, obtém bom desempenho quando determinada classe possui diversos objetos e a superfície de decisão é irregular. Esse método é baseado em instâncias (*instance based learning*), pois ele determina a classe de um objeto com base nas classes de outros objetos (instâncias). Também é chamado de algoritmo preguiçoso (*lazy learning*), pois ele não é treinado a priori, uma vez que sua saída é calculada apenas quando se deseja saber a classe de algum novo objeto. O único parâmetro a ser definido no KNN é o valor do k , ou seja, o número de vizinhos mais próximos a serem considerados para se definir a classe de X . Embora não exista

uma regra para se definir k , valores grandes reduzem o efeito de ruídos na classificação, mas tornam fronteiras de classes menos definidas. Geralmente o valor de k é definido por heurística ou por tentativa e erro (ZHANG *et al.*, 2017; TANG *et al.*, 2011; SONG *et al.*, 2007; COVER e HART, 1967).

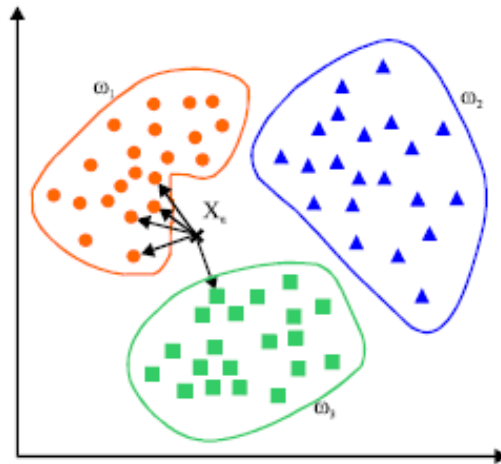


Figura 2.5: Classificação de um objeto pelo método KNN

- **Máquina de Vetor Suporte (*support vector machine* – SVM):** pertence a família dos classificadores lineares generalizados. Diferente de outros métodos que tentam otimizar o desempenho sobre o conjunto de treinamento, a SVM busca minimizar o risco estrutural no reconhecimento de padrões, isto é, a probabilidade de classificar de forma errada padrões ainda não vistos pela distribuição de probabilidade dos dados (NOBLE, 2006, TAN *et al.*, 2006). O objetivo de classificação por SVM é elaborar uma forma computacionalmente eficiente de aprender hiperplanos de separação, que otimizam os limites de generalização, em um espaço de características de alta dimensão. Existem diferentes limites de generalização, o que motiva o desenvolvimento de algoritmos diferentes: otimizar a margem máxima; ou a distribuição da margem; ou ainda o número de vetores suporte (CRISTIANINI e SCHÖLKOPF, 2002). Como diferencial, destaca-se que as SVMs simultaneamente minimizam erros de classificação empírica e maximizam a margem geométrica. Dois hiperplanos paralelos são construídos em

cada lado do hiperplano de separação para maximizar a separação dos objetos/dados. Os hiperplanos de separação maximizam a distância que separa os dados do hiperplano central. Os modelos gerados pela SVM só dependem de um subconjunto de dados de treino, que são aqueles que estão na margem formada pelos dois hiperplanos paralelos (figura 2.6). Os demais dados serão descartados da solução do problema. Esta característica torna esta técnica especialmente interessante para utilização em situações onde a confiabilidade dos dados de entrada é duvidosa ou incompleta (JOACHIMS, 2006; DECOSTE e SCHOELKOPF, 2002).

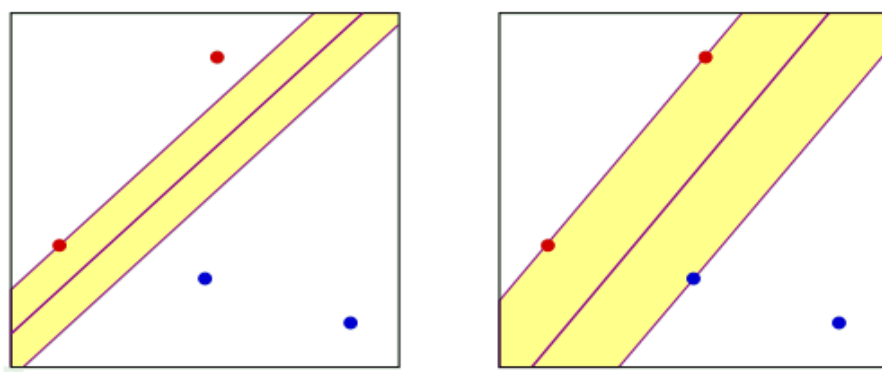


Figura 2.6: Hiperplanos de separação de objetos - método SVM
(Adaptado de ABU-MOSTAFA, 2019)

- **Árvores de Decisão:** o método funciona como um fluxograma em forma de árvore, onde cada nó (não folha) indica um teste a ser feito sobre os valores dos atributos (por exemplo, idade > 20). As ligações entre os nós representam os valores possíveis do teste do nó superior, e as folhas indicam a classe (categoria) a qual o registro pertence. Após a árvore de decisão ser montada, para classificar um novo registro, basta seguir o fluxo na árvore (mediante os testes nos nós não-folhas) começando no nó raiz até chegar a uma folha (figura 2.7). Pela estrutura que formam, as árvores de decisões podem ser convertidas em Regras de Classificação (CAMILO E SILVA, 2009). Importante destacar que as Árvores de decisão abrangem uma família de algoritmos/técnicas, ou implementações

computacionais, os quais possuem particularidades que definem sobre a organização estrutural da árvore (nós, folhas e ramificações), maiores detalhes sobre essas implementações serão abordados em seção subsequente.

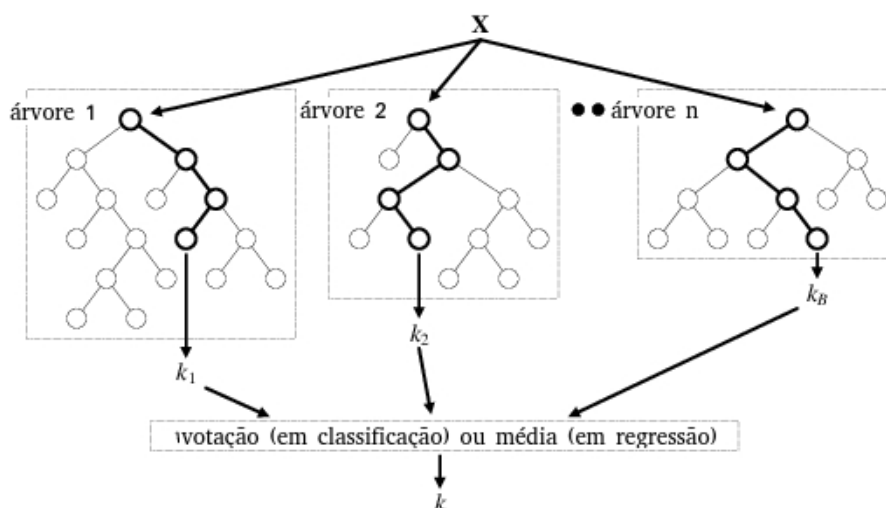


Figura 2.7: Classificação de amostras pelo método de Árvores de decisão
(Adaptado de: VERIKAS *et al*, 2016)

- **Programação Genética – PG:** são métodos de busca e otimização que trabalham baseando-se nos princípios de genética e seleção natural (GOLDBERG, 1989). São algoritmos probabilísticos com forte aplicação em otimização e são considerados robustos por serem aplicáveis a vários tipos de problemas nas mais diversas áreas (BORGES, 1999). Esses algoritmos combinam o conceito de sobrevivência por meio de testes de classificação com o de operadores genéticos para formar um mecanismo eficiente de busca. Os testes de classificação têm o objetivo de selecionar indivíduos promovendo a distinção entre melhores e piores, ou seja, fornecendo a aptidão do indivíduo de interagir com o meio (figura 2.8). Em geral, requerem pouco conhecimento do problema tratado, são de fácil implementação e apresentam bom desempenho em comparação com métodos convencionais. Entretanto, uma característica dos AG é a necessidade de realizar numerosas simulações até que uma

solução satisfatória seja encontrada, resultando em alta demanda de recursos computacionais (MALHOTRA *et al*, 2011; FONSECA, 2009; MAN *et al.*, 2001; RASHEED, 2000; STOCKWELL, 1999).

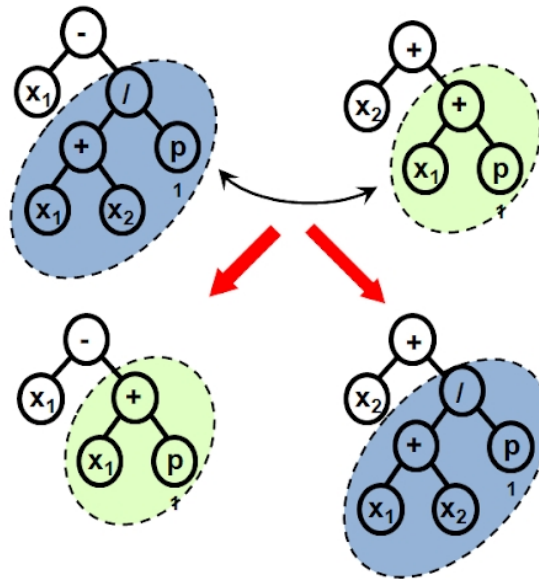


Figura 2.8: Geração de novas árvores de classificação a partir da aplicação de um operador de recombinação

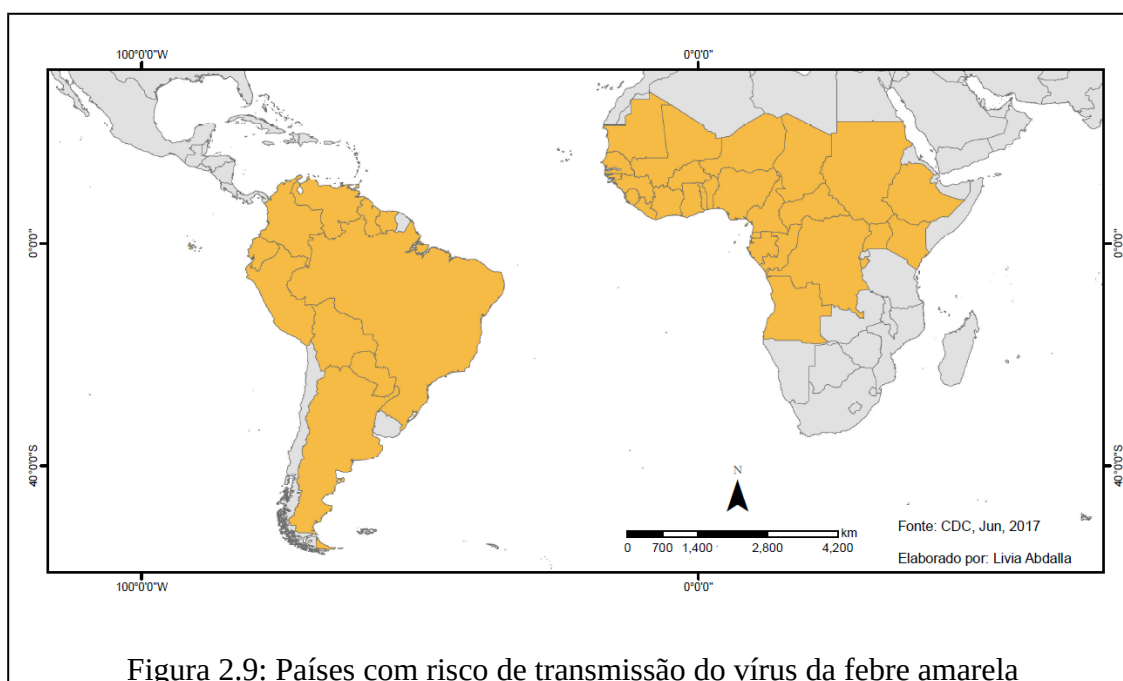
Os métodos Redes Neurais, KNN e SVM retornam modelos com acurácia preditiva, porém não permitem a clara interpretação sobre o fenômeno em questão. Já os métodos Programação Genética e Árvores de Decisão permitem a obtenção de modelos com acurácia e interpretabilidade (duas principais características de interesse da pesquisa).

Os métodos que possuem como característica a interpretabilidade destacam-se tanto na etapa de construção por um especialista quanto na etapa de utilização, para que o usuário possa compreender porque foi tomada determinada decisão, de modo que suas decisões possam ser explicadas pelas hipóteses e dados. Esse tipo de resposta permite abrir novas perspectivas de busca e estudo aos próprios especialistas, propondo muitas vezes associações desconhecidas. A acurácia preditiva também é fundamental para a construção de modelos que sirvam de instrumento estratégico para as políticas de saúde pública.

2.7 Febre amarela

2.7.1 Aspectos gerais

Febre amarela (FA) é uma doença infecciosa, de grande relevância epidemiológica, principalmente nas regiões tropicais do globo terrestre (figura 2.9). Historicamente, casos autóctones foram reportados somente entre os paralelos 12°N e 12°S, e a doença se mantém endêmica ou enzoótica nas florestas tropicais e áreas rurais da América Central, América do Sul e África (MONATH e VASCONCELOS, 2015), causando periodicamente surtos isolados ou epidemias de maior ou menor impacto na saúde pública. Em números, a febre amarela é responsável por 29.000 a 60.000 mortes anualmente na América do Sul e África (FARIA *et al.*, 2018).



As regiões tropicais reúnem condições ecológicas favoráveis à ocorrência da FA e de outras doenças conhecidas como “doenças tropicais” (Malária, Dengue, Doença de Chagas, Leishmanioses, Esquistossomose, Tuberculose, Hanseníase, entre outras) (CAMARGO, 2008). Dentre as características das regiões tropicais que facilitam a ocorrência da FA, destacam-se: a grande concentração e diversidade de espécies de primatas não humanos; condições climáticas e pluviométricas (clima quente e úmido) que favorecem a reprodução e

proliferação de mosquitos transmissores do vírus; e o desenvolvimento do parasito. Esses elementos combinados colaboram para a ampla transmissão do agente etiológico às espécies de hospedeiros.

A FA é exclusivamente propagada pela picada de insetos hematófagos infectados pelo vírus amarelo, arbovírus do gênero *Flavivirus* e família *Flaviviridae*. O ciclo epidemiológico da doença pode ocorrer sob duas modalidades: silvestre e urbano (figura 2.10). A diferença entre essas está relacionada com a espécie dos insetos transmissores, dos hospedeiros vertebrados e com local da ocorrência (BRASIL/MS/SVS, 2004; VASCONCELOS, 2002).

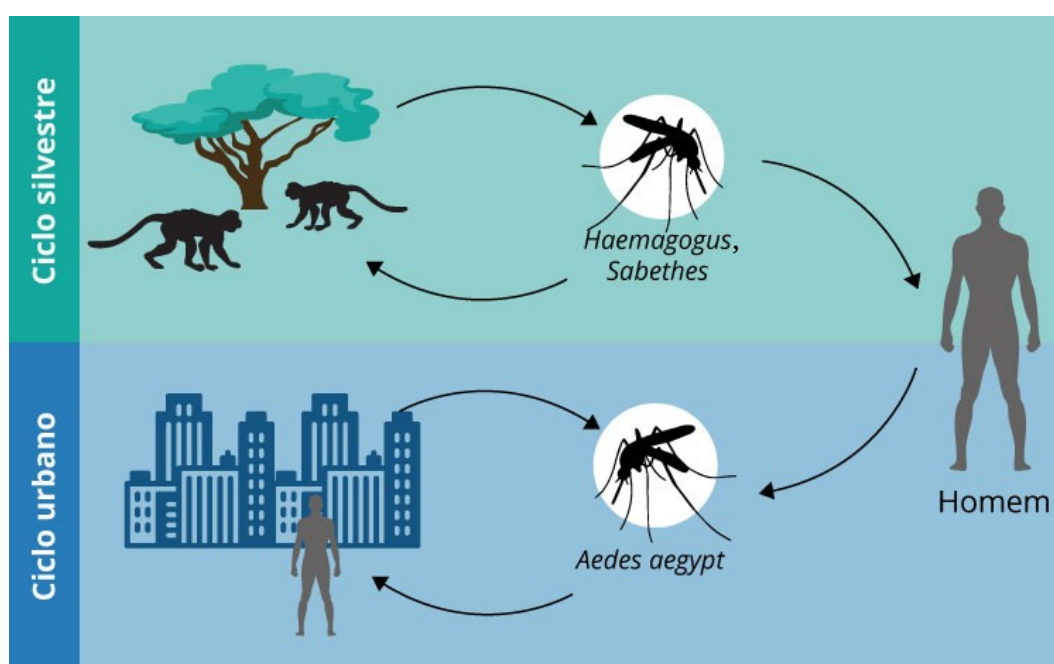


Figura 2.10: Ciclo de transmissão da febre amarela

Os humanos, após a picadura do mosquito, levam de 3 a 6 dias para apresentar os primeiros sintomas da doença, sendo importante destacar que 60% dos humanos infectados não apresentam sintomas. Os aspectos clínicos apresentam diferentes graus de gravidade. Nos primeiros dias, geralmente são detectados febre alta, dor de cabeça, mal-estar geral, náuseas, vômitos, dores musculares, prostração e, às vezes, calafrios. No terceiro ou quarto dia pode ocorrer ausência de febre e melhora clínica, podendo o quadro clínico evoluir para a cura ou para a forma mais grave da doença, que se caracteriza por aumento da febre, diarreia e o

reaparecimento do vômito. Surgem também icterícia, dor abdominal, manifestações hemorrágicas, insuficiência renal, que podem evoluir até o coma e a morte. As populações que residem nas áreas de circulação do vírus devem ser vacinadas para o controle eficiente da doença. Os viajantes com destino às áreas de risco devem receber a vacinação pelo menos 10 dias antes da viagem. De acordo com a alteração do Anexo 7 do Regulamento Sanitário Internacional (2005), uma única dose da vacina é suficiente para conferir imunidade e proteção ao longo da vida. Enquanto que as pessoas com contraindicações à vacina (crianças abaixo de 9 meses, mulheres grávidas ou amamentando, pessoas com hipersensibilidade grave aos antígenos dos ovos e imunodeficiência grave ou com mais de 60 anos) devem consultar profissional de saúde para melhor orientação (WHO, 2017).

No Brasil, o vírus no ciclo silvestre (febre amarela silvestre – FAS) é principalmente transmitido pelos mosquitos dos gêneros *Haemagogus* (*Hg. janthiomy*s, *Hg. albomaculatus*, entre outros) e *Sabethes* (*Sa. chloropterus*, *Sa. soperi*, etc) que infectam diferentes espécies de primatas não humanos (PNHs) e os humanos, sendo a espécie *Hg. janthinomys* o principal transmissor do vírus. No ciclo de transmissão urbano (febre amarela urbana – FAU), o vírus é o mesmo, o que difere é a participação do *Aedes aegypti* como único vetor transmissor, e os humanos como os únicos hospedeiros com importância epidemiológica (BRASIL/MS/SVS, 2017).

Na África, várias espécies de mosquitos do gênero *Aedes* (*Ae. africanus*, *Ae. simpsoni*, entre outros) têm sido associados à transmissão silvestre. E ainda tem sido registrado o isolamento viral a partir de carrapatos da espécie *Amblyoma variegatum* em áreas secas, o que pode indicar o papel secundário desses vetores no ciclo de transmissão da doença (VASCONCELOS, 2003).

2.7.2 Breve histórico da Febre amarela no Brasil

A primeira descrição de epidemia de febre amarela no Brasil ocorreu em 1685, em Recife, e permaneceu por pelo menos 10 anos, apresentando-se de maneira esporádica. O vírus teria sido trazido de barco, oriundo da cidade de São Tomé, na África, com escala em Santo Domingo, nas Antilhas, local onde a enfermidade dizimava a população. No ano seguinte (1686) emergiu em Salvador (BA), havendo relatos de sua presença até meados de

1692 (COSTA, 2005).

Em 1691 foi realizada a primeira campanha profilática, que utilizava ações direcionadas para a: desinfecção de navios; quarentena; cuidados especiais com os enterros e velórios; medidas sanitárias coletivas que incluíam aterramento de valas e limpeza de esgotos; segregação dos doentes; purificação do ar das casas; cemitérios; portos; limpeza de ruas e outras. Durante todo o século XVIII não houve relatos de infecção, o que sugere seu desaparecimento sob a forma epidêmica (TEIXEIRA, 2001).

Já em 1849 a epidemia ressurgiu em Salvador, atribuída à chegada de navio americano sem o cumprimento de rigorosas medidas sanitárias recomendadas pela Coroa portuguesa. A partir desse episódio, a febre amarela se disseminou para diversas outras cidades portuárias, atingindo em 1850 o Rio de Janeiro, configurando-se como grave problema de saúde pública no país. Foi organizada pelo então governo imperial a defesa sanitária do país (Lei nº 598, de 14 de setembro de 1850), na qual foram criadas Comissões de Higiene Pública de Provedores de Saúde Pública em todas as Províncias, que realizavam medidas semelhantes à campanha de 1691 (COSTA *et al*, 2011).

Já no século XX o médico cubano Juan Carlos Finlay trouxe como hipótese a participação do mosquito *Stegomyia fasciata* (denominado *Aedes aegypti* a partir de 1920) no processo de transmissão. A teoria foi testada com êxito em Cuba (TEIXEIRA, 2001). Com base nessa teoria, Oswaldo Cruz liderou medidas de controle da doença e direcionamento das ações de vigilância e controle sobre o mosquito transmissor. As medidas foram regulamentadas em 1903 pelo Serviço de Profilaxia de Febre Amarela. As campanhas, inspiradas na disciplina militar, contavam com os exércitos de “mata mosquitos” que entravam nas moradias à procura de focos do vetor (COSTA, 2011).

Após a “Era Oswaldiana” (1903 a 1913), a FA perdeu o seu caráter epidêmico nas áreas urbanas, deixando de ser prioridade da política de saúde, com escassez de recursos e enfraquecimento de ações nos períodos extra-epidêmicos. Com isso, os mosquitos voltaram a proliferar e novas epidemias desencadearam-se nos grandes centros urbanos, inclusive no Rio de Janeiro, entre 1928 e 1929. O modelo profilático de Oswaldo Cruz foi então reproduzido e as epidemias foram novamente controladas (COSTA, 2005).

Entre 1923 a 1939 a Fundação Rockefeller trouxe contribuições relevantes sobre o

conhecimento da febre amarela no país. Foram realizadas uma série de pesquisas de campo, como os estudos epidemiológicos realizados no Vale do Canaã, Estado do Espírito Santo, que identificaram a existência do ciclo silvestre da FA (BENCHIMOL, 2001).

A descoberta do ciclo silvestre deu início a intensa investigação por parte da Fundação, na busca de possíveis vetores e hospedeiros, e também na tentativa de conhecer a área geográfica de abrangência da doença no território brasileiro. Assim foi comprovado o papel dos primatas não humanos no ciclo epidemiológico da doença, além de identificadas várias espécies de mosquitos silvestres com infecção natural. Concluiu-se ainda que a febre amarela silvestre passou a ser endêmica/enzoótica e que a modalidade urbana demonstrava-se como forma exótica de manifestação (BENCHIMOL, 2001; TEIXEIRA, 2001).

A partir das pesquisas realizadas pela Fundação Rockefeller foi produzido extenso acervo, com ampla descrição epidemiológica, ambiental e social das viagens pelo interior do país. Nesse acervo estão incluídas 15 mil fotografias das expedições e das moradias dos indivíduos infectados, imagens aéreas, croquis, representações cartográficas, registros de atividades laboratoriais, entre outros. Todas essas informações foram utilizadas para caracterizar os cenários de manifestação da doença e estabelecer a provável rota de disseminação da onda epizootico-epidêmica de FA no Brasil, entre 1932 e 1942 (figura 2.11) (LACERDA, 2001; BENCHIMOL, 2001).

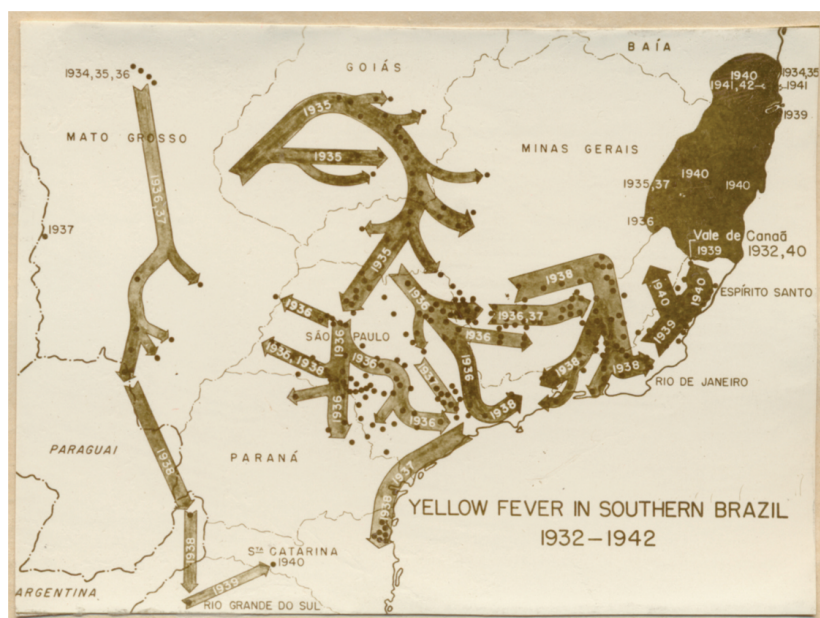


Figura 2.11: Provável rota da disseminação da febre amarela no Brasil (1932-1942) Fonte: COSTA, 2011

A partir do conhecimento científico produzido, a Fundação contribuiu com as políticas de intervenção técnica e organização das práticas sanitárias com o conceito “focos-chaves” para a eliminação das larvas de mosquito *Aedes aegypti* (LACERDA, 2001; BENCHIMOL, 2001). O trabalho da Fundação Rockefeller contribuiu ainda com pesquisas para avanços na formulação da vacina com vírus vivo de febre amarela, visto a identificação de suscetibilidade de camundongos e de macacos *Rhesus* ao vírus. Em 1937, a primeira vacina eficaz contra a febre amarela (cepa 17D) foi testada e o Instituto Oswaldo Cruz iniciou a fabricação e vacinação em larga escala (VASCONCELOS *et al*, 2001; ENGEL *et al*, 2006; MONATH *et al*, 2008). Assim, desde 1942, a transmissão urbana da doença é considerada controlada, com o último registro no estado do Acre (BRASIL/SVS/MS, 2004).

Em 1990 foi promulgada a Lei Orgânica de Saúde (nº 8080), que estabeleceu uma nova política de saúde para o Brasil. Essa lei instituiu o Sistema Único de Saúde – SUS, que descentralizou da esfera federal a formulação e execução dos programas de saúde, permitindo que estados e municípios assumissem responsabilidades sobre as ações de vigilância e controle (PORTO, 1994).

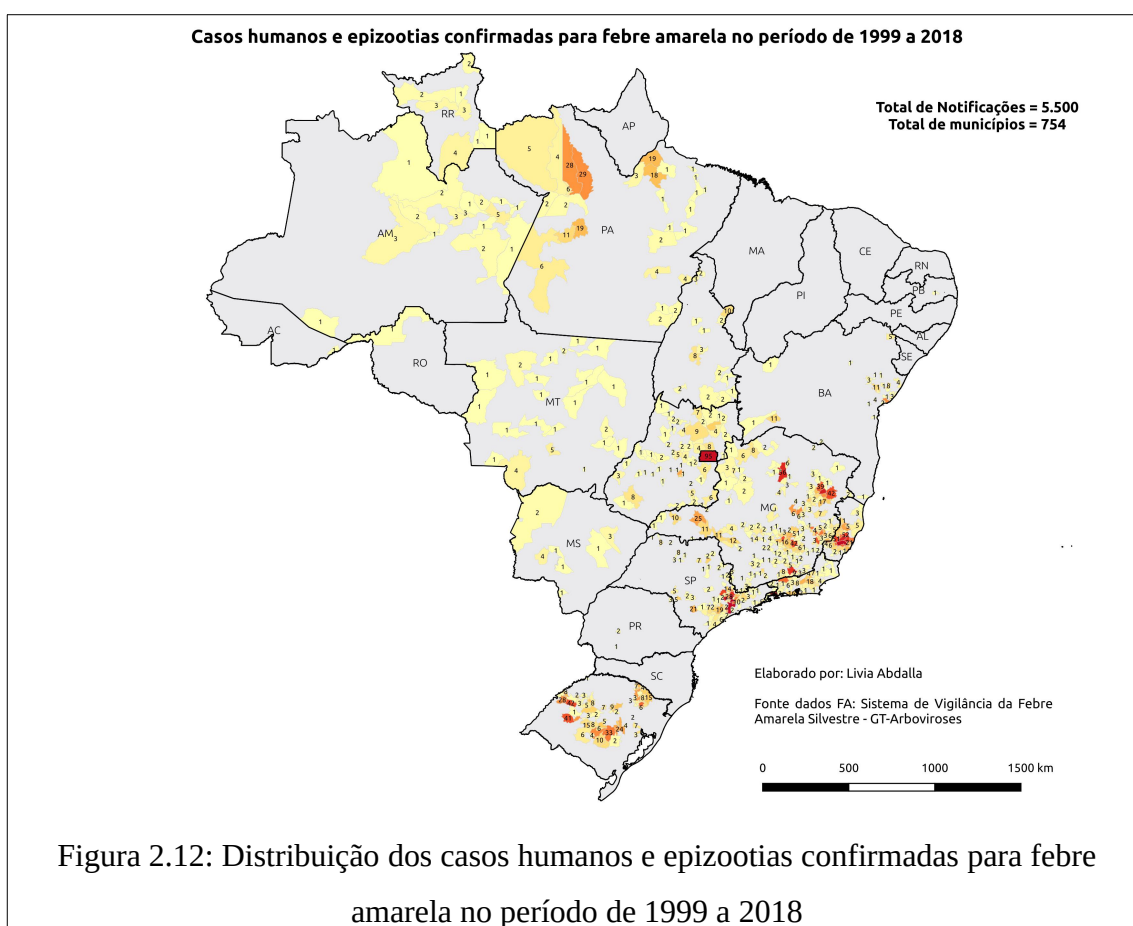
A partir da reorganização da vigilância da FA no país (e a extinção da Superintendência de Campanhas de Saúde Pública - SUCAM e do Departamento Nacional de Endemias Rurais), foram realizadas capacitações dos profissionais das secretarias estaduais de saúde de todas as Unidades Federativas, que atuaram como multiplicadores dos treinamentos para os profissionais da esfera municipal de saúde. O diagnóstico laboratorial também contou com avanços na padronização dos métodos e técnicas de sorologia (MAC-ELISA) e agilidade com a descentralização dos locais de realização dos exames para detecção e confirmação do diagnóstico para FA. Outro ponto importante no fortalecimento da vigilância, como estratégia de detecção precoce da circulação viral, foi a implantação da vigilância de epizootias em primatas, desde de 1999 (HAMMANN e LAGUARDIA, 2000), tornando a notificação compulsória a partir de 2006 .

2.7.3 Sistema de vigilância da Febre Amarela Silvestre (FAS) – 1999 a 2018

A partir de janeiro de 1999 houve uma mudança no perfil epidemiológico da FAS, quando a maioria dos casos passou a ser registrada fora das regiões Norte e Centro-Oeste, consideradas focos naturais da doença no Brasil. A partir do mesmo ano, com a observação de mortes de PNHs em vários municípios dos estados de Tocantins e Goiás e com a subsequente emergência da doença na população humana, tais eventos passaram a ser vistos como indicadores de risco (evento sentinela) e alerta para ocorrência de casos humanos de febre amarela silvestre. Assim, com o propósito de aumentar a sensibilidade do sistema de vigilância e a oportunidade de resposta dos serviços de Saúde Pública, foi implantada de forma gradativa, em todo o país, a notificação e investigação de epizootias em primatas, visando a detecção precoce da circulação do vírus (ARAÚJO *et al*, 2011).

A partir da observação da base de dados verifica-se que a doença se dispersou por diferentes regiões do país em diferentes eventos epidêmicos ao longo do referido período (1999 a 2018) e se estabeleceu na forma endêmica na Amazônia. Nesta região, em razão da alta cobertura vacinal, se observa casos isolados em indivíduos residentes ou visitantes não vacinados (ROMANO, 2012). Na região extra-amazônica, a febre amarela se apresenta em surtos precedidos de epizootias, em ciclos irregulares, e se dispersa entre as regiões Centro-Oeste, Sudeste e Sul do País.

No histórico completo (1999 a 2018) de monitoramento da febre amarela no Brasil, 754 municípios apresentaram 5.500 notificações confirmadas (figura 2.12) referentes a casos humanos e/ou epizootias, distribuídos em 18 Unidades Federativas: Acre; Rondônia; Amazonas; Roraima; Pará; Mato Grosso; Mato Grosso do Sul; Goiás; Distrito Federal; Tocantins; Bahia; Paraíba; Espírito Santo; Minas Gerais; Rio de Janeiro; São Paulo; Paraná e Rio Grande do Sul. Os casos de febre amarela também ocorreram de maneira heterogênea no tempo, existindo três surtos bem definidos e reconhecidos: 1999 a 2003; 2007 a 2009 e 2015 a 2018 (figura 2.13).



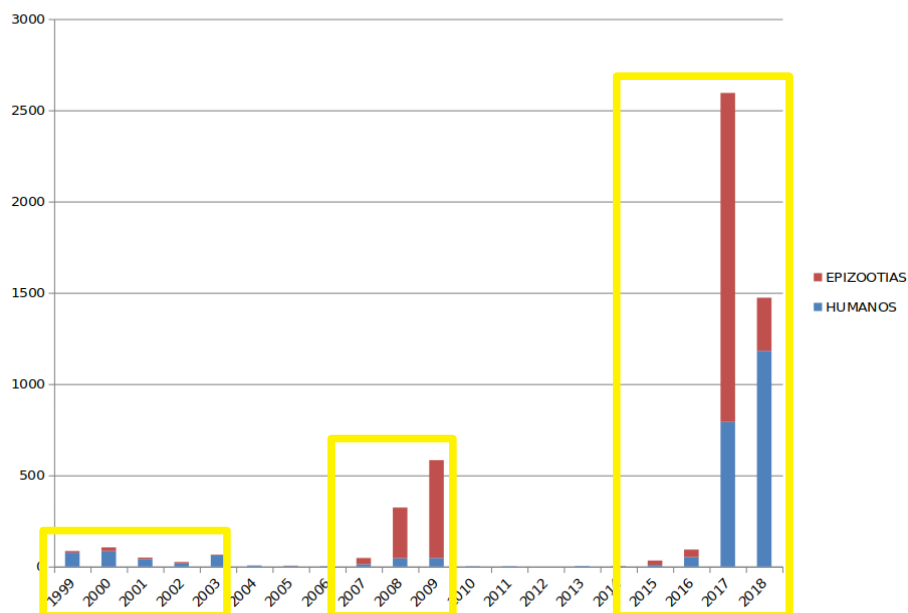
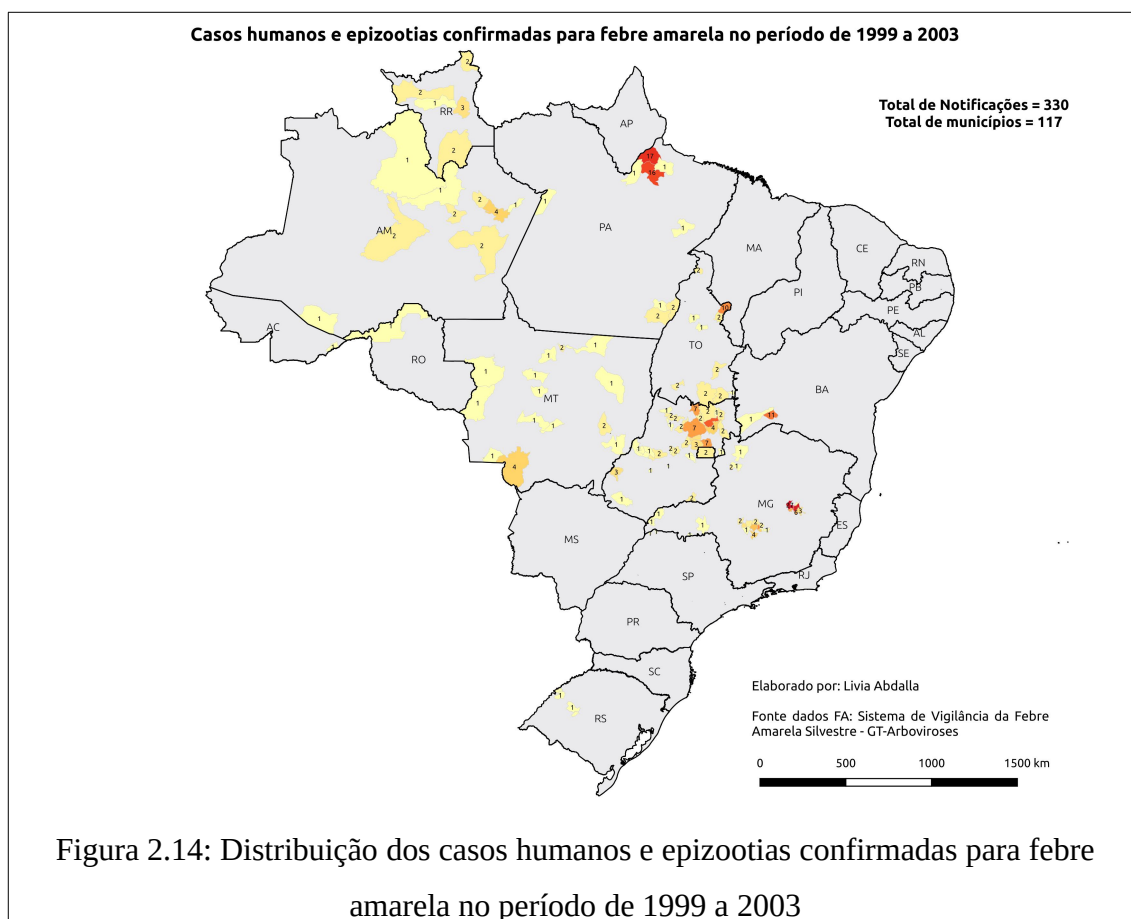


Figura 2.13: Casos humanos e epizootias confirmadas para febre amarela no período de 1999 a 2018 (Fonte: Sistema de vigilância da Febre Amarela Silvestre - GT-Arboviroses)

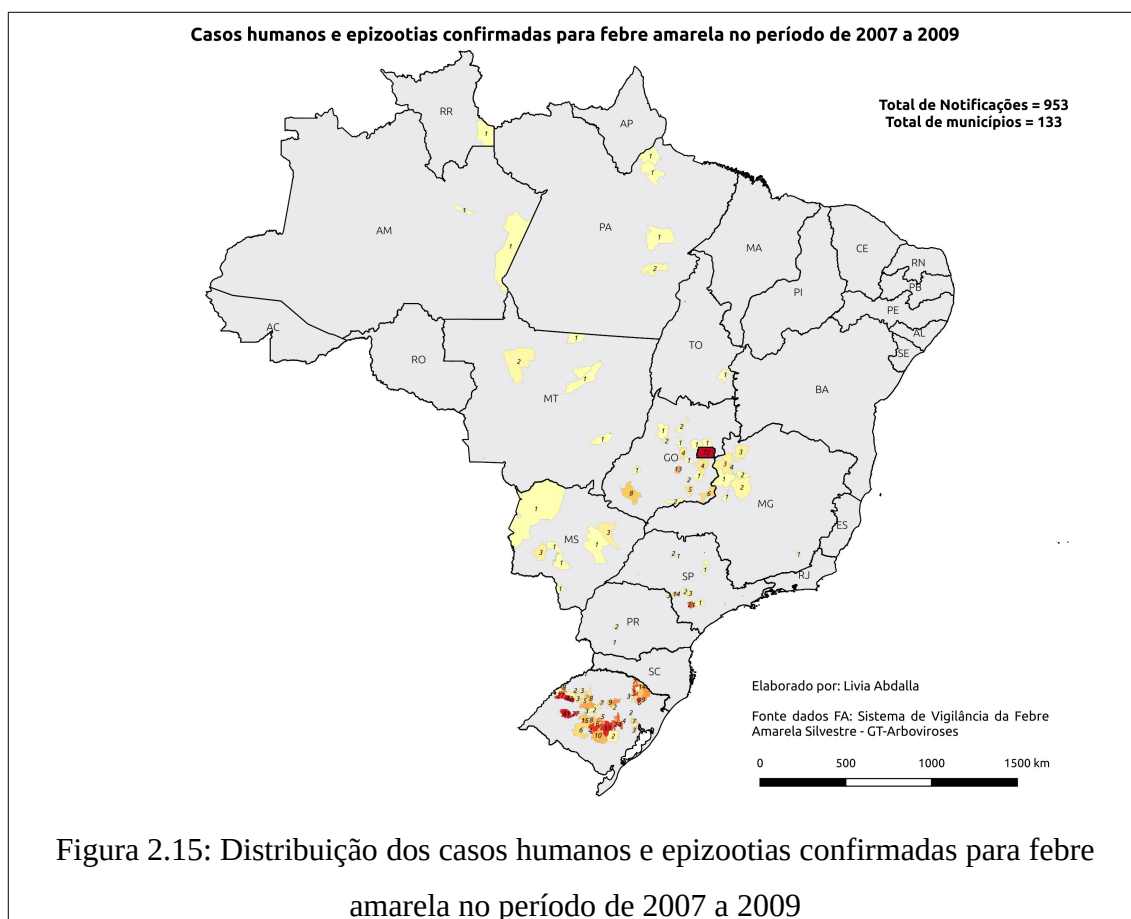
No primeiro surto (1999 a 2003) ocorreram 330 notificações confirmadas, distribuídas em 117 municípios localizados em 12 UFs (figura 2.14): Acre; Amazonas; Pará; Rondônia; Roraima; Mato Grosso do Sul; Mato Grosso; Goiás; Tocantins; Bahia; Minas Gerais e Rio Grande do Sul.

Segundo Costa (2005) durante essa epidemia, curiosamente observou-se uma explosão de múltiplos focos de atividade do vírus da febre amarela em vários pontos do país, em lugares que se encontravam silenciosos há mais de quatro décadas. O surto iniciou no Centro-Oeste, passando pelo sudoeste da Bahia, norte de Minas Gerais, noroeste de São Paulo, emergindo também, em 2001, em estados da Região Sul (Paraná e Rio Grande do Sul), repetindo o caminho percorrido pelo vírus na década de 1930 (BENCHIMOL, 2001).



No segundo surto (2007 a 2009) foram registradas 953 notificações em 133 municípios das seguintes Unidades Federativas (figura 2.15): Amazonas; Roraima; Pará; Mato Grosso do Sul; Mato Grosso; Goiás; Tocantins; Minas Gerais; São Paulo; Paraná e Rio Grande do Sul.

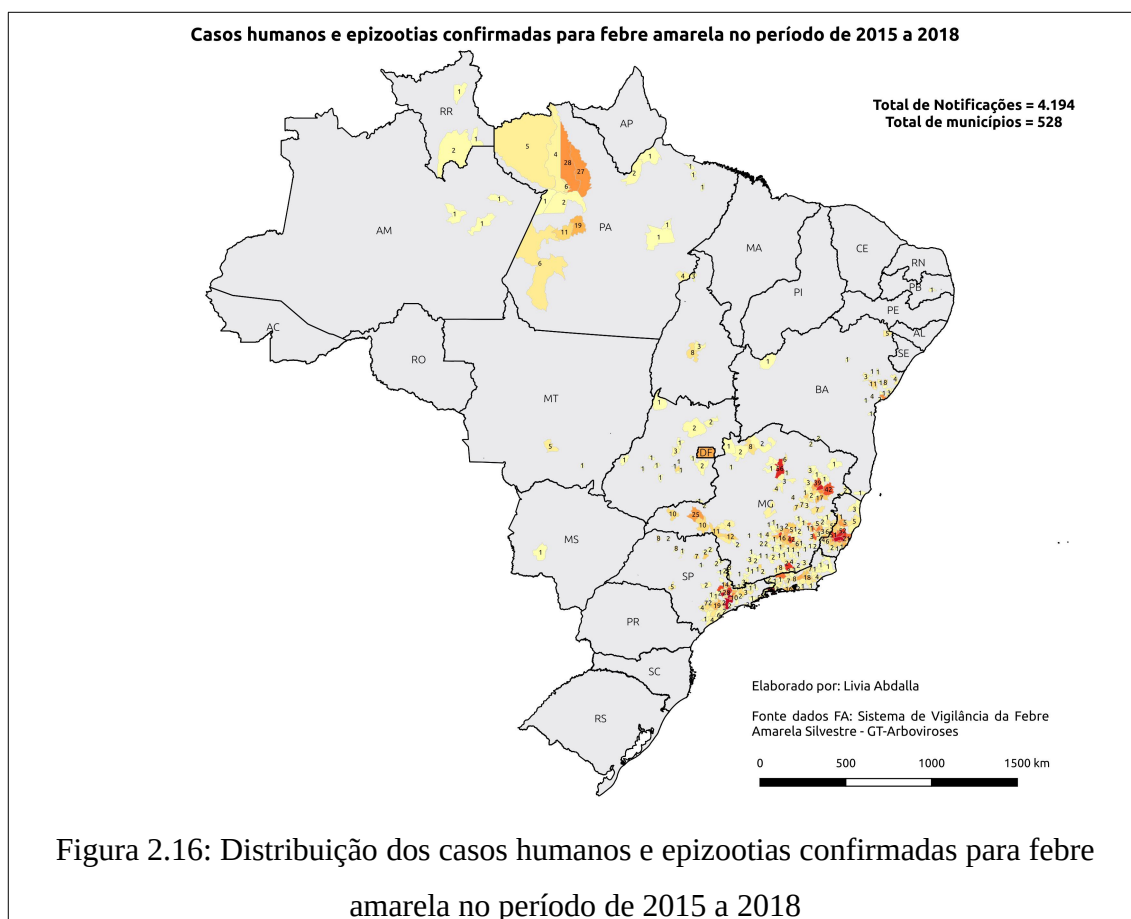
Com base nos períodos epidêmicos de manifestação da doença era eminente a preocupação sobre o risco de retomada da transmissão urbana, dada a proximidade dos locais prováveis de infecção com os grandes centros urbanos, infestados por *Aedes aegypti*, onde a maior parte da população não é vacinada. Esse fato suscitou a adoção de novas estratégias de vigilância, prevenção e controle (intensificação da vigilância em locais com transmissão, adotando os conceitos de área afetada e área ampliada) a partir de 2009, especialmente em relação a sazonalidade da doença (definidos três períodos epidemiológicos distintos: período de baixa ocorrência; pré sazonal e sazonal).



Segundo Romano (2012):

“As características da doença e a situação epidemiológica atual apontam para uma preocupação relevante, considerando os momentos de aumento do número de casos humanos e o risco de dispersão do vírus para novas áreas que, no Brasil, avança sentido leste e sul do país, justamente para regiões densamente povoadas e altamente infestadas por *Aedes aegypti*”

A preocupação se tornou real com o terceiro e mais recente surto (2015 a 2018), que já acumularam 4.194 notificações em 528 municípios, distribuídos em 13 UFs: Amazonas; Pará; Roraima; Mato Grosso do Sul; Mato Grosso; Tocantins; Goiás; Distrito Federal; Paraíba; Bahia; Minas Gerais; Espírito Santo; Rio de Janeiro e São Paulo (figura 2.16).



Este evento epidêmico teve como ponto de dispersão as regiões de Goiás e Tocantins, com registros a partir de julho de 2014, seguindo nos sentidos sul e sudeste do país, quando afetou as áreas de fragmentos florestais e matas de galerias do centro-oeste de Minas Gerais, com registros oficiais de epizootias a partir de outubro de 2016. Os registros foram progressivamente, e em curto período de tempo, acrescentados de novos casos atingindo em ordem cronológica: Minas Gerais, Espírito Santo, São Paulo e Rio de Janeiro. Esses estados tiveram a maior representatividade (99% dos casos confirmados – Informe COES – Febre Amarela nº 43/2017) no recente evento epidêmico, que alcançou o *status* de maior surto de febre amarela nas últimas décadas.

3. METODOLOGIA PARA PREVISÃO DA EMERGÊNCIA DA FEBRE AMARELA SILVESTRE NO BRASIL POR MODELAGEM BASEADA EM DADOS

Os métodos de modelagem baseada em dados apresentam, entre suas vantagens, a derivação de um modelo cujas suposições estejam concentradas nas propriedades intrínsecas dos dados e na existência de um princípio físico, ou ainda, na existência de correlação estatística que estabeleça uma relação entre as variáveis de entrada e as de saída. Portanto, a modelagem é independente de equações ou classificações predefinidas (lineares ou não lineares, empíricos ou sistêmicos, dinâmicos ou estáticos, determinísticos ou estocástico ou probabilístico). Os métodos se adaptam aos dados para construir os modelos, podendo esses serem constituídos de diferentes classificações e com equações não viáveis/conhecidas, visto que o fenômeno abordado é tipicamente complexo (ESTRADA-PEÑA *et al*, 2014; OSTFELD *et al*, 2008). Dessa forma, caracteriza-se como uma modelagem desafiadora, que no caso dessa tese, transcende a modelagem de nicho ecológico e os estudos epidemiológicos.

As doenças infecto parasitárias compartilhadas entre homens e animais (zoonoses) são fenômenos tipicamente complexos e, por muitas vezes, incógnitos para os especialistas da área. Esta tese traz como principal e original contribuição científica a investigação de diversas relações entre diferentes elementos, possivelmente envolvidos no processo de distribuição de zoonoses, por meio da abordagem de modelagem computacional baseada em dados e seus métodos de aprendizagem de máquina. Na construção dos modelos de distribuição de zoonoses, a investigação será livre de premissas e restrições, visto que nela o fenômeno (zoonose) será implicitamente descrito pelos dados e o modelo computacional será construído a partir desses dados.

A pesquisa demandou ainda a construção de uma base de dados, que foi concebida para tentar minimizar/compensar as vulnerabilidades no processo de modelagem provenientes do não conhecimento do posicionamento geográfico específico dos dados epidemiológicos do Brasil. A opção adotada foi produzir a caracterização mais diversificada possível dos municípios, reunindo estatísticas descritivas sobre as feições naturais e artificiais que compõem as paisagens dos municípios, a qual poderá contribuir relevantemente para as

pesquisas epidemiológicas, e para outras diversas pesquisas de modelagem baseada em dados. Importante destacar ainda que a metodologia referida é totalmente adaptável a outras unidades espaciais de referência (localidades, bairros, distritos), ou como áreas de abrangência representadas por áreas de circulação (*buffers*) das espécies de hospedeiros e vetores.

Dessa forma, considera-se que muitas das relações indiretas ou não óbvias, mas que influenciam significativamente no processo de distribuição de doenças, possam contribuir para esclarecer muitas das questões/relações/suposições ainda não plenamente conhecidas no complexo contexto socioambiental brasileiro (com dimensões continentais, enorme diversidade ambiental/ecológica, desigualdades sociais e com intensas alterações ambientais). As etapas metodológicas e as fases de desenvolvimento desta tese são representadas na figura 3.1.

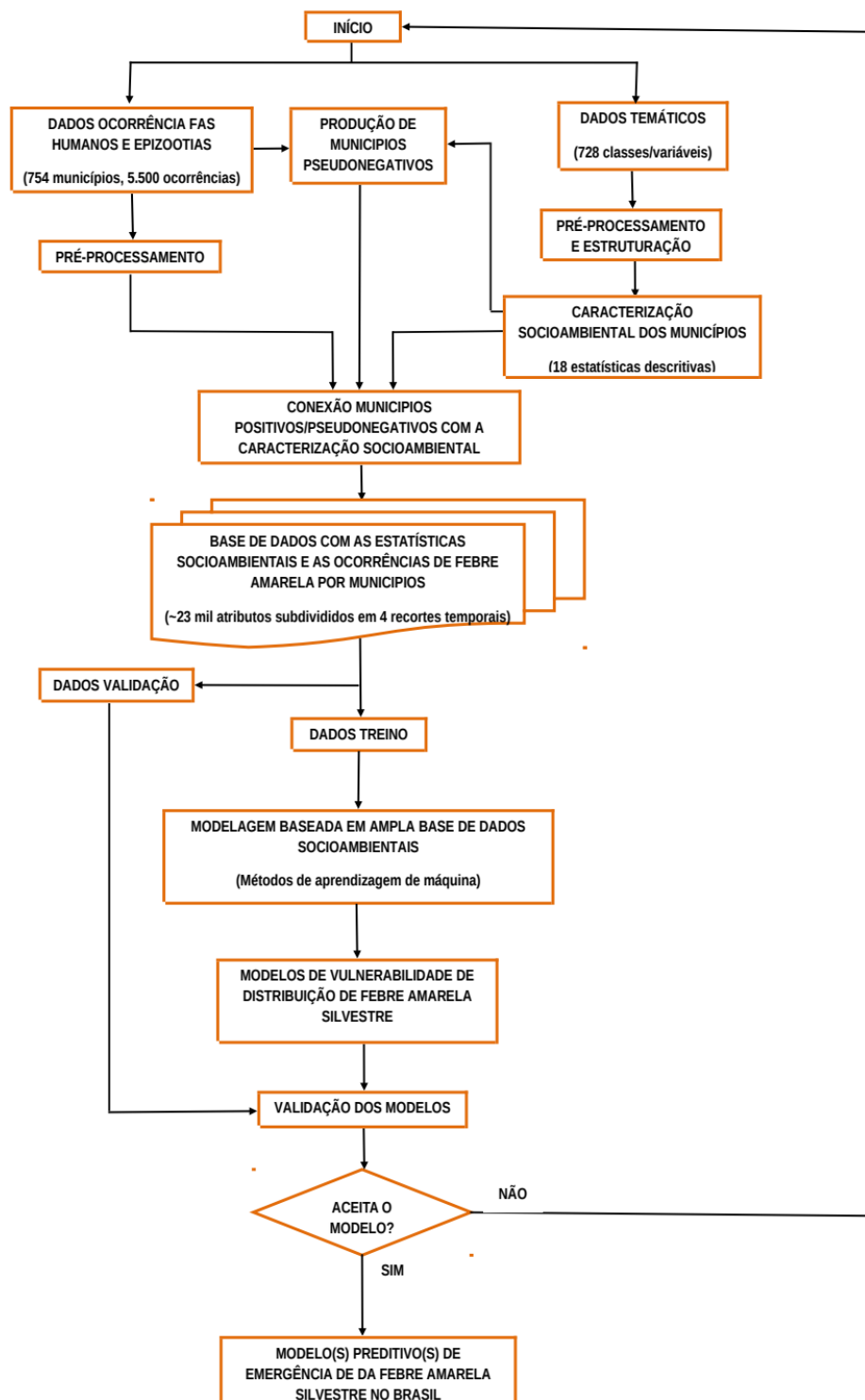
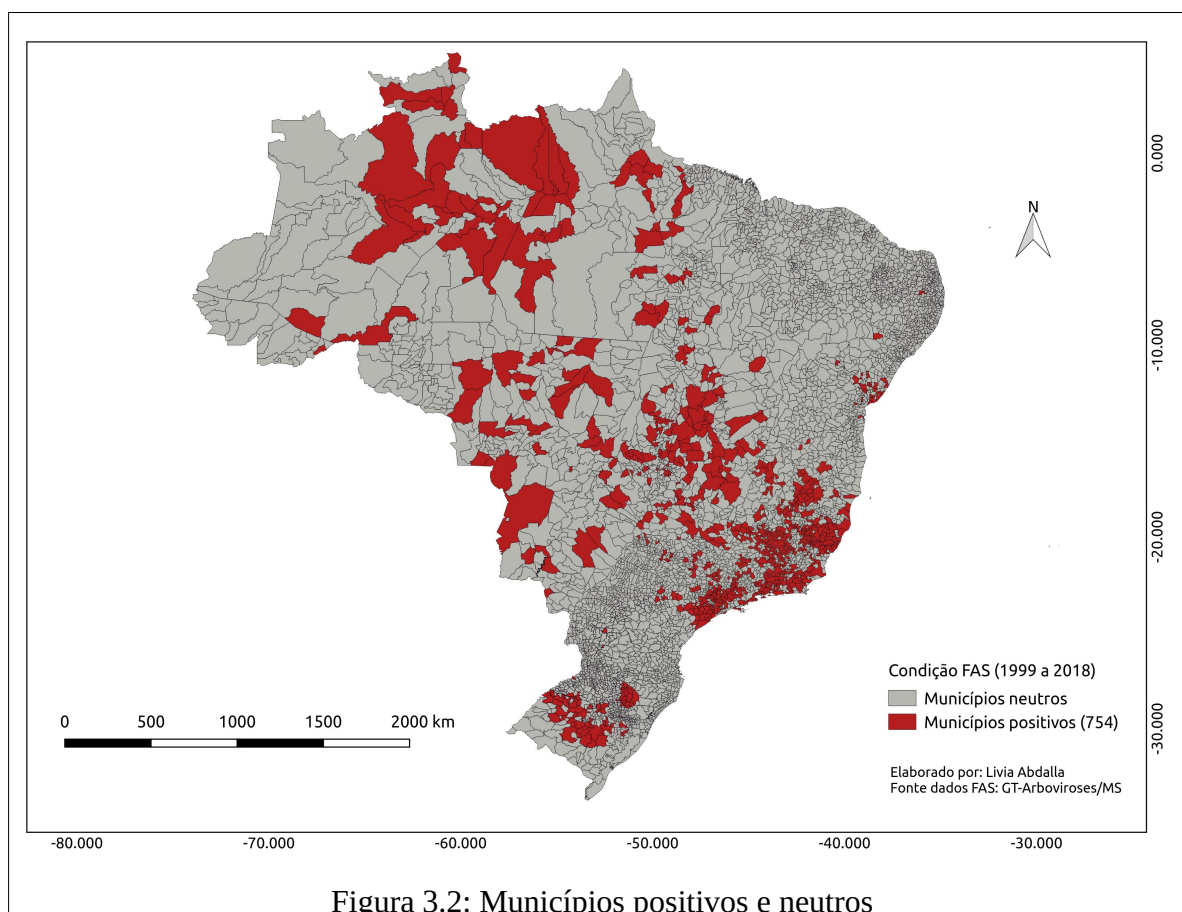


Figura 3.1: Fluxograma da modelagem de favorabilidade de emergência da febre amarela silvestre

3.1 Aquisição das bases de dados

Primeiramente foram adquiridas 5.500 ocorrências confirmadas de febre amarela, agrupados em 754 municípios (figura 3.2), referentes aos 2.441 casos humanos confirmados (44,38%) e as 3.059 indicações de epizootias em primatas não humanos (PNH) (55,62%), para o período de janeiro de 1999 a maio de 2018. Os dados utilizados nesta tese foram concedidos pelo Grupo Técnico de Vigilância das Arboviroses (GT-Arboviroses), pertencente a seguinte estrutura hierárquica: Ministério da Saúde (MS); Secretaria de Vigilância em Saúde (SVS); Departamento de Vigilância das Doenças Transmissíveis (DEVIT); Coordenação-Geral de Doenças Transmissíveis (CGDT) e Unidade Técnica de Vigilância das Doenças de Transmissão Vetorial (UVTV).



Os dados referidos podem ser consultados também por meio do Sistema de Informação de Agravos de Notificação (SINAN), que permite cadastro e acesso somente para profissionais de saúde designados pelas esferas federal, estadual e municipal, ou ainda por meio do sistema TABNET do Departamento de Informática do SUS (DATASUS), que é de acesso público, mas que disponibiliza somente casos humanos (DATASUS, 2019).

Os limites municipais, principal unidade de agregação empregada na tese, foram obtidos por meio da base de dados geográficos do Instituto Brasileiro de Geografia e Estatística - IBGE, correspondente à malha territorial de 2015, com 5.571 municípios brasileiros. (IBGE, 2016).

As variáveis explicativas correspondem ao conjunto de dados que caracterizam o fenômeno. Para essa tese utilizou-se um diverso conjunto de dados temáticos, para todos os períodos de tempo (anual, atemporal ou sem regularidade temporal) disponíveis. No total foram utilizadas 19 macro camadas temáticas, produzidas por diferentes instituições governamentais brasileiras e estrangeiras (Tabela 3.1). Essas camadas possuem uma ou mais classes temáticas ou variáveis, dependendo da natureza do tema, totalizando 728 classes ou variáveis temáticas (Dicionário de dados – APÊNDICE A). Para cada classe/variável foram calculadas dezoito estatísticas descritivas (a serem detalhadas no subcapítulo 3.4), no escopo dos 5.571 municípios, para todos os anos disponíveis, totalizando 22.869 atributos.

Tabela 3.1: Camadas temáticas que compõem a base de dados

ID	Temporalidade	Camadas Temáticas	Quantitativo de classes/variáveis	Quantitativo de atributos (n° classes/variáveis X total de anos X n° de estatísticas descritivas)
1	Anuais (2000 a 2016)	Uso do solo e cobertura vegetal (Mapbiomas coleção 2.3)	16	16 (classes) x 17 (anos) x 18 (estatísticas) = 4.896
2		Temperatura e precipitação (Global Precipitation Mission)	19	19 (classes) x 16 (anos) x 18 (estatísticas) = 5.472
3	Quinquenal (2000, 2005, 2010, 2015)	Contagem da população (SEDAC - NASA)	1	1 (variável) x 4 (anos) x 18 (estatísticas) = 72
4		Densidade da população (SEDAC - NASA)	1	1 (variável) x 4 (anos) x 18 (estatísticas) = 72
5	Atemporais	Normais climatológicas da temperatura e precipitação (Worldclim)	48	48 (variáveis) x 1 (ano) x 18 (estatísticas) = 864
		Altimetria (Aster GDEM - NASA)	1	1 (variável) x 1 (ano) x 18 (estatísticas) = 18
6		Geomorfologia (IBGE)	10	10 (classes) x 1 (ano) x 18 (estatísticas) = 180
7		Solos (IBGE)	67	67 (classes) x 1 (ano) x 18 (estatísticas) = 1.206
8		Fitofissionomiais vegetacionais (IBGE)	53	53 (classes) x 1 (ano) x 18 (estatísticas) = 954
9		Limites biomas (IBGE)	6	6 (classes) x 1 (ano) x 18 (estatísticas) = 108
10	Sem Regularidade Temporal	Áreas de exploração mineral (IBGE)	345	345 (classes) x 1 (ano) x 18 (estatísticas) = 6.210
11		Rodovias (IBGE)	1	1 (variável) x 1 (ano) x 18 (estatísticas) = 18
12		Ferrovias (IBGE)	1	1 (variável) x 1 (ano) x 18 (estatísticas) = 18
13		Hidroviias (IBGE)	1	1 (variável) x 1 (ano) x 18 (estatísticas) = 18
14		Hidrelétricas (IBGE)	1	1 (variável) x 1 (ano) x 18 (estatísticas) = 18
15		Barragens (IBGE)	1	1 (variável) x 1 (ano) x 18 (estatísticas) = 18
16		Áreas de unidades de conservação (IBGE)	1	1 (variável) x 1 (ano) x 18 (estatísticas) = 18
17		Terras indígenas (IBGE)	1	1 (variável) x 1 (ano) x 18 (estatísticas) = 18
18		Climas zonais e suas subunidades regionais (IBGE)	51	51 (classes) x 1 (ano) x 18 (estatísticas) = 918
19		Área de distribuição de PNHs do Brasil (IUCN)	100	100 (classes) x 1 (ano) x 18 (estatísticas) = 1.800
Total =			728	22869

As camadas temáticas anuais de uso do solo e cobertura vegetal reúnem 16 classes temáticas no período de 2000 a 2016 para todo o território brasileiro, com resolução espacial de 30 x 30 m. Essas camadas foram produzidas e disponibilizadas pela plataforma *online* MapBiomias (MAPBIOMAS, 2018), coleção 2.3. Os mapas anuais de uso do solo e cobertura vegetal foram produzidos a partir de processos de classificação automática aplicada em imagens do satélite Landsat. O Projeto MapBiomias é uma iniciativa multi-institucional coordenada pelo Sistema de Estimativas de Emissões de Gases de Efeito Estufa do Observatório do Clima (SEEG/OC) e é composta por uma rede colaborativa de cocriadores formado por Organizações Não Governamentais (ONGs), universidades e empresas. Tem como objetivo a produção de mapas anuais da cobertura e uso do solo do Brasil, desde 1985 até os dias atuais.

As camadas anuais de temperatura e precipitação reúnem 19 diferentes tipos de dados no período de 2000 a 2015, para toda a superfície terrestre, com resolução espacial de 1 km². Essa base de dados denominada *Climate Hazards Group Infrared Precipitation with Stations* (CHIRPS) da iniciativa *Global Precipitation Mission* (GPM), parceria entre as agências espaciais dos EUA (*National Aeronautics and Space Administration* – NASA) e Japão (*Japan Aerospace Exploration Agency* – JAXA) (FUNK *et al*, 2014).

As camadas quinquenais de Contagem da População e Densidade Populacional foram obtidas da base de dados *Socioeconomic Data and Applications Center* – SEDAC – da NASA (<http://sedac.ciesin.columbia.edu/data/collection/gpw-v4/sets/browse>). Esta trata-se de um centro de dados da *Earth Observing System Data and Information System*, hospedado no *Center for International Earth Science Information Network* (CIESIN) da Universidade de Columbia. Este conjunto de dados estima a contagem de população nos anos 2000, 2005, 2010 e 2015, com base em censos nacionais e registros de população, no formato raster, com resolução espacial de 1km². Esse conjunto de dados foi adotado por permitir a aplicação da metodologia de cálculo das estatísticas por municípios, o que não seria possível com conjunto de dados censitários do IBGE que fornece valores já consolidados, únicos por município, o que inviabilizaria a obtenção de estatísticas (por exemplo: mínimo, máximo, desvio padrão, entre outros) sobre o quantitativo da população por município.

Os dados atemporais reúnem os seguintes temas: Normais Climatológicas da

Temperatura; Altimetria; Geomorfologia; Solos; Fitofisionomias Vegetacionais e Limite Biomas. A base altimétrica utilizada foi disponibilizada pelo portal ASTER GDEM, (<https://asterweb.jpl.nasa.gov/gdem.asp>), fruto da ação conjunta da NASA com o Ministério da Economia, Indústria e Comércio do Japão. O modelo Digital de elevação (GDEM) foi criado a partir do processamento e correlação estéreo de 1,3 milhão de imagens óticas individuais, cobrindo 99% da superfície do planeta, entre as latitudes 83 graus norte e 83 graus sul. As imagens que deram origem ao mapa foram obtidas pelo Radiômetro de Reflexão e Emissão Termal (*Aster*), com resolução espacial de 30 x 30 metros, um sensor japonês a bordo do satélite Terra.

As normais climatológicas da temperatura correspondem a dados de observação, representativos de 1950 a 2000, e que foram interpolados para a resolução 1 x 1 km. Os valores de temperatura são fornecidos em °C x 10. Esse conjunto de dados é oriundo da plataforma Worldclim, que é uma organização de pesquisas norte-americana reconhecida internacionalmente (HIJMANS *et al.*, 2005). Os dados utilizados de temperatura média, mínima e máxima compreendem informações de diferentes sensores remotos, a bordo dos satélites MODIS e NOAA, que operam conjuntamente para captação dos valores de temperatura de superfície e de umidade atmosférica. Importante destacar que se optou também por considerar os dados de normais climatológicas, além dos dados anuais de temperatura, pois as Normais Climatológicas fornecem valores médios mensais de temperatura. Esses valores complementam as informações anuais (bastante influenciados por eventos climáticos como *El Niño* e *La Niña*) e representam uma importante referência sobre os padrões de variações estacionais da temperatura, fator que influencia diretamente na dinâmica de reprodução e sobrevivência de espécies como os vetores.

Os demais temas – Geomorfologia, Solos, Fitofisionomias Vegetacionais e Limites Biomas - são disponibilizados pelo Instituto Brasileiro de Geografia e Estatística – IBGE. Foram construídos a partir da interpretação de imagens de satélites e de diversos trabalhos de campo (com início em 1990) por todo o Brasil. A escala cartográfica adotada de 1:250.000 permite um nível de detalhamento regional e possui semelhança metodológica e informacional dos projetos RADAMBRASIL (Radar da Amazônia), iniciado nos anos 1970 e encerrado em 1985, e SIVAM (Sistema de Vigilância da Amazônia) (LIMA, 2008).

As camadas sem regularidade temporal reúnem: Áreas de exploração mineral; Rodovias; Ferrovias; Hidrovias; Hidrelétricas; Barragens; Áreas de Unidades de Conservação - UCs; Terras Indígenas - TIs; Climas zonais e suas subunidades; e Áreas de distribuição de Primatas Não Humanos (PNHs). A camada Áreas de exploração mineral possui 345 classes, que representam os diferentes tipos de minerais explorados no território brasileiro. As demais camadas tratam-se de classes únicas de rodovias, ferrovias, hidrovias, hidrelétricas, barragens e limites de UCs e Tis, as quais foram obtidas a partir da base cartográfica contínua do IBGE (<http://www.geoservicos.ibge.gov.br>), na escala 1:250.000.

A camada Climas Zonais e suas subunidades representa as diferentes zonas climáticas do território brasileiro agrupadas pela temperatura e umidade. Esta camada também identifica os tipos climáticos, caracterizado por tons e sobretons: quente, subquente, mesotérmico brando e mesotérmico mediano, na escala 1:5.000.000 (IBGE, 2002).

As áreas de distribuição dos Primatas não Humanos no Brasil (hospedeiros silvestres principais) foram obtidas a partir da plataforma *Red List da International Union for Conservation of Nature* – IUCN (IUCN, 2018). Foram obtidas as áreas de distribuição de todos os 101 Primatas Não Humanos – PNHs - com ocorrência no Brasil. Além dos PNHs com diagnóstico positivo para febre amarela silvestre (gêneros *Callithrix*, *Alouatta*, *Cebus*, *Ateles*, *Aotus*, *Saimiri*, *Sapajus*), não existe investigação suficiente que comprove que outros gêneros, como os *Pithecia*, *Brachyteles*, *Lagothrix*, entre outros, não estejam acometidos pelo vírus. Portanto, todas as espécies de PNHs com ocorrência no Brasil foram incorporados na base de dados.

Dados de ocorrência de espécies de vetores transmissores do vírus (gêneros *Haemagogus* e *Sabethes*) poderiam ser obtidos a partir da plataforma speciesLink (www.splink.cria.org.br). A plataforma reúne dados oriundos dos principais museus, herbários e coleções microbiológicas do Brasil e de outros países. Porém, optou-se pela não utilização desses dados de ocorrência por considerar que o enviesamento e descontinuidade amostral dos dados disponibilizados na plataforma poderiam agregar maior imprecisão e ruído à modelagem.

Além das camadas temáticas utilizadas por essa pesquisa, outros níveis de informação ainda poderiam ser avaliados e agregados para a construção dos modelos (por exemplo: dados relativos ao monitoramento sistemático de focos de queimadas, desmatamento de florestas ou

degradação de pastagens; identificadores sobre o risco geológico para ocorrência de desastres naturais; limites de bacias hidrográficas e suas subdivisões, entre outros). Essa, sem dúvida, é uma oportunidade de continuidade dessa pesquisa, de modo que, a partir da inclusão de novas bases de dados sejam realizadas análises comparativas com os resultados obtidos por essa tese.

3.2 Produção de municípios pseudonegativos

Para o processo de modelagem aqui adotados, tão relevante quanto os dados sobre as ocorrências são os dados de não ocorrências. Para a identificação dos dados de ausência de febre amarela seria necessário uma investigação ampla e profunda, com coleta em campo para diagnóstico laboratorial do material biológico recolhido de diversos humanos e primatas, para a comprovação de que o vírus não estaria presente no recorte espacial (município) e temporal (1999 a 2018) adotado nesta pesquisa.

Em vista da indisponibilidade de dados confiáveis sobre regiões comprovadamente negativas à ocorrência da doença e, principalmente, em razão de alguns métodos de aprendizagem de máquina (métodos discriminativos²) requererem dados positivos e negativos para a etapa do treinamento, foi definida como estratégia a seleção de municípios que não tenham histórico de ocorrência de FAS, definidos aqui como neutros, segundo a base de dados do Ministério da Saúde, e que possuam as maiores distâncias euclidianas, com relação aos municípios positivos, no espaço de atributos. A figura 3.3 ilustra e exemplifica como é realizado o cálculo da distância euclidiana no espaço de atributos, entre 1 município neutro e 3 municípios positivos. Como contraponto, a figura 3.4 apresenta a distribuição geográfica

²Métodos discriminativos: modelam a distribuição de probabilidade a posteriori (condicional) $P(Y/X)$, de modo que, dado X , retornam a distribuição de probabilidade para Y . Os modelos discriminativos focam apenas na geração de um modelo que consiga definir os resultados de acordo com os dados observados, como é o caso das Redes Neurais e das Máquinas de Vetores de Suporte (SVM) (SMOLA *et al*, 2000). Outra categoria é formada pelos modelos Generativos, que possuem a preocupação em determinar o mecanismo que gerou os dados e, dessa forma, tendem a incorporar informações estruturais dos dados no classificador, como *Isolation Forest*.

desses mesmos municípios no espaço geográfico. A partir da observação de ambas as figuras verifica-se que o valor das distâncias euclidianas no espaço de atributos e no geográfico são distintos para o mesmo conjunto de municípios.

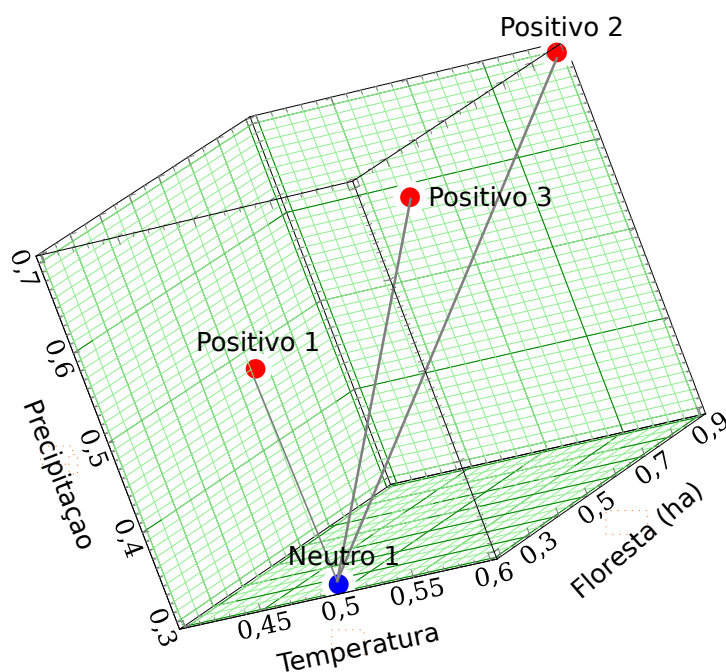


Figura 3.3: Distância euclidiana no espaço de atributos entre municípios

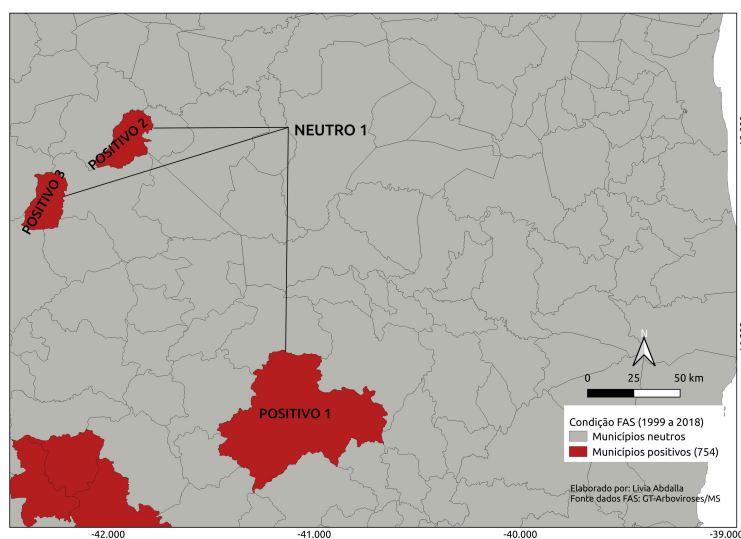


Figura 3.4: Distância euclidiana no espaço geográfico entre municípios

Para indicação dos municípios pseudonegativos, baseados no cálculo das distâncias euclidianas no espaço de atributos, foi utilizada a biblioteca “*parallelDist*” (<https://cran.r-project.org/web/packages/parallelDist/parallelDist.pdf>) disponibilizada no software *R project* (<https://cran.r-project.org/>). O cálculo das distâncias euclidianas no espaço de atributos utilizou os arquivos de caracterização produzidos por essa pesquisa, somente com valores normalizados, de todos os municípios brasileiros (os arquivos de caracterização serão descritos adiante no subcapítulo 3.4), havendo até então o conhecimento somente sobre municípios positivos e neutros.

Para a indicação dos municípios pseudonegativos, baseada nas distâncias euclidianas no espaço de atributos, cumpriram-se as seguintes etapas:

- 1) cálculo das distâncias entre todos os municípios neutros contra todos os municípios positivos;
- 2) listagem das distâncias existentes entre cada um dos municípios neutros com todos os municípios positivos;
- 3) seleção da menor distância existente entre cada um dos municípios neutros com um município positivo;
- 4) cálculo do valor da mediana das menores distâncias selecionadas na etapa 3;
- 5) seleção dos municípios neutros que possuem distâncias maiores do que o valor da mediana calculada na etapa 4;
- 6) seleção aleatória de parte dos municípios neutros selecionados na etapa 5 (o que reduziu ou minimizou o possível viés da estratégia adotada);
- 7) definição dos municípios neutros como pseudonegativos, a partir da seleção aleatória na etapa 6.

Para o cumprimento das etapas 1 e 2 evidencia-se o código a seguir:

```
#Ler dados da base de caracterização  
base = "01_caracterizacao_final_mun_1999_2017.csv"  
rawdata = read.csv(base)  
last = length(rawdata[,]) - 1 #exclude the FA attribute  
  
#Aplique o filtro para selecionar apenas atributos normalizados
```

```
filtdata = rawdata[, grep('normalized',names(rawdata))]
```

```
#Normalize os dados
```

```
q = apply(filtdata, 2, max)
```

```
for (i in 1:length(q)) { if (q[i] == 0) { q[i] = 1 } }
```

```
filtdata = t(apply(filtdata, 1, function(x){ x/q } ) )
```

```
#Importar biblioteca parallel Dist
```

```
library(parallelDist)
```

```
#Calcule a distância com 15 threads (uso paralelo)
```

```
distances = parallelDist(as.matrix(filtdata), threads=15)
```

```
mdistances = as.matrix(distances)
```

```
normalized = TRUE
```

Para execução da etapa 3 evidenciam-se os comandos a seguir:

```
size = length(mdistances[,1])
```

```
#Encontre o índice dos casos positivos
```

```
pindex = which(rawdata[,last+1] == 1)
```

```
#Encontre o índice dos casos pseudonegativos
```

```
nindex = which(rawdata[,last+1] == 0)
```

```
#Matriz para armazenar o geocódigo da área pseudonegativa e sua menor distância para uma área positiva
```

```
mdisttopositive = matrix(ncol=2, nrow=length(nindex))
```

```
#Encontre a distância mínima entre uma área negativa e todas as áreas positivas
```

```
k = 1
```

```
for (i in nindex ) {
```

```
  mdisttopositive[k, 1] = rawdata[i, 1]
```

```
  mdisttopositive[k, 2] = min(mdistances[i, pindex])
```

```
  k = k + 1
```

```
}
```

```
orderednegatives = mdisttopositive[order(mdisttopositive[,2]),]
```

```
if (normalized == TRUE){
```

```
write.csv(orderednegatives, "distancepositives_normalized.csv", quote=FALSE, row.names=FALSE)
```

```
} else {
```

```
write.csv(orderednegatives, "distancepositives.csv",quote=FALSE,row.names=FALSE)
```

```
}
```

Conforme descrito, as etapas anteriores foram executadas utilizando biblioteca disponibilizada pelo *software* R (R DEVELOPMENT CORE TEAM, 2019). Já as etapas 4 a 6 foram realizadas utilizando funções disponíveis na plataforma PostgreSQL. Os comandos utilizados para o cumprimento das etapas 4 a 6 são:

```
-- Acrescenta tabela com o valor de distancias dos positivos

CREATE TABLE distancepositives_01_normalized (geocode integer, distance double precision)

COPY distancepositives_01_normalized FROM 'distancepositives_normalized_01.csv' CSV HEADER
DELIMITER ',';

ALTER TABLE distancepositives_01_normalized ADD COLUMN distance_01_numeric numeric;

UPDATE distancepositives_01_normalized SET distance_01_numeric = distance;

-- Calcula a mediana das distancias

SELECT median(distance_01_numeric) AS median_value FROM distancepositives_01_normalized;

-- Faz o corte de 572 municípios pseudonegativos baseada na mediana calculada no comando anterior

CREATE TABLE municipios_negativos_from_distancepositives_01_normalized AS
SELECT * FROM distancepositives_01_normalized WHERE distance_01_numeric >= 5.60 ORDER
BY random () LIMIT 572;

SELECT * FROM municipios_negativos_from_distancepositives_01_normalized;

SELECT * FROM fa_positivo_1999_2017

-- Faz o Join da tabela com 5571 municípios com os positivos e em seguida com os pseudonegativos

CREATE TABLE municipios_geocode_join_fa_positivo_1999_2017 AS
SELECT A.*, B.fa_pos_1999_2017 FROM municipios_geocode A LEFT JOIN
fa_positivo_1999_2017 B ON A.geocode = B.geocode;

CREATE TABLE municipios_geocode_join_fa_positivo_pseudonegativos_1999_2017 AS
SELECT A.*, B.distance_01_numeric FROM municipios_geocode_join_fa_positivo_1999_2017 A
LEFT JOIN municipios_negativos_from_distancepositives_01_normalized B ON A.geocode =
B.geocode;

COPY municipios_geocode_join_fa_positivo_pseudonegativos_1999_2017 TO
'municipios_geocode_join_fa_positivo_pseudonegativos_1999_2017.csv' CSV HEADER
DELIMITER ',';
```

Após a exemplificação dos comandos executados para o cumprimento das etapas, se faz necessário algumas complementações:

(i) a estratégia de inicialmente selecionar das menores distâncias existentes entre neutros e positivos, para posterior seleção dos municípios mais distantes entre os similares, foi necessária, pois somente assim houve a garantia de que havia conhecimento sobre os municípios neutros mais diferentes do universo dos positivos mais similares. Caso a primeira seleção (menores distâncias existentes entre neutros e positivos) fosse sobre as maiores distâncias entre neutros e positivos, não haveria a garantia de que a segunda seleção contemplaria municípios mais distantes possíveis, abrindo margem, portanto, para a indicação de um determinado neutro que fosse ambigualmente o mais distante de um determinado positivo, mas que tivesse grande similaridade com outro positivo;

(ii) a definição de municípios pseudonegativos foi realizada de maneira independente para cada um dos quatro recortes temporais adotados nesta tese (1999 a 2017; 1999 a 2003; 2007 a 2009 e 2014 a 2017 - a ser detalhado no subcapítulo 3.5), ou seja, as etapas de 1 a 6 foram integralmente cumpridas para cada recorte. Isto foi necessário e fundamental para garantir que a indicação de municípios pseudonegativos estaria ajustada às especificidades e particularidades de cada um dos surtos (que contaram com quantidade e arranjo de municípios positivos diferentes e, portanto, com diferentes distâncias calculadas), havendo assim a possibilidade de um município ser pseudonegativo em um recorte e manter-se neutro em outro recorte. Importante destacar que a escolha pelo cálculo das distâncias euclidianas, em espaço de atributos com alta dimensionalidade, pode representar um problema. Sendo assim, a indicação de pseudonegativos por recorte temporal das bases de dados representa uma precaução para minimizar este problema, já que os quatro recortes utilizaram quantitativos menores de atributos. Caso a dimensionalidade do espaço de atributos não fosse reduzida conforme os recortes adotados, seria prudente a utilização de métodos de seleção de atributos, como CFS ou PCA, antes do cálculo das distâncias.

Considerando o entendimento dessas complementações, indica-se na tabela 3.2 o valor da mediana calculado (utilizado para seleção aleatória dos pseudonegativos) para cada uma das quatro listas de distâncias produzidas (APÊNDICE B) em função de cada recorte temporal da tese:

Tabela 3.2: Valor da mediana das menores distâncias calculadas entre municípios positivos e neutros no espaço de atributos

Recorte temporal	Lista de distâncias	Valor mediana
1999 a 2017	distance_positives_normalized_01	5.60
1999 a 2003	distance_positives_normalized_02	7.07
2007 a 2009	distance_positives_normalized_03	6.92
2015 a 2017	distance_positives_normalized_04	6.68

Outras soluções poderiam ser utilizadas para definir as regiões negativas, como por exemplo: i) escolha de municípios negativos baseada no conhecimento de especialistas; ii) escolha aleatória (e balanceada com os 754 municípios positivos) de parte dos 4.817 municípios “neutros”, ou seja, que não possuem registros positivos, segundo a base de dados do Ministério da Saúde. Porém, a identificação dos municípios com maior distanciamento das características sociais e ambientais dos municípios positivos demonstrou ser a solução mais coerente, posto que está alinhada com a metodologia de modelagem baseada em dados e com o contexto biogeográfico e o histórico do fenômeno de emergência da febre amarela.

As soluções exemplificadas foram inicialmente pensadas como opções para contornar a carência de dados negativos. Contudo, foram identificados alguns contra-argumentos que fundamentam e direcionam esta pesquisa para a escolha de municípios com características socioambientais distantes das características dos municípios positivos. A primeira opção torna-se incompatível, pois o nível de conhecimento por parte dos especialistas sobre o fenômeno é restrito, de modo que não se conhecem suficientemente, até a presente data, todos os fatores/características socioambientais que determinam se a febre amarela emergirá ou não em um determinado lugar. Logo, não seria confiável, ou coerente, a indicação de áreas negativas de manifestação da doença. A segunda opção também apresenta-se como conflitante com o propósito desta tese, pois à medida que determina-se, mesmo que de maneira aleatória, municípios pseudonegativos, esta ação impossibilita que o modelo extrapole a previsão de ocorrência para esses municípios, justamente porque já foram determinados como pseudonegativos, mas que na realidade não existe critério para essa inelegibilidade ou investigação científica suficiente que garanta a negatividade desses municípios para emergência da doença.

Nesse contexto, a fim de garantir a imparcialidade, evitar enviesamento por parte dos especialistas e permitir flexibilidade no processo de predição de novos municípios, optou-se pelo cálculo das distâncias euclidianas no espaço de atributos entre os municípios positivos e neutros para a indicação dos municípios pseudonegativos.

Por tudo isso, considera-se que os municípios pseudonegativos darão as condições necessárias para que os métodos discriminativos identifiquem padrões consistentes nos positivos. E assim, tenham flexibilidade suficiente para extrapolar a previsão aos municípios que ainda não manifestaram registros de ocorrência (neutros), mas que possuem condições/características socioambientais favoráveis/viáveis para a manifestação da doença em seu território (futuros positivos). Importante destacar que esse processo de produção de municípios pseudonegativos também pode representar oportunidade de continuidade da pesquisa. Pode haver a investigação de outras estratégias de produção de municípios pseudonegativos que não foram pensadas ou discutidas por essa pesquisa, como por exemplo: a geração de limites artificiais para simulação de municípios pseudonegativos.

3.3 Pré-processamento e estruturação

Considerando a diversidade das instituições produtoras dos dados utilizados e as particularidades estruturais das camadas temáticas adquiridas, foi fundamental realizar a etapa de pré-processamento e estruturação das bases de dados para a realização da etapa seguinte de cálculo das estatísticas. Todas as tarefas de pré-processamento, reunindo 189GB de dados, foram realizadas nos softwares ArcGIS, versão 10.0, e QGIS, versão 2.18. Nessa etapa foi necessário padronizar as características cartográficas dos dados geoespaciais, corrigir erros topológicos, eliminar duplicidade de informações e uniformizar a estrutura da tabela de atributos. De modo geral, os dados foram organizados em dois grandes grupos: dados vetoriais e dados matriciais.

Para os dados matriciais de Uso do Solo e Cobertura Vegetal foi necessário realizar a conversão do formato *raster* para *shapefile*, por meio da ferramenta *raster2shape* (QGIS Development Team, 2019) do software QGIS 2.18, totalizando aproximadamente 168 horas de processamento computacional em um computador de alto desempenho (Memória RAM de 64 GB DDR4, Processador com 8 núcleos físicos, Armazenamento com SSD de 128 GB e

HD de 4 TB). A conversão foi necessária, pois o formato vetorial (*shapefile*) permitiu o cálculo das estatísticas dos polígonos de todas as feições de uso e cobertura, o que no formato matricial (*raster*) não foi possível com as técnicas/funções utilizadas (a serem descritas no subcapítulo 3.4). Antes da conversão do formato foi necessário realizar o processo de reamostragem do tamanho dos pixels dos dados de uso e cobertura, pois antes da reamostragem (com a resolução espacial de 30m) o processo de conversão bloqueava em um limite do próprio formato *shapefile*, não sendo possível criar arquivos *shapefiles* com tamanho maior do que 2GB. Dessa forma, a solução pensada para redimensionar o *raster*, e consequentemente, viabilizar a conversão para vetor, foi a reamostragem do tamanho do pixel para 300m, o que resultou em arquivos *shapefiles* com 1,7GB. Apesar da perda na precisão do dado, esta foi a melhor solução encontrada pela pesquisa para viabilizar a utilização desses dados. Pensar em outras soluções sem redimensionar também pode ser tema para continuidade (por exemplo: adotar formato *geodatabase *.gdb* no processo de conversão; realizar recorte de cada um dos 17 *rasters* de uso e cobertura anual em função dos limites territoriais de cada um dos 5.571 municípios, para posterior conversão de cada um desses 94.707 novos *rasters* para o formato *shp*).

Os dados de Temperatura, Precipitação, Contagem/Densidade da população, Altimetria e Normais Climatológicas, também fornecidas no formato matricial, foram convertidos para o formato ponto, pois tratam-se de dados de natureza pontual que foram interpolados por suas fontes provedoras antes de serem disponibilizados.

Inicialmente, para os dados vetoriais foi necessário homogeneizar os referenciais cartográficos dos dados, utilizando a Projeção Equivalente de Albers (EPSG 102033) para dados que demandam o cálculo de áreas (polígonos), Projeção Equidistante (EPSG 102032) para dados que demandam cálculo de distâncias (linhas) e o Referencial Geodésico SIRGAS 2000 (EPSG 4674) para dados com localização restrita (pontos) (IBGE, 1999). Foi necessário também a correção de alguns erros topológicos dos dados vetoriais nos formatos de linhas e polígonos, os quais tratam-se de erros inerentes à etapa de construção/vetorização dos dados.

Os dados vetoriais correspondem aos seguintes temas: Geomorfologia; Solos; Fitofisionomias vegetacionais; Limites biomas; Áreas de exploração mineral; Rodovias; Ferrovias; Hidrovias; Hidrelétricas; Barragens; Áreas de unidades de conservação; Terras

indígenas; Climas zonais e suas subunidades; Área de distribuição de PNHs.

Com relação a eliminação de duplicidades e uniformização das informações, foi realizada a simplificação das informações da tabela de atributos de todos os dados vetoriais, sendo adotado o seguinte padrão de colunas (atributos) para todos os dados: *datetime* (ano de referência do dado); *geometry* (formato WKT); *variable* (nome da classe/variável temática) e *value* (valor correspondente a intervalos específicos dos dados convertidos em pontos).

3.4 Descrição estatística das características socioambientais dos municípios

Essa etapa teve como objetivo produzir o perfil estatístico descritivo das características socioambientais dos municípios brasileiros, com base nas camadas temáticas descritas anteriormente na seção 3.1. Na prática, significa que para cada camada, no escopo de um município, foram calculadas medidas de tendência central, como média mediana e moda, e medidas de dispersão, como valor/tamanho máximo e mínimo, desvio padrão e percentis.

Importante destacar que para cada uma das camadas foram calculados os valores originais de área e os valores normalizados em função da área total dos municípios. A normalização dos valores deve-se à grande variabilidade territorial das áreas dos municípios brasileiros. Como exemplo, destacam-se Altamira – PA, como o maior município brasileiro, com 159.533 km², e Santa Cruz de Minas – MG, como o menor, com 3.612 km² (IBGE, 2016). Essa grande variabilidade territorial poderia direcionar a modelagem para a identificação de correlações distorcidas, como por exemplo, a identificação de relações entre maiores porções de feições naturais ou antrópicas com a maior concentração de casos, que apenas se deve a maior dimensão territorial do município.

A partir da estruturação dos dados gráficos e não gráficos foi executada a rotina de cruzamento dos dados com os limites municipais, por meio de diferentes funções espaciais da extensão PostGIS do sistema gerenciador de banco de dados PostgreSQL (QUEIROZ e FERREIRA, 2006), versão 1.22.2.

Inicialmente todas as camadas temáticas e a malha municipal foram importadas para a base de dados criada no Sistema Gerenciador de Banco de Dados PostgreSQL/PostGIS por meio do *plugin* shppgsql-gui (<https://www.postgresonline.com/journal/archives/145-PgAdmin-III-Plug-in-Registration-PostGIS-Shapefile-and-DBF-Loader.html>). Posteriormente

foram construídos *scripts* para cálculo das estatísticas. Nesses scripts foram utilizadas funções próprias da linguagem SQL e as funções espaciais ST_AREA e ST_INTERSECTION. Também foi utilizado o *plugin* aggs_for_arrays 1.3.2 (https://pgxn.org/dist/aggs_for_arrays/), que trata-se de uma extensão que oferece várias funções para computar diferentes estatísticas, de maneira otimizada, a partir de números agrupados em vetores. Esta etapa de cálculo das estatísticas descritivas das 728 classes/variáveis temáticas acumularam mais de 336 horas de processamento computacional em um computador de alto desempenho (Memória RAM de 64 GB DDR4, Processador com 8 núcleos físicos, Armazenamento com SSD de 128 GB e HD de 4 TB). Como *script* de exemplo é apresentado a seguir o cálculo da área máxima de floresta para os municípios brasileiros:

```
-- Habilitar extensão para cálculo de estatísticas de um array:

CREATE EXTENSION aggs_for_arrays;

-- Cria um array para cada município com a área de todas as classes temáticas pertencentes a cada município:

SELECT geocodigo,
array_to_max(ARRAY( SELECT a.area FROM unnest(resultado_area) WITH ORDINALITY a(area)
GROUP BY a.area ORDER BY a.area DESC)) as max_classe_temática_1
from
  (select
    m.geocodigo,
    (
      ARRAY(
        select
          ST_AREA(ST_INTERSECTION(m.geom, u.geom)) as area
        from
          camada_tematica u
        where
          u.classe_temática_1 = 'Floresta' and st_intersects(m.geom, u.geom)
      )
    ) as resultado_area
  from
    municípios_2015_albers m
  where
    m.gid > 0)
  as tabela
```

Dessa forma, no escopo de cada um dos 5.571 municípios brasileiros foram calculadas nove estatísticas: Quantidade de feições/geometrias; Valor/Área mínima; Valor/Área máxima; Valor/Área média; Valor/Área Total; Desvio Padrão dos Valores/Áreas; 25º Percentil; 50º

Percentil (mediana); 75° Percentil (em inglês *Count, Min, Max, Mean, Sum, Std, 25th percentile, 50th percentile e 75th percentile*). Para todas as estatísticas calculadas foram calculados também seus valores normalizados, havendo portanto o cálculo de dezoito estatísticas (nove estatísticas sobre valores reais e nove estatísticas sobre valores normalizados) (Tabela 3.3).

Tabela 3.3: Estatísticas calculadas das feições/variáveis no escopo dos municípios

count	Quantidade de feições/geometrias de cada classe ou variável das camadas temáticas que estão contidos(as) em cada município
min	Valor, área ou distância da menor feição/geometria ou valor de cada classe ou variável das camadas temáticas que está contido(a) em cada município
max	Valor, área ou distância da maior feição/geometria ou valor de cada classe ou variável das camadas temáticas que está contido(a) em cada município
mean	Área média das feições/geometrias ou dos valores de cada classe ou variáveis das camadas temáticas que estão contido(a)s em cada município
sum	Soma das áreas, trechos ou valores de cada classe ou variável das camadas temáticas que estão contidos(as) em cada município
sd	Desvio padrão das áreas, distâncias ou valores de cada classe ou variável das camadas temáticas que estão contido(a)s em cada município
25th percentile	Primeiro quartil das áreas, distância ou valores de cada classe ou variável das camadas temáticas que estão contido(a)s em cada município
50th percentile	Mediana das áreas, distâncias ou valores de cada classe ou variável das camadas temáticas que estão contidas em cada município
75th percentile	Terceiro quartil das áreas, distâncias ou valores de cada classe ou variável das camadas temáticas que estão contidas em cada município
Sufixo <i>_normalized</i>	Significa que a estatística que precede o sufixo foi dividida pela área do município

As estatísticas foram calculadas de acordo com a natureza/formato do dado de origem. No formato polígono foram calculadas as áreas (m²) de cada classe (ex: uso do solo). No formato ponto foram identificados os valores (milímetros, K e °C) de cada camada (ex: temperatura, precipitação). No formato linha foram calculados os tamanhos dos trechos (m) de cada camada (ex: rodovias, ferrovias).

Todas as estatísticas foram armazenadas em tabelas temporárias, as quais foram

agregadas por meio da função INNER JOIN (utilizando como campo de junção o *geocode* de cada município) em uma tabela final. Ao final da execução da rotina de cruzamento dos dados temáticos com os limites municipais, foi organizada uma base de dados integrada e estruturada, contendo 22.869 atributos, que representam a pluralidade, no tempo e espaço, das principais características socioambientais dos 5.571 municípios brasileiros.

Nesse sentido, é relevante destacar que esta mesma base de dados poderá ser utilizada em uma ampla diversidade de outros estudos ecológicos e de emergência de doenças, que tenham como objetivo a identificação de padrões consistentes ou relacionamentos sistemáticos entre as condições sociais e ambientais do espaço geográfico com as ocorrências de diferentes doenças zoonóticas. Importante considerar a necessidade de atualização da base de dados produzida, considerando a inserção e a consequente caracterização estatística de novos dados de uso e cobertura ou temperatura e precipitação dos anos subsequentes aos utilizados (2017, 2018, 2019). Mantendo-se assim a base de dados útil para a modelagem e a investigação de emergência ou reemergência de outras zoonoses ou até mesmo para ocorrência de novos vírus e doenças infecciosas associadas.

3.5 Base de dados com as estatísticas socioambientais e as ocorrências de febre amarela por município

A partir da base de dados completa contendo a descrição estatística das características socioambientais dos municípios (atributos descritivos), também foi estabelecida a conexão, por meio da função INNER JOIN, com os municípios de ocorrência (positivos) e os municípios sem histórico de ocorrência e mais distantes no espaço de atributos (pseudonegativos), utilizando o campo *geocode* como condição para junção das tabelas.

Considerando a problemática da variabilidade territorial das áreas dos municípios brasileiros (apresentada na seção 3.2), foram selecionados somente os atributos normalizados (todos atributos com sufixo “*_normalized*”) para a construção dos modelos. Os atributos não normalizados (todos os atributos sem o sufixo “*_normalized*”) foram utilizados de maneira complementar, somente nas etapas de interpretação dos modelos para ajudar na indicação de valores reais de determinados atributos que separam os municípios positivos dos pseudonegativos, ou seja, quais condições e valores favoreceram ou não a emergência da

doença.

Importante destacar que o quantitativo de ocorrências (casos humanos ou epizootias) por município não foi utilizado para composição da base de dados, pois a abordagem utilizada pretende minimizar os vieses amostrais cometidos em função da variabilidade de esforços e recursos (humanos, financeiros, hospitalares, laboratoriais, etc) utilizados e disponíveis para a realização de exames, diagnóstico e notificação de casos humanos. Além das questões que envolvem as amostras dos casos humanos, procurou-se também minimizar o viés amostral correspondente à investigação de epizootias, que além de ser dependente da variabilidade de recursos financeiros, é também dependente da disposição física da equipe, da acessibilidade às regiões de matas (que podem estar associadas a relevos muito íngremes, escarpados, fundos de vale, etc) para a realização de trabalhos de campo de captura e coleta de amostras de primatas para a investigação da positividade para FA. Portanto, compreendendo o viés amostral existente sobre o total de casos humanos e epizootias, esse trabalho optou por adotar somente a classificação de positivos “1” para os municípios com ocorrências, e “0” para os municípios pseudonegativos (previamente descrito na seção 3.2).

Em razão da variabilidade espaço temporal dos surtos epidêmicos e epizooticos das FAS no Brasil, apresentados na seção 2.7.3, esse trabalho organizou 4 bases de dados diferentes para construção, compreensão e avaliação dos modelos: (i) 1999 a 2017; (ii) 1999 a 2003; (iii) 2007 a 2009 e (iv) 2015 a 2017. Cada base de dados teve um quantitativo diferente de atributos descritivos, pois foram selecionados somente os atributos correspondentes ao período do recorte temporal em questão (por exemplo: no recorte 02 foram utilizados somente os atributos correspondentes às estatísticas das classes/variáveis dos anos de 1999 a 2003). Nesse sentido, também variaram o total de positivos e negativos, pois cada recorte corresponde a um surto epidemiológico da doença, que possui maior ou menor magnitude territorial (por exemplo: no recorte temporal 03 foram utilizados 133 positivos e 133 pseudonegativos).

Situação idêntica – no que se refere à diferença de quantitativo entre as bases – também ocorreu para as amostras tomadas para validação da capacidade preditiva do modelo, pois variaram em função da dimensão do evento no período/ano considerado (por exemplo: no recorte 01 e 04 foram utilizados todos os 256 positivos referentes ao ano subsequente, pois

tratam-se dos municípios com ocorrência de FA em 2018; no recorte 02 foram utilizados os 133 municípios positivos com ocorrência da doença no período subsequente – 2007 a 2009).

A avaliação da qualidade dos modelos construídos deve ser analisada sob um conjunto de dados não presente em seu treinamento, minimizando, assim, o risco de sobreajuste. Nesta tese utilizamos dois conjuntos independentes para esta avaliação, os quais aqui denominamos teste e validação (a serem detalhados no subcapítulo 3.7), assim definidos:

(i) teste: conjunto de amostras não contido no treinamento, mas pertencentes ao mesmo recorte temporal. Utilizado para avaliar a acurácia interna (dentro do surto em análise) do modelo;

(ii) validação: conjunto de amostras pertencentes ao recorte temporal subsequente. Utilizado para avaliar a acurácia preditiva do modelo.

Assim, para todos os recortes temporais adotados, o conjunto de validação utilizado foi composto pelos municípios positivos no período do surto seguinte. Por exemplo: para o recorte 02 (1999 a 2003) o conjunto de validação corresponde aos dados do surto 03 (2007 a 2009) (ver tabela 3.4). Essa decisão de utilizar dados do surto consecutivo contribui para verificação da capacidade de generalização e extrapolação do modelo e da não especialização do modelo nos dados de treinamento. A escolha sobre o modelo final de cada recorte foi orientada pela combinação do desempenho das métricas de acurácia interna e de acurácia preditiva, contextualizadas nesse subcapítulo.

Tabela 3.4: Quantitativo de atributos, dados de treino, teste e validação para cada recorte temporal

N	Recorte temporal	Atributos descritivos	Positivos	Pseudo Negativos	Treino	Teste	Validação
1	1999 a 2017	11619	572	572	458	114	256 (2018)
2	1999 a 2003	7642	117	117	94	23	133 (2007 a 2009)
3	2007 a 2009	9549	133	133	106	27	601 (2015 a 2018)
4	2015 a 2017	11619	345	345	276	69	256 (2018)

A base de dados contendo o recorte temporal 01 reúne 11.619 atributos descritivos

(somente atributos normalizados), possuindo 572 municípios positivos e 572 municípios pseudonegativos. Desse total de municípios, foram separados 458 para treino e 114 para teste, tanto considerando as ocorrências positivas quanto as pseudonegativas. Dessa forma, a base foi dividida em 80% para construção (treino) do modelo e 20% para avaliação da coerência interna (teste) do modelo. Os 256 municípios positivos em 2018 foram utilizados exclusivamente para validação da acurácia preditiva, ou seja, após a construção do modelo, verifica-se qual a capacidade de generalização desse modelo (e do seu conjunto de regras estabelecido) para previsão de novos municípios favoráveis à emergência da doença.

O recorte temporal 02 (1999 a 2003) corresponde a base com 7.642 atributos, 117 registros positivos e 117 negativos, tendo sido utilizados 94 e 23 para treino e teste, respectivamente. Para a validação da capacidade preditiva utilizam-se 133 municípios que apresentaram-se como positivos no surto seguinte de 2007 a 2009. O recorte temporal 03 (2007 a 2009) formou a base com 9.549 atributos, 133 instâncias positivas e 133 negativas, com 106 separadas em treino e 27 para teste. Para verificação da capacidade de extrapolação utilizaram-se 601 municípios positivos correspondentes ao surto mais recente iniciado em 2015 e que se estende até 2018. Finalmente, o recorte temporal 04 também compôs a base com 11.619 atributos, possuindo 345 municípios positivos e 345 negativos, sendo que 276 foram utilizados para treino e 69 para teste. Para a validação da capacidade preditiva foram separados 256 municípios que demonstraram ser positivos em 2018.

Considerando a amplitude de atributos descritivos e o pequeno/mediano conjunto amostral de municípios positivos e pseudonegativos, a pesquisa utilizou o algoritmo *Correlation based Feature Selection* - CFS para a seleção de atributos, que, consequentemente, reduziu a dimensionalidade horizontal das 4 bases de dados (quantidade de atributos selecionados em cada recorte pelo método a ser descrito nos subcapítulos 4.1.1, 4.1.2, 4.1.3 e 4.1.4), tornando a matriz mais equilibrada ao quantitativo de instâncias positivas e pseudonegativas. Foi adotado o método CFS porque a literatura demonstra que sua utilização é recorrente no histórico de publicações e também por apresentar os melhores resultados experimentais para os conjuntos de dados considerados por estas publicações (GNANAMBAL *et al*, 2018; HALL e HOLMES, 2003; LEE, 2005).

O CFS avalia o valor de um subconjunto de atributos considerando a capacidade preditiva

individual de cada atributo, juntamente com o grau de correlação entre eles. Os coeficientes de redundância são usados para estimar a correlação entre o subconjunto de atributos e as classes, bem como as intercorrelações entre os recursos. A relevância de um conjunto de atributos aumenta com a correlação entre os atributos e as classes e diminui com o aumento da intercorrelação entre si (HALL, 1999; KAREGOWDA *et al*, 2010).

Além dos recortes temporais em razão dos surtos epidemiológicos da FAS, ainda foram adotados sub-recortes em função da temporalidade das camadas temáticas, visto que entre os 19 temas utilizados, existem 4 tipos de temporalidades: anuais; quinquenais; atemporais; sem regularidade temporal. Os dados temáticos do tipo anuais correspondem aos temas com intensa dinâmica espaço-temporal e, portanto, possuem o mapeamento sistemático das áreas naturais/antrópicas e dos valores de temperatura/precipitação específicos para cada ano. Os dados quinquenais referem-se aos valores sobre a distribuição e concentração populacional apresentados em intervalos de 5 em 5 anos. Os dados temáticos atemporais correspondem às áreas ou valores sobre as feições naturais ou antropogênicas que não possuem variações ou modificações representativas ao longo do tempo considerado. Os temas sem regularidade temporal representam dados que possuem modificações pontuais e menos intensas no espaço-tempo, e por isso não possuem periodicidade regular de atualizações.

Em função da temporalidade dos dados temáticos, optou-se por avaliar, em caráter experimental e investigativo, sobre os ganhos e perdas da adoção desses sub-recortes, bem como avaliar a competência de conjuntos de dados mais restritos (nos seus aspectos temporais e temáticos) para a construção de modelos e descrição de um fenômeno complexo. Nesse sentido, a partir das 4 bases de dados já descritas, foram organizadas 4 novas bases de dados para cada recorte temporal (tabela 3.5). As 4 novas bases possuem as mesmas instâncias de positivos, negativos e o mesmo conjunto de treino, teste e validação, dos seus respectivos recortes em função dos surtos. A diferença refere-se somente a quantidade de atributos descritivos, já que foram realizados sub-recortes em função dos temas, reduzindo assim o quantitativo de atributos, e consequentemente reduzindo de dimensionalidade das bases.

Tabela 3.5: Quantitativo de atributos, dados de treino, teste e validação para cada recorte temporal e seus respectivos sub-recortes temáticos

1) Recorte temporal: 1999 a 2017							
N	Sub-recorte temporal dos dados	Atributos descritivos	Positivos	Pseudo Negativos	Treino	Teste	Validação (2018)
1.0	Anuais	5186	572	572	458 + / 458 -	114 + / 114 -	256
1.1	Quinquenais	74					
1.2	Atemporais	1838					
1.3	Sem regularidade temporal	4529					

2) Recorte temporal: 1999 a 2003							
N	Sub-recorte temporal dos dados	Atributos descritivos	Positivos	Pseudo Negativos	Treino	Teste	Validação (2007 a 2009)
2.0	Anuais	1262	117	117	94 + / 94 -	23 + / 23 -	133
2.1	Quinquenais	20					
2.2	Atemporais	1838					
2.3	Sem regularidade temporal	4529					

3) Recorte temporal: 2007 a 2009							
N	Sub-recorte temporal dos dados	Atributos descritivos	Positivos	Pseudo Negativos	Treino	Teste	Validação (2015 a 2018)
3.0	Anuais	3152	133	133	106 + / 106 -	27 + / 27 -	601
3.1	Quinquenais	38					
3.2	Atemporais	1838					
3.3	Sem regularidade temporal	4529					

4) Recorte temporal: 2015 a 2017							
N	Sub-recorte temporal dos dados	Atributos descritivos	Positivos	Pseudo Negativos	Treino	Teste	Validação (2018)
3.0	Anuais	5186	345	345	276 + / 276 -	69 + / 69 -	256
3.1	Quinquenais	74					
3.2	Atemporais	1838					
3.3	Sem regularidade temporal	4529					

Por fim, como última alternativa de base de dados empregada nessa pesquisa, em razão dos surtos e dos sub-recortes considerando a temporalidade dos dados temáticos, destaca-se ainda que foram selecionados os atributos indicados pelo método de aprendizagem de máquina J48 em cada sub-recorte e que foram novamente reunidos em 4 novas bases de dados, correspondente a cada um dos 4 recortes temporais. Essa alternativa foi testada para avaliar se a seleção de atributos e a consequente redução de dimensionalidade baseada nos próprios métodos de aprendizagem de máquina foi capaz de trazer contribuições relevantes.

Ao todo foram preparadas 28 bases de dados independentes (APÊNDICE C³), sendo:

- i) 4 bases de dados (1999 a 2017, 1999 a 2003, 2007 a 2009, 2015 a 2017) contendo todos os atributos descritivos;
- ii) 4 bases de dados (1999 a 2017, 1999 a 2003, 2007 a 2009, 2015 a 2017) contendo atributos selecionados pelo método CFS;
- iii) 16 bases de dados correspondentes aos 4 sub-recortes temático temporal (anuais, quinquenais, atemporais, sem regularidade temporal) das 4 bases de dados (1999 a 2017, 1999 a 2003, 2007 a 2009, 2015 a 2017);
- iv) 4 bases de dados (1999 a 2017, 1999 a 2003, 2007 a 2009, 2015 a 2017) correspondentes aos atributos selecionados pelos modelos construídos (método J48) com os sub-recortes temático temporal.

3.6 Modelagem baseada em dados

A modelagem baseada em dados utilizando os métodos de aprendizagem de máquina foi etapa fundamental da pesquisa, na qual foram identificados os fatores e analisadas as possíveis relações dos casos da FAS com as características socioambientais dos municípios.

Os métodos de aprendizagem de máquina e suas técnicas computacionais permitem a construção de modelos complexos e robustos, baseados em grandes bases de dados. Isso possibilita que a escolha dos dados seja oriunda de análises exploratórias, fundamentadas em

³ O Apêndice C é composto por 28 base de dados, disponíveis em formato ARFF (Attribute-Relation File Format - <https://www.cs.waikato.ac.nz/~ml/weka/arff.html>) (HALL et al., 2009), que podem ser consultadas/manipuladas utilizando o software WEKA (<https://www.cs.waikato.ac.nz/ml/weka/>). Sendo que cada uma dessas bases dividi-se em novas base com valores normalizados (“normalized”) e não normalizados (“no_normalized”).

premissas estatísticas, e não somente oriunda de conhecimento prévio de pesquisadores. Como consequência, podem ser explorados os efeitos diretos ou indiretos que os fatores socioambientais podem exercer sobre as espécies de hospedeiros e parasitas, abordagem inédita na investigação da distribuição de zoonoses.

Com as bases de dados organizadas, foi possível a busca por modelos computacionais que determinem correlações, temporais e espaciais, entre as feições naturais e antrópicas dos municípios e os casos humanos e epizootias confirmadas para febre amarela silvestre.

Considerando a complexidade inerente ao fenômeno, explicitada pela amplitude de dados e estatísticas descritivas e pelos diferentes recortes temporais e sub-recortes temáticos, bem como a limitação do tempo para desenvolvimento da tese e escopo/propósito deste documento, optou-se pela utilização de um único método de classificação. Dentre os métodos descritos no subcapítulo 2.6 foi priorizado o método Árvore de Decisão, pois ele atende aos dois principais objetivos da pesquisa (interpretabilidade e acurácia) e também por sua flexibilidade, robustez e velocidade de processamento. A escolha pela utilização de apenas um método deve-se ao objetivo maior dessa pesquisa de construção do conhecimento e obtenção de cenários preditivos acurados. Com certeza a experimentação de diferentes métodos tornaria inevitável a avaliação e a discussão sobre o desempenho dos métodos na modelação do problema, distanciando a pesquisa do aprofundamento necessário sobre a compreensão deste complexo fenômeno. Entretanto, esta discussão torna-se aberta para trabalhos futuros.

Nessa fase da pesquisa utilizaram-se métodos baseados em Árvore de Decisão implementado no *software* de aprendizagem de máquina WEKA (*Waikato Environment for Knowledge Analysis*), desenvolvido pela Universidade de Waikato – Nova Zelândia, e que é gratuito e de código aberto (FRANK *et al*, 2016).

Árvore de Decisão (AD) são uma forma de estruturar uma classificação de dados; elas estabelecem regras para classificação com base em um conjunto de dados que foram supervisionados, observados. Os métodos baseados em árvores de decisão (*decision tree*) criam modelos de classificação ou regressão sob a forma de uma estrutura de árvore na qual cada nó interno corresponde ao teste de um ou mais atributos, cada ramo representa um resultado do teste e os nós folhas representam classes ou distribuições de classes (figura 3.5).

O nó mais elevado da árvore é conhecido como nó raiz, e cada caminho da raiz até um nó folha corresponde a uma regra de classificação. A regra destacada representa um possível caminho que, neste caso, leva à classe 4.

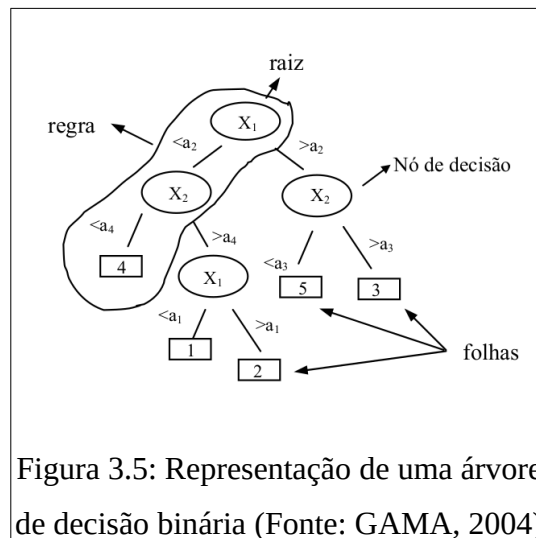


Figura 3.5: Representação de uma árvore de decisão binária (Fonte: GAMA, 2004)

A partir da construção da árvore, ela pode ser usada para classificar um objeto de classe ainda desconhecida. Devem ser testados os valores dos atributos na árvore e percorrê-la até atingir um nó folha, que corresponde a classe predita para aquele objeto. As árvores de decisão possuem vantagens de serem de fácil visualização e compreensão. A facilidade de explicar as classificações propostas está na observação do caminho percorrido de um objeto classificado, sendo possível explicitar porque o modelo fornece cada resposta. (KOHAVI e QUINLAN, 2002).

Foram utilizados os algoritmos J48 (QUINLAN, 1993; QUINLAN, 1996; WU *et al.*, 2007) e *Random Forest* (BREIMAN, 2001; BREIMAN e GHAHRAMANI, 2004), ambos baseados em árvores de decisão. É importante destacar que esses métodos permeiam, respectivamente, duas importantes características de interesse da pesquisa: a interpretabilidade e a acurácia preditiva. Os métodos indicados, (como o algoritmo J48, que possuem como característica a interpretabilidade) destacam-se tanto na etapa de construção por um especialista quanto na etapa de utilização, permitindo que o usuário possa compreender porque foi tomada determinada decisão, de modo que suas decisões possam ser

explicadas pelas hipóteses e dados. Esse tipo de resposta permite abrir novas perspectivas de busca e estudo aos próprios especialistas, propondo muitas vezes associações coerentes, porém desconhecidas.

Importante destacar que, apesar de o ajuste dos parâmetros ser etapa relevante nos experimentos que utilizam métodos de Aprendizagem de Máquina, esta não foi adotada, pois compreende-se que não é o foco da tese. O ajuste de parâmetros exige um abrangente e diverso conjunto de testes e esforços amplos e específicos, devendo-se destacar nesta tese como um ponto importante para evolução em trabalhos futuros.

Assim, foram utilizados os parâmetros padrão (*default*) dos métodos. O método J48 (QUINLAN, 1993) possui os seguintes parâmetros e valores/opções de configuração:

- a) *BinarySplits*: False (para usar divisões binárias em atributos nominais ao construir as árvores);
- b) *ConfidenceFactor*: 0.25 (o fator de confiança usado para a poda; valores menores incorrem em mais podas);
- c) *Debug*: False (se definido como true, o classificador pode gerar informações adicionais para o console);
- d) *MinNumObj*: 2 (o número mínimo de instâncias por folha);
- e) *NumFolds*: 3 (determina a quantidade de dados usados para a remoção com erro reduzido. Uma dobra é usada para poda, o restante para o crescimento da árvore);
- f) *ReducedErrorPruning*: False (se a remoção com erro reduzido é usada em vez da remoção do C 4.5);
- g) *SaveInstanceData*: False (se deseja salvar os dados de treinamento para visualização);
- h) *Seed*: 1 (a semente usada para randomizar os dados quando a remoção com erro reduzido é usada);
- i) *SubtreeRaising*: True (se considerar a subtree aumentando a operação durante a poda);
- j) *Unpruned*: False (se a poda é executada);
- k) *UseLaplace*: False (se as contagens de folhas são suavizadas com base no Laplace).

O único parâmetro modificado foi o *MinNumObj*, tendo sido elevado seu valor de 2 para 10, pois com o valor 2 os nós das árvores de decisão buscam por regras muito específicas, que

em muitos nós correspondem somente as condições particulares de amostras específicas. A fim de evitar a construção de árvores de decisão muito profundas e que retratassem condições específicas ou exceções, optou-se por ampliar a quantidade de amostras por nós para dez, contribuindo assim para construção de modelos mais generalistas e representativos sobre o fenômeno, e não restritos a especificidades sobre o conjunto de dados amostrais.

Para o método *Random Forest* foi utilizada a configuração padrão. O método possui os seguintes parâmetros e valores de configuração (BREIMAN, 2001):

- a) *debug*: False (se definido como true, o classificador pode gerar informações adicionais para o console).
- b) *MaxDepth*: 0 (a profundidade máxima das árvores, 0 para ilimitado).
- c) *NumFeatures*: 0 (o número de atributos a serem usados na seleção aleatória).
- d) *NumTrees*: 100 (o número de árvores a serem geradas).
- e) *Seed*: 1 (a semente numérica aleatória a ser usada).

A implementação J48, implementação WEKA do método C4.5 (QUINLAN, 1993), gera árvores denominadas univariantes. Isso deve-se ao conceito que ele utiliza para testar sua decisão que é avaliar um único atributo por nó. Para selecionar o teste a ser executado em um nó, o algoritmo se orienta em um índice, chamado *information gain ratio* (taxa de ganho de informação), que é calculado através das equações 1, 2, 3 e 4:

$$info(T) = - \sum_{j=1}^k \left[\frac{freq(C_j, T)}{|T|} \times -\log_2 \left(\frac{freq(C_j, T)}{|T|} \right) \right] bits, \quad (1)$$

$$gain(X) = info(T) - \sum_{i=1}^p \left(\frac{|T_i|}{|T|} \times info(T_i) \right) bits, \quad (2)$$

$$split \ info(X) = - \sum_{i=1}^p \left[\frac{|T_i|}{|T|} \times -\log_2 \left(\frac{|T_i|}{|T|} \right) \right] bits, \quad (3)$$

$$gain \ ratio(X) = \frac{gain(X)}{split \ info(X)}, \quad (4)$$

Nas equações 1 a 4, T é o conjunto de dados de treinamento; C são as classes, indo de j até k ; X é uma referência ao atributo testado; e p é o número de subconjuntos gerados pelo teste X . O cálculo tem por objetivo selecionar o teste que utilize a menor quantidade de informação necessária para a classificação. Ou seja, o nó é construído pelo atributo que detém a maior taxa de ganho, e tal atributo será usado no nó para dividir o conjunto de dados de treinamento. Quando o subconjunto a ser testado só contiver exemplos da mesma classe ou todos os exemplos nele contidos obterem a mesma taxa de ganho, nenhum teste é proposto e, então, uma folha é criada.

O *Random Forest* utiliza a técnica de decisão por comitê⁴ (BREIMAN, 2001, FREUND e SCHAPIRE, 1996), dividindo de forma aleatória o conjunto de treinamento em n subconjuntos distintos, sendo cada subconjunto utilizado para construir uma árvore de decisão. Assim sendo, diversas árvores são construídas e, então, uma floresta é formada. Para a construção de suas árvores, ele seleciona aleatoriamente um número de k atributos, considerando k menor que o número total de atributos. Para a seleção do teste a ser realizado no nó, o método utiliza a metodologia empregada no algoritmo CART (BREIMAN *et al*, 1984), que constrói sua árvore de decisão utilizando um critério de seleção que é aplicado recursivamente até que cada nó seja estabelecido na árvore. O critério de seleção utilizado é o Gini, que do mesmo modo que o J48, utiliza a taxa de ganho de informação para reduzir a quantidade de informação necessária para sua classificação. E isso se dá porque normalmente o CART cria uma folha quando não consegue mais determinar um novo teste capaz de satisfazer seus critérios de seleção, assim como já foi visto no J48.

O conjunto de treinamento é dividido gerando os subconjuntos, denotados de i , como já visto anteriormente. A árvore de decisão correspondente a esse subconjunto i é construída nó a nó, de forma que o conjunto de exemplos de i utilizados para a construção do nó é definido pelos k atributos aleatórios, em vez de serem construídos pelos atributos do conjunto original. Por fim, tendo as árvores de decisão treinadas e o comitê pronto, a classe que for definida pela maioria é escolhida como classificação final.

⁴O comitê trata-se de uma comissão formada por árvores treinadas com os mesmos algoritmos e conjuntos de treinamento, sendo assim, a classificação final de um objeto é obtida pela combinação das classificações de cada árvore de decisão participante do comitê.

3.7 Validação dos modelos

Para garantir a confiabilidade dos resultados, os modelos devem ser testados e aprimorados como parte do seu desenvolvimento e aplicação. Dependendo da concordância entre o observado e o processado, o modelo será aceito, rejeitado ou modificado, para novamente ser testado.

Para avaliar a qualidade do modelo gerado é preciso possuir um conjunto independente de dados, podendo ser por meio da coleta de novos dados (trabalho de campo ou levantamento de literatura) ou pela divisão do conjunto de dados em conjuntos de treinamento, teste e validação antes da modelagem propriamente dita. A separação de dados em conjuntos de teste e treinamento é uma parte importante da avaliação de modelos baseado em dados. Em ambos os casos existirão dois conjuntos de dados diferentes, um que gerará o modelo (conjunto de treinamento) e outro que irá testá-lo (conjunto de teste).

Normalmente, quando se particiona um conjunto de dados em um conjunto de treinamento e um conjunto de teste, a maior parte dos dados é usada para treinamento e uma parte menor dos dados é usada para teste. Uma maneira comum de dividir o conjunto de dados (na área de aprendizagem de máquina) é 80% dos dados para gerar o modelo (treino) e 20% para testar o modelo (teste) (HIRZEL e GUIBAN, 2002; FIELDING e BELL, 1997). Deve-se ter claro que a divisão do conjunto de dados dependerá da quantidade do número de parâmetros, pois quanto maior o número de parâmetros maior deverá ser o número de dados para treino.

Diferentes técnicas de validação são utilizadas e descritas em uma grande diversidade de trabalhos que objetivam a construção de modelos, sobretudo nas pesquisas que utilizam métodos de aprendizagem de máquina. Para esta tese destacam-se: Estatística Kappa; Erro Absoluto Médio; Raiz do Erro Quadrático Médio; Erro Relativo Médio; Raiz do Erro Quadrático Relativo; Matriz de confusão; *Precision*; *Recall*; Curva ROC; F-measure. Para todas essas métricas, a obtenção de valores próximos de 1 indicam melhor avaliação de desempenho.

Segundo Landis e Koch (1977), a estatística Kappa é a medida que indica o grau de concordância entre dois classificadores (equação 5), levando em consideração as probabilidades de as concordâncias terem acontecido ao acaso. No caso de haver somente um

classificador, a concordância é calculada entre o conjunto de treinamento, considerado um “classificador infalível”, e o conjunto de teste. Na equação 5 $Pr(a)$ é a concordância observada e $Pr(e)$ é a concordância esperada ao acaso. Ao obter o valor de K , é possível estabelecer uma interpretação semântica acerca do valor obtido de K , conforme explícito na Tabela 3.6.

$$K = \frac{Pr(a) - Pr(e)}{1 - Pr(e)} \quad (5)$$

Tabela 3.6: Interpretação dos Valores da Estatística Kappa
(Fonte: LANDIS e KOCH, 1977)

K	Interpretação
< 0	Nenhuma concordância
0 a 0,2	Leve concordância
0,21 a 0,4	Concordância Regular
0,41 a 0,6	Concordância Moderada
0,61 a 0,8	Concordância Substancial

Para estimar a qualidade de um classificador tem-se, dentre outras, o Erro Absoluto Médio, que fornece uma média do afastamento de todos os valores fornecidos pelos classificadores e o seu valor real (equação 6), onde n é o número de amostras, x_i é o valor fornecido pelo classificador e \bar{x} é a média dos valores de todas as amostras (WEBB, 2011; WILLMOTT e MATSUURA, 2005).

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (6)$$

Outra medida frequentemente utilizada é a Raiz do Erro Quadrático Médio (equação 7). Nessa equação, n é o número de amostras, x_i é o valor fornecido pelo classificador e \bar{x} é a média dos valores de todas as amostras (WILLMOTT e MATSUURA, 2005).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (7)$$

De forma análoga temos o Erro Relativo Médio (equação 8), onde n é o número de amostras, x_i é o valor fornecido pelo classificador, \bar{x} é a média dos valores de todas as amostras e \hat{x} é o valor correto que deve ser fornecido pelo classificador para a amostra em questão (HILL, 2012).

$$RAE = \frac{\sum_{i=1}^n |x_i - \hat{x}|}{\sum_{i=1}^n |x_i - \bar{x}|} \quad (8)$$

Por fim, para estimar a qualidade de um classificador utiliza-se a Raiz do Erro Quadrático Relativo (equação 9). Na equação 9, n é o número de amostras, x_i é o valor fornecido pelo classificador, \bar{x} é a média dos valores de todas as amostras e \hat{x} é o valor correto que deve ser fornecido pelo classificador para a amostra em questão (HILL, 2012).

$$RRSE = \sqrt{\frac{\sum_{i=1}^n (x_i - \hat{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \quad (9)$$

A Matriz de Confusão apresenta o esquema que reúne as possíveis formas de acerto e erro em relação ao que o modelo previu e a distribuição “real”. Considere, como exemplo, um teste laboratorial para a detecção de uma doença. Quando o resultado for positivo, o indivíduo pode ter a doença (verdadeiro positivo) ou pode não tê-la (falso positivo). Da mesma forma, quando o resultado for negativo, o indivíduo pode não ter a doença (verdadeiro negativo) como pode tê-la (falso negativo), que seria algo sério, pois seria informado ao paciente que ele não tem doença, quando na realidade ele a possui. A Matriz de Confusão (tabela 3.7) traz esses valores: verdadeiro positivo; falso positivo; verdadeiro negativo; e o falso negativo, para que assim se saiba os tipos de erros de predição que ocorreram (TING, 2011).

Tabela 3.7: Estrutura de uma Matriz de Confusão

	Positivo	Negativo
Positivo	Verdadeiro Positivo	Falso Negativo
Negativo	Falso Positivo	Verdadeiro Negativo

O *Precision* (P) é a proporção de casos positivos que estão classificados como corretos dentre todas que são positivas (equação 10) (POWERS, 2011).

$$P = \frac{VerdadeiroPositivo}{VerdadeiroPositivo + FalsoPositivo} \quad (10)$$

Já o *Recall* (R), é a proporção de casos positivos que estão classificados como corretos dentre todas que realmente são positivas (equação 11) (POWERS, 2011).

$$R = \frac{VerdadeiroPositivo}{VerdadeiroPositivo + FalsoNegativo} \quad (11)$$

A Curva ROC é uma aproximação gráfica mostrada entre a taxa de verdadeiro positivo (TPR) e a taxa de falso positivo (FPR), onde o falso positivo é representada no eixo x e o eixo y representa o verdadeiro positivo (FAWCETT, 2006). Ela tem por objetivo medir o desempenho de um teste, sendo que um valor acima de 0,7 pode-se considerar como um desempenho satisfatório (Figura 3.6).

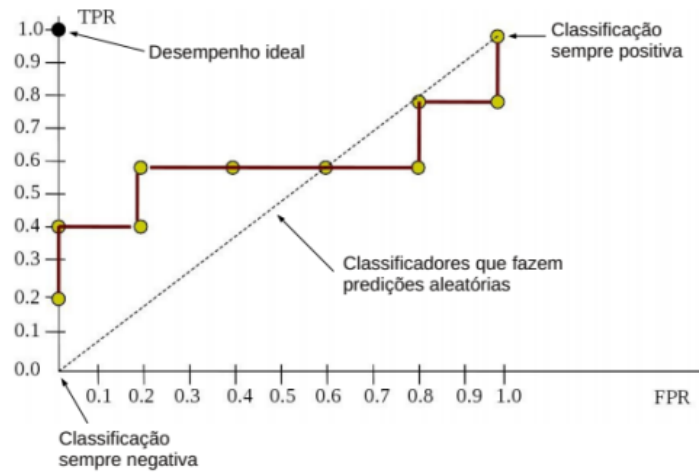


Figura 3.6: Geração da Curva ROC de um classificador
(MERSCHMANN, 2018)

Por fim, temos o F-Measure, que combina as métricas do *Recall* e do *Precision*, expressando a média harmônica. Seu resultado é um indicativo que quanto mais próximo de 1 mais previsões foram assertivas, e quanto mais próximo de 0 mais previsões foram equivocadas (equação 12) (POWERS, 2011).

$$\frac{1}{F} = \frac{1}{2} \left(\frac{1}{R} + \frac{1}{P} \right) = \frac{P + R}{2PR} \quad (12)$$

A validação de modelos é uma área de pesquisa ativa e estritamente necessária métodos alternativos aos testes estatísticos também podem ser bastante eficazes. Aplicar o modelo de volta no campo é o teste mais robusto que se pode fazer com os resultados de modelagem. Outra importante forma de validação é a avaliação do especialista com conhecimento aprofundado do fenômeno/evento em questão. É o que se pode chamar de avaliação biológica, feita através de interpretação visual do especialista (PHILLIPS *et al.*, 2006).

4. RESULTADOS E DISCUSSÃO

4.1 Modelos experimentais

Para obtenção dos modelos experimentais foi utilizado o método J48, sobre todas as 28 bases de dados, e o método *Random Forest*, sobre 12 bases de dados (todas descritas no subcapítulo 3.5). Destaca-se que em concordância com o objetivo definido por esta pesquisa, foram utilizados exclusivamente técnicas de Modelagem Baseada em Dados e seus métodos de Aprendizagem de Máquina. Porém, como possibilidade de continuidade da pesquisa, sugere-se a construção de modelos de previsão para a FAS utilizando as técnicas da Modelagem de Distribuição de Espécies, e a comparação dos resultados obtidos com estas duas metodologias distintas de modelagem.

Ao todo foram produzidos 40 modelos experimentais, cada um contendo 14 métricas de acurácia interna que precisam ser avaliadas, seja para a avaliação da construção do modelo (acurácia interna) quanto da aplicação do modelo para predição (acurácia preditiva), para posterior escolha e indicação dos modelos finais.

O método *Random Forest* não foi empregado sobre as 16 bases de dados correspondentes aos sub-recortes temático temporal dos atributos, pois essas bases foram preparadas com o objetivo de compreender qual a relação, entre a dinâmica de modificação espaço-temporal da paisagem com a emergência da febre amarela, o que só é possível com a utilização de métodos interpretativos, como o J48.

Inicialmente deve-se destacar sobre a maior relevância das métricas que consideram os valores de verdadeiros positivos, falsos positivos e falsos negativos, a saber: *Recall*; *Precision*; *F-Measure* e Curva ROC. Justifica-se esse destaque porque essas métricas ponderam: (i) os Verdadeiros Positivos, municípios onde já ocorreram a doença e o modelo foi capaz de estimar; (ii) os Falsos Negativos, municípios onde ocorreram a doença mas o modelo não foi capaz de estimar, ou seja, deveriam ter sido apontadas pelo modelo (erro crítico de previsão do modelo); (iii) os Falsos Positivos, que são as regiões que nunca ocorreram a manifestação da doença mas que reúnem condições socioambientais favoráveis à manifestação da doença (nesse caso não é possível considerar esse indicativo como erro por

serem justamente os municípios de maior interesse para antecipar a realização de ações de vigilância e vacinação).

Ressalta-se ainda que as métricas indicadas, o *Recall* possui importância ainda maior para avaliação dos modelos, principalmente para o modelos preditivos. Esta métrica considera tanto os Verdadeiros Positivos quanto os Falsos Negativos, ou seja, essa métrica sintetiza justamente o que há de mais relevante para a proposta dessa tese: maior acerto sobre municípios positivos reais e menor erro sobre municípios negativos.

Visto que a modelagem proposta poderá servir de instrumento para direcionar as políticas de vigilância epidemiológica e, principalmente, de priorização de áreas para vacinação, destaca-se que a maior assertividade sobre o número de Verdadeiros Positivos pressupõe maior confiabilidade sobre os novos municípios potenciais de manifestação da doença (Falsos positivos), locais estes que devem ser prioritários para o envio de lotes de vacina para imunização das populações locais, o que representa melhor utilização dos recursos financeiros e ampliação dos esforços das campanhas. Ao mesmo tempo, deve-se minimizar o maior erro sobre o número de Falsos negativos, pois indica que municípios vulneráveis à manifestação da doença não foram identificados pelos modelos e logo não seriam priorizados na distribuição das vacinas, o que pode representar a exposição ao risco da população daquele local.

Além das classes Verdadeiro Positivos, Falso Negativo e Falso Positivo (fundamentais para avaliar a acurácia interna e preditiva dos modelos), ainda existe a classe Verdadeiro Negativo, a qual será indicado em cinza nos mapas (APÊNDICE D). Estes tratam-se dos municípios sem registro de ocorrência da doença e que o modelo também não identificou condições favoráveis para a manifestação da doença, ou seja, regiões que não possuem histórico de manifestação da doença e o modelo sugere que não haverá emergência da doença. Portanto, não devem ser prioritárias as ações de vigilância e vacinação.

As métricas de Erro Absoluto Médio, Raiz do Erro Quadrático Médio, Erro Absoluto Relativo e Raiz do Erro Quadrático Relativo demonstram ser mais ajustadas a outras aplicações, como projeção financeira, por exemplo, onde tanto o falso positivo quanto o falso negativo representam erros graves. Dessa maneira, não serão priorizadas essas quatro métricas para avaliação dos modelos.

4.1.1 Modelos recorte 1999 a 2017

No recorte temporal de 1999 a 2017, que contém todo o histórico de surtos e manifestações da doença, as métricas de acurácia (tabela 4.1) indicam melhor desempenho para os modelos 1.1, 1.4, 1.5 (método J48) e 1.8 e 1.9 (método *Random Forest*). Importante relembrar que no conjunto de treinamento foram utilizadas 916 amostras, igualmente divididas entre positivas e negativas. No conjunto de teste foram utilizadas 228, também igualmente divididas. O número de atributos varia de acordo com o sub-recorte adotado ou com estratégia de seleção de atributos empregada: 1.0 e 1.7 (11.621); 1.1 e 1.8 (78); 1.2 (5.186); 1.3 (74); 1.4 (1.838); 1.5 (4.529); 1.6 e 1.9 (69).

Tabela 4.1: Métricas de validação dos modelos experimentais (J48 e *Random Forest*) do recorte temporal 1999 a 2017

1) Recorte temporal: 1999 a 2017	ACURÁCIA INTERNA											
	Instâncias classificadas corretamente	Instâncias classificadas incorretamente	Estatística Kappa	Erro absoluto médio	Raiz do erro quadrático médio	Erro absoluto relativo	Raiz do erro quadrático relativo	Número total de instâncias	Precision	Recall	F-Measure	Curva ROC
1.0) Todos os atributos (J48)	83,77%	16,23%	0,6754	0,186	0,3484	37,20%	69,68%	228	0,838	0,838	0,838	0,896
1.1) Seleção atributos pelo método CFS (J48)	85,09%	14,91%	0,7018	0,1798	0,3378	35,97%	67,57%	228	0,852	0,851	0,851	0,893
1.2) Sub-recorte temporal dos dados: Anuais (J48)	79,39%	20,61%	0,5877	0,2368	0,4212	47,37%	84,24%	228	0,795	0,794	0,794	0,808
1.3) Sub-recorte temporal dos dados: Quinquenais (J48)	66,67%	33,33%	0,3333	0,4191	0,4544	83,82%	90,89%	228	0,725	0,667	0,644	0,708
1.4) Sub-recorte temporal dos dados: Atemporais (J48)	85,09%	14,91%	0,7018	0,1933	0,3504	38,65%	70,09%	228	0,854	0,851	0,851	0,907
1.5) Sub recorte temporal dos dados: Sem regularidade temporal (J48)	89,91%	10,09%	0,7982	0,1485	0,2925	29,69%	58,50%	228	0,899	0,899	0,899	0,926
1.6) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (J48)	79,39%	20,61%	0,5877	0,2102	0,3685	42,04%	73,71%	228	0,795	0,794	0,794	0,897
1.7) Todos os atributos (Random Forest)	87,72%	12,28%	0,7544	0,2417	0,3142	48,34%	62,83%	228	0,877	0,877	0,877	0,944
1.8) Seleção de atributos pelo método CFS (Random Forest)	93,42%	6,58%	0,8684	0,1715	0,2546	34,31%	50,92%	228	0,934	0,934	0,934	0,976
1.9) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (Random Forest)	93,86%	6,14%	0,8772	0,1718	0,2499	34,37%	49,99%	228	0,94	0,939	0,939	0,978

Entre os modelos construídos utilizando o método J48, o modelo 1.5 (sub-recorte dos dados Sem regularidade temporal) possui o melhor resultado para todas as métricas, tendo classificado corretamente 89,91% (205 instâncias), com valor de 0,899 para as métricas de *Precision*, *Recall*, e F-Measure, e com maior curva ROC, no valor de 0,926. Os modelos 1.1 (atributos selecionados pelo método CFS) e 1.4 (sub-recorte dos dados Atemporais) possuem valores idênticos para maior parte das métricas. Ambos obtiveram 85,09% das instâncias classificadas corretamente, valor de 0,7018 para Estatística Kappa, *Recall* e F-Measure de 0,851. Houve pequena variação (0,02) entre os valores de *Precision* e maior valor de Curva ROC para o modelo 1.4, com 0,907, e 0,893 para o modelo 1.1.

O método *Random Forest* retornou modelos com valores de acurácia elevados, sendo os modelos 1.8 e 1.9 os melhores avaliados. O modelo 1.9 (atributos selecionados pelos modelos construídos com sub-recortes dos dados temáticos) retornou maior percentual de instâncias classificadas corretamente, com 93,86% , maior estatística Kappa dentre todos os 10 modelos obtidos, com 0,8772, assim como para *Precision* (0,94), *Recall* e F-Measure (0,939) e Curva ROC (0,978). O modelo 1.8 (atributos selecionados pelo método CFS) obteve valores bastante próximos do anterior, sendo eleito como o segundo melhor modelo obtido para o recorte de 1999 a 2017, com 93,42% das instâncias classificadas corretamente, Estatística Kappa de 0,8684, e curva ROC de 0,976.

O modelo 1.3 (sub-recorte dados Quinquenais) retornou os piores valores de acurácia, o que pode indicar uma menor relação entre a ocorrência da doença com os padrões de ocupação humanas, já que o histórico de manifestação da doença demonstra sua flexibilidade em emergir em regiões com menor ocupação humana (como estados de Roraima, Amapá, Acre), como em regiões que possuem cidades com as maiores concentrações urbanas do país (São Paulo, Minas Gerais, Rio de Janeiro).

Os demais modelos construídos (1.0, 1.2, 1.6 e 1.7) demonstraram desempenho intermediário, com valores das métricas menores, mas não tão distantes do que as métricas dos descritos como piores, o que não permite descartá-los previamente. Portanto, a partir da verificação de valores de acurácia próximos, a decisão sobre os modelos finais de cada um dos métodos (J48 e *Random Forest*) dependerá da capacidade de generalização, ou seja, a partir do modelo construído (conjunto de regras aprendidas) com a base nos dados de treinamento (e avaliado internamente pelo conjunto de teste), esse modelo será utilizado para predição de novos municípios brasileiros (não empregados na base de treino) com potencialidade para emergência da febre amarela silvestre. Conforme já descrito no subcapítulo 3.7, foi utilizado um conjunto independente e exclusivo de dados para validação da capacidade preditiva do modelo. Neste recorte (1999 a 2017), o conjunto de validação correspondem aos 256 municípios positivos para febre amarela em 2018 (tabela 4.2).

Nessa etapa de avaliação da capacidade preditiva dos modelos é importante ressaltar o significado das classes Verdadeiro positivo e Verdadeiro negativo. Verdadeiro positivo é a classe de concordância entre o real e o previsto, ou seja, que indica o quantitativo de municípios com notificação de ocorrência da doença e que foram identificados pelo modelo como positivos. A classe Verdadeiro negativo indica os municípios sem registro de ocorrência da doença e que o modelo também não identificou condições favoráveis para a manifestação da doença.

Tabela 4.2: Métricas de validação da acurácia preditiva utilizando positivos de 2018 para o recorte temporal 1999 a 2017

1) Recorte temporal: 1999 a 2017	ACURÁCIA PREDITIVA							
	Instâncias classificadas corretamente	Instâncias classificadas incorretamente	Precision	Recall	Falso negativo	Falso positivo	Verdadeiro negativo	Verdadeiro positivo
1.0) Todos os atributos (J48)	56,61%	43,39%	0,0791	0,7930	53	2364	2951	203
1.1) Seleção atributos pelo método CFS (J48)	57,84%	42,16%	0,0822	0,8047	50	2299	3016	206
1.2) Sub-recorte temporal dos dados: Anuais (J48)	57,53%	42,47%	0,0780	0,7617	61	2305	3010	195
1.3) Sub-recorte temporal dos dados: Quinquenais (J48)	79,61%	20,39%	0,0626	0,2461	193	943	4372	63
1.4) Sub-recorte temporal dos dados: Atemporais (J48)	56,11%	43,89%	0,0837	0,8594	36	2409	2906	220
1.5) Sub-recorte temporal dos dados: Sem regularidade temporal (J48)	58,75%	41,25%	0,0839	0,8047	50	2248	3067	206
1.6) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (J48)	93,61%	6,39%	0,0098	0,0039	255	101	5214	1
1.7) Todos os atributos (Random Forest)	53,29%	46,71%	0,0790	0,8594	36	2566	2749	220
1.8) Seleção de atributos pelo método CFS (Random Forest)	53,44%	46,56%	0,0822	0,8984	26	2568	2747	230
1.9) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (Random Forest)	52,38%	47,62%	0,0860	0,9727	7	2646	2669	249

A partir da verificação da acurácia interna dos modelos e da indicação dos modelos com as melhores métricas, realizou-se a avaliação sobre a capacidade preditiva desses modelos. Dentre os modelos indicados com maior acurácia interna, para o método J48, o modelo 1.4 indicou maior valor para as métricas *Precision* (0,08) e *Recall* (0,86), havendo a classificação equivocada de 36 municípios como negativos, mas que tiveram ocorrências de febre amarela (falso negativo), e classificação correta para 220 positivos (verdadeiros positivos). Os modelos 1.1 e 1.5 obtiveram métricas de *Precision* pouco diferentes, 0,082 e 0,084, respectivamente. Ambos retornaram valor de *Recall* igual a 0,8 e indicação de 50 falsos negativos e 206 verdadeiros positivos.

O modelo 1.6, apesar do elevado percentual de instâncias classificadas corretamente (93,61%), retornou valores insignificantes de *Precision* (0,0098) e *Recall* (0,0039). O elevado valor de Verdadeiros negativos (5.214) refletiu no elevado percentual de instâncias classificadas corretamente; porém demonstrou a baixa capacidade desse modelo acertar áreas verdadeiramente positivas, com o elevado valor de Falsos negativos (255) e o ínfimo valor de Verdadeiros positivos (1).

O modelo 1.6 é um exemplo notório sobre a importância de avaliar conjuntamente as métricas (não destacando previamente modelos em função de uma única métrica), mas observar todos os valores e compreender o que cada métrica pode trazer de relevante sobre os modelos. No caso do modelo 1.6, o elevado valor “Instâncias classificadas corretamente” deve-se à ausência de conhecimento sobre municípios verdadeiramente negativos no processo de validação. Destaca-se que estes não estão contemplados pois não existem garantias de que os municípios que não tenham registrado a ocorrência da doença durante o período utilizado para validação (neste caso 2018), não existam subnotificações ou epizootias em locais remotos ou de difícil acesso que não tenham sido verificadas pelas equipes de vigilância. Assim, como temos a certeza somente sobre municípios positivos em 2018, considera-se que essa base de referência está desbalanceada para validação preditiva. Portanto, neste momento de análise da acurácia preditiva, não é coerente indicá-lo como modelo final, pois além do viés existente no percentual de acerto, este não foi destaque dentro da análise de acurácia interna (tabela 8).

Apesar de o modelo 1.6 não atender aos critérios previamente estabelecidos por essa

pesquisa para sua indicação como modelo final (desempenho equilibrado sobre os dois cenários de avaliação – Acurácia interna e Acurácia preditiva; Análise do desempenho de todas as métricas em conjunto, mas com a priorização das métricas *Recall* e *Precision*) sugere-se, como oportunidade de continuidade da pesquisa, que os modelos experimentais não sejam descartados de caráter imediato, mas que sejam sobrepostos/combinados entre si para verificação da recorrência de previsão dos municípios como positivos, permitindo assim avançar na obtenção de um modelo final contendo a hierarquização de municípios como mais favoráveis a serem positivos.

Entre os modelos construídos pelo método *Random Forest*, destaca-se que o modelo 1.9 obteve maiores valores para as métricas de *Precision* (0,086), *Recall* (0,97), com a indicação de apenas 7 falsos negativos e indicação de 249 verdadeiros positivos. O modelo 1.8 teve desempenho inferior, retornando valores inferiores para *Precision* (0,082), *Recall* (0,89) e classificação de 26 municípios como falsos negativos. Dessa maneira, determina-se que modelo 1.4 (método J48) e 1.9 (método *Random Forest*) são os modelos com os melhores desempenhos, para o recorte 1999 a 2017, à medida que foram capazes de combinar relevantes valores de acurácia interna com melhores indicadores de acurácia preditiva.

4.1.2 Modelos recorte 1999 a 2003

No recorte temporal de 1999 a 2003, correspondente ao primeiro surto da febre amarela silvestre, as métricas de acurácia (tabela 4.3) indicaram resultados consistentes para todos os recortes, sendo destaque os modelos 2.0, 2.1, 2.5, 2.6 (método J48) e 2.7, 2.8 e 2.9 (método *Random Forest*). Deve-se destacar que no conjunto de treinamento foram utilizados 188 instâncias, divididas em positivas e negativas, no conjunto de teste foram utilizadas 46, igualmente divididas entre positivas e negativas. O número de atributos também é diferente em função do sub-recorte temático temporal adotado ou do método de seleção de atributos: 2.0 e 2.7 (7.643); 2.1 e 2.8 (50); 2.2 (1.262); 2.3 (20); 2.4 (1.838); 2.5 (4529); 2.6 e 2.9 (10).

Tabela 4.3: Métricas de validação dos modelos experimentais (J48 e *Random Forest*) do recorte temporal 1999 a 2003

2) Recorte temporal: 1999 a 2003	ACURÁCIA INTERNA											
	Instâncias classificadas corretamente	Instâncias classificadas incorretamente	Estatística Kappa	Erro absoluto médio	Raiz do erro quadrático médio	Erro absoluto relativo	Raiz do erro quadrático relativo	Número total de instâncias	Precision	Recall	F-Measure	Curva ROC
2.0) Todos os atributos (J48)	93,48%	6,52%	0,8696	0,088	0,2541	17,60%	50,83%	46	0,942	0,935	0,935	0,929
2.1) Seleção atributos pelo método CFS (J48)	93,48%	6,52%	0,8696	0,0894	0,2574	17,88%	51,48%	46	0,942	0,935	0,935	0,929
2.2) Sub-recorte temporal dos dados: Anuais (J48)	86,96%	13,04%	0,7391	0,135	0,3391	27,00%	67,81%	46	0,881	0,87	0,869	0,888
2.3) Sub-recorte temporal dos dados: Quinquenais (J48)	84,78%	15,22%	0,6957	0,2305	0,354	46,11%	70,81%	46	0,865	0,848	0,846	0,848
2.4) Sub-recorte temporal dos dados: Atemporais (J48)	91,30%	8,70%	0,8261	0,1169	0,2865	23,38%	57,30%	46	0,926	0,913	0,912	0,924
2.5) Sub-recorte temporal dos dados: Sem regularidade temporal (J48)	93,48%	6,52%	0,8696	0,0894	0,2574	17,88%	51,48%	46	0,942	0,935	0,935	0,929
2.6) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (J48)	93,48%	6,52%	0,8696	0,0894	0,2574	17,88%	51,48%	46	0,942	0,935	0,935	0,929
2.7) Todos os atributos (Random Forest)	93,48%	6,52%	0,8696	0,1741	0,2559	34,83%	51,17%	46	0,942	0,935	0,935	0,985
2.8) Seleção de atributos pelo método CFS (Random Forest)	93,48%	6,52%	0,8696	0,1133	0,2645	22,65%	52,90%	46	0,942	0,935	0,935	0,967
2.9) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (Random Forest)	93,48%	6,52%	0,8696	0,098	0,2553	19,61%	51,05%	46	0,942	0,935	0,935	0,965

Dentre os modelos construídos utilizando o método J48, os modelos 2.0 (todos os atributos), 2.1 (atributos selecionados pelo método CFS), 2.5 (sub-recorte dos dados Sem regularidade temporal) e 2.6 (atributos selecionados pelos modelos construídos com sub-recortes dos dados temáticos) retornaram valores idênticos para as seguintes métricas: 93,48% das instâncias classificadas corretamente; Estatística Kappa de 0,08696; *Recall* e *F-Measure* com valor de 0,935; *Precision* 0,942 e Curva ROC de 0,929.

O método *Random Forest* retornou modelos com os mesmos valores de acurácia dos modelos obtidos com o método J48, com exceção das métricas de erro (mas que não são determinantes para essa análise) e da Curva ROC, que retornou valores maiores para os modelos 2.7, 2.8 e 2.9, correspondendo, respectivamente aos valores: 0,985; 0,967 e 0,965.

Os demais modelos construídos (2.2, 2.3, 2.4) demonstraram desempenho razoável e, portanto, também não foram descartados antecipadamente. Nesse recorte (1999 a 2003) o conjunto de validação da acurácia preditiva também é independente aos dados de treinamento e teste, e corresponde aos 133 municípios positivos do surto seguinte, ou seja, no período de 2007 a 2009. A avaliação das métricas de acurácia preditiva (tabela 4.4) que direciona a escolha dos modelos finais.

Tabela 4.4: Métricas de validação da acurácia preditiva utilizando positivos de 2007 a 2009 para o recorte temporal 1999 a 2003

2) Recorte temporal: 1999 a 2003	ACURÁCIA PREDITIVA							
	Instâncias classificadas corretamente	Instâncias classificadas incorretamente	Precision	Recall	Falso negativo	Falso positivo	Verdadeiro negativo	Verdadeiro positivo
2.0) Todos os atributos (J48)	74,40%	25,60%	0,0544	0,5940	54	1372	4066	79
2.1) Seleção atributos pelo método CFS (J48)	75,12%	24,88%	0,0437	0,4511	73	1313	4125	60
2.2) Sub-recorte temporal dos dados: Anuais (J48)	76,38%	23,62%	0,0317	0,3008	93	1223	4215	40
2.3) Sub-recorte temporal dos dados: Quinquenais (J48)	78,42%	21,58%	0,0497	0,4436	74	1128	4310	59
2.4) Sub-recorte temporal dos dados: Atemporais (J48)	73,16%	26,84%	0,0392	0,4361	75	1420	4018	58
2.5) Sub-recorte temporal dos dados: Sem regularidade temporal (J48)	75,12%	24,88%	0,0437	0,4511	73	1313	4125	60
2.6) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (J48)	73,74%	26,26%	0,0376	0,4060	79	1384	4054	54
2.7) Todos os atributos (Random Forest)	50,69%	49,31%	0,0365	0,7744	30	2717	2721	103
2.8) Seleção de atributos pelo método CFS (Random Forest)	51,68%	48,32%	0,0432	0,9098	12	2680	2758	121
2.9) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (Random Forest)	46,98%	53,02%	0,0256	0,5714	57	2897	2541	76

A avaliação da acurácia preditiva indicou que para o método J48, o modelo 2.0 retornou maior valor de *Precision* ($\sim 0,05$) e *Recall* ($\sim 0,59$), menor erro no total de falsos negativos (54) e maior acerto no total de verdadeiros positivos (79). Enquanto que os modelos 2.1 e 2.5 retornaram valores de *Precision* ($\sim 0,04$), *Recall* ($\sim 0,45$) e Verdadeiros Positivos (60) e maior quantidade de falsos negativos (73) (quando comparados com o modelo 2.0). Os modelos 2.3 e 2.4 demonstram pequenas variações, quando comparadas com as métricas dos modelos 2.1 e 2.5, (*Precision* de $\sim 0,05$ e $\sim 0,04$; *Recall* de $\sim 0,44$ e $\sim 0,43$; Falso negativo de 74 e 75 e Verdadeiros positivos com 59 e 58). Enquanto que os modelos 2.2 e 2.6 apresentam-se com o menor valor de *Precision* ($\sim 0,03$) e Verdadeiros positivos (40 e 54), e maior erro na indicação de negativos, com 93 e 79 falsos negativos.

Já entre os modelos que utilizam o método *Random Forest*, o modelo 2.8 obteve maior valor de *Precision* (0,043), *Recall* ($\sim 0,9$) e verdadeiros positivos (121) e menor erro sobre negativos, com apenas 12 falsos negativos. Os modelos 2.7 e 2.9 retornaram valores inferiores de *Precision* (0,036 e 0,025), *Recall* ($\sim 0,77$ e $\sim 0,57$) e Verdadeiros positivos (103 e 76), e maior erro para falsos negativos (30 e 57). Assim, determina-se que os modelos 2.0 (método J48) e 2.8 (*Random Forest*) obtiveram melhor desempenho, no comparativo entre os modelos do recorte 1999 a 2003.

4.1.3 Modelos recorte 2007 a 2009

No recorte temporal de 2007 a 2009, correspondente ao segundo surto da febre amarela silvestre, as métricas de acurácia (tabela 4.5) demonstram melhores resultados para os modelos 3.4 e 3.5 (método J48) e 3.7 e 3.8 (método *Random Forest*). No conjunto de treinamentos foram utilizadas 212 amostras (positivas e negativas) e para teste 54, sendo 27 positivas e 27 negativas. O número de atributos nos modelos 3.0 e 3.7 e de 9.551, nos modelos 3.1 e 3.8 (seleção de atributos pelo método CFS) utilizaram-se 54 atributos, no modelos 3.6 e 3.9 (atributos selecionados pelos modelos construídos com sub-recorte dos dados) utilizaram 17, modelo 3.2 utilizou 3.152, modelo 3.3 com 38, modelo 3.4 com 1.838 e modelo 3.5 com 4.529.

Tabela 4.5: Métricas de validação dos modelos experimentais (J48 e *Random Forest*) do recorte temporal 2007 a 2009

3) Recorte temporal: 2007 a 2009	ACURÁCIA INTERNA											
	Instâncias classificadas corretamente	Instâncias classificadas incorretamente	Estatística Kappa	Erro absoluto médio	Raiz do erro quadrático médio	Erro absoluto relativo	Raiz do erro quadrático relativo	Número total de instâncias	Precision	Recall	F-Measure	Curva ROC
3.0) Todos os atributos (J48)	81,48%	18,52%	0,6296	0,1759	0,3556	35,19%	71,13%	54	0,815	0,815	0,815	0,917
3.1) Seleção atributos pelo método CFS (J48)	81,48%	18,52%	0,6296	0,2046	0,3783	40,93%	75,67%	54	0,831	0,815	0,813	0,849
3.2) Sub-recorte temporal dos dados: Anuais (J48)	81,48%	18,52%	0,6296	0,2046	0,3783	40,93%	75,67%	54	0,831	0,815	0,813	0,849
3.3) Sub-recorte temporal dos dados: Quinquenais (J48)	81,48%	18,52%	0,6296	0,2915	0,382	58,31%	76,39%	54	0,831	0,815	0,813	0,875
3.4) Sub-recorte temporal dos dados: Atemporais (J48)	87,04%	12,96%	0,7407	0,1786	0,3582	35,72%	71,64%	54	0,875	0,87	0,87	0,871
3.5) Sub-recorte temporal dos dados: Sem regularidade temporal (J48)	85,19%	14,81%	0,7037	0,1779	0,3572	35,59%	71,43%	54	0,854	0,852	0,852	0,87
3.6) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (J48)	81,48%	18,52%	0,6296	0,2046	0,3783	40,93%	75,67%	54	0,831	0,815	0,813	0,849
3.7) Todos os atributos (Random Forest)	88,89%	11,11%	0,7778	0,1809	0,2599	36,19%	51,99%	54	0,898	0,889	0,888	0,988
3.8) Seleção de atributos pelo método CFS (Random Forest)	88,89%	11,11%	0,7778	0,1291	0,2432	25,81%	48,64%	54	0,898	0,889	0,888	0,981
3.9) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (Random Forest)	87,04%	12,96%	0,7407	0,1507	0,3065	30,15%	61,30%	54	0,875	0,87	0,87	0,957

O modelo 3.4 indicou 87,04% de instâncias classificadas corretamente, ~0,74 para Estatística Kappa, ~0,87 para *Precision*, *Recall*, F-Measure e Curva ROC com ~0,87. O modelo 3.5 teve 85,19% das instâncias classificadas corretamente, Kappa de ~0,7, valores de *Precision*, *Recall* e F-Measure de ~0,85 e Curva ROC de ~0,87. Para o método *Random Forest* os modelos 3.7 e 3.8 retornaram valores idênticos para todas as métricas: 88,89% de instâncias classificadas corretamente, Kappa de 0,778, valores de *Precision*, *Recall*, F-Measure de 0,8. Os demais modelos construídos (3.0, 3.1, 3.2, 3.3 e 3.6) demonstraram desempenhos inferiores para o método J48, mas com valores idênticos entre si (com exceção das métricas de *Precision* e F-Measure para o 3.0 e Curva ROC diferente entre todos). O modelo 3.9 (método *Random Forest*) também demonstrou desempenho inferior, quando comparado com os modelos 3.7 e 3.8, para todas as métricas. A decisão sobre os modelos finais para esse recorte seguiram a mesma lógica dos recortes anteriores, e dependem da avaliação da acurácia preditiva. Nesse recorte (2007 a 2009) o conjunto de validação da acurácia preditiva é independente

aos dados de treinamento e teste, e corresponde aos 601 municípios positivos do recente surto, período de 2015 a 2018 (tabela 4.6) .

Tabela 4.6: Métricas de validação da acurácia preditiva utilizando positivos de 2015 a 2018 para o recorte temporal 2007 a 2009

3) Recorte temporal: 2007 a 2009	ACURÁCIA PREDITIVA							
	Instâncias classificadas corretamente	Instâncias classificadas incorretamente	Precision	Recall	Falso negativo	Falso positivo	Verdadeiro negativo	Verdadeiro positivo
3.0) Todos os atributos (J48)	57,15%	42,85%	0,0946	0,4110	311	2076	2967	217
3.1) Seleção atributos pelo método CFS (J48)	61,39%	38,61%	0,0423	0,1420	453	1698	3345	75
3.2) Sub-recorte temporal dos dados: Anuais (J48)	59,56%	40,44%	0,0390	0,1383	455	1798	3245	73
3.3) Sub-recorte temporal dos dados: Quinquenais (J48)	54,42%	45,58%	0,0716	0,3182	360	2179	2864	168
3.4) Sub-recorte temporal dos dados: Atemporais (J48)	58,36%	41,64%	0,0778	0,3125	363	1957	3086	165
3.5) Sub-recorte temporal dos dados: Sem regularidade temporal (J48)	53,98%	46,02%	0,0484	0,2064	419	2145	2898	109
3.6) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (J48)	79,97%	20,03%	0,0664	0,0852	483	633	4410	45
3.7) Todos os atributos (Random Forest)	46,51%	53,49%	0,0612	0,3239	357	2623	2420	171
3.8) Seleção de atributos pelo metodo CFS (Random Forest)	45,97%	54,03%	0,0549	0,2898	375	2635	2408	153
3.9) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (Random Forest)	40,82%	59,18%	0,0368	0,2083	418	2879	2164	110

As métricas de acurácia preditiva indicam que para o método J48 o modelo 3.4 obteve segundo melhor desempenho, com *Precision* de 0,078, *Recall* de ~0,31, 363 falsos negativos e 165 Verdadeiros positivos. O melhor desempenho foi obtido com o modelo 3.0, quando considerados os valores de *Precision* (~0,09), *Recall* (~0,41), 311 Falsos negativos e 217 Verdadeiros positivos. Importante destacar que apesar do melhor desempenho nas métricas de acurácia preditiva, o modelo 3.0 obteve acurácia interna (tabela 4.5) significativamente inferior ao modelo 3.4 e 3.5. Portanto, por este não atender ao compromisso de equilibrar coerência interna e capacidade preditiva, este não foi selecionado como modelo final para o recorte 2007 a 2009.

Situação idêntica – no que se refere à seleção do modelo final para o recorte 01 – também ocorreu para o modelo 3.6 que apresentou elevado percentual de acerto (também influenciado pelo viés do número elevado de municípios negativos), porém igualmente apresentou métricas de *Recall* e *Precision* significativamente inferiores aos demais modelos. Assim, este não foi selecionado como modelo final.

Os modelos do método *Random Forest* não retornaram valores consistentes de acurácia preditiva, o que evidencia a forte alteração no padrão de ocorrência da doença, já que o método em questão é reconhecidamente robusto na obtenção de modelos com maior acurácia preditiva. No período utilizado para treinamento (2007 a 2009) as ocorrências de febre amarela concentraram-se no Centro-Oeste, Noroeste de Minas Gerais, Sul de São Paulo e, principalmente, no Rio Grande do Sul, já no período utilizado para validação preditiva (2015 a 2018), as ocorrências ampliaram-se sobre todos os estados do Sudeste, ocorreram de forma mais pontual no Centro Oeste, e não houveram registros da doença no Rio Grande do Sul. Sabendo que os métodos aprendem com os dados de treinamento e estimam novas regiões com base nos padrões identificados, compreende-se a perda de acurácia preditiva desses métodos, já que os padrões socioambientais (tipo de vegetação, ocupação antrópica, geomorfologia, etc) possuem diferenças expressivas entre os locais de ocorrência entre um surto e outro.

Visto que todos os modelos, para ambos os métodos, não obtiveram métricas relevantes de acurácia preditiva, considera-se que o mais coerente é indicar como modelos finais para esse recorte, os modelos 3.4 e 3.7. O modelo 3.4 demonstrou maior acurácia interna, o que

permite avançar na interpretação sobre os padrões do próprio surto, sem a perspectiva de prospecção futura.

Já no caso do método *Random Forest*, justifica-se a escolha do modelo 3.7 pelo menor erro de falsos negativos, o que decide a situação de empate de acurácia interna com o modelo 3.8. Apesar da escolha do modelo 3.7, deve-se assumir de antemão a deficiência deste cenário preditivo.

4.1.4 Modelos recorte 2015 a 2017

O recorte temporal de 2015 a 2017, correspondente ao recente surto, obteve maiores valores de acurácia para os modelos 4.4, 4.0, 4.2, 4.5 e 4.6 (método J48) e para o modelo 4.8 (método *Random Forest*) (tabela 4.7). O treinamento foi realizado com 552 amostras e o teste 138. A quantidade de atributos também variou de acordo com o recorte, sendo para os modelos: 4.0 e 4.7 11621 atributos; 4.1 e 4.8 = 64; 4.2 = 5.186; 4.3 = 74; 4.4 = 1.838 e 4.5 = 4.529.

Tabela 4.7: Métricas de validação dos modelos experimentais (J48 e *Random Forest*) do recorte temporal 2015 a 2017

4) Recorte temporal: 2015 a 2017	ACURÁCIA INTERNA											
	Instâncias classificadas corretamente	Instâncias classificadas incorretamente	Estatística Kappa	Erro absoluto médio	Raiz do erro quadrático médio	Erro absoluto relativo	Raiz do erro quadrático relativo	Número total de instâncias	Precision	Recall	F-Measure	Curva ROC
4.0) Todos os atributos (J48)	86,23%	13,77%	0,7246	0,1763	0,3338	35,25%	66,75%	138	0,862	0,862	0,862	0,895
4.1) Seleção atributos pelo metodo CFS (J48)	84,06%	15,94%	0,6812	0,1912	0,3576	38,23%	71,52%	138	0,842	0,841	0,84	0,886
4.2) Sub-recorte temporal dos dados: Anuais (J48)	86,23%	13,77%	0,7246	0,1942	0,3402	38,84%	68,04%	138	0,866	0,862	0,862	0,898
4.3) Sub-recorte temporal dos dados: Quinquenais (J48)	61,59%	38,41%	0,2319	0,437	0,494	87,40%	98,81%	138	0,622	0,616	0,611	0,643
4.4) Sub-recorte temporal dos dados: Atemporais (J48)	88,41%	11,59%	0,7681	0,1624	0,3057	32,47%	61,13%	138	0,885	0,884	0,884	0,915
4.5) Sub-recorte temporal dos dados: Sem regularidade temporal (J48)	86,23%	13,77%	0,7246	0,19	0,3318	38,00%	66,36%	138	0,862	0,862	0,862	0,889
4.6) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (J48)	86,23%	13,77%	0,7246	0,1728	0,3385	34,56%	67,70%	138	0,863	0,862	0,862	0,885
4.7) Todos os atributos (Random Forest)	89,86%	10,14%	0,7971	0,2216	0,2949	44,32%	58,97%	138	0,902	0,899	0,898	0,961
4.8) Seleção de atributos pelo metodo CFS (Random Forest)	91,30%	8,70%	0,8261	0,1343	0,2518	26,87%	50,35%	138	0,916	0,913	0,913	0,98
4.9) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (Random Forest)	90,58%	9,42%	0,8116	0,1391	0,2474	27,83%	49,47%	138	0,906	0,906	0,906	0,975

O modelo 4.4 retornou melhor desempenho entre os modelos produzidos pelo método J48, sendo 88,41% das instâncias classificadas corretamente, Kappa de 0,7681, *Precision*, *Recall* e F-Measure de ~0,88 e curva ROC com 0,915. Os modelos 4.0, 4.2, 4.5 e 4.6 retornaram valores idênticos para maior parte das métricas de interesse, sendo 86,23% das instâncias classificadas corretamente, Kappa de 0,7246, *Precision*, *Recall*, F-Measure com ~0,86 e Curva ROC de ~0,8 com pequenas variações decimais. Entre os modelos produzidos pelo método *Random Forest* o modelo 4.8 obteve pequena vantagem, sobre os modelos 4.7 e 4.9, nos valores de acurácia, com: 91,30% de acerto na classificação das instâncias; Kappa de 0,8261; *Precision*, *Recall* e F-Measure com ~0,91 e Curva ROC de ~0,98.

O modelo 4.1 demonstrou desempenho pouco inferior aos demais produzidos pelo método J48, enquanto que o modelo 4.3 retornou valores

baixos em todas as métricas. Essa perda de acurácia no modelo 4.3 deve-se a baixa capacidade de representação dessas variáveis (densidade e contagem da população), de maneira isolada, na compreensão dos padrões de manifestação da doença nesse período, já que ocorreram epidemias e epizootias da doença tanto em municípios interioranos quanto em municípios metropolitanos dos estados de Rio de Janeiro, São Paulo e Minas Gerais. Os modelos finais para esse recorte também serão determinados pela avaliação da acurácia preditiva, e utilizaram os 256 municípios positivos no ano de 2018 (tabela 4.8).

Tabela 4.8: Métricas de validação da acurácia preditiva utilizando positivos de 2018 para o recorte temporal 2015 a 2017

4) Recorte temporal: 2015 a 2017	ACURÁCIA PREDITIVA							
	Instâncias classificadas corretamente	Instâncias classificadas incorretamente	Precision	Recall	Falso negativo	Falso positivo	Verdadeiro negativo	Verdadeiro positivo
4.0) Todos os atributos (J48)	60,29%	39,71%	0,0935	0,8789	31	2181	3134	225
4.1) Seleção atributos pelo metodo CFS (J48)	62,32%	37,68%	0,0924	0,8164	47	2052	3263	209
4.2) Sub-recorte temporal dos dados: Anuais (J48)	70,85%	29,15%	0,1162	0,8086	49	1575	3740	207
4.3) Sub-recorte temporal dos dados: Quinquenais (J48)	39,42%	60,58%	0,0606	0,8398	41	3334	1981	215
4.4) Sub-recorte temporal dos dados: Atemporais (J48)	60,73%	39,27%	0,0958	0,8945	27	2161	3154	229
4.5) Sub-recorte temporal dos dados: Sem regularidade temporal (J48)	56,20%	43,80%	0,0892	0,9258	19	2421	2894	237
4.6) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (J48)	77,74%	22,26%	0,0802	0,3672	162	1078	4237	94
4.7) Todos os atributos (Random Forest)	53,47%	46,53%	0,0870	0,9609	10	2582	2733	246
4.8) Seleção de atributos pelo metodo CFS (Random Forest)	54,30%	45,70%	0,0908	0,9922	2	2544	2771	254
4.9) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos (Random Forest)	53,72%	46,28%	0,0839	0,9141	22	2556	2759	234

As métricas de acurácia preditiva indicam que, para o método J48, o modelo 4.5 obteve melhor desempenho, visto que retornou maior valor de *Recall* (~0,92) e de Verdadeiros positivos (237). O valor do *Precision* (0,089) foi menor do que nos modelos 4.0, 4.1 e 4.4, o que é reflexo do elevado número de falsos positivos (2.421). O modelo 4.5 é indicado como final, pois para essa pesquisa é prioritária a obtenção de um número menor de falsos negativos. Pois a indicação de áreas negativas, que não são verdadeiramente negativas, é pior do que a indicação equivocada de áreas positivas, já que não existem garantias de que num futuro próximo possa ocorrer a emergência da doença. Para o método *Random Forest*, o modelo 4.8 obteve melhor desempenho para todas as métricas, com destaque para o menor erro de Falsos negativos, com apenas 2 municípios, e maior acerto de Verdadeiros positivos (254).

Por fim, no que se refere sobre a indicação de modelos finais, destacam-se como oportunidades de continuidade da pesquisa duas importantes propostas: (i) validação baseada não somente no surto seguinte, mas baseada em todos os municípios positivos em todos os surtos seguintes (visto que a previsão proposta não se limita ao ano seguinte específico); (ii) elaboração de um modelo final que contemple a sobreposição de todos os municípios previstos como positivos, para produção de um modelo final que contenha a hierarquização sobre os municípios mais favoráveis à emergência da doença.

4.2 Principais aspectos sobre os dados/variáveis temáticas indicadas nos modelos experimentais

Conforme descrito no subcapítulo 4.1, o objetivo da construção de todos esses modelos é tanto compreender os fatores ambientais e/ou sociais relacionados com a emergência da febre amarela quanto indicar, antecipadamente, com maior acurácia interna e preditiva possível, novas regiões oportunas para a manifestação da doença. Por isso a escolha dos métodos J48 e *Random Forest*.

Para Sabroza *et al* (1992) há de se considerar que o sucesso do controle da doença está em identificar fatores ambientais e comportamentais que explicam como vão sendo produzidos os processos endêmicos e/ou epidêmicos, em conhecer a distribuição dos vetores,

reservatórios e hospedeiros, em identificar comportamentos de riscos, e em identificar e proteger o grupo de suscetíveis. Assim, para a compreensão dos elementos favoráveis ou desfavoráveis à ocorrência da doença foram discutidos, principalmente, o impacto/influência desses elementos sobre a manutenção ou migração de espécies de primatas, vetores e vírus.

Nesse sentido, após a indicação dos modelos finais, baseada na combinação de acurácia interna e preditiva, considera-se relevante nesse momento discutir, numa abordagem mais geral, sobre os dados indicados pelos modelos (método J48) como mais relevantes e relacionados com a explicação/compreensão do fenômeno. As métricas de acurácia interna dos modelos interpretativos (método J48) foram, em sua maioria, semelhantemente consistentes, o que nos dá certo fundamento para avançar na avaliação sobre essas relações dos dados temáticos com a doença.

A partir da observação das 28 árvores de decisão correspondentes a cada um dos 7 modelos experimentais do método J48, para cada um dos 4 recortes temporais (APÊNDICE D), verificou-se que existe recorrência na indicação de alguns atributos descritivos. A fim de evitar a construção de um texto repetitivo, por consequência exaustivo, pela discriminação de cada um dos 216 atributos descritivos indicados pelas 28 árvores, optou-se por analisar e discutir os atributos de acordo com as 68 classes ou variáveis temáticas dos dados aos quais pertencem e com a frequência de indicação nas árvores de decisão (tabela 4.9). Considerando a multidisciplinaridade desta pesquisa, justifica-se ainda a necessidade de contextualizar cada variável evidenciada e indicar possíveis os interrelacionamentos, proporcionando assim uma análise abrangente e integrada da pesquisa. Podendo esta análise servir ainda como um filtro de seleção dos dados para continuidade da pesquisa.

Portanto, neste primeiro momento, não serão discutidas as especificidades sobre o valor e o tipo de estatística, nem tampouco a particularidade temporal de cada um dos atributos. Considera-se que este aprofundamento sobre o valor de separação, o tipo de informação estatística associada ao dado e a discretização temporal (ano ou período) de cada atributo seja mais adequado para as árvores de decisão dos modelos finais indicados para cada um dos respectivos recortes.

Tabela 4.9: Frequência de indicação dos dados ou variáveis temáticas nas árvores de decisão

		Recorte 1999 a 2017	Recorte 1999 a 2003	Recorte 2007 a 2009	Recorte 2015 a 2017	
Id	Dado ou Variável temática	Frequência	Frequência	Frequência	Frequência	Frequência acumulada
1	Contagem população	2	1	4	5	12
2	Callithrix jacchus	5	4	1	1	11
3	Formações Florestais	10			1	11
4	Pastagem	9			2	11
5	Agricultura	5		4	1	10
6	Densidade população	4		3	3	10
7	Formações Savânicas	5		4	1	10
8	Sazonalidade da Temperatura – bio4	2	1	1	5	9
9	Precipitação total mensal – Outubro – prec10	8				8
10	Corpos d’água	5			2	7
11	Callithrix penicillata	2			4	6
12	Precipitação do trimestre mais frio – bio19	6				6
13	Unidades de conservação	2			4	6
14	Alouatta caraya	4		1		5
15	Florestas Plantadas	3			2	5
16	Precipitação total mensal – Dezembro – prec12	5				5
17	Alouatta guariba	1	3			4
18	Alouatta macconnelli	1			3	4
19	Callicebus personatus	4				4
20	Mineração – Água mineral	3				3

21	Precipitação Anual – bio12	2	1			3
22	Precipitação total mensal – Fevereiro – prec2	1			2	3
23	Sazonalidade da Precipitação – bio15	3				3
24	Áreas úmidas naturais não florestais	2				2
25	Infraestrutura Urbana	1		1		2
26	Isotermalidade – bio3	1			1	2
27	Latossolo bruno Distrófico	2				2
28	Mineração – Bauxita	1			1	2
29	Neossolo litólico Eutrófico	2				2
30	Patamares	2				2
31	Precipitação do trimestre mais chuvoso – bio16	2				2
32	Precipitação do trimestre mais quente – bio18	2				2
33	Precipitação total mensal – Agosto – prec8	1	1			2
34	Precipitação total mensal – Novembro – prec11	1			1	2
35	Serras	2				2
36	Temperatura média mensal – Agosto – tmean8	2				2
37	Tropical Brasil Central mesotérmico brando média entre 10 e 15 C úmido e 3 meses secos	2				2

38	Vegetação Campestre	2				2
39	Agricultura ou Pastagem	1				1
40	Alouatta belzebul	1				1
41	Altitude	1				1
42	Amplitude térmica anual – bio7			1		1
43	Aotus trivirgatus	1				1
44	Bioma Amazônia	1				1
45	Bioma Caatinga		1			1
46	Bioma Mata Atlântica			1		1
47	Callicebus moloch	1				1
48	Callicebus nigrifrons	1				1
49	Equatorial quente média 18 C em todos os meses úmido e de 1 a 2 meses secos	1				1
50	Estepe Gramíneo Lenhosa	1				1
51	Ferrovias	1				1
52	Latossolo vermelho Distrófico			1		1
53	Mineração - Areia	1				1
54	Mineração – Caulim	1				1
55	Mineração – Ferro	1				1
56	Mineração – Granito	1				1
57	Mineração – Manganês	1				1
58	Mineração – Saibro	1				1
59	Planaltos	1				1
60	Praias e dunas	1				1
61	Precipitação do mês mais chuvoso – bio13	1				1
62	Precipitação total mensal – Julho – prec7	1				1

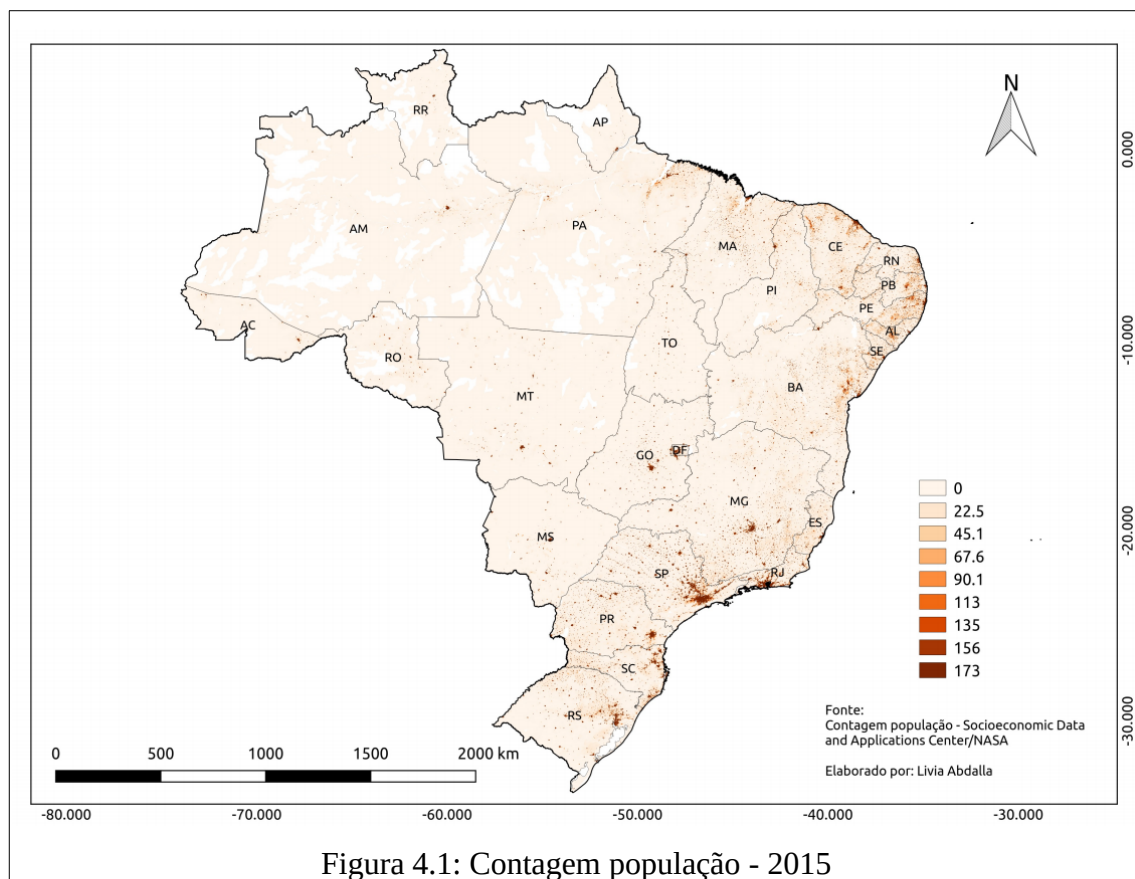
63	Precipitação total mensal – Junho – prec6		1			1
64	Temperado mesotérmico brando média entre 10 e 15 C super úmido sem seca			1		1
65	Temperatura máxima mensal – Agosto – tmax8				1	1
66	Temperatura média do trimestre mais úmido – bio8	1				1
67	Tropical Brasil Central subquente média entre 15 e 18° C em pelo menos 1 mês super úmido subseca			1		1
68	Variação Diurna Média de Temperatura – bio 2		1			1

A partir da verificação da frequência de indicação dos 68 dados/variáveis temáticas nas árvores de decisão, optou-se por analisar esses dados, e suas prováveis relações com a emergência da febre amarela, de modo agrupado. Foi identificado que esses dados/variáveis pertencem a dez macrogrupos temáticos: Perfil demográfico; Primatas não humanos; Biomas; Unidades de conservação; Tipos de uso e cobertura da terra; Temperatura e precipitação; Mineração; Ferrovias; Tipos de solos e Tipos de relevos. Assim, foram organizados os dez subcapítulos seguintes, que indicam e discutem os padrões e as relações existentes entre os dados/variáveis (ou ao menos aos temas aos quais pertencem) com a emergência da febre amarela silvestre no Brasil.

4.2.1 Perfil demográfico

Os dados correspondentes ao perfil demográfico da população brasileira foram evidenciados nas árvores de decisão dos modelos experimentais (método J48), tendo sido indicado doze vezes os atributos sobre Contagem da população e dez vezes os atributos de

Densidade da população (figuras 4.1 e 4.2).



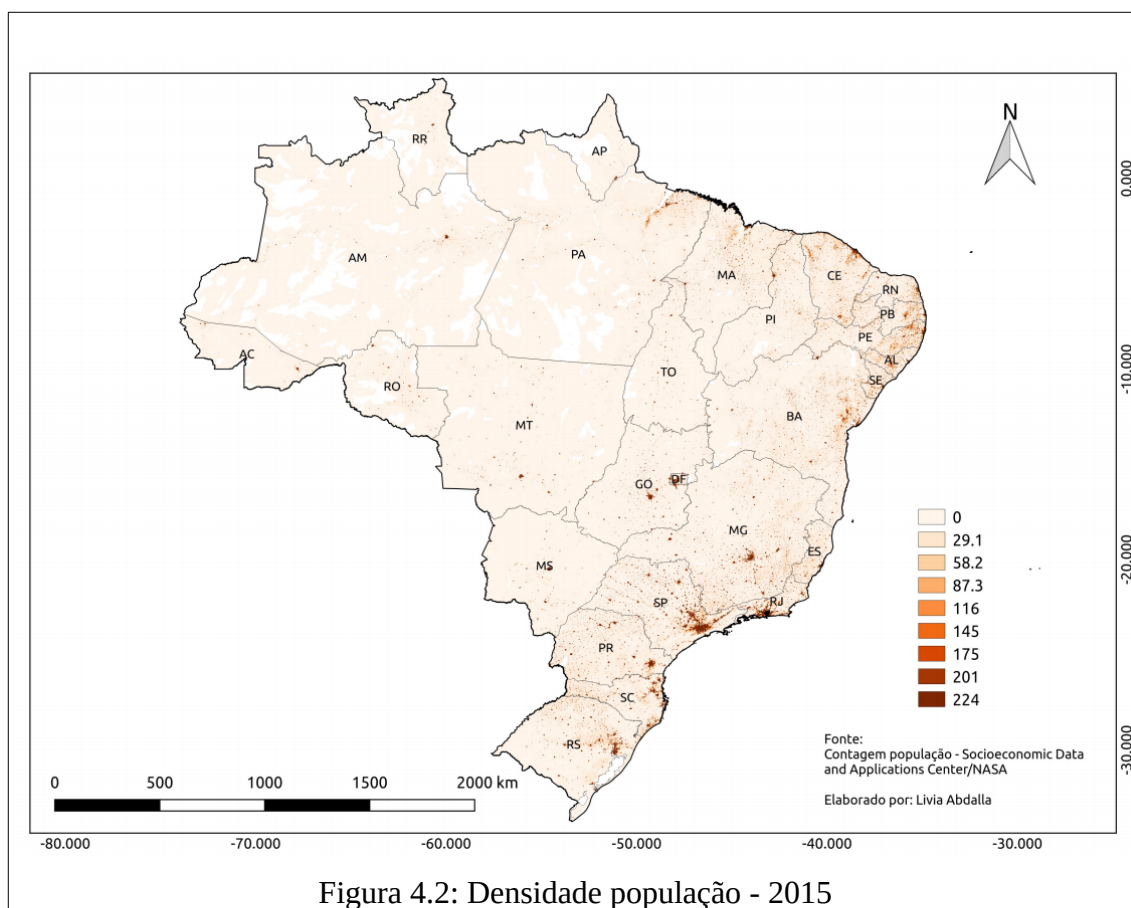


Figura 4.2: Densidade população - 2015

A projeção da população do Brasil para 2018 corresponde a 209,1 milhões de habitantes (IBGE, 2018), sendo a densidade populacional de 25 pessoas por km² e 82% da população localizada em áreas urbanas. A taxa anual de crescimento populacional varia entre 0,77% e 0,82% (FAO, 2018; BANCO MUNDIAL, 2017).

Inicialmente, para compreender as características do perfil demográfico brasileiro é fundamental analisar e compreender os dados oficiais oriundos dos Censos Demográficos do Brasil (1991, 2000 e 2010), pois estes demonstram como ocorreu, no tempo e no espaço, a dinâmica de distribuição da população brasileira no território.

De acordo com o Censo Demográfico de 2010, a distribuição geográfica da população se mantém desigual entre as Regiões, de maneira que as Regiões mais populosas do país são o Sudeste, Nordeste e Sul, compreendendo respectivamente a 42,13%, 27,83% e 14,36%. Já as regiões Norte e Centro-Oeste são as menos populosas, com 8,32% e 7,37%, respectivamente.

A heterogeneidade da distribuição da população também é verificada nos dados de densidade populacional por Unidade da Federação. A distribuição espacial das cidades, assim como seu distanciamento, são também um elemento fundamental do processo de ocupação, divisão política e, conseqüentemente, a diferenciação regional dos estados brasileiros.

Além da diferenciação de ocupação entre regiões, as formas de ocupação dos espaços rurais e urbanos nos estados também apresentam particularidades interessantes. Nas regiões com menores densidades populacionais, destaca-se que nos domínios territoriais de mais baixa densidade populacional da Região Norte existe um predomínio do formato linear da ocupação ribeirinha, que caracteriza a forma de povoamento da porção oeste dessa Grande Região. Existe também o predomínio de densidades lineares ao longo de estradas, traço este que caracteriza a porção ocidental tanto das Regiões Norte e Centro-Oeste, onde coexistem áreas com baixa densidade e com densas manchas urbanas, representadas pelas capitais estaduais (IBGE, 2011).

Interessante complementar que, na Região Centro-Oeste, a produção agropecuária industrializada conferiu nova especificidade ao processo de crescimento e ordenamento da população no território. Embora a produção agropecuária tradicionalmente seja responsável por um espaço rural caracterizado por baixas densidades demográficas, o avanço da monocultura da soja e do algodão acarretou no crescimento demográfico de inúmeros municípios dessa região, sobretudo nos seus centros urbanos. Características territoriais semelhantes às da Região Centro-Oeste ocorreram nas áreas do oeste da Bahia, assim como as do sul do Maranhão e do Piauí, as quais apresentam padrões de ocupação semelhantes aos dos estados que compõem aquela macrorregião, justamente em razão da expansão da agricultura mecanizada (IBGE, 2013).

Os estados da região Nordeste possuem o padrão de maior densidade populacional, que se apresenta, de modo geral, do litoral em direção ao interior. Em contraponto a esse padrão, destaca-se a Zona da Mata e do Agreste, que possuem ocupação territorial mais densa, e o sertão, que apesar de possuir áreas de ocupação menos densas, na atualidade convivem com novos padrões de ocupação derivados de projetos públicos e privados, que alteram o uso e o modelo tradicional de ocupação territorial do interior dessa macrorregião. A construção de hidrelétricas no entorno deste domínio (sertão), de iniciativa estatal, também foi a estrutura

responsável pelo desenvolvimento da economia e da sociedade nordestina (IBGE, 2011).

As Regiões Sudeste e Sul do país são compostas por um domínio de maior adensamento populacional conjugada a uma rede viária urbana igualmente densa. O sudeste comporta parcela significativa da população e dos centros urbanos que comandam a vida econômica, política e cultural do País. Nessa porção do País partem grandes eixos rodoviários, portos e aeroportos que interligam as diferentes regiões brasileiras e que facilitam a movimentação de mercadorias, pessoas e recursos financeiros. A região concentra as principais cidades, cuja rede de influência se irradia, em diferentes graus de intensidade, por todo o País, além dos principais complexos industriais, polos tecnológicos e centros de ensino e pesquisa (IBGE, 2013).

Destaca-se nesse contexto o estado de São Paulo, o qual caracteriza-se por uma rede urbana com vários núcleos, além de uma distribuição de cidades e formas de ocupação de seu território bem mais distribuída que a dos demais estados brasileiros. Essa estrutura urbano-regional, com interiorização pioneira, deve-se ao adensamento da rede ferroviária durante o ciclo do café, principal produto de exportação entre o final do século XIX e início do século XX.

No surto mais recente de FAS, a densidade populacional dos estados que concentraram maior número de casos humanos (São Paulo; Rio de Janeiro; Minas Gerais e Espírito Santo) é substantivamente maior do que nas duas situações anteriores. Alguns aspectos do surto mais recente chamam atenção e merecem destaque. O primeiro deles é o fato do ciclo silvestre ter ocorrido em área onde a população era predominantemente urbana, inclusive com a detecção dos vetores *Haemagogus janthinomys* e *Leucocelaenus*. Já o vetor *Aedes albopictus* foi encontrado em áreas limítrofes entre o meio urbano e o rural, em municípios periféricos às principais regiões metropolitanas (GÓMEZ *et al*, 2018; COUTO-LIMA, 2017). Essas ocorrências podem estar indicando um novo cenário de instalação do padrão epizootico-epidêmico da febre amarela silvestre no Brasil. Enquanto no Brasil são reconhecidos os ciclos silvestre e urbano, na África existem três ciclos de transmissão da febre amarela (figura 4.3). Segundo a WHO (2016), são eles:

- i. “Febre amarela selvática ou silvestre: nas florestas tropicais, os primatas, que são o principal reservatório da febre amarela, são picados por mosquitos selvagens que passam o vírus para outros primatas. Ocasionalmente, os humanos que trabalham ou viajam pela

floresta são picados por mosquitos e contraem a febre amarela”;

ii. “Febre amarela intermédia: neste tipo de transmissão, os mosquitos semi-domésticos (os que proliferam tanto na selva como junto das casas) infectam tanto os primatas como os humanos. O maior contato entre as pessoas e os mosquitos infectados gera uma maior transmissão e muitas aldeias, separadas numa determinada zona, podem desenvolver surtos em simultâneo. Este é o tipo mais comum de surtos em África”;

iii. “Febre amarela urbana: as grandes epidemias ocorrem quando pessoas infectadas introduzem o vírus em zonas densamente povoadas, com elevada densidade de mosquitos, e onde a cobertura vacinal é baixa. Nestas condições, os mosquitos infectados transmitem o vírus de pessoa para pessoa”.



Figura 4.3: Ciclos de transmissão da febre amarela na África (Fonte: WHO, 2016)

Nos três ciclos de transmissão o vírus e a evolução clínica são idênticos. A diferença refere-se apenas aos vetores e ao ambiente (natural ou antropizado) de aquisição da infecção. O surto mais recente de febre amarela no Brasil parece indicar uma adaptação do ciclo silvestre da doença para regiões urbanas e periurbanas, o que traz como possibilidade o

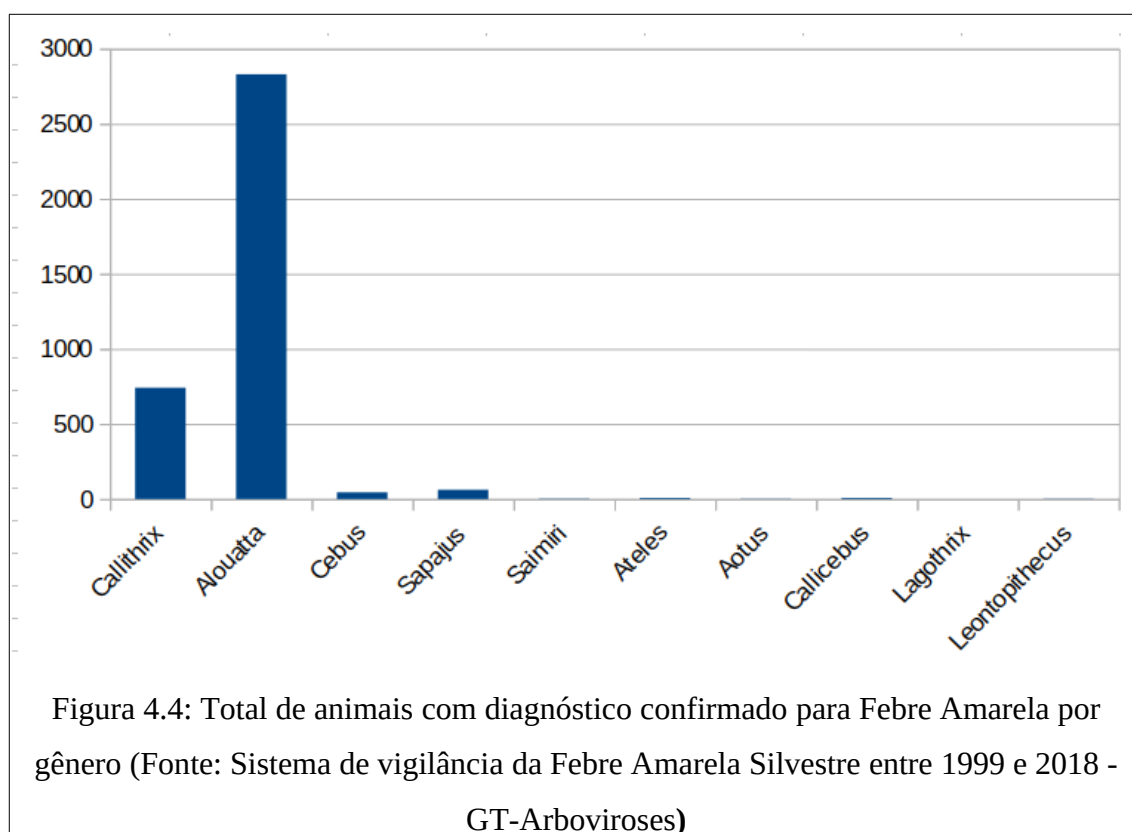
estabelecimento de um ciclo intermediário também no Brasil, ou ainda, o alerta para o risco de reurbanização da doença. Para tais hipóteses é fundamental a investigação sobre as espécies de vetores participantes do recente ciclo de transmissão.

As pressões ambientais exercidas pelo crescimento da população humana são um dos principais moduladores do *habitat* de primatas e do seu declínio populacional (CRIST *et al.*, 2017). No Brasil, as áreas de alta densidade populacional humana estão concentradas perto da costa do Atlântico, que perdeu quase 88% de sua floresta original (RIBEIRO *et al.*, 2009). Projeções sobre o crescimento da população humana para 2050 indicam que ocorrerá crescimento progressivo da população urbana e declínio das populações rurais (ESTRADA, 2013; ESTRADA *et al.*, 2017). Assim, espera-se que com o aumento projetado da população também aumentem exponencialmente os impactos e pressões em *habitats* de primatas. Esses impactos negativos resultarão das crescentes demandas por energia, espaço, alimentos, água, minerais, petróleo, material de construção, produtos florestais e transporte, bem como dos danos ambientais causados pela poluição e pela expansão de redes rodoviárias e ferroviárias. Considera-se, portanto, que com o aumento dos impactos sobre as áreas naturais e, consequente redução das populações e dos *habitats* de espécies de primatas, ampliam-se os riscos de acometimento da febre amarela sobre a própria população humana. Vale lembrar que os primatas são sentinelas para a ocorrência da FA e, portanto, servem de alerta sobre a circulação da doença, para que medidas protetivas sejam tomadas antes que elas acometam humanos.

4.2.2 Primatas Não Humanos

No mundo existem descritas 376 espécies de Primatas Não Humanos (PNHs). O Brasil destaca-se nesse contexto por ser o país com a maior diversidade de primatas, com 102 espécies conhecidas, das quais 69 são endêmicas, incluindo espécies e subespécies (RYLANDS *et al.*, 2000). A proximidade evolutiva dos primatas não humanos com os seres humanos, além de suas semelhanças físicas e comportamentais, ampliam as possibilidades de compartilhamento de doenças infecciosas, como a febre amarela. Em regiões com habitações humanas a morte de primatas, como os dos gêneros *Callithrix*, *Alouatta*, *Cebus* e *Sapajus*

(gêneros com maior número de animais envolvidos no histórico de epizootias notificadas no Brasil) (figura 4.4), é um alerta para a saúde pública. Nesse contexto, justifica-se a importância de monitoramento do grau de conservação desses animais, já que a erradicação da doença no seu ciclo silvestre é impossível (REIS *et al.*, 2008).



O descobrimento do ciclo silvestre da FAS estimulou investigações em vários países que, entre outras contribuições, levaram ao reconhecimento do padrão cíclico envolvendo macacos e mosquitos silvestres. A suscetibilidade dos macacos ao vírus da febre amarela foi testada em laboratório e comprovada em todas as espécies estudadas por Soper em 1942 e Strode em 1951 (BENCHIMOL, 2001), em maior ou menor grau.

Nesse contexto, destaca-se que foram evidenciadas pelas árvores de decisão onze espécies de primatas, com diferentes frequências de indicação. A compreensão sobre sua distribuição geográfica, *habitat* e tipos de ameaças podem contribuir na busca de relações mais específicas dessas espécies, e suas particularidades, com os surtos de febre amarela.

Portanto, ao longo desse subcapítulo serão inicialmente descritas, para cada uma das onze espécies, as principais características sobre o *habitat*, distribuição geográfica original e recente, bem como sobre as principais ameaças à manutenção de seus *habitats* e de suas populações (as quais podem ter tornado mais vulneráveis e expostas ao acometimento da doença). Assim, somente após a descrição dessas características básicas, considera-se que é possível avançar para a indicação e compreensão sobre as prováveis relações/conexões entre as epizootias de febre amarela com o *habitat*, distribuição geográfica e/ou principais ameaças aos gêneros pertencentes a estas espécies.

As espécies de PNHs com maior frequência de indicação nas árvores de decisão dos modelos experimentais (método J48) pertencem a família *Callitrichidae* e ao gênero *Callithrix* (figura 4.5). Foram indicados onze vezes o *Callithrix jacchus* e seis vezes o *Callithrix penicillata*.

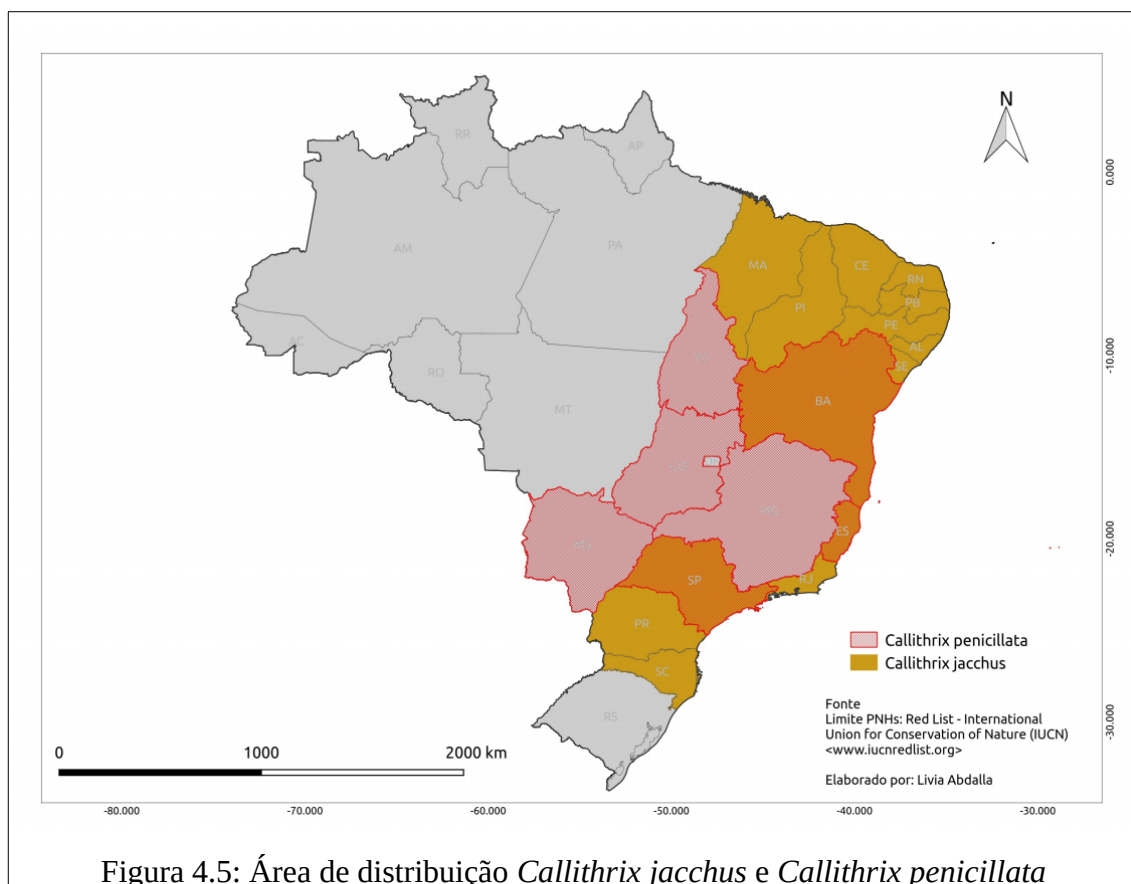


Figura 4.5: Área de distribuição *Callithrix jacchus* e *Callithrix penicillata*

A espécie *Callithrix jacchus* (sagui-de-tufos-brancos) é abundante e amplamente

distribuída nos biomas Caatinga e Mata Atlântica. Distribui-se originalmente no Nordeste, ocorre nos estados de Alagoas, Ceará, Maranhão, Paraíba, Pernambuco, Piauí e Rio Grande do Norte; foi introduzido nos estados da Bahia, Sergipe, Espírito Santo, Rio de Janeiro, São Paulo, Paraná e Santa Catarina (IUCN, 2018). Seu *habitat* é a floresta semidecidual, a floresta decidual, a floresta ombrófila densa, as florestas ripárias, as manchas de caatinga arbórea e a caatinga arbustiva (RYLANDS *et al.* 2008). Estudos indicam que essa espécie não é restrita a *habitats* primários e possui grande capacidade de adaptação, tanto em áreas rurais e urbanas (capazes de viver em pequenas porções florestais como parques, jardins urbanos e aldeias rurais), e de manutenção nas regiões em que foi introduzido, estabelecendo competição com outras espécies de primatas nativos dessas regiões (IUCN, 2018).

O *Callithrix penicillata* (sagui-de-tufos-pretos) também tem ampla distribuição no Cerrado, Mata Atlântica e Caatinga. A espécie é residente e nativa nos estados da Bahia, Goiás, Mato Grosso do Sul, Minas Gerais, São Paulo, Tocantins, sudoeste do Maranhão e Piauí; foi introduzida no Espírito Santo (IUCN, 2018). Segundo Rylands e Mendes (2008) foi introduzida também nos estados do Rio de Janeiro, Paraná e Santa Catarina, habitando áreas de florestas ripárias, floresta decidual, floresta semidecidual, cerrado e cerradões. Assim como outras espécies de saguis, esta não é restrita a ambientes naturais primários e tem preferência por florestas secundárias e florestas alteradas (FONSECA e LACHER JÚNIOR, 1984; LACHER JÚNIOR *et al.*, 1984; RYLANDS, 1984; SEABRA *et al.*, 1991; RYLANDS e FARIA, 1993). Apesar de ser considerada generalista, resistente e capaz de sobreviver em *habitats* extremamente degradados, as populações desta espécie desapareceram ou estão diminuindo em muitas regiões. A principal suspeita refere-se a captura para o comércio ilegal desses como bichos de estimação (IUCN, 2018).

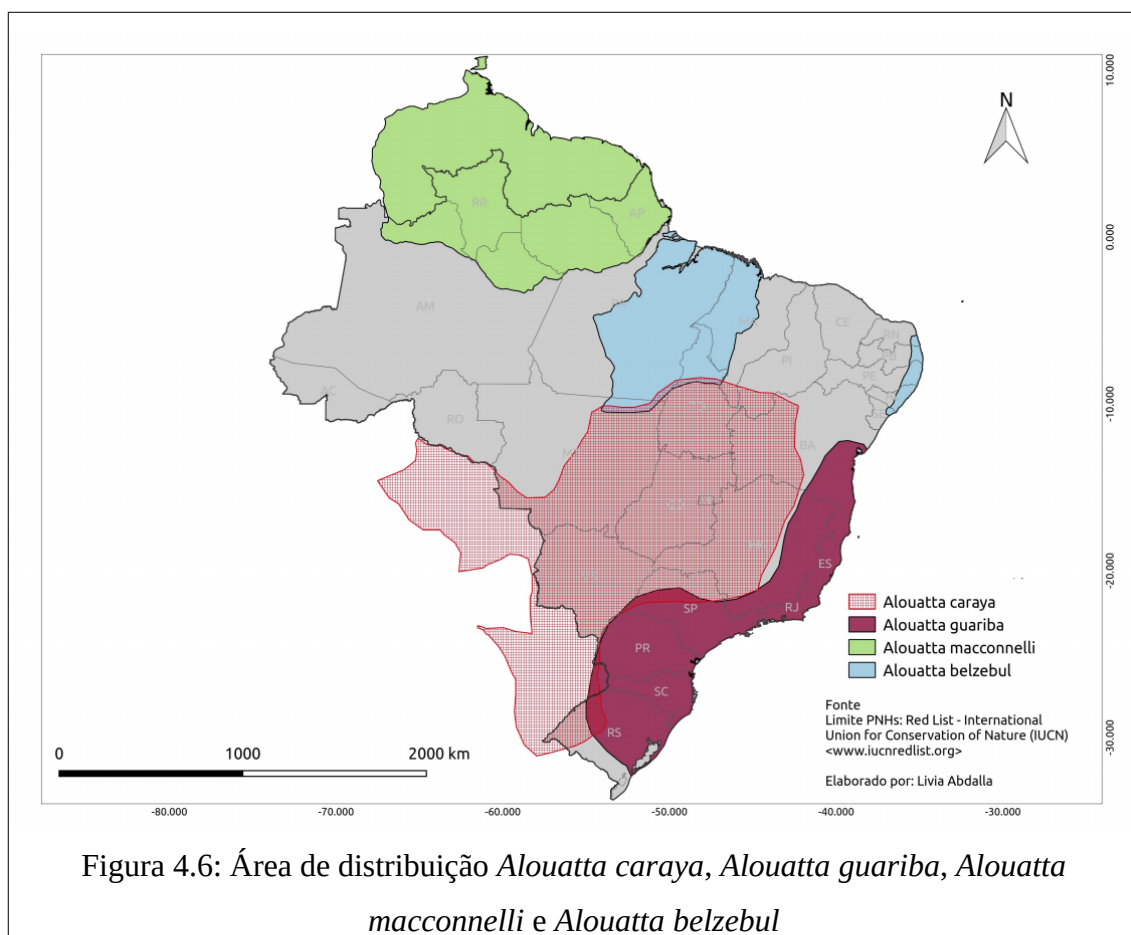
As principais ameaças identificadas para as espécies de *Callithrix* estão relacionadas com: incêndios; agricultura; pecuária; desmatamento; desconexão de *habitat*; fragmentação florestal; vulnerabilidade a epidemias como zika (TERZIAN *et al.*, 2018), febre amarela (ARAÚJO *et al.*, 2011) e herpes humana (MATZ-RESING *et al.*, 2003; HATT *et al.*, 2004; LEFAUX *et al.*, 2004; CASAGRANDE *et al.*, 2014); além da apreensão para comercialização ilegal (OLIVEIRA e GRELLE, 2012).

No que se refere a participação desse gênero no processo enzoótico-epizoótico da febre

amarela existem muitas lacunas para investigação. Conforme visto, este primata tem ampla disseminação nos Estados indicados, sendo visto habitualmente nas praças e nos quintais das casas em busca de alimento. Sua participação no ciclo da doença é vista com cautela e preocupação, pois se tem observado um aumento de populações desses animais em áreas urbanas em virtude da disponibilidade de alimento (os primatas são limitados na fertilidade pela comida e pelo espaço geográfico, portanto, quanto mais alimentos disponíveis, mais proliferam). Além disso, seus predadores naturais estão praticamente extintos nesses ambientes modificados (COSTA, 2005).

De modo geral, o sagui está presente em todos os estados do Nordeste, em Goiás, Tocantins e nos estados do Sudeste, tendo sido uma das espécies introduzidas no Estado do Rio de Janeiro por intermédio do tráfico de animais no Estado do Rio de Janeiro. Sabe-se que são animais arborícolas, encontrados em vários tipos de vegetação, como ambientes de cerrado, matas de galeria e matas secundárias, além de pomares e áreas de reflorestamentos (COSTA, 2005). Entretanto, o conhecimento de suas características ecológicas e comportamentais, principalmente no contexto dos fragmentos florestais remanescentes, é pouco conhecido (ICMBIO, 2014).

As espécies com segunda maior frequência de indicação pertencem a família *Atelidae* e ao gênero *Alouatta* (figura 4.6). Foram indicados cinco vezes o *Alouatta caraya*, quatro vezes o *Alouatta guariba* e o *Alouatta macconnelli*, e uma vez o *Alouatta belzebul*.



Alouatta caraya (bugio-preto) é uma espécie com ampla distribuição no Brasil, ocorrendo nos biomas Amazônia, Cerrado, Pantanal, Caatinga, Mata Atlântica e Pampa, e nas Unidades Federativas (UFs) de Mato Grosso, Mato Grosso do Sul, Goiás, Distrito Federal, Tocantins, Rondônia (sudoeste), Pará (sudeste), Maranhão (sudoeste), Piauí (sudoeste), Bahia (oeste), Minas Gerais (centro-oeste), São Paulo (noroeste) e nas porções oeste do Paraná, Santa Catarina e Rio Grande do Sul.

Sua distribuição ocorre em florestas tropicais secas, florestas inundadas, floresta estacional semidecidual e decidual (PI, oeste da BA, MG, PR e RS), florestas contínuas, matas de galeria e remanescentes florestais dos biomas Pampa, Cerrado, Pantanal, Mata Atlântica e Caatinga (MOURA, 2007; AGUIAR *et al.*, 2011). A espécie ocorre ainda na formação florestal “cerradão” no bioma Cerrado (STAVIS, 2013), em matas de galeria nas áreas de transição Cerrado-Amazônia (IWANAGA e FERRARI, 2002; ALVES, 2006) e em

floresta de igapó no bioma Amazônia, particularmente no estado de Rondônia (ALVES, 2013). Vivem em estratos arbóreos de 10m a 20m nas florestas montanhosas úmidas, ou em vegetação mais aberta como a caatinga, o cerrado, o babaçual ou de araucária, em altitudes que variam entre 0 e 1.200m (AURICCHIO, 1995). O *Alouatta* também apresenta tolerância a modificações/perturbações no ambiente, conseguindo sobreviver em ambientes fragmentados e alterados (RÍMOLI *et al.*, 2012; LÁZARO, 2012; MUHLE, 2008; PRATES, 2007; BICCA-MARQUES, 2003). A área de vida dessa espécie pode variar de pequenos fragmentos florestais menores que 1 hectare, a pomares com aproximadamente 20ha (LUDWIG, 2006; BRAVO e SALLENAVE, 2003).

Alouatta guariba (bugio-marrom) apresenta distribuição restrita ao bioma Mata Atlântica, mais especificamente no sul da Bahia, nordeste de Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná, Santa Catarina e Rio Grande do Sul (IUCN, 2018). Tem preferência pela Floresta Atlântica de terras baixas, submontana e montana, estendendo-se até florestas sazonais semidecíduas e decíduas no estado de Minas Gerais (MELO, 2004; CHIARELLO, 1999; MENDES, 1989; RYLANDS *et al.*, 1988). Também apresenta tolerância a modificações/perturbações no ambiente (MENDES *et al.*, 2008). Segundo ICMBIO (2014) estima-se uma perda populacional maior que 25% em suas populações causadas perda e fragmentação de *habitat*. Suspeita-se que o tamanho populacional atual é extremamente reduzido, não havendo mais do que 50 indivíduos maduros em qualquer subpopulação em decorrência da severa fragmentação.

Alouatta macconnelli (bugio-vermelho-das-Guianas) possui ampla distribuição na Amazônia, principalmente na porção norte do bioma, onde é residente e nativa nos estados do Amazonas, Pará, Amapá e Roraima. Seu *habitat* são áreas de floresta primária tropical de planície, floresta ciliar e floresta decidual seca (BOUBLI *et al.*, 2008). Não foram identificados estudos sobre a restrição ou tolerância da espécie com áreas modificadas/perturbadas (ICMBIO, 2018).

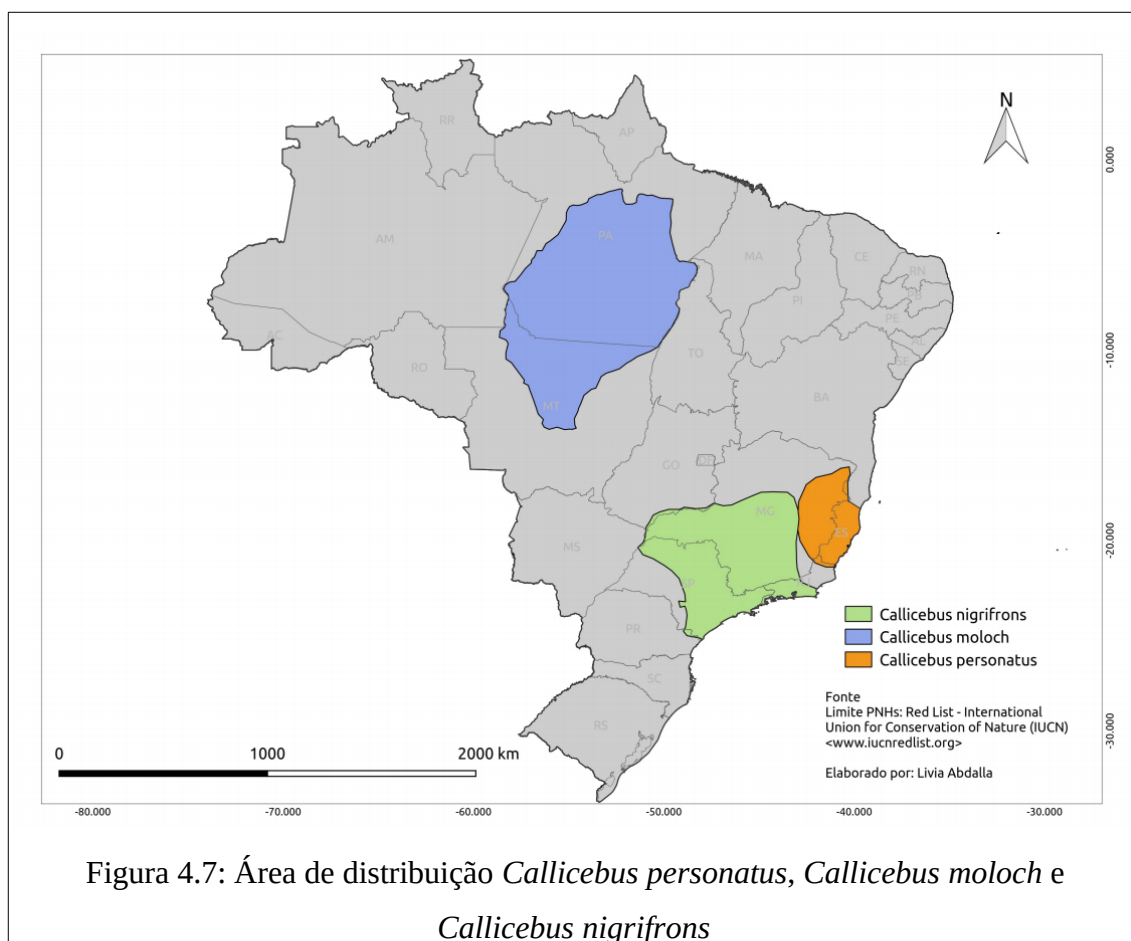
Alouatta belzebul (guariba-de-mãos-ruivas) é endêmico do Brasil, com distribuição em parte da Amazônia (Pará, Maranhão e norte do Mato Grosso) e em parte da Mata Atlântica nordestina (Rio Grande do Norte, Paraíba, Pernambuco, Alagoas e Sergipe) (IUCN, 2018). A espécie ocorre em Floresta Tropical Amazônica de Planície, Floresta de Várzea de Marajó e

fragmentos de Floresta Tropical Atlântica do Nordeste (VEIGA *et al.*, 2008). Possui tolerância a modificações/perturbações no ambiente (grande parte da distribuição da espécie está incluída no arco do desmatamento na região Amazônica). Ocupa ainda fragmentos da Mata Atlântica de tamanho reduzido e em regeneração, e é também encontrado em áreas perturbadas pela extração mineral (CARVALHO, 2010).

Conforme visto, três das quatro espécies de *Alouatta* apresentam tolerância a modificações/perturbações no ambiente, entretanto, estas espécies vêm sofrendo declínio populacional. As principais ameaças sobre essas espécies são a perda e a fragmentação do *habitat* provocados pela expansão das áreas agrícolas monoculturas e de pecuária, por incêndios florestais, pela expansão urbana e das matrizes energética e rodoviária, pela caça (consumo e tráfico) e a vulnerabilidade a epidemias de doenças infecciosas, especialmente a febre amarela (FREITAS e BICCA-MARQUES, 2013; ALMEIDA *et al.*, 2012; FOGUESATTO, 2011; VEIGA *et al.*, 2008).

Primatas da família *Ateliade*, como o *Ateles* (macaco aranha) e, em especial, o *Alouatta*, conhecido como “macaco berrador” ou “guariba”, desempenham papel importante na manutenção da febre amarela silvestre, por serem altamente suscetíveis ao vírus. Segundo publicação de Strano *et al* em 1975 (COSTA, 2005), sua função no ciclo silvestre da FAS é tão dominante e evidente que o repentino silêncio nas matas serve de sinal aos habitantes sobre a circulação do vírus amarílico.

As espécies com terceira maior frequência de indicação pertencem a família *Pitheciidae* e ao gênero *Callicebus* (figura 4.7). Foram indicados quatro vezes o *Callicebus personatus*, e uma vez o *Callicebus moloch* e o *Callicebus nigrifrons*.



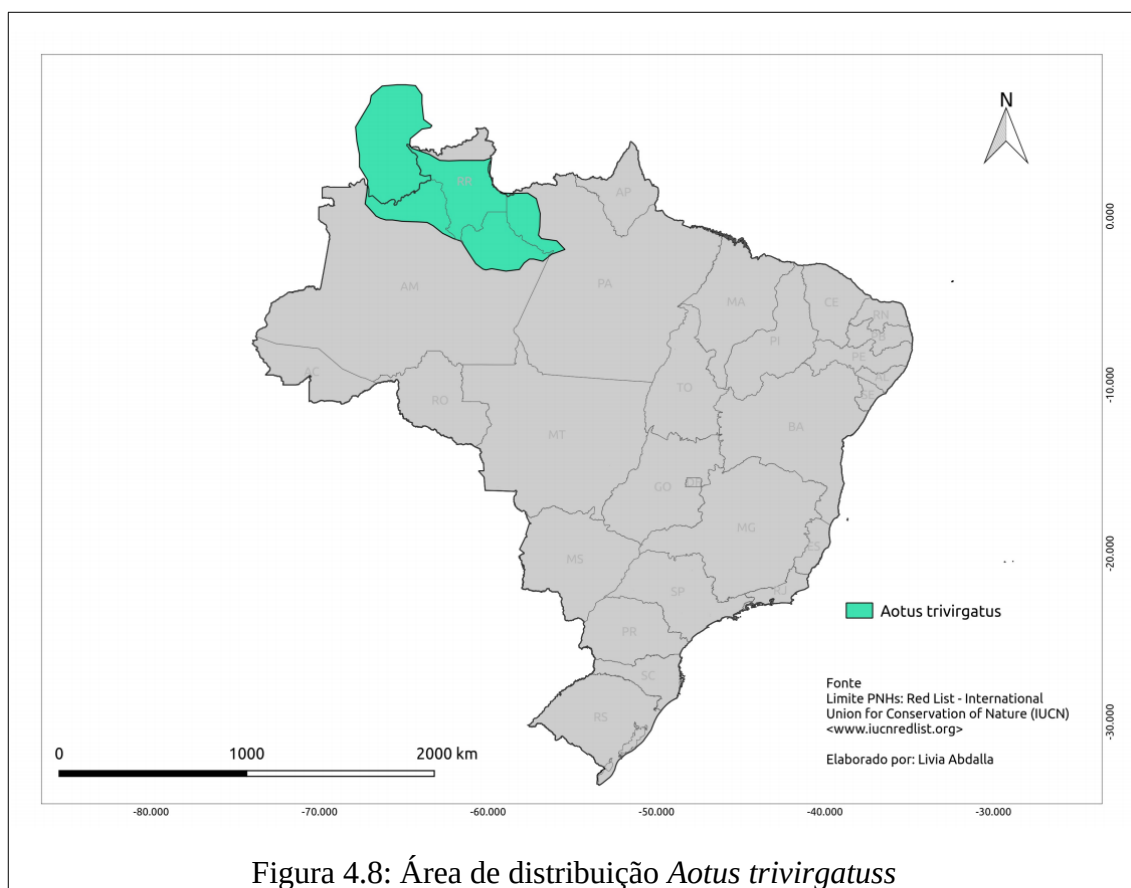
Callicebus personatus (Guigó) é endêmica da Mata Atlântica, ocorrendo no centro e norte do Espírito Santo e no nordeste de Minas Gerais (IUCN, 2018). Pesquisas realizadas nos últimos anos sugerem que esta espécie tem se tornado cada vez mais rara e de difícil observação. Ocupa formações florestais da Mata Atlântica e, apesar de não estar restrita a *habitats* primários, suspeita-se que a espécie apresenta baixa tolerância a modificações/perturbações no ambiente (ICMBIO, 2018).

Callicebus moloch (Zogue-zogue) é endêmico do Brasil, ocorre na Amazônia brasileira, mais especificamente no sul do rio Amazonas nos estados do Pará e Mato Grosso (VEIGA e FERRARI, 2008). Ocorre em floresta ripária e floresta ombrófila sazonalmente inundada (KINZEY, 1981). Segundo Ferrari *et al.* (2003) e Veiga e Ferrari (2008), de uma maneira geral as espécies desse grupo são consideradas tolerantes a habitats com distúrbios provocados por atividade humana e também áreas sazonalmente inundadas.

Callicebus nigrifrons (Guigó) ocorre em uma área relativamente ampla no sudeste do Brasil, incluindo o norte do estado de São Paulo, sul de Minas Gerais e sudoeste do Rio de Janeiro (VEIGA *et al*, 2008; PRINTES *et al*, 2013). Esta espécie é capaz de sobreviver em pequenos fragmentos florestais, com ampla capacidade de dispersão pela matriz, além de apresentar tolerância a perturbações ambientais. Segundo Veiga *et al* (2008) habita áreas de Mata Atlântica e de transição com o Cerrado, ocupando tanto áreas conservadas quanto florestas com distúrbios antropogênicos e em regeneração.

As espécies de *Callicebus* descritas também tem tolerância (baixa ou média) com alterações/distúrbios no seu ambiente natural, sendo principalmente ameaçadas por assentamentos rurais, agricultura, pecuária, desmatamento, aumento da matriz energética, aumento da matriz rodoviária, desconexão e redução de *habitat* e extração de minério (VEIGA e FERRARI, 2008).

A espécie de PNH *Aotus trivirgatus* (Macaco da noite de pescoço cinza), pertencente a família *Aotidae*, teve uma frequência de indicação (recorte 1999 a 2017), (figura 4.8). Essa espécie ocorre na Venezuela e no Brasil, sua distribuição ocorre nos estados brasileiros do Amazonas (norte), Pará (noroeste) e Roraima (FERNANDEZ-DUQUE *et al*, 2013; IUCN, 2018).



A espécie habita regiões de floresta tropical/subtropical úmida de planície, floresta tropical/subtropical inundada e, úmida de montana (HERNÁNDEZ-CAMACHO e COOPER, 1976; DEFLER, 2003). Apresenta tolerância a modificações/perturbações no ambiente, ocorrendo em florestas alteradas, incluindo as que tiveram corte seletivo de madeira (DEFLER, 2003). As principais ameaças identificadas para a espécie são: agricultura; pecuária; desmatamento; desconexão e redução de *habitat*.

O *Aotus* (macaco da noite), habita as bacias dos rios Amazonas, Orinoco e Madalena (os grandes focos sul-americanos do vírus amarelo) tem hábitos noturnos e dorme de dia em ocos de árvores, o que dificulta o contato com os mosquitos, cujas atividades são exclusivamente diurnas. No entanto, por ser muito suscetível ao vírus, pode contribuir para a manutenção do ciclo em áreas onde não existam outros primatas (STRANO *et al.*, 1975).

No contexto geral sobre ameaças às espécies de primatas, destaca-se que, além da destruturação dos ambientes naturais por ação antrópica, outros aspectos comportamentais do ser humano podem contribuir para o processo de transmissão da doença, como é o caso de

práticas de caça de macacos (quer seja com a finalidade de aquisição de animal de estimação, quer seja por hábitos de consumo), o que implica na redução do número de hospedeiros e logo possibilita a entrada accidental do homem no ciclo natural de transmissão da doença.

O tráfico de animais silvestres, principalmente das espécies de *Callithrix* e *Allouata*, pode ser um caminho para a compreensão da emergência da febre amarela em ambientes antropizados. O governo brasileiro estima que cerca de 12 milhões de animais (entre macacos, jaguatiricas, araras, papagaios, tartarugas, serpentes, borboletas e peixes tropicais) são capturados (em florestas, cerrados e outros ambientes naturais do Brasil) e comercializados em feiras livres no mercado interno ou exportados irregularmente para vários países, movimentando bilhões de dólares por ano (RENCTAS, 2004). Entre os mamíferos indicados estão incluídas algumas espécies de macacos comercializadas para *pet shops*, como o *Callithrix jacchus* e o *penicillata*.

Ainda segundo RENCTAS (2004), destaca-se que as principais rotas utilizadas pelos traficantes são aeroportos e rodovias, seguindo um fluxo entre a região Centro-Oeste, com o Estado de Minas Gerais. A publicação indica que as “rotas mineiras” de tráfico de animais silvestres, sem nenhum controle sanitário, criam condições locais para a introdução, sobrevivência e disseminação de hospedeiros da febre amarela, trazendo como consequência, uma explosão da doença como a que ocorreu nessas regiões durante o mais recente surto da doença. Possivelmente inúmeras outras rotas ainda desconhecidas, especialmente terrestres, estejam sendo usadas para driblar o monitoramento da RENCTAS e a fiscalização do IBAMA, facilitando assim o escoamento e a introdução dos animais em áreas distantes de seu *habitat* original. Essa dinâmica de circulação do tráfico possibilita a movimentação de patógenos de uma paisagem geográfica para outra, o que favorece o aparecimento da doença em novos locais (BEJARANO, 1971).

Destaca-se que somente no estado do Rio de Janeiro, de 2003 a 2010, aproximadamente 457 espécies de primatas foram entregues ao CETAS (Centro de Triagem de Animais Silvestres), enquanto outros 438 foram registrados em reclamações de cidadãos que denunciaram animais de estimação ilegais (Linha Verde). A maioria dos animais apreendidos foram dos gêneros *Callithrix* (335), *Cebus* (86) e *Sapajus* (17), embora outros gêneros (*Leontopithecus*, *Callicebus*, *Lagothrix* e *Alouatta*) também tenham sido registrados. Embora

o impacto de tais números nas populações dessas espécies possa ser pouco representativo, o impacto de animais liberados fora de sua área original pode ameaçar espécies nativas por hibridização, como é o caso do *Callithrix aurita* no Rio de Janeiro (OLIVEIRA e GRELLE, 2012).

A caça é outra principal ameaça que promove a redução da população de primatas, que traz prejuízos diretos aos *habitats* dos primatas e por consequência, resulta na mudança da dinâmica de dispersão de sementes, no tamanho e na distribuição das manchas florestais, bem como na redução da diversidade genética das plantas e no recrutamento de mudas (CAUGHLIN *et al.*, 2015; PACHECO e SIMONETTI, 2000; BRODIE *et al.*, 2009). Consequentemente também ficam reduzidas as fontes de alimento disponíveis para as comunidades locais de mamíferos e aves (ABERNETHY *et al.*, 2013; NUÑEZ-ITURRI *et al.*, 2008). Na Amazônia brasileira, a caça excessiva de primatas tem impactado ainda na regeneração de espécies de árvores centenárias e de alta densidade, que possuem grande capacidade de armazenar carbono (PERES *et al.*, 2016; STEVENSON e ALDANA, 2008).

Por fim, os impactos antrópicos também podem resultar na degradação dos *habitats* de primatas nativos, o que pode refletir em declínios populacionais desses (CEBALLOS e EHRLICH, 2002). À medida que se amplia o contato entre primatas, humanos e animais domésticos, aumenta-se também a ameaça de exposição a doenças infecciosas emergentes e a surtos epizooticos periódicos (HOFFMANN *et al.*, 2017; NUNN e ALTIZER, 2006; NUNN e GILLESPIE, 2016).

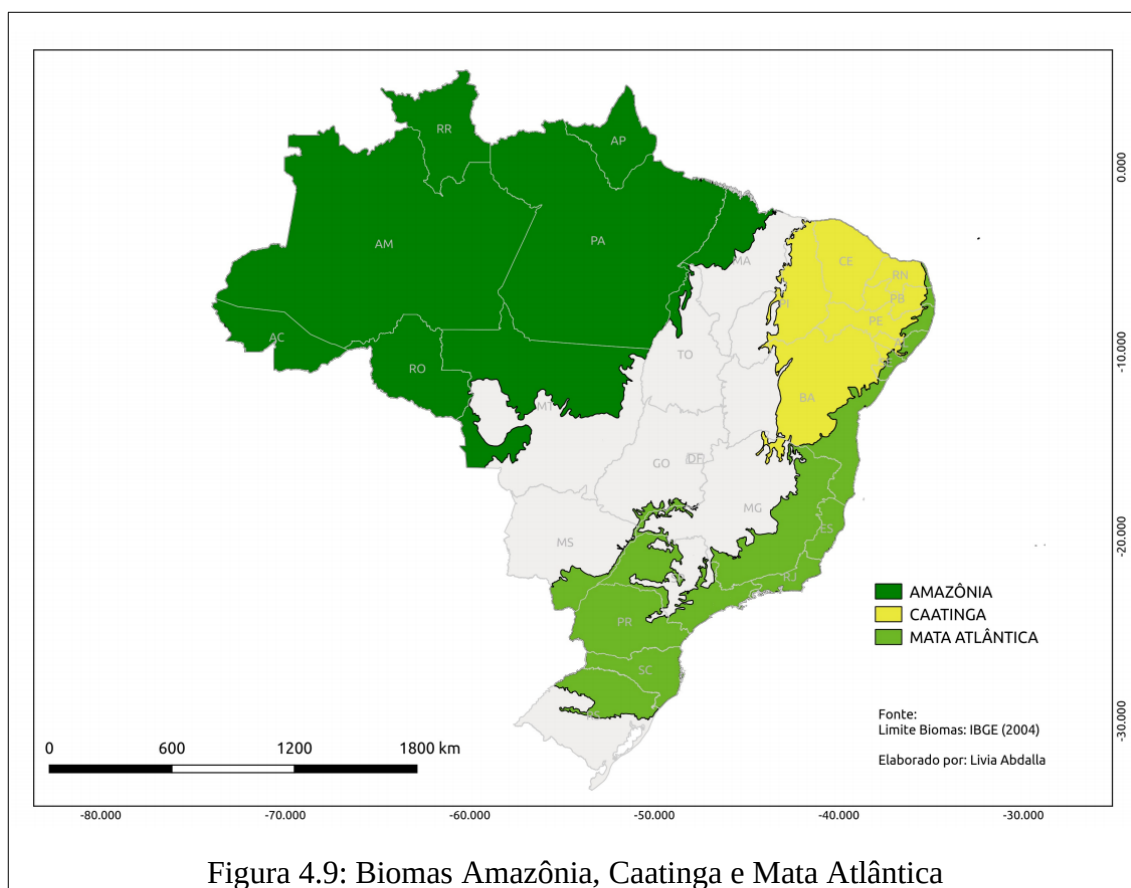
Como exemplo da ameaça de doenças infecciosas sobre as populações de primatas, destaca-se que entre 2008 e 2009, em duas áreas no estado do Rio Grande do Sul, cerca de 80% das populações isoladas de bugios (*Alouatta caraya* e *Alouatta guariba clamitans*) foram eliminadas em razão de grande evento epizootico de Febre Amarela (ALMEIDA *et al.*, 2011; FREITAS e BICCA-MARQUES, 2011; VASCONCELOS, 2017; VEIGA *et al.*, 2014), incluindo populações que habitavam áreas protegidas (FIALHO *et al.*, 2012). Segundo Bicca-Marques *et al.* (2017) a partir de 2016, o surto de FA em curso no Sudeste do Brasil causou a morte de milhares de primatas, incluindo espécies ameaçadas, como o *Callicebus personatus* e o *Alouatta guariba*. Em muitos desses casos, a desinformação sobre os vetores de transmissão da febre amarela resultou no extermínio de populações de primatas por parte de

indivíduos humanos que residem em locais próximos.

4.2.3 Biomas

O Brasil é formado por seis biomas (Amazônia, Caatinga, Cerrado, Mata Atlântica, Pampa e Pantanal) com características particulares (macroclima, fitofisionomia, solo, altitude, entre outros) e proporções distintas. Segundo IBGE (2004), bioma é um conjunto de vida vegetal e animal, constituído pelo agrupamento de tipos de vegetação que são semelhantes e que podem ser identificados em nível regional, com condições de geologia e clima semelhantes e que, geologicamente sofreram os mesmos processos de formação da paisagem, resultando em uma diversidade de flora e fauna próprias. Segundo Costa *et al.* (2011) há que se considerar ainda que a diversidade de ecossistemas existentes no Brasil refletem padrões de chuvas diversos e que determinam a abundância e distribuição dos mosquitos vetores e dos hospedeiros vertebrados. Segundo Ab'Sáber (2003), a paisagem constitui uma herança de processos fisionômicos e biológicos, constituindo um patrimônio coletivo das sociedades que se apropriaram territorialmente. A compreensão dos principais processos modeladores da paisagem perpassa pela formação geológica e pelas mudanças no clima global, com destaque ao Quaternário, que foi o período-chave da formação das principais paisagens brasileiras.

Nesse contexto, destaca-se que foram evidenciadas pelas árvores de decisão três tipos de biomas brasileiros (Amazônia, Caatinga e Mata Atlântica) (figura 4.9), cada um com 1 indicação nos modelos experimentais produzidos pelo método J48. A compreensão sobre as principais características e inter-relações geoclimáticas, vegetacionais, faunísticas e seus históricos de alterações por atividades geológicas e/ou humanas podem indicar caminhos para a compreensão das relações existentes entre os três surtos de FAS e os domínios de natureza no Brasil.



A Amazônia é a maior reserva de diversidade biológica (abriga mais de 2,5 mil espécies de árvores e 30 mil de plantas – um terço de todas as espécies do planeta) e a maior rede hidrográfica (escoa cerca de 1/5 do volume de água doce) do planeta. É o maior bioma brasileiro em extensão, com 4.196.943 km², corresponde a 49,29% do território nacional, onde localizam-se integralmente cinco UFs (Acre, Amapá, Amazonas, Pará e Roraima) e parcialmente os estados de Rondônia (98,8%), Mato Grosso (54%), Maranhão (34%) e Tocantins (9%) (IBGE, 2004).

A vegetação predominante na Amazônia é a Floresta Ombrófila Densa, que corresponde a 42% do bioma, Floresta Estacional decidual e semidecidual, campinarana, savanas (inclusive estépica), além de Formações pioneiras, refúgios vegetacionais e áreas de tensão ecológica. Possui flora variada, rica em plantas, e vasta fauna, incluindo ampla diversidade de fungos, bactérias e outros micro-organismos. Pela transpiração de suas florestas, grandes quantidades de vapor d'água e calor são transferidas para a atmosfera, fatores estes

fundamentais para a manutenção das características atuais do clima regional da América do Sul e também global.

Segundo Ab'Sáber (2003) a Amazônia se destaca pela grandeza da malha hidrográfica e pequenas variações de seus ecossistemas regionais de altitude, além da intensa entrada de energia solar e massa de ar úmido. A Amazônia brasileira é dominada por chuvas de verão austral (de janeiro a março) ao sul, enquanto que ao norte ocorrem precipitações maiores durante o verão boreal (de maio a julho). Mesmo se deparando com aspectos diferenciados em diferentes porções do domínio amazônico, o bioma continua sendo um dos mais homogêneos e de ritmo anual habitual mais constante de todo Brasil intertropical. A exuberância hidrográfica Amazônica é resultado da pluviosidade que atinge a grande depressão topográfica que variam entre 30 e 120 metros de profundidade.

Segundo Mondet (2001), a Bacia Amazônica é o grande reservatório endêmico do vírus amarílico no Brasil, mas reconhece uma área epizootica – atualmente denominada de transição – fora da bacia Amazônica, na qual o vírus circula de forma epizootica e ocasionalmente entre primatas não humanos e se manifesta de forma epidêmica na população humana, em função do estado imunológico e da abundância de vetores. Em caráter esporádico, o vírus apareceria nos estados da Bahia, Minas Gerais e São Paulo, mudando para outra bacia hidrográfica. As epidemias nessas regiões seriam diretamente ligadas à área endêmica, precedidas de epizootias, geralmente despercebidas.

Devido ao caráter pouco sistemático desses surtos virais para além da Amazônia, as áreas de risco para febre amarela silvestre são reavaliadas periodicamente pelo Ministério da Saúde com a finalidade de estabelecer políticas de intervenção distintas. A epizootica ou de transição é a que tem sofrido maiores alterações, com ampliação de seus limites. Atualmente esta área abrange uma faixa que vai da região centro-sul do Piauí, oeste da Bahia, noroeste de Minas Gerais, São Paulo, Rio de Janeiro, Paraná, Santa Catarina e Rio Grande do Sul.

A Mata Atlântica é o terceiro maior bioma (13% do território nacional), com área de 1.110.182 km² (Cerrado é o segundo maior, com um pouco mais de 2 milhões de quilômetros quadrados) (IBGE, 2004). Ocupa toda a faixa continental atlântica leste brasileira e se estende para o interior no Sudeste e Sul do País. É definido pela vegetação florestal predominante e relevo diversificado (serras, platôs, vales e planícies). O bioma ocupa inteiramente três

estados – Espírito Santo, Rio de Janeiro, Santa Catarina – e 98% do Paraná, além de porções de outras 11 UFs. Possui grande diversidade de fitofisionomias e complexidade de aspectos bióticos, como grande amplitude altimétrica, regime pluviométrico intenso e temperaturas médias elevadas, compreendendo uma importante parcela da biodiversidade do país. Estende-se por uma diversidade de formas de relevo, abrangendo cadeias de montanhas, platôs, vales e planícies de toda a faixa continental atlântica leste brasileira (GOMES *et al.*, 2009).

A Mata Atlântica caracteriza-se como um diversificado mosaico de ecossistemas florestais, diferenciados em suas estruturas e composições florísticas, tendo como elemento comum a influência dos ventos úmidos oriundos do Oceano Atlântico (BOHRER, 2012). Predominam diversos tipos florestais como Florestas Ombrófila Densa, Estacional Semidecidual e Mista, conforme a classificação da vegetação brasileira desenvolvida pelo Instituto Brasileiro de Geografia e Estatística (IBGE), a partir dos levantamentos e mapeamentos feitos pelo Projeto RADAMBRASIL (LIMA, 2008). Abriga muitas espécies endêmicas e a maior parte das espécies ameaçadas (84%) de extinção da fauna e flora brasileira. Seus remanescentes englobam um diversificado conjunto de ecossistemas florestais, com estrutura e composição florística diferenciadas.

Originalmente, a Mata Atlântica recobria grande parte da região costeira do Brasil (cerca de 12% do território nacional). Aproximadamente 70% da população brasileira ocupa a porção do território situada no domínio da Mata Atlântica e, em consequência, apenas uma pequena parte da sua área (entre 11% a 20%) permanece com sua cobertura vegetal original, representada por fragmentos florestais de diferentes tamanhos, a maior parte com área inferior a 1000ha (BOHRER, 2012). Segundo Ab’Sáber (2003), as florestas atlânticas abrangiam aproximadamente um milhão de quilômetros quadrados, tendo a sua vegetação origem entre o Pleistoceno superior e o Holoceno.

A paisagem atual desta região encontra-se bastante fragmentada e desconectada, representada em sua quase totalidade por manchas florestais, isoladas e impactadas, circundadas por extensas matrizes antrópicas como pastos, monoculturas e áreas de desenvolvimento urbano (FUNDAÇÃO SOS MATA ATLÂNTICA/INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS, 2002). Como consequência da fragmentação, tem-se a perda de área do *habitat* natural, a modificação da dinâmica biológica, a maior

dificuldade na manutenção da diversidade biológica e na passagem da fauna que cruzam os espaços entre os fragmentos. Nesse sentido, a recuperação de fragmentos florestais em ambientes naturais intensamente impactados tem se destacado cada vez mais como uma forma de reverter a atual situação de degradação ambiental, promovendo a conservação da biodiversidade e dos recursos naturais para uso das gerações presentes e futuras (WUETHRICH, 2007; BENAYAS *et al.*, 2009).

A Caatinga é o único bioma exclusivamente brasileiro e corresponde ao quarto maior bioma, possui área de 844.453 km², correspondendo a 9,9% do território. O Bioma se estende pela totalidade do estado do Ceará, na maior parte dos estados do Rio Grande do Norte (95%), da Paraíba (92%), de Pernambuco (83%), mais da metade dos estados do Piauí (63%) e da Bahia (54%), e de quase metade de Alagoas (48%) e Sergipe (49%), além de pequenas porções de Minas Gerais (2%) e do Maranhão (1%) (IBGE, 2004). Ab'Sáber (2003) apresenta a Caatinga como um domínio marcante no contexto climático, hidrológico e ecológico, com aproximadamente 23 milhões de habitantes que ocupam os territórios dos sertões secos, de clima quente, chuvas escassas, periódicas e irregulares. Esse domínio possui precipitações médias anuais 268 e 800 mm, no Nordeste seco. Um dos fatores responsáveis pela paisagem desértica dos sertões a estrutura geológico-litológica em certas áreas, os chamados “altos pelados”, constituídos de colinas desnudas, fragmentos de quartzo e a presença de rocha metamórfica argilosa (filitos), que apresenta-se pela extensão da estrutura topográfica.

O clima, muito quente e sazonalmente seco, tem fortes reflexos sobre a hidrologia do bioma. A grande variabilidade de deflúvios é decorrente das irregularidades interanuais nas chuvas e produz efeitos muitas vezes catastróficos, que se manifestam tanto por secas prolongadas como por inundações. As elevadas taxas de evaporação contribuem enormemente para esgotar os recursos hídricos, principalmente os superficiais. O clima árido também influenciou na formação original de solos com horizontes superficiais pouco espessos, com baixos teores de matéria orgânica e presença significativa de cascalhos e matacões em superfície, havendo expressivas áreas de ocorrência dos Luvisolos Crômicos. As elevadas taxas de evaporação tornam os solos potencialmente vulneráveis aos processos de salinização, sendo comum a presença de crostas de sais cristalizados (eflorescências) em leitos secos de córregos e rios. Apesar das adversidades climáticas, hidrológicas e pedológicas, a flora e

fauna são bastante ricas e adaptadas a essas condições, com diversas espécies endêmicas (IBGE, 2004).

O histórico de monitoramento da febre amarela silvestre demonstra que os biomas Amazônia, Cerrado, Pampa e Mata Atlântica concentraram o maior número de casos humanos e epizootias. Considera-se que a indicação do bioma Caatinga nas árvores de decisão deve-se justamente por ser uma feição de separação sobre o local de baixa ocorrência da doença (o próprio bioma) e os dois biomas de contato (Cerrado e Mata Atlântica) que acumulam eventos de emergência da doença.

Como a vegetação é um dos elementos mais importantes da biota, seu estado de conservação e de integridade determinam a existência ou não de *habitats* de diversas espécies de animais silvestres, a manutenção de serviços ambientais e o fornecimento de bens essenciais à sobrevivência de populações humanas. A agricultura para exportação (agronegócio), promove implicações nos ecossistemas, sendo a principal responsável pelo desmatamento, uso e apropriação inadequada do solo, entre outras ações que alteram a biota e a qualidade de vida humana. As atividades agropecuárias já resultaram na eliminação de uma expressiva porção da cobertura vegetal nativa, principalmente dos biomas Amazônia e Cerrado, e na fragmentação da maioria dos seus *habitats*, acarretando na perda de biodiversidade e aumento da erosão dos solos, com o consequente assoreamento de mananciais. Sabendo que a desestruturação de ambientes naturais é muitas vezes “gatilho” para surtos epidêmicos-epizoóticos, foram verificados os históricos de conservação ou degradação desses biomas. O recente relatório da WWF (2018) apontou para um desmatamento intenso nos biomas brasileiros, que reduziu, de 1970 a 2018, 20% da Floresta Amazônica e 50% do Cerrado. Dados divulgados pelo Ministério do Meio Ambiente (2018) indicaram que 80 mil km² de terras foram devastadas no Cerrado e 50 mil km² na Amazônia.

O mapeamento sistemático anual da Plataforma Mapbiomas (MAPBIOMAS, 2018), utilizado na base de dados dessa pesquisa, também confirma, num recorte temporal menor (1985 e 2017), o indiscutível processo de desmatamento desses biomas. Os gráficos oriundos dos mapas de uso e cobertura da terra, indicam que a classe Floresta (que hierarquicamente engloba as classes Florestas Naturais, Formações Florestais, Formações Savânicas, Mangue e Florestas Plantadas) apresentou diminuição ascendente na área total de dois (Amazônia e

Caatinga) dos três biomas indicados pelas árvores de decisão. Segundo estatísticas publicadas na plataforma Mapbiomas, entre 1985 e 2017, o bioma Amazônia perdeu aproximadamente 40,8 milhões de hectares de floresta (figura 4.10), o que equivale a 11,8% da cobertura. Já o bioma Caatinga perdeu 5 milhões (figura 4.11), correspondente a 10,5% de decréscimo da classe.

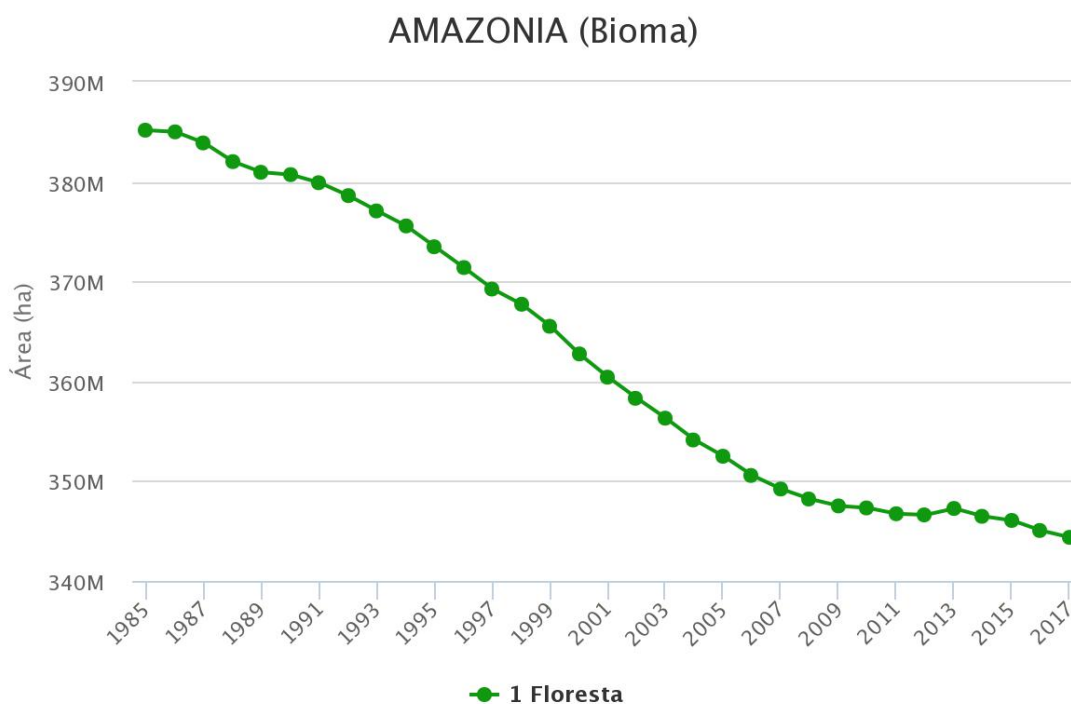


Figura 4.10: Área da classe Floresta no bioma Amazônia entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018)

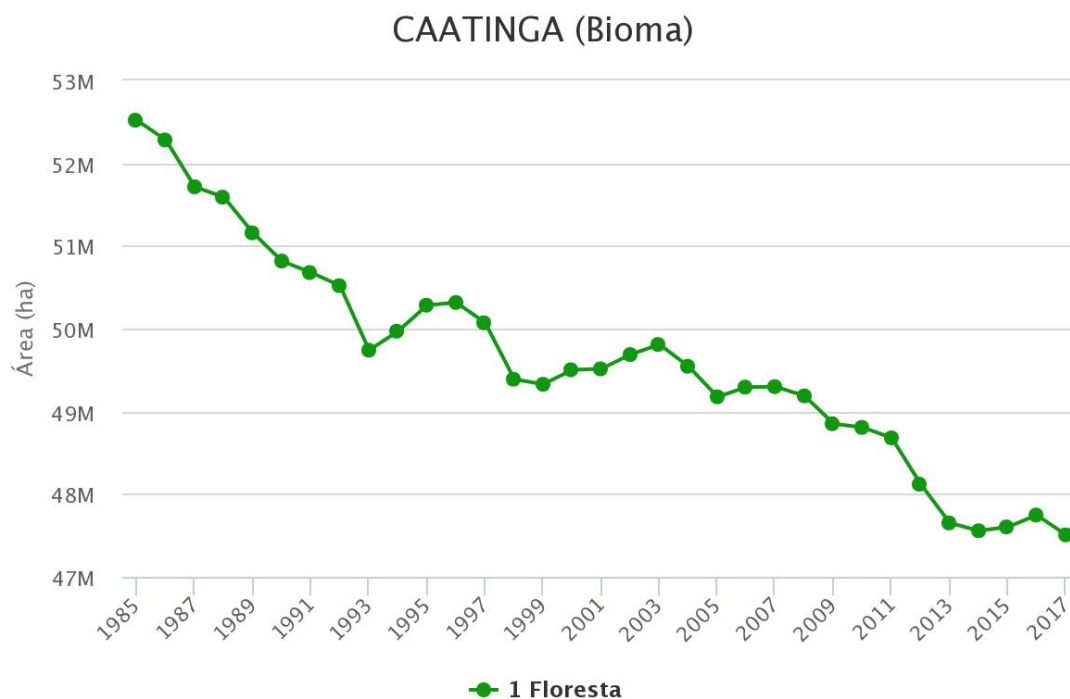


Figura 4.11: Área da classe Floresta no bioma Caatinga entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018)

O processo é inverso para a classe Agropecuária (que hierarquicamente inclui as classes Agricultura, Pastagem e Agricultura ou Pastagem) em ambos os biomas. A Amazônia teve o incremento exponencial da classe, com o aumento de 39,6 milhões de hectares (figura 4.12), ou seja, um crescimento de 74,6% da área de Agropecuária. Na Caatinga a classe Agropecuária teve incremento de 5,3 milhões de hectares (figura 4.13), o que corresponde ao aumento de 17,5 % da classe.

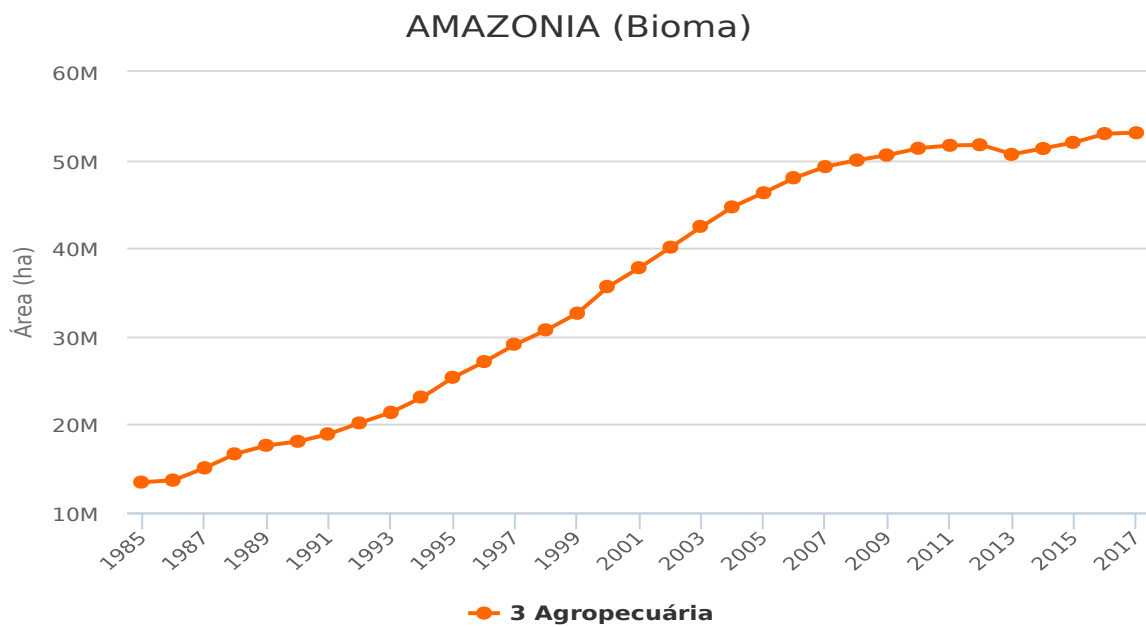


Figura 4.12: Área da classe Agropecuária no bioma Amazônia entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018)

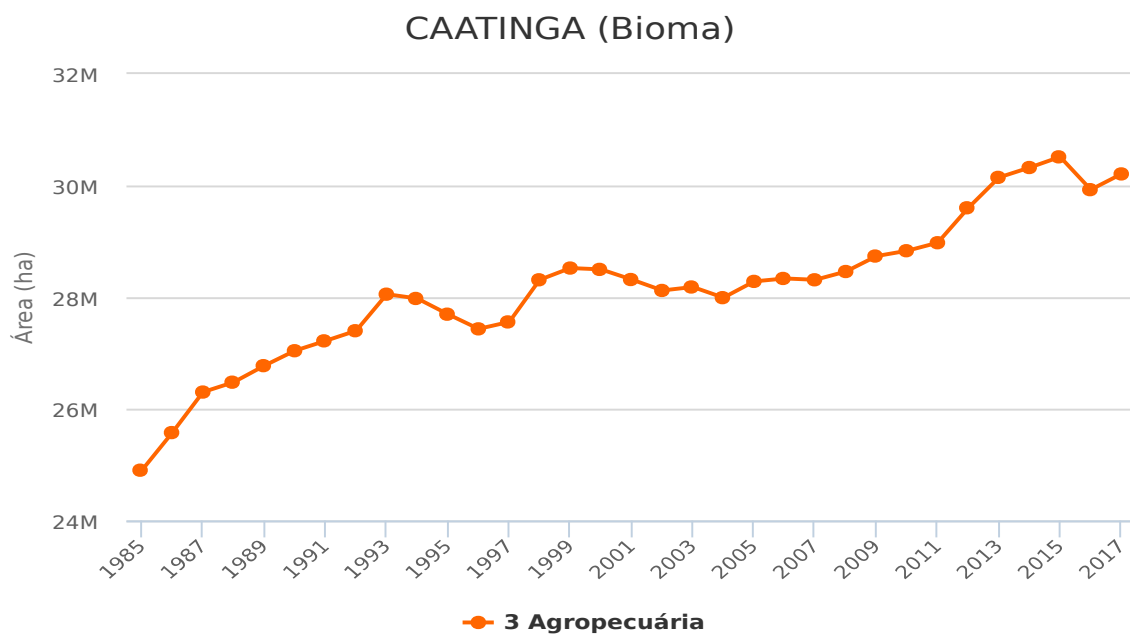


Figura 4.13: Área da classe Agropecuária no bioma Caatinga entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018)

Interessante notar que os valores de perdas de florestas e ganho de Agropecuária são quase que equivalentes para ambos os biomas. No bioma Amazônia houve perda de 40,8 milhões de florestas e incremento de 39,6 milhões de áreas de uso agropecuário. Já o bioma Caatinga ocorreu perda de 5 milhões de áreas florestais e aumento de 5,3 milhões de áreas de uso agropecuário.

No bioma Mata Atlântica o padrão de decréscimo ascendente de florestas não ocorreu da mesma maneira que nos biomas Amazônia e Caatinga (figura 4.14). Verifica-se que entre 1985 e 2003 ocorreu perda de 1,2 milhão de hectares de florestas, porém ocorreu o processo de compensação dessas perdas entre 2003 e 2017, com o aumento de 1,7 milhão de hectares florestais recuperados. Essa informação sobre áreas recuperadas confirma-se no gráfico sobre a classe Florestas Plantadas na Mata Atlântica (figura 4.15), no qual se verificou, entre 1985 e 2017, ganho de 2,3 milhões de hectares de florestas plantadas.

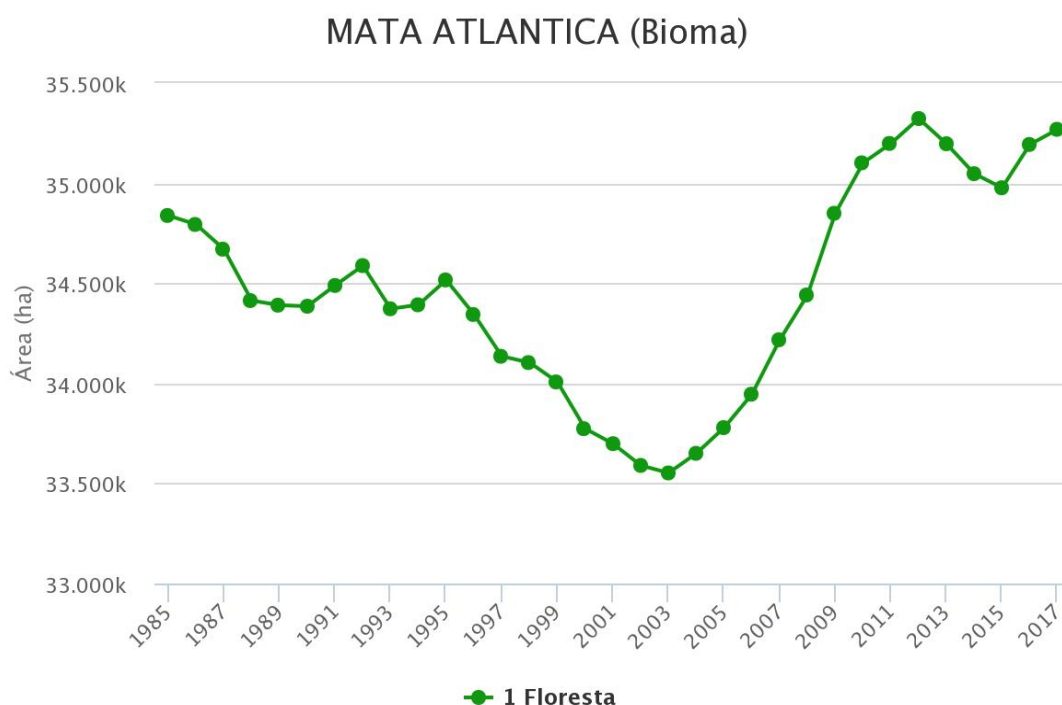


Figura 4.14: Área da classe Floresta no bioma Mata Atlântica entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018)

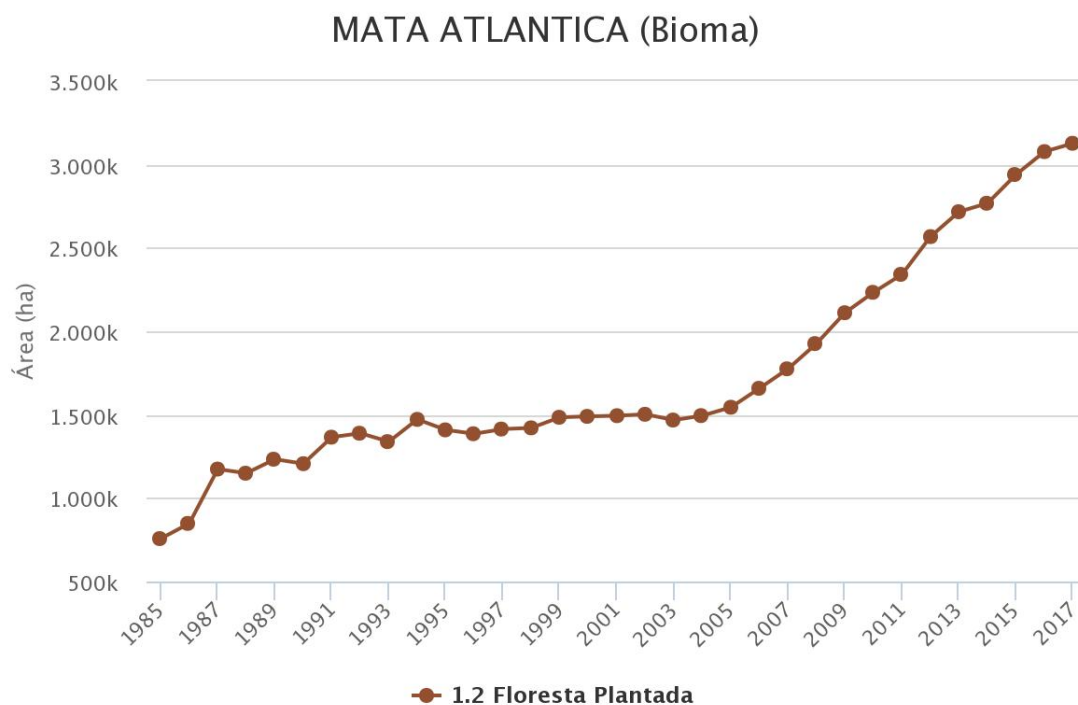


Figura 4.15: Área da classe Floresta Plantada no bioma Mata Atlântica entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018)

Taubert *et al.* (2018) destacam que o reflorestamento é uma importante ferramenta de conservação para mitigar a extinção de espécies devido a perda de *habitat*, fragmentação e isolamento, especialmente se envolver a proteção de grandes áreas florestais. Assim, a restauração de *habitats* pode facilitar, por exemplo, o aumento do tamanho da população das espécies e conectar fragmentos e áreas protegidas, isso se a floresta restaurada puder promover a imigração e o fluxo gênico de populações anteriormente isoladas (HYLANDER e EHRLÉN, 2013). Segundo Newmark *et al.* (2017), a restauração de *habitats*, em áreas de florestas anteriormente conectadas, deve utilizar corredores de 1 km de largura entre os fragmentos florestais mais extensos, intactos e mais próximos, pois esses corredores podem ter um efeito positivo sobre a expansão da população de animais silvestres. Embora a regeneração completa desses corredores provavelmente ocorra entre 10 e 40 anos, espécies de primatas com tendência à extinção, como micos-leões-dourados (*Leontopithecus rosalia*, *Leontopithecus chrysomelas*) podem contribuir na dispersão de mudas e sementes (DOSEN *et*

al., 2017).

A classe Agropecuária também experimentou oscilações entre ganhos e reduções de áreas da classe no bioma Mata Atlântica (figura 4.16). Entre 1985 e 2003 ocorreu uma tendência de acréscimo dessas áreas, com 991,6 milhares de hectares. Já entre 2003 e 2017 ocorreu redução vertiginosa dessas áreas, com a diminuição de 1,3 milhão de hectares. As variações no tamanho das áreas destinadas a classe Agropecuária parecem estar bastante alinhadas com a tendência de recuperação de florestas, tendo sido evidenciado o ganho de 1,7 milhão entre 2003 e 2017. A manutenção ou a recuperação de zonas de APPs e a criação de Reserva Legal em propriedades rurais certamente contribuíram tanto para a recuperação das áreas, com o plantio de árvores nativas, quanto para a regulação sobre os limites máximos de exploração da terra, devendo-se respeitar margens dos rios, mananciais, nascentes, topo de morros e proteção da erosão do solo. Nesse contexto destaca-se ainda a Lei da Mata Atlântica nº 11.428, sancionada em 2006, que dispõe sobre a utilização e proteção da vegetação nativa do biomas, trouxe dentre outras providências a efetivação dos Planos Municipais de Conservação e Recuperação e a regulação do Fundo de Restauração.

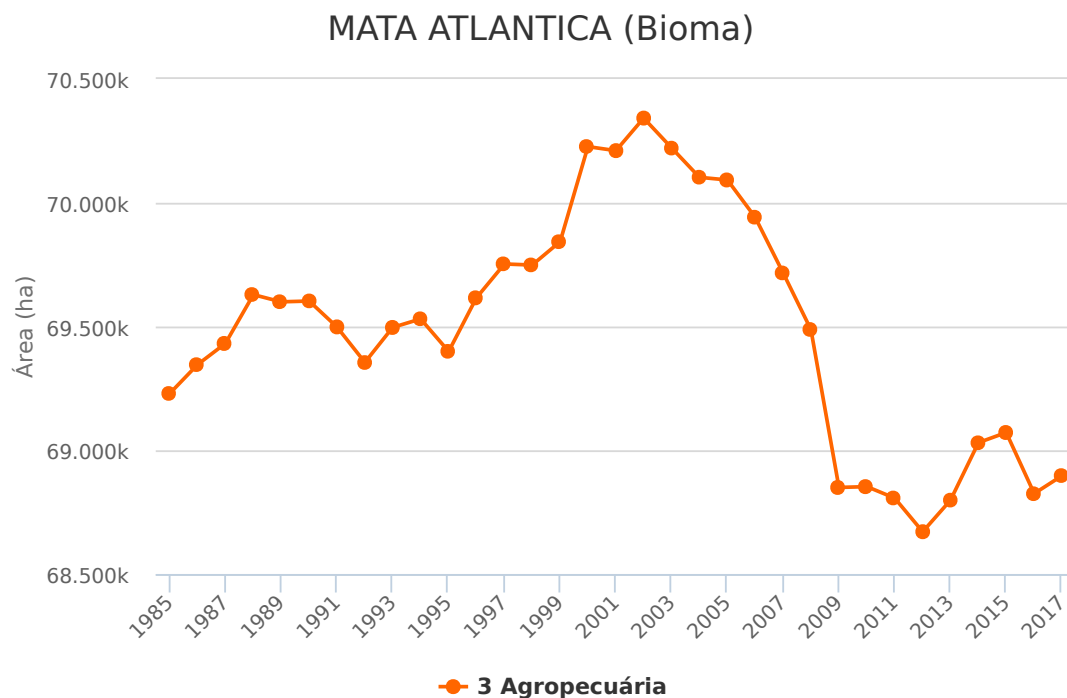


Figura 4.16: Área da classe Agropecuária no bioma Mata Atlântica entre 1985 e 2017
(Fonte: MAPBIOMAS, 2018)

De maneira pontual, complementar e excepcional à prerrogativa da modelagem baseada em dados, na qual a discussão deve ser concentrada e orientada pelos dados evidenciados pelos modelos, considera-se pertinente para esta discussão destacar sobre o acentuado decréscimo de áreas florestais ocorrida no bioma Cerrado (WWF, 2018) (figura 4.17). Esta exceção deve-se por este se tratar de um bioma com amplo e recorrente histórico de ocorrência de febre amarela silvestre, além de ser o bioma de separação entre a Amazônia e a Mata Atlântica, evidenciados nos modelos.

A classe floresta no bioma Cerrado apresentou grande decréscimo de área, da ordem de 20,7 milhões de hectares, o que representa 22,3% de perda, enquanto que a classe Agropecuária obteve um acréscimo de 22,1 milhões de hectares, o que representa 25,9% de ampliação das áreas agropecuárias no bioma (figura 4.18).

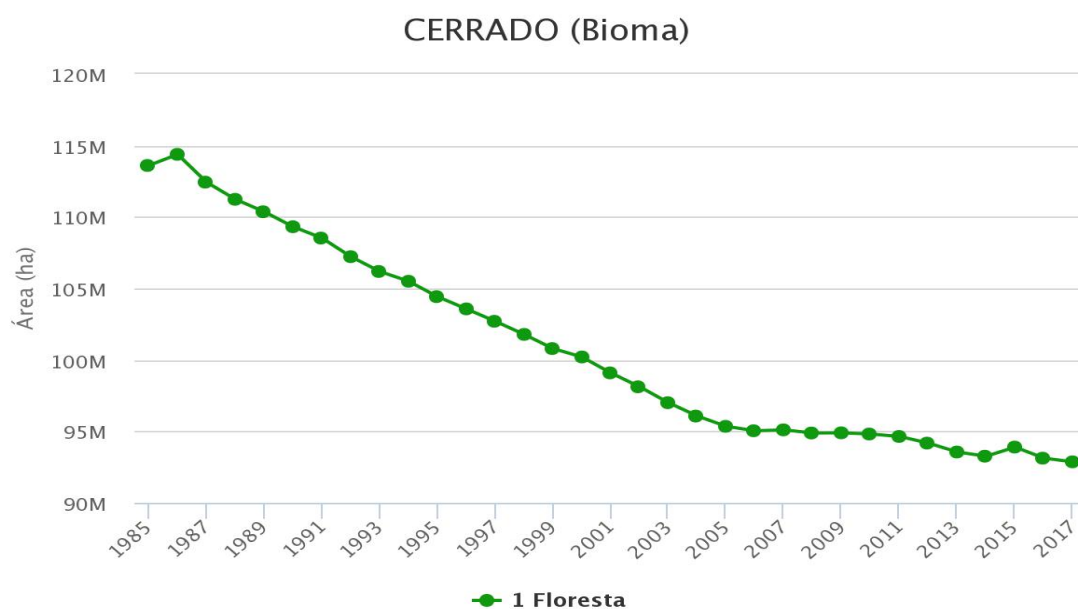


Figura 4.17: Área da classe Floresta no bioma Cerrado entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018)

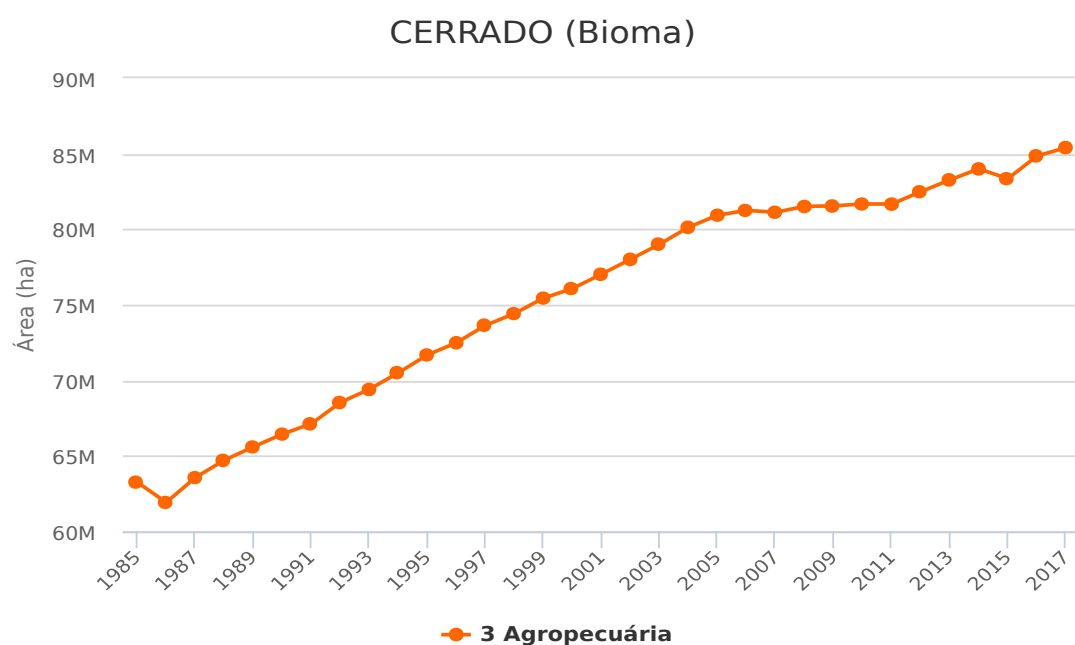


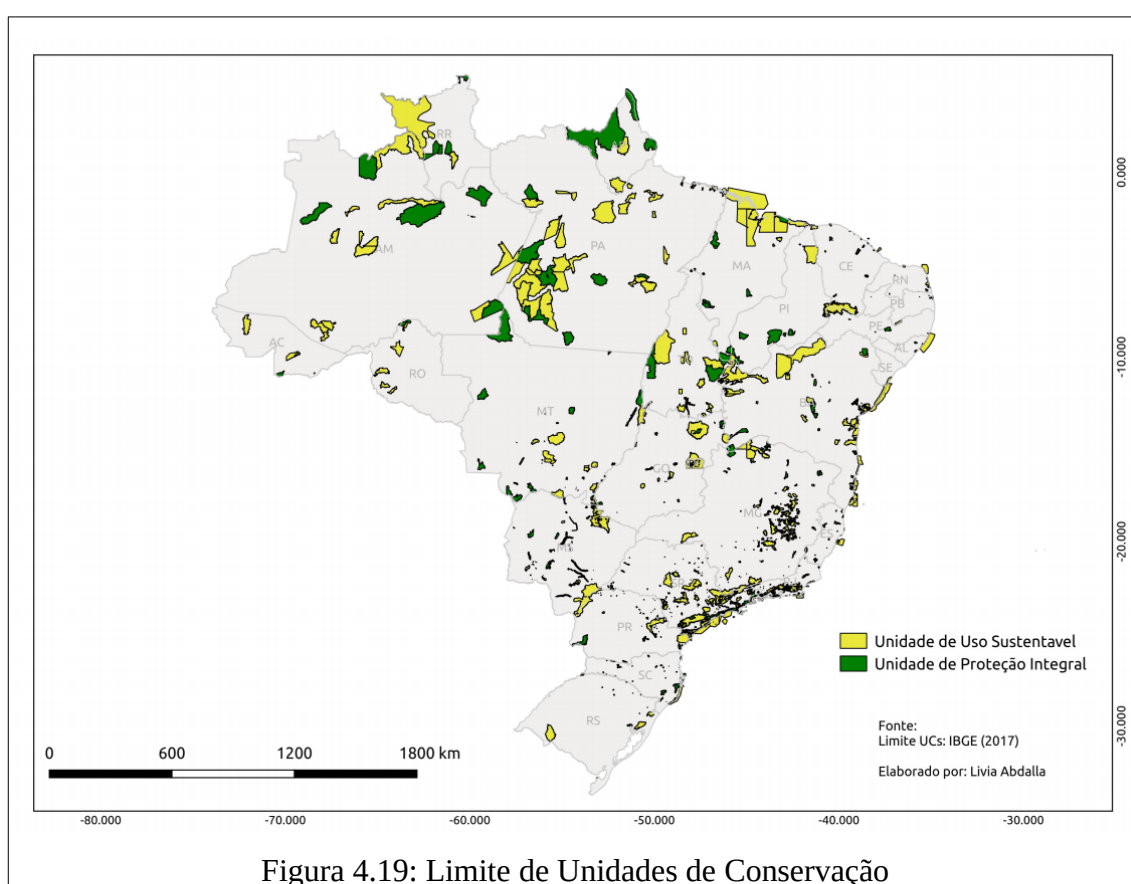
Figura 4.18: Área da classe Agropecuária no bioma Cerrado entre 1985 e 2017 (Fonte: MAPBIOMAS, 2018)

Embora algumas espécies de primatas possam sobreviver temporariamente em florestas exploradas, tanto a extração legal quanto ilegal de madeira resultam em uma diminuição da cobertura do dossel, redução de umidade no subsolo e aumento da incidência de incêndios sobre a vegetação e o solo. Esses efeitos tornam-se evidentes com a redução de crescimento da floresta e impacta negativamente a regeneração de grandes espécies de árvores que fornecem alimento, locais de abrigo e refúgio para primatas (ALISJAHBANA e BUSCH, 2017; LEWIS *et al.*, 2015; PERES, 2001; BANCO MUNDIAL, 2016). O desmatamento gera impactos negativos para os primatas (dos gêneros *Callithrix*, *Alouatta*, *Cebus*, *Ateles*, *Aotus*, *Saimiri* e *Sapajus*) que sobrevivem nessas pequenas porções florestais, como por exemplo: a concentração de indivíduos, o que favorece a transmissão do vírus pelos vetores; a competição intra e interespecífica, além do estresse, promovido pela limitação de alimentos, espaço, entre outros recursos vitais. Esses impactos atrelados podem promover a alteração da imunidade individual, tornando-os suscetíveis às infecções; a dispersão para outras áreas em busca de melhores condições, o que pode aumentar o estresse e a dispersão viral.

A partir dos dados verificados na Plataforma Mapbiomas (2018) e das publicações recentes do WWF (2018) e MMA (2018), verifica-se que, principalmente nos biomas Amazônia e Cerrado, foram grandes as perdas sobre as áreas de vegetação, enquanto que o bioma Mata Atlântica demonstrou uma tendência de recuperação das áreas florestais. Importante destacar ainda que os atributos correspondentes às feições/categorias de vegetação (Formações florestais, Formações savânicas, Florestas plantadas, Vegetação Campestre) foram frequentemente indicados pelos modelos dessa tese (tabela 16), o que reforça a pertinência da discussão e dos apontamentos realizados neste subcapítulo. Dessa forma considera-se a possibilidade de ter ocorrido um movimento de migração, ou mesmo fuga, das espécies de primatas (especialmente do gênero *Callithrix*, bastante adaptado a ambiente de matas em regeneração) que habitam as áreas florestais, principalmente no Cerrado, para os fragmentos em regeneração na Mata Atlântica. Essa conjectura pode indicar um caminho para a compreensão sobre a magnitude da emergência do recente surto nos domínios de Mata Atlântica dos estados de Minas Gerais, Espírito Santo, São Paulo e Rio de Janeiro.

4.2.4 Unidades de conservação

Os dados correspondentes aos limites de Unidades de Conservação foram indicados seis vezes nas árvores de decisão dos modelos experimentais que utilizaram o método J48 (figura 4.19). Seguindo a mesma estrutura lógica dos subcapítulos anteriores, será inicialmente apresentado o contexto geral e as principais características sobre as Unidades de Conservação no Brasil e, em seguida, serão discutidos os prováveis vínculos entre áreas protegidas com os eventos epidêmicos/epizoóticos passados e recentes.



As Unidades de Conservação (UCs) têm a função de preservar o patrimônio biológico existente no país, por meio da proteção de espaços naturais compostos por espécies, *habitats* e ecossistemas, e também das águas jurisdicionais do território brasileiro. As UCs, de acordo com a categoria, garantem a utilização sustentável dos recursos naturais, permitindo que comunidades locais desenvolvam atividades econômicas sustentáveis no seu interior ou no

entorno. O Sistema Nacional de Unidades de Conservação (SNUC) foi criado em 18 de junho de 2000, por meio da Lei nº 9.985 do Governo Federal. O SNUC é o responsável pelo estabelecimento de critérios e normas para a criação, implantação e gestão das Unidades de Conservação do Brasil. As Unidades de Conservação (federais, estaduais, municipais ou particulares) integrantes do SNUC dividem-se em dois grupos, com características específicas (forma de proteção e usos permitidos): Unidades de Proteção Integral e Unidades de Uso Sustentável.

O objetivo principal das Unidades de Proteção Integral é preservar a natureza, sendo permitido somente o uso indireto dos seus recursos naturais, com exceção para casos que devem estar previstos na Lei. As seguintes categorias pertencem a este grupo: Estação Ecológica; Reserva Biológica; Parque Nacional; Monumento Natural e Refúgio da Vida Silvestre.

O grupo de Unidade de Uso Sustentável tem como objetivo principal combinar a conservação dos ecossistemas com o uso sustentável de parcela dos seus recursos naturais. As categorias que pertencem a esse grupo são: Área de Proteção Ambiental; Área de Relevante Interesse Ecológico; Floresta Nacional; Reserva Extrativista; Reserva de Fauna; Reserva de Desenvolvimento Sustentável e Reserva Particular do Patrimônio Natural.

Segundo MMA (2014) as unidades de proteção integral exigem de maiores cuidados por sua fragilidade e particularidades ambientais, e as unidades de uso sustentável, possibilitam a utilização de recursos naturais de forma direta e sustentável e, ao mesmo tempo em que são conservados. Assim, as unidades de conservação formam uma estrutura, na qual cada categoria contribui de uma forma específica para a conservação dos recursos naturais (MMA, 2014).

Desde de 1934 que as unidades de conservação promovem a proteção do patrimônio ambiental do Brasil. Atualmente, 1.538.888 km² de área continental (18% do território brasileiro) são protegidas por unidades de conservação e 959.307 km² de áreas marinha (26% do Mar territorial e Zona Econômica exclusiva – ZEE). Essas áreas estão protegidas por 2.180 unidades de conservação continentais (Proteção Integral e Uso sustentável) e 177 UCs em área marinha. O bioma Amazônia é o que mais abrange áreas protegidas, sendo detentor de 1,2 milhões de km² (28% da área do bioma). Os demais biomas protegem percentuais

menores de suas áreas, em ordem decrescente de área total (km²): Cerrado 170 mil km² (8,34%); Mata Atlântica 105 mil km² (9,4%); Caatinga 72,9 mil km² (8,8%); Pampa 5 mil km² (2,8%) e Pantanal 6,8 mil km² (4,5%) (MMA, 2018).

Em números absolutos, o Brasil ocupa o 3º lugar em quantidade de área continental destinada a unidades de conservação, ficando atrás somente dos Estados Unidos (2.607.132 km²) e Rússia (1.543.466 km²) (MMA, 2018). As florestas (naturais ou plantadas) representam cerca de 516 milhões de hectares ou 60,7% do território nacional, o que nos eleva ao status de país com a flora mais rica e com a maior biodiversidade do planeta. A principal maneira de preservação de florestas, e proteção da diversidade biológica abrigada por elas, é por meio de Unidades de Conservação (MMA, 2014). Apesar desse cenário favorável de conservação, o Brasil enfrenta muitos desafios para consolidação, manutenção ou ampliação das áreas de Unidades de Conservação. No intuito de formalizar esse desafio o Brasil definiu em 2006 as metas nacionais para conservação de pelo menos 30% da Amazônia e 10% dos outros biomas. Essas metas têm sido utilizadas como objeto norteador da política da biodiversidade no Brasil, e que foram definidas em resposta à meta 1 do Plano Estratégico da Convenção sobre Diversidade Biológica das Nações Unidas - CDB.

Estudos evidenciam que as áreas protegidas são ferramentas fundamentais para a conservação de primatas ameaçados, pois fornecem um *habitat* central sustentável. Segundo Estrada *et al* (2018), os primatas no Brasil tem 38% de sua abrangência dentro das áreas protegidas. Strier *et al* (2017) destacam que aproximadamente 80% do total de localidades da Mata Atlântica, onde os muriquis (*Brachyteles spp.*) atualmente habitam, são áreas protegidas (unidades privadas ou governamentais – estaduais e federais). Joppa e Pfaff (2009) destacam que, globalmente, as áreas protegidas têm baixo valor comercial e pequena demanda por conversão de terras, enquanto que as áreas florestais de terras baixas e ricas em primatas, encontram-se sob risco, pelo interesse no corte raso para comercialização de madeira, ou pelas oportunidades lucrativas que as florestas de terras baixas oferecem por serem adequadas à agricultura mecanizada (VENTER *et al.*, 2014). Portanto, os indivíduos fora das áreas protegidas estão sob maior risco (DAVENPORT *et al.*, 2017), sendo cada vez mais ameaçadas pela expansão agrícola, exploração madeireira e caça ilegal (GOUVEIA *et al.*, 2017; MASCIA *et al.*, 2014; ROVERO *et al.*, 2015; SPRACKLEN *et al.*, 2015; WAEBER *et*

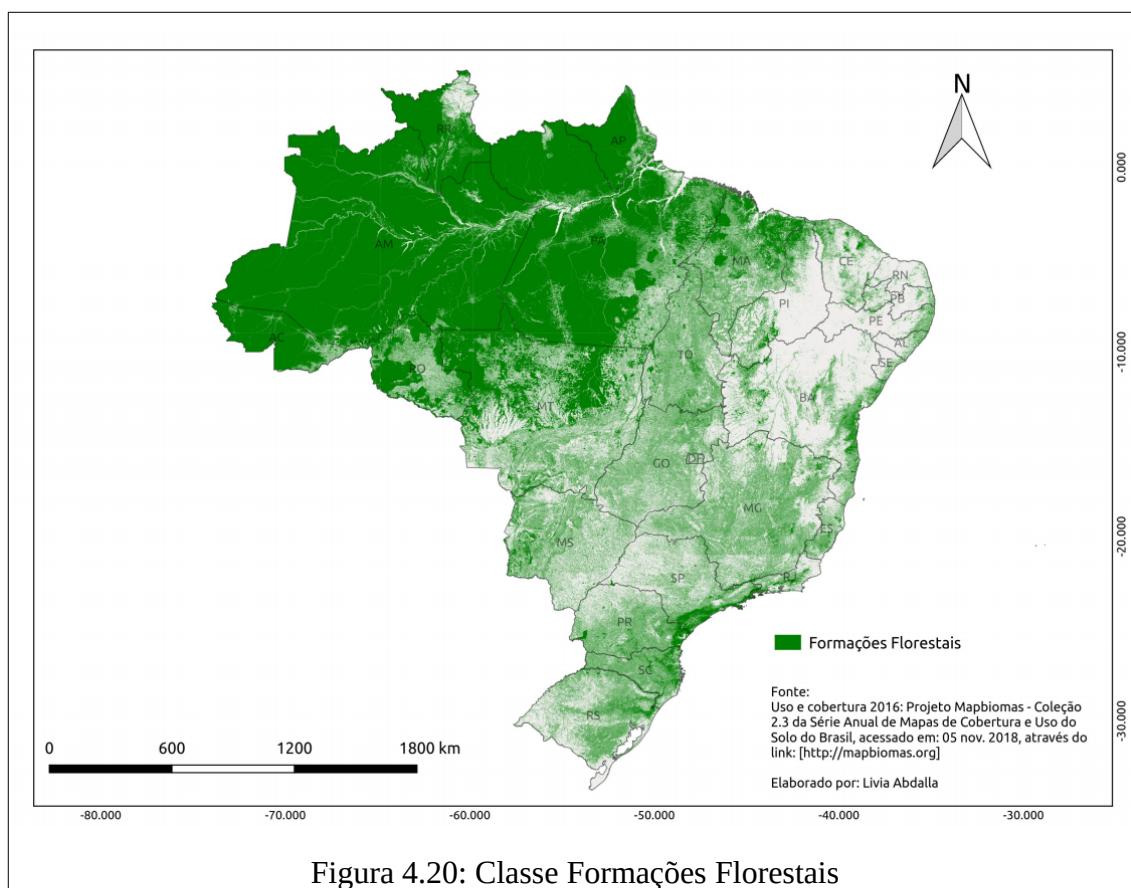
al., 2016).

Destaca-se ainda que o aumento da densidade populacional humana dentro de um raio de distância de até 50 km de uma área protegida é fator que ameaça a área pela prática de atividades ilegais (por exemplo: degradação de *habitats*, caça esportiva e pesca ilegal). Assim, considerando tantos interesses comerciais/financeiros e suas ameaças, Butchart *et al* (2015) destacam que, mesmo com o cumprimento das metas globais para a manutenção de áreas protegidas, isto será insuficiente para proteger a biodiversidade, a menos que essas áreas sejam bem gerenciadas e devidamente delimitadas. Assim, verifica-se que a gestão e a preservação de UCs fazem parte de uma estratégia fundamental para a manutenção de populações saudáveis de primatas, o que pode representar uma redução no risco de grandes eventos epizooticos de febre amarela, pois a elevada quantidade de indivíduos promove o efeito diluição e o ciclo da doença se ocorre de maneira controlada, e com a alta imunidade desses, alguns indivíduos infectados são capazes de desenvolver baixas taxas de viremia e, portanto, não promovem a ampliação da circulação do vírus, quando em contato com vetores ate então não infectados. Ao contrário, ou seja, a perda ou abandono de áreas protegidas aumenta os riscos de sobrevivência para diferentes espécies de primatas, ampliando-se assim as condições de acometimento à doenças como a febre amarela.

4.2.5 Tipos de uso e cobertura da terra

Dentre as dezesseis classes temáticas que compõem a camada uso e cobertura da terra, foi frequente a indicação de sete classes: Formações florestais; Agricultura; Pastagem; Formações savânicas; Vegetação campestre; Áreas úmidas naturais não florestais; Corpos d'água. Neste subcapítulo também serão indicadas as definições de cada classe, principais características e suas extensões geográficas, para em sequência apontar/discutir as possíveis relações com os eventos da FAS.

A classe Formações florestais foi indicada onze vezes pelas árvores de decisão dos modelos experimentais utilizando o método J48. Essa classe corresponde ao agregado de formações florestais naturais, podendo tratar-se de ampla diversidade de fitofisionomias florestais (MAPBIOMAS, 2018) (figura 4.20).



Segundo IBGE (2004), as formações florestais são compostas pelas florestas ombrófilas e estacionais situadas tanto na região amazônica quanto nas áreas extra-amazônicas, mais precisamente na Mata Atlântica. Na Amazônia, predominam as florestas ombrófilas densas e abertas, com árvores de médio e grande porte, com ocorrência frequente de cipós, bromélias e orquídeas. As florestas extra-amazônicas correspondem as formações florestais que compõem a Mata Atlântica, onde predominam as florestas estacionais semidecíduais (em que 20 a 50% das árvores perdem as folhas no período seco do ano) e as florestas ombrófilas densas e mistas (incluindo araucária). Em ambos os conjuntos florestais ocorrem ainda as florestas estacionais decíduais (em que mais de 50% das árvores perdem folhas no período seco). Estudos indicam que nas regiões de floresta alta, como a amazônica, as espécies podem se especializar em diferentes *habitats* na floresta, inclusive em diferentes estratos da vegetação (WARNER, 2002; SCHWARZKOPF e RYLANDS, 1989), aumentando a densidade demográfica local. Já em formações vegetais mais baixas, como as do Cerrado, a diversidade

de *habitats* é menor e os grupos de primatas precisam competir de forma mais intensa para manter o seu território (LAZARO-PEREA *et al.*, 1999; DECANINI e MACEDO, 2008).

As áreas mínimas de fragmentos florestais, por parte dos primatas, podem variar de acordo com a espécie. Por exemplo: o *Callithrix penicillata* possui área de vida estimada em 3,5 ha; já o *Alouatta belzebul* necessita de uma área estimada entre 13,5 a 18,05 ha na região amazônica (PINA *et al.* 2002) e entre 4,75 a 9,5 ha na Floresta Atlântica (BONVICINO, 1989);

As espécies de vetores transmissores da FAS também possuem como *habitat* regiões florestais, sendo que as espécies do gênero *Haemagogus* demonstram capacidade de deslocamento de até 6km (GOMES *et al.*, 2008), logo, são espécies com grande capacidade de circulação entre muitas porções de matas habitadas por diferentes espécies de PNHs.

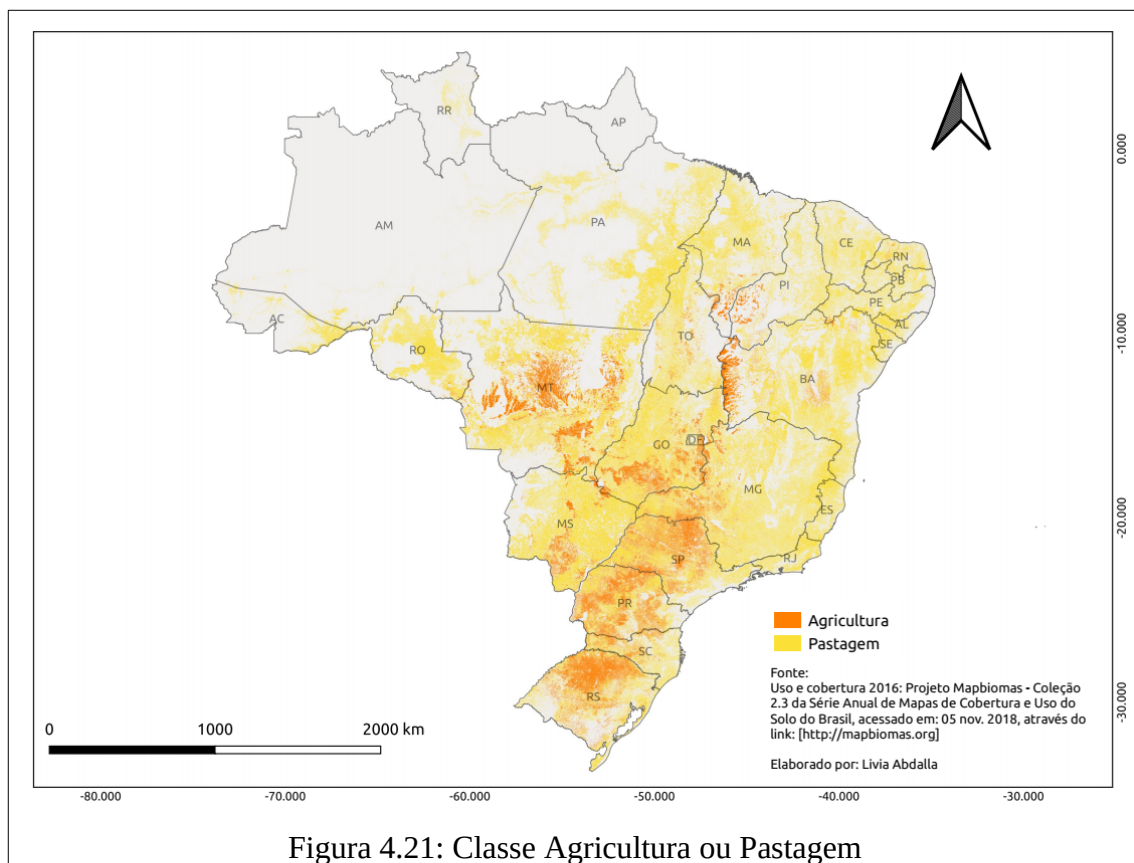
Segundo Fioravanti (2018) em estudo recente realizado pela Superintendência de Controle de Endemias - SES-SP foram indicadas áreas prioritárias para vacinação baseadas na demarcação de corredores ecológicos ativos e potenciais. O estudo utilizou dados (georreferenciamento e data) de primatas infectados com o vírus da FAS no estado de São Paulo, durante o mais recente surto, e investigou o padrão (direção e velocidade) de dispersão espacial e temporal do vírus entre os fragmentos florestais. A partir da verificação do padrão de dispersão foi calculada a velocidade média diária de deslocamento do vírus (~2,7 km em dias quentes e ~0,5 em dias frios, na direção norte-sul), e assim, foram estimadas as datas e áreas prováveis de chegada do vírus no estado de São Paulo. Dessa maneira foi possível priorizar adequadamente as ações de imunização para áreas que foram previamente indicados pelo estudo. Portanto, o conhecimento sobre o cenário de fragmentação florestal do estado (mais os dados do local/data exatos da epizootia) foi base para avançar na previsão do caminho de disseminação do vírus.

Por tudo isso, a relação da classe de formações florestais com todo histórico de manifestação da doença demonstra ser clara e direta, já que ecossistemas florestais tratam-se do meio natural e preferencial das espécies de primatas. Segundo Costa (2005) a análise espacial dos casos humanos de FAS ocorridos no Brasil nos últimos dez anos mostra um padrão de ocorrência em áreas com vegetação do tipo savana e de florestas ombrófilas.

Segundo dados da *Global Forest Watch* – GFW (2018), o Brasil reduziu 46,4 milhões de

hectares de florestas entre 2001 e 2016. Em biomas como o Cerrado, as florestas secas continuam com elevadas taxas de desmatamento (STRASSBURG *et al*, 2017). Na contramão das iniciativas de redução do desmatamento, Brancalion *et al* (2016) destacam que a aprovação da Lei de Proteção da Vegetação Nativa nº 12.651, em 2012, enfraqueceram as regras de proteção da vegetação nativa e, portanto, verificou-se aumento acentuado do desmatamento em 2016. Segundo Potapov *et al* (2017), as perdas de Florestas Primárias no Brasil corresponderam a 6,2%, entre 2001 e 2016. As práticas de desmatamento, principalmente em pequenas áreas, tornam difícil a identificação e o monitoramento por sensoriamento remoto (KALAMANDEEN *et al*, 2018). Como exemplo, destaca-se que, na Amazônia, o número de novas pequenas clareiras (< 1ha) aumentaram 34% entre 2001 e 2007, e que a perda de florestas de baixa densidade expandiu-se acentuadamente entre 2008 e 2014. Entre 2001 e 2014, as porções florestais desmatadas com menos de 6,25ha representaram 34% do total da perda de floresta amazônica (KALAMANDEEN *et al*, 2018). Nesse contexto, Hanski *et al* (2013) destacam que o desmatamento promove a redução ou eliminação de áreas florestais, o que ameaça a sobrevivência de espécies de primatas, e leva a maiores taxas de extinção nas populações locais.

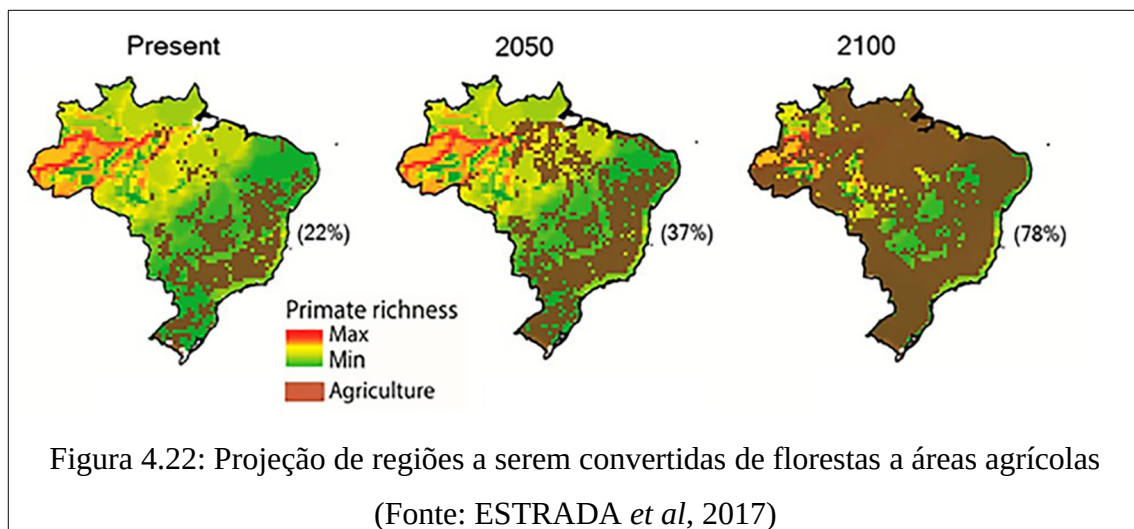
As classes Pastagem e Agricultura também demonstraram notória importância, com frequência de aparição de seis e três vezes, respectivamente, nas árvores de decisão do método J48 (figura 4.21). A classe Pastagem corresponde a áreas de campos naturais ou outras pastagens relacionadas com a atividade pecuária. Já a classe Agricultura corresponde a áreas de agricultura anual, semiperene (cana de açúcar) ou mosaico de cultivos (MAPBIOMAS, 2018).



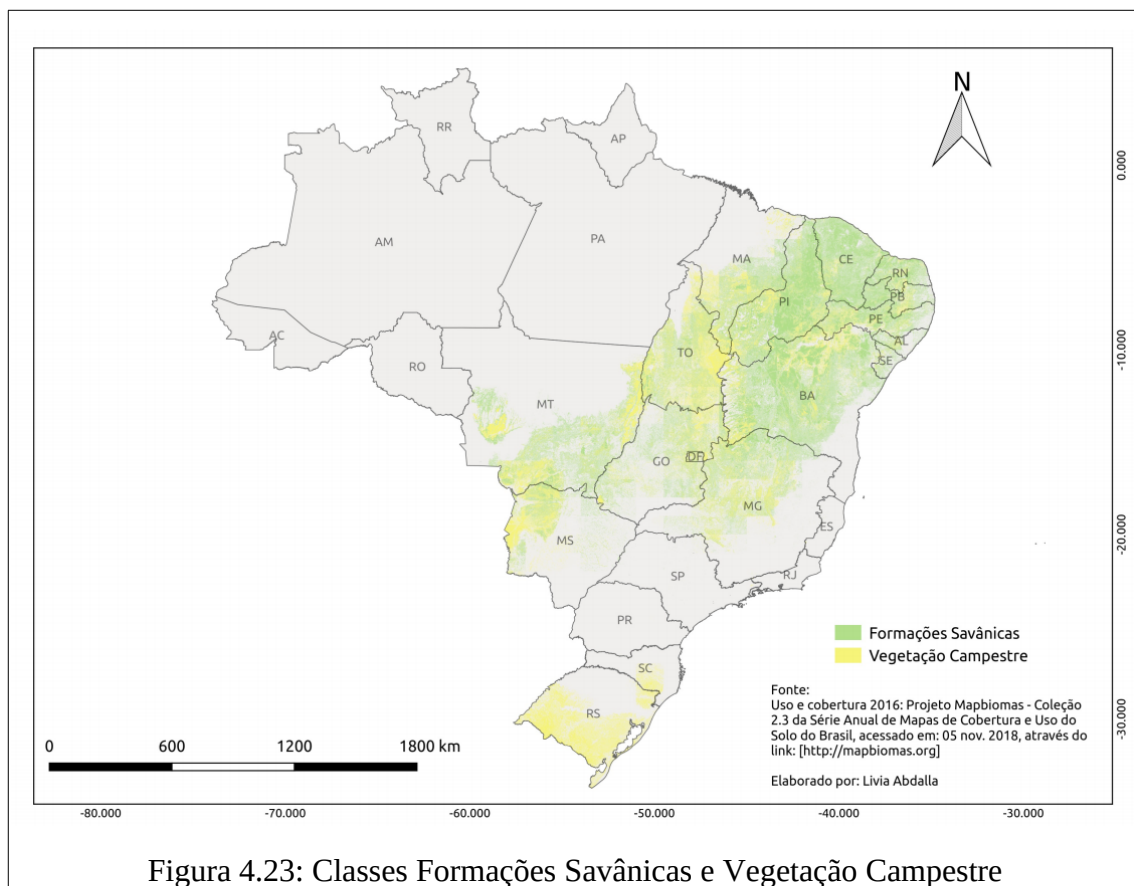
A produção agropecuária no Brasil é de expressiva importância no cenário nacional e internacional, sendo atualmente o terceiro maior produtor e exportador agrícola do mundo (FAO, 2018). A monocultura é prática comum para a produção de alimentos em larga escala, na qual se explora o solo para um único tipo de cultura. No Brasil, 30% do desmatamento entre 2000 e 2010 foi impulsionado pelas demandas globais por exportações de carne bovina e soja (KARSTENSEN, 2013). A eliminação de diferentes tipos de coberturas florestais originais para o monocultivo deprecia a diversidade de espécies de plantas e animais silvestres naturalmente associados a estes ambientes, além de trazer prejuízos ao solo, tornando-o empobrecido, sendo necessário, portanto, a aplicação de adubos e fertilizantes químicos para reciclagem de nutrientes. Para o controle de insetos e pragas nas propriedades monoculturas, é comum a utilização de agrotóxicos e pesticidas, que são aplicados diretamente e em grandes quantidades, prejudicando as espécies nativas. A produção agropecuária é atividade geralmente desenvolvida em regiões rurais, que coexistem ou possuem proximidade com

ambientes naturais florestais, onde é comum a prática de substituição de diversas porções de matas em pastos, por meio de queimadas ou corte raso para preparação de novas áreas de plantio e/ou pecuária extensiva de criação de animais para comercialização. O desmatamento promovido pela agroindústria é ameaçador (principalmente nas regiões Norte e Centro-Oeste) para as diversas espécies de animais silvestres, especialmente muitas espécies de PNHs (*Alouatta caraya*, *Alouatta guariba*, *Callithrix sp.*, *Callicebus nigrifrons*, entre outras) que sobrevivem em porções florestais reduzidas, confinadas e cercadas por extensas matrizes agrícolas. São múltiplas as ameaças sobre a sobrevivência e saúde dessas espécies, tornando-as extremamente vulneráveis e ameaçadas a diversos tipos de riscos diretos e indiretos (estresse, confinamento, escassez de alimentos, competição interespecífica, redução da imunidade ampliando assim os riscos a doenças infecciosas).

Segundo dados da Organização das Nações Unidas para Alimentação e Agricultura – FAO (2018), ocorreu um aumento de 19,1 milhões de hectares de terras agrícolas, no período de 2001 a 2015. Deve-se destacar que estas informações são fornecidas à FAO pelos próprios governos, os quais podem – ou não – fornecer dados incompletos, o que pode sugerir um valor ainda maior de conversão de áreas naturais em terras agrícolas. Estrada *et al* (2017) previram em um estudo de abrangência global, que as regiões a serem convertidas de florestas para áreas agrícolas correspondem a 68% da área utilizada por primatas, sendo que no Brasil é prevista uma redução de 78% (figura 4.22). Assim, a expansão agrícola prevista globalmente poderá promover conflito espacial com 75% das espécies de primatas no mundo. O autor destaca ainda que, além da conversão de áreas florestais em áreas de expansão agrícola, outras ameaças, como caça, exploração madeireira, mineração, extração de combustíveis fósseis, infecções virais e mudanças climáticas também deverão resultar na contração da faixa de primatas. Considera-se que esse estudo fornece projeção com caráter pessimista, para esta previsão seria fundamental incluir os limites de unidades de conservação, os quais limitariam o avanço de áreas agrícolas sobre as áreas florestais preservadas.



As classes Formações Savânicas e Vegetação Campestre (figura 4.23) tiveram frequência de indicação dez e dois, respectivamente, nas árvores de decisão dos modelos experimentais. As classes Formações Savânicas e Vegetação Campestre também são oriundas dos mapas de uso e cobertura do solo da plataforma Mapbiomas. A classe Formações Savânicas corresponde aos tipos de vegetação com predomínio de espécies de dossel semicontínuo (Savana-Estépica Arborizada, Savana Arborizada). A classe Vegetação Campestre corresponde aos tipos de vegetação com predomínio de espécies herbáceas – Savana-Estépica Parque, Savana-Estépica Gramíneo-Lenhosa, Savana Parque, Savana Gramíneo-Lenhosa (MAPBIOMAS, 2018).



As classes Formações Savânicas e Vegetação Campestre são tipos de vegetação encontradas nas regiões pertencentes principalmente aos biomas Cerrado, Caatinga e Pampa. O Cerrado tem a savana como vegetação predominante, em seu estrato gramíneo-lenhoso, composto predominantemente por árvores esparsas e vegetais rasteiros, apresentando extensas áreas antropizadas. Ao longo dos cursos d'água, as florestas de galeria fazem parte da fisionomia da savana, em terrenos relativamente férteis. Segundo IBGE (2004), as vegetações do tipo campestres são constituídas pela vegetação do tipo abertas, identificadas como: (i) savana, correspondente ao Cerrado, que predomina no Brasil central, ocorrendo também em pequenas áreas em outras regiões do país, inclusive na Amazônia; (ii) savana estépica, que inclui a caatinga nordestina, os campos de Roraima, o Pantanal mato-grossense e uma pequena ocorrência no extremo oeste do Rio Grande do Sul; (iii) estepe, que corresponde aos campos, do planalto e da campanha, do extremo sul do Brasil; (iv) campinarana, um tipo de

vegetação decorrente da falta de nutrientes minerais no solo e que ocorre na Amazônia, na bacia do rio Negro.

A fitofisionomia campestre (Estepe), própria do bioma Pampa, abrange cerca de 63% do Estado do Rio Grande do Sul. Esse tipo de vegetação recobre, em grande parte, superfícies com formas de relevo aplainadas ou suavemente onduladas, conhecidas como “coxilhas” ou Pradarias (Domínio Morfoestrutural e Morfoclimático definido, segundo Ab’Sáber, 1967). Essa paisagem com pastagens naturais favoreceu o desenvolvimento da pecuária extensiva. Outras atividades vêm sendo desenvolvidas na região, como o plantio de florestas comerciais (silvicultura), por parte de empresas do setor de papel e celulose, que ocupam grandes extensões de terra para implantação da monocultura de árvores exóticas, principalmente eucalipto, acácia-negra e pinus (CARVALHO, 2018). As atividades de monoculturas consomem quantidades elevadas de água, o que ampliam as preocupações com a perda dos recursos hídricos. Lavouras de arroz tem promovido ainda o armazenamento de água por meio do barramento de rios, arroios e sangas, sendo que o cultivo ocorre muitas vezes nas cabeceiras de drenagens, em áreas legalmente protegidas, que são importantes para reprodução e alimentação de várias espécies animais.

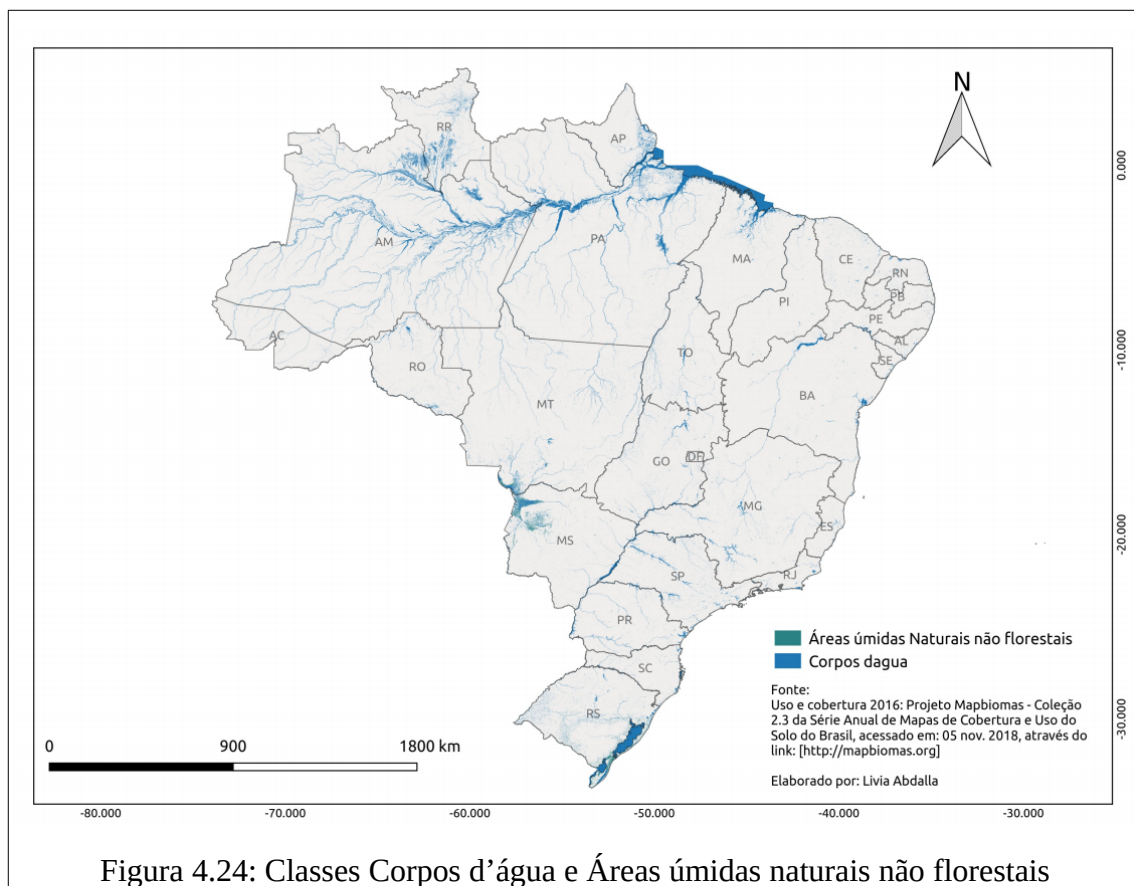
A relação demonstrada pelas árvores de decisão com a classe remete ao surto de febre amarela ocorrido entre 2007 e 2009, o qual concentrou maior número de epizootias no Rio Grande do Sul. Durante o surto milhares de bugios (830 *A. caraya* e 1.183 *A. guariba clamitans*) morreram entre outubro de 2008 a junho de 2009 no estado do Rio Grande do Sul (ALMEIDA *et al.*, 2011), levando à reclassificação do status de conservação de *Alouatta caraya* de Vulnerável para Ameaçada. As epizootias do Rio Grande do Sul ocorreram em áreas com vegetação do tipo estepe, caracterizada por campos com formações gramíneo-lenhosas, intercaladas com capões de matas.

Ambas as classes (Formações Savânicas e Vegetação Campestre) correspondem também à paisagem observada em grande parte dos estados de Goiás, Tocantins, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Bahia, Maranhão, Piauí, Rondônia, Paraná, São Paulo e Distrito Federal. Em todas essas regiões (exceto MA e PI) ocorreram manifestações de surtos e epizootias de FAS. Estudos apontam para ocorrência de 15 táxons de primatas neotropicais no Cerrado, sendo elas: *Alouatta caraya*, *A. guariba*, *A. belzebul*, *Aotus azarae azarae*, *A. a.*

infulatus, *Callicebus nigrifrons*, *C. personatus*, *Callithrix geoffroyi*, *C. jacchus*, *C. penicillata*, *Mico melanurus*, *Sapajus cay*, *S. libidinosus*, *S. nigritus* e *S. xanthosternus* (HIRSCH *et al.* 2002; OLIVEIRA, 2015).

Interessante ainda destacar que, apesar de o Cerrado ser composto em sua maior parte por savanas e campos abertos, a fauna desse bioma tem uma associação muito forte com as florestas, de modo que mais da metade dos mamíferos e aves presentes no Cerrado é dependente dos ambientes florestais (REDFORD e FONSECA, 1986; SILVA e BATES, 2002). Isso teria ocorrido em função do avanço da Floresta Amazônica e Atlântica nos períodos úmidos e da Caatinga e do Chaco nos períodos mais secos (variações climáticas ocorridas ao longo do Quaternário). Esses fatos levaram as espécies da fauna dependentes de florestas a se refugiarem nos remanescentes florestais (SILVA e BATES, 2002).

As classes Corpos d'água e Áreas úmidas naturais não florestais (figura 4.24) tiveram frequência de três e duas aparições nas árvores de decisão. A classe Corpos d'água correspondem a rios, lagos, represas, reservatórios e outros corpos d'água e a classe Áreas úmidas naturais não florestais correspondem a Áreas inundáveis, localizadas ao longo dos cursos de água, em áreas de depressões que acumulam água, ou a uma rede de lagoas interligadas (MAPBIOMAS, 2018).

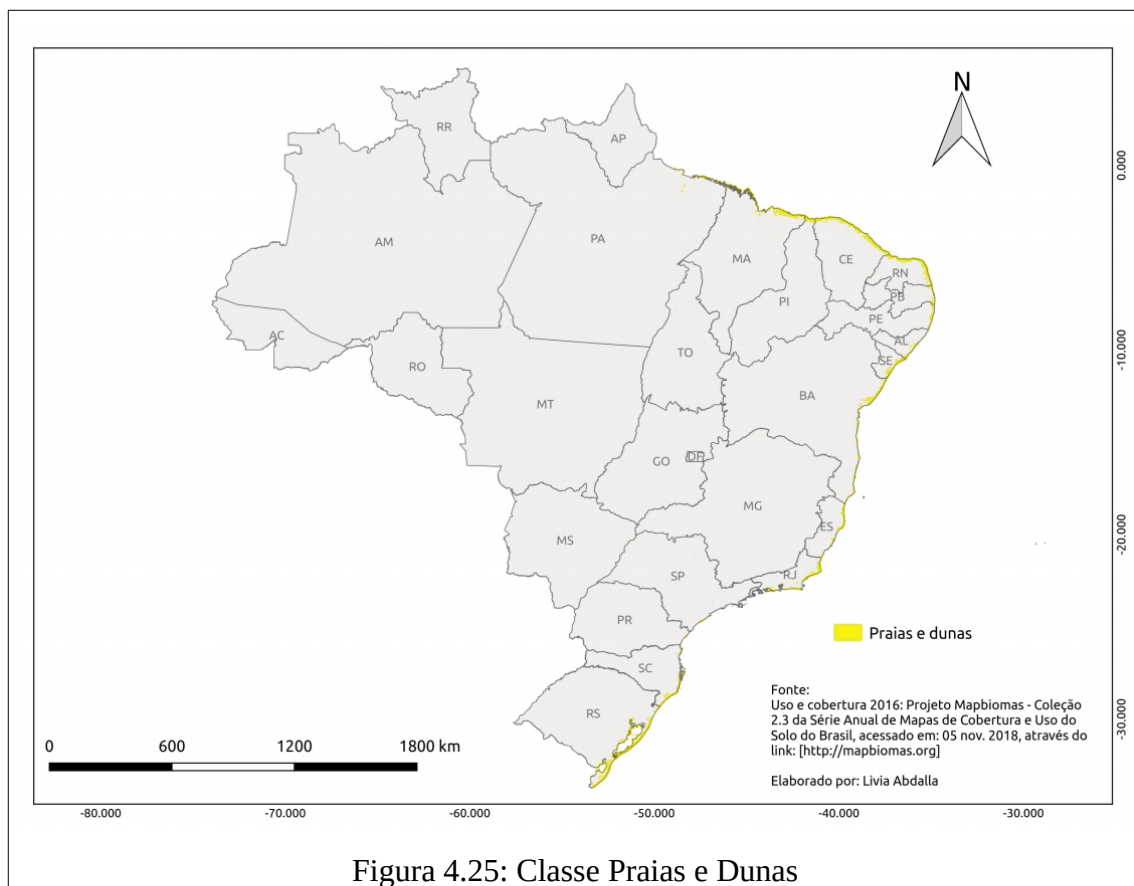


Oliveira-Filho e Ratter (1995) indicaram que espécies de primatas podem utilizar manchas e corredores florestais como rotas de dispersão entre biomas diferentes, já que foi identificado grande número de espécies arbóreas que ocorrem desde a Floresta Amazônica até a Atlântica, atravessando o Cerrado pelos cursos d'água na diagonal noroeste sudeste. Nesse sentido, as classes de corpos d'água podem estar associadas a esses corredores de espécies arbóreas ou matas ciliares (utilizados como meio de deslocamento ou *habitat* de primatas) que conectam as áreas endêmicas com as não-endêmicas. Nesse contexto o Cerrado é considerado o berço das águas brasileiras, de modo que todos os biomas (com exceção do bioma Pampa) possuem nascentes no Cerrado, que representa fortes influências hidrológicas e ecológicas. A existência desses cursos de água contribui para a biodiversidade da região, pois os vales dos rios são rotas preferenciais de migração e dispersão para vários tipos de organismos. A existência de grande quantidade de nascentes e corpos hídricos deve-se ao fato de o Planalto Central Brasileiro ser um grande divisor de bacias hidrográficas e ter altitude média elevada

quando comparado ao restante do País. As principais bacias do bioma são dos rios Tocantins, Araguaia, São Francisco, Paraná e Parnaíba. Além dessas, outras bacias apresentam nascentes neste bioma e drenam áreas em outros biomas (rios Madeira, Tapajós e Xingu, que drenam para o rio Amazonas; bacia dos rios Doce, Jequitinhonha e Pardo, que drenam para o oceano Atlântico; e a bacia dos rios Itapecuru, Mearim, Pindaré e Munim, que drenam para o Golfão Maranhense).

Nesse contexto, importante ainda destacar que as atividades agropecuárias representam grande ameaça à integridade desses ecossistemas hídricos, as quais têm no Cerrado seu principal polo de expansão, ameaçando assim a biodiversidade local e tornando este bioma um possível local de fuga de espécies de primatas para outros biomas.

A classe Praias e Dunas (figura 4.25) teve uma aparição na árvore de decisão do recorte 1999 a 2017. A classe Praias e Dunas corresponde a cordões/faixas arenosas, de coloração predominantemente branca (MAPBIOMAS, 2018), localizadas ao longo de toda a faixa litorânea brasileira.



Considera-se que essa classe demonstrou relevância, justamente pela recente emergência da doença nas regiões litorâneas (próximas a matas) do Sudeste, como nas cidades de Macaé, Angra dos Reis e Paraty, e das baixadas litorâneas, como Casimiro de Abreu. Espécies como o muriqui (*Brachytheles aracnoides*), o guariba (*Alouatta fusca*), o guigó (*Callicebus personatus*), o sagui (*Callithrix aurita*; *C. flaviceps*), o macaco-prego (*Cebus nigritus*), e o mico-leão-dourado (*Leontopithecus rosalia*) habitam regiões de matas costeiras com as fitofisionomias de Restinga arbustivo-arbórea e Restinga arbórea madura (INEA, 2012).

4.2.6 Temperatura e precipitação

As variáveis de temperatura e precipitação evidenciadas pelas árvores de decisão correspondem tanto aos dados anuais da base de dados *Climate Hazards Group Infrared Precipitation with Stations* (CHIRPS) da *Global Precipitation Mission* (GPM) (2000 a 2015), quanto às Normais Climatológicas da base *Worldclim*. Conforme justificado no subcapítulo 4.1.1, neste momento não serão as discutidas especificidades sobre ano ou sobre a estatística vinculada às variáveis indicadas. O foco neste momento é abordar os dados numa análise macro regional sobre as prováveis relações existentes entre o dado/variável temática com o fenômeno de emergência da febre amarela. Portanto, para a apresentação das variáveis de temperatura e precipitação nos mapas deste subcapítulo foram utilizados os *rasters* de Normais Climatológicas do *Worldclim*, pois este fornece um panorama sobre os padrões climatológicos e os valores médios calculados para um período relativamente longo e uniforme (1985 a 2000), que sugerem uma tendência e uma continuidade para os valores médios, tornando-se assim possível alcançar o objetivo de discussão e análise macro espacial e temporal deste subcapítulo.

Sete variáveis de temperatura foram representativas aos modelos, com a frequência de indicação em parênteses: Sazonalidade da Temperatura – bio4 (9); Isotermalidade – bio3 (2); Temperatura média mensal – Agosto – tmean8 (2); Temperatura máxima mensal – Agosto – tmax8 (1); Amplitude térmica anual – bio7 (1); Temperatura média do trimestre mais úmido – bio8 (1); Variação Diurna Média de Temperatura – bio 2 (1).

Treze variáveis de precipitação foram representativas aos modelos, em ordem de frequência: Precipitação total mensal – Outubro – prec10 (8); Precipitação do trimestre mais frio – bio19 (6); Precipitação total mensal – Dezembro – prec12 (5); Precipitação total mensal – Fevereiro – prec2 (3); Precipitação Anual – bio12 (3); Sazonalidade da Precipitação – bio15 (3); Precipitação total mensal – Novembro – prec11 (2); Precipitação do trimestre mais chuvoso – bio16 (2); Precipitação do trimestre mais quente – bio18 (2); Precipitação total mensal – Agosto – prec8 (2); Precipitação total mensal – Junho – prec6 (1); Precipitação total mensal – Julho – prec7 (1); Precipitação do mês mais chuvoso – bio13 (1).

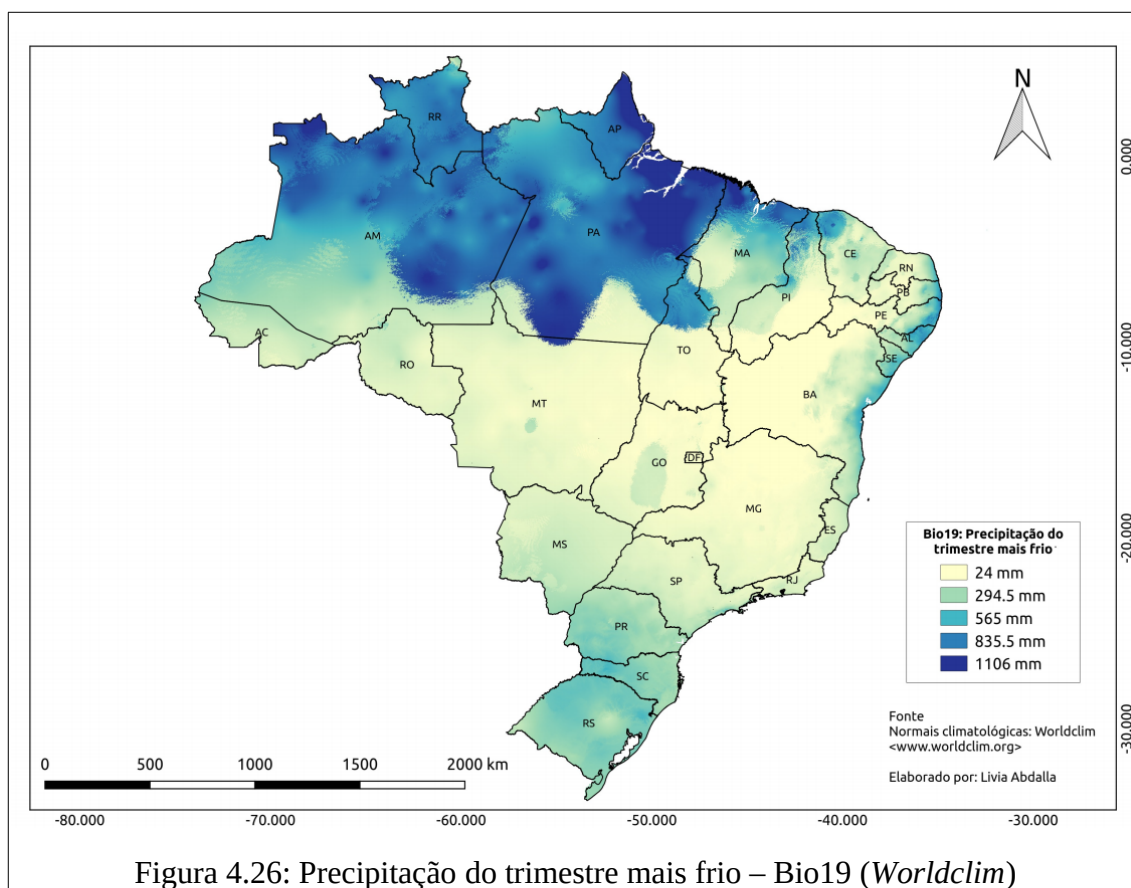
Para vigilância, prevenção e controle da FA no Brasil, o Ministério da Saúde utiliza três períodos epidemiológicos distintos para priorização das suas ações: Período de baixa

ocorrência (entre as semanas epidemiológicas 20^a e 37^a); Período pré-sazonal (38^a e 51^a semana epidemiológica); Período sazonal (52^a e 19^a semana epidemiológica). A definição dos diferentes períodos levou em conta um estudo realizado a partir da análise da série histórica de casos e óbitos humanos, que resultou em três períodos claramente distintos (ROMANO *et al*, 2011).

Com base na organização da Vigilância Epidemiológica da FA, e em razão da diversidade de variáveis climáticas relevantes aos modelos, destaca-se que as variáveis de temperatura e precipitação não serão apresentadas de acordo com a ordem de frequência, mas serão descritas de acordo com a ordem dos períodos epidemiológicos aos quais pertencem, e com as estações do ano que o período compreende. As variáveis que não correspondem a valores médios de um mês ou de um trimestre serão apresentadas e discutidas no aspecto mais amplo sobre suas prováveis relações com a doença, sem detalhamento sobre determinada estação ou sobre a sazonalidade da doença. Ao descrever detalhadamente o clima do Brasil, convencionou-se dividir o ano em 4 trimestres, em que cada trimestre corresponde a uma estação do ano: Verão (Dezembro, Janeiro e Fevereiro); Outono: (Março, Abril e Maio); Inverno (Junho, Julho e Agosto); Primavera (Setembro, Outubro e Novembro).

O Período de baixa ocorrência da FA corresponde a meado de Maio até meado de Setembro. Dessa maneira incorporam-se o final do Outono, todo o Inverno e o início da Primavera. As variáveis de temperatura e precipitação relacionadas ao período de baixa ocorrência da doença são: Precipitação do trimestre mais frio – bio19 (6); Precipitação total mensal – Junho – prec6 (1); Precipitação total mensal – Julho – prec7 (1); Precipitação total mensal – Agosto – prec8 (2); Temperatura média mensal – Agosto – tmean8 (2); Temperatura máxima mensal – Agosto – tmax8 (1).

A variável Bio19 (figura 4.26) corresponde à Precipitação do trimestre mais frio, ou seja, o total acumulado de chuvas nos três meses da estação inverno (Junho, Julho e Agosto). Os maiores volumes de precipitação no trimestre mais frio concentram-se na faixa centro-norte do bioma Amazônia (1106 a 835mm). Os Estados que concentraram os maiores volumes de chuvas no inverno foram: Amazonas (porções norte e leste); Pará (porções nordeste e sul) e Amapá (nordeste).



As variáveis Prec6, Prec7 e Prec8 (figuras 4.27, 4.28 e 4.29) correspondem à Precipitação total mensal dos meses de Junho, Julho e Agosto (período correspondente ao inverno). Os biomas que concentram maior volume de chuvas são Amazônia e Mata Atlântica (Zona da Mata e porção Sul), com destaque para os Estados do Amazonas (centro-norte), Roraima, Paraná, Santa Catarina, Rio Grande do Sul e litoral dos Estados nordestinos da Zona da Mata (Paraíba, Pernambuco, Alagoas, Sergipe e Bahia).

Segundo o Centro de Previsão de Tempo e Estudos Climáticos (CPTEC, 2018):

“no trimestre Junho, Julho e Agosto os maiores acumulados de chuva ocorrem sobre o extremo norte do Amazonas e norte de Roraima, associados principalmente à atuação da Zona de Convergência Intertropical (ZCIT) e à formação de Linhas de Instabilidade (LI's). No leste do Nordeste os acumulados de precipitação correspondem a valores em torno de 500mm entre o litoral do Rio Grande do Norte e Alagoas. As chuvas de maneira escassa no semiárido nordestino, com expansão das áreas de estiagem para o Tocantins, norte de Goiás e leste do Mato Grosso, onde a precipitação total no trimestre

costuma ser inferior a 25mm. Na Região Sul, os totais de chuva variam entre 400mm, no Rio Grande do Sul, e 100mm, no norte do Paraná.”

Destaca-se que esse período demonstra ser o menos favorável a manifestação da febre amarela, visto que a redução do volume de chuva em grande parte do país e nas principais regiões de ocorrência da doença em regiões como Minas Gerais, Tocantins, Goiás, Mato Grosso e centro-sul dos estados do Amazonas e Pará. A redução do volume de chuvas implica na redução da atividade reprodutiva dos vetores, os quais dependem do acúmulo de água em cavidades de troncos de árvores e em bromélias para deposição e eclosão de seus ovos (SILVA *et al*, 2004).

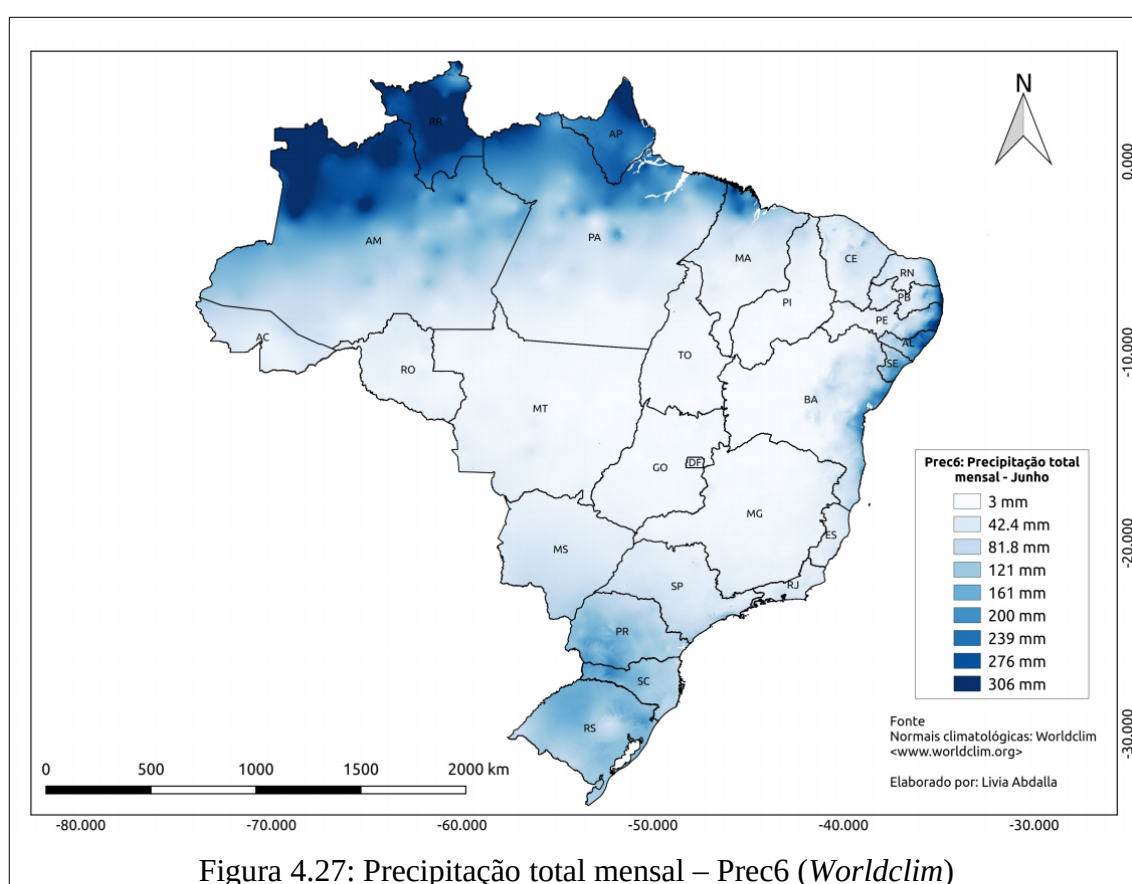
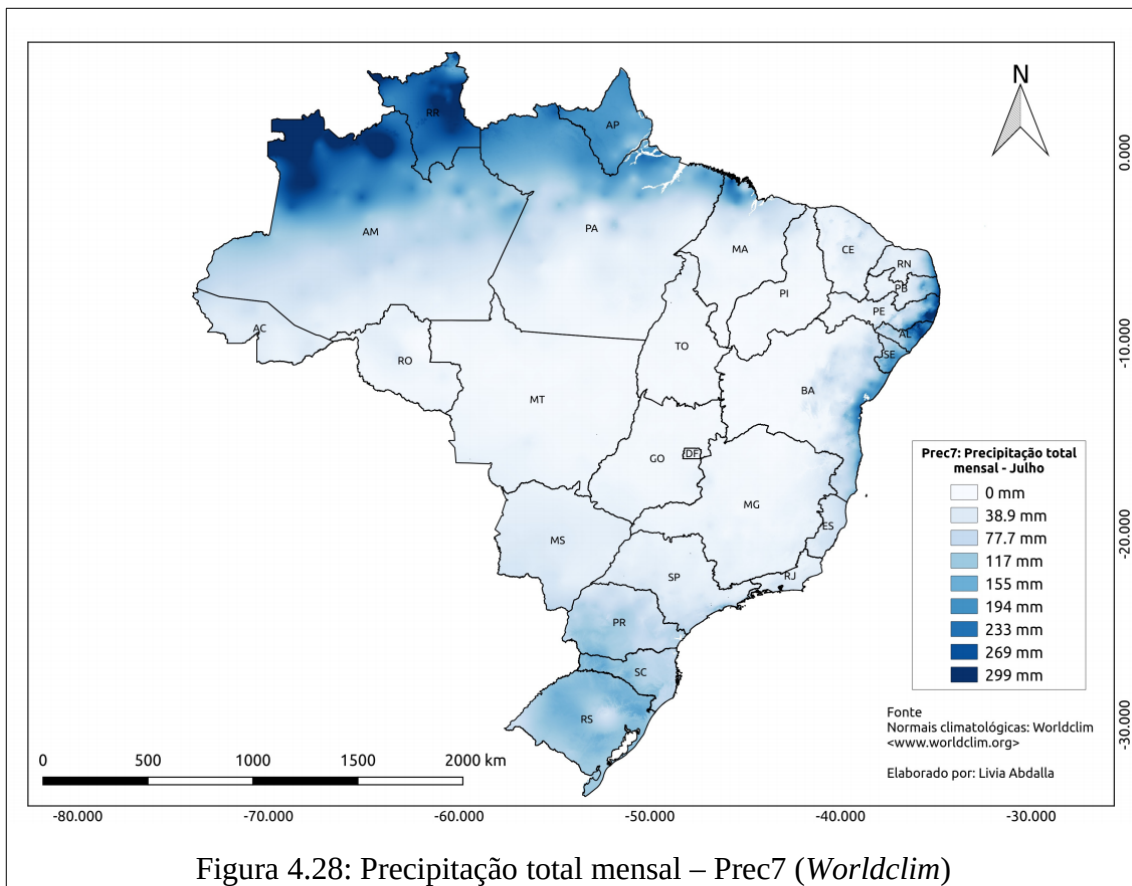
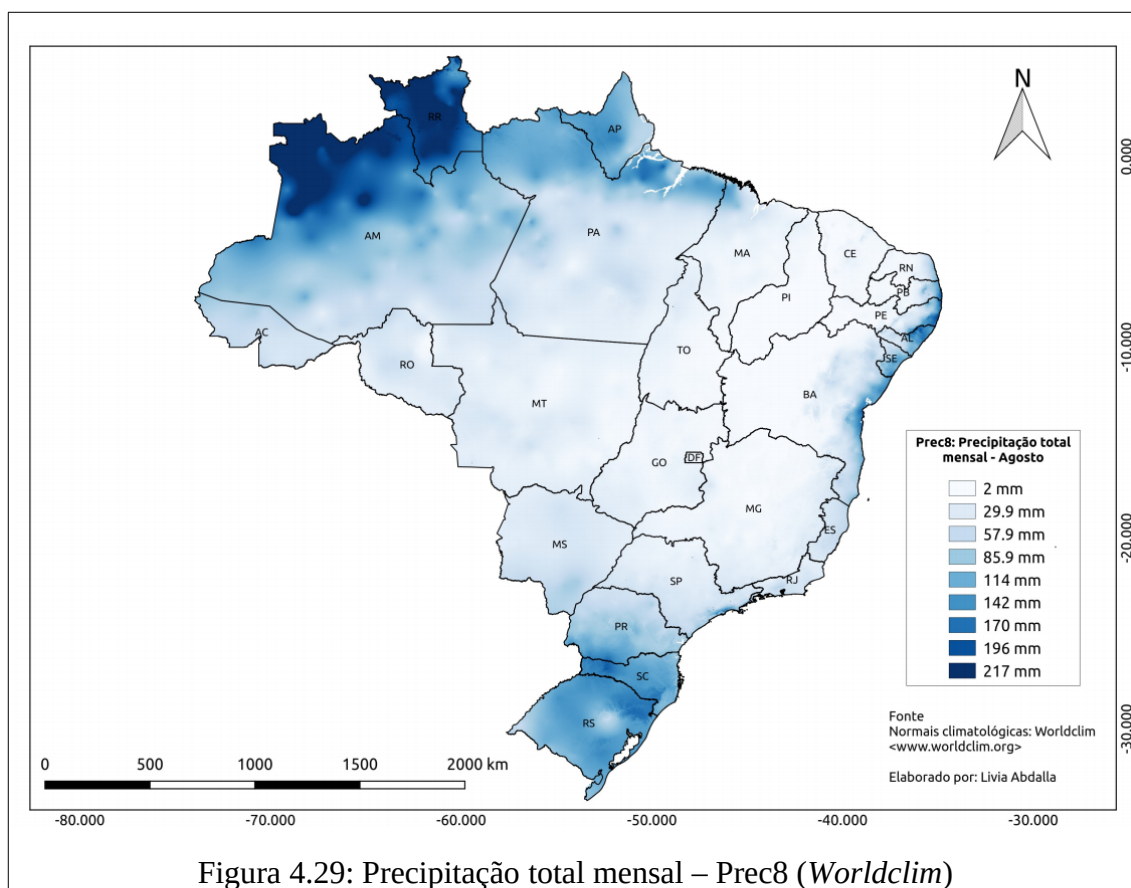
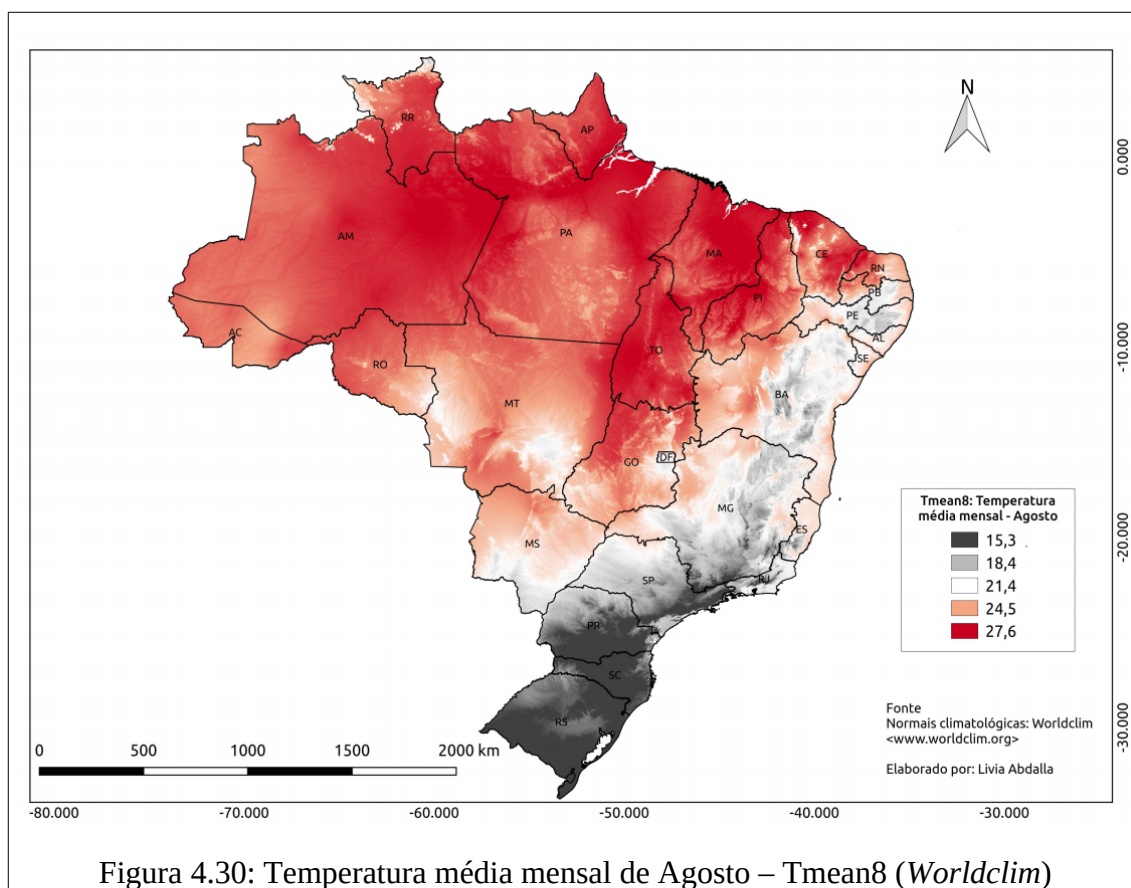


Figura 4.27: Precipitação total mensal – Prec6 (Worldclim)





A variável Tmean8 (figura 4.30) corresponde a Temperatura média mensal do mês de Agosto. Os biomas Amazônia, porção centro-norte do Cerrado e porção norte da Caatinga registraram maiores valores dessa variável (28 a 24°C). Os Estados da Região Norte concentraram os maiores valores médio de temperatura no mês de agosto. Além desses, os Estados do Maranhão, Piauí, Ceará, noroeste do Rio Grande do Norte, centro-norte do Mato Grosso e Goiás também concentram maiores valores médios de temperatura.



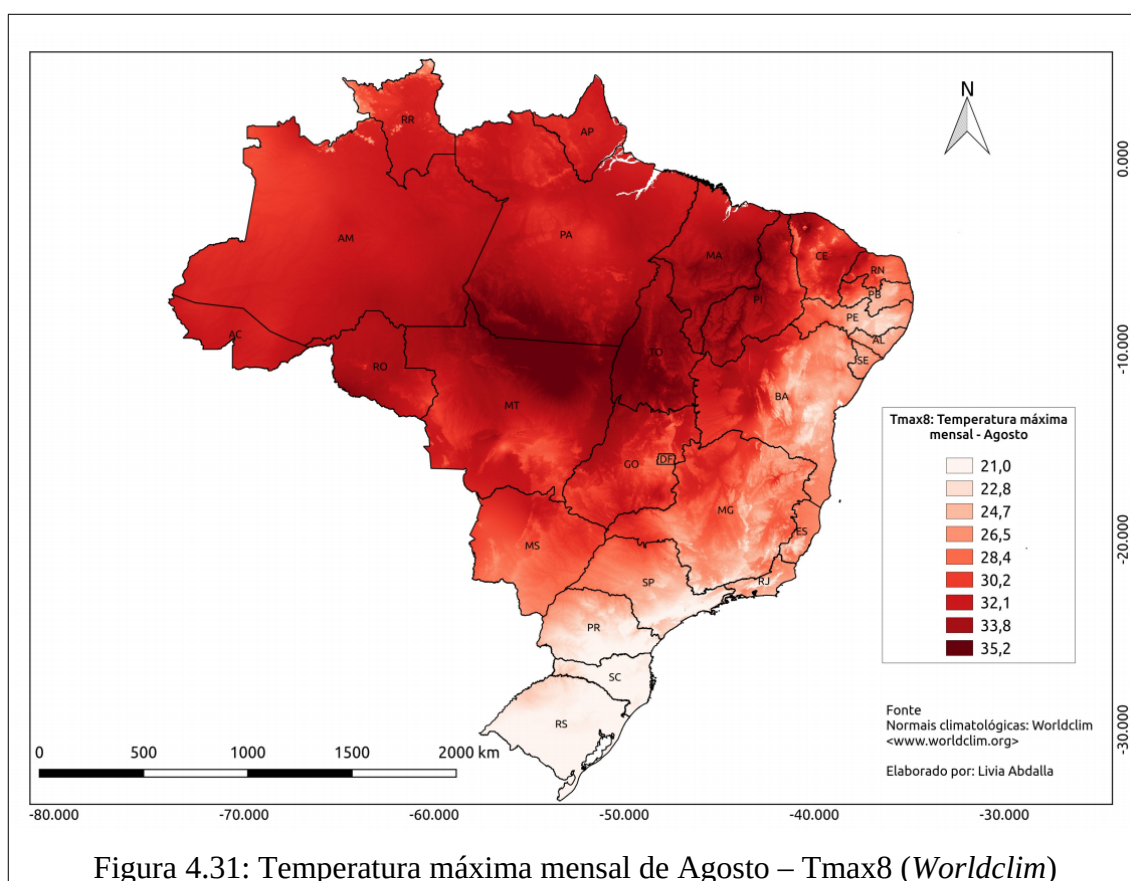
A variável Tmax8 (figura 4.31) corresponde à temperatura máxima mensal no mês de agosto. O bioma Cerrado e as suas faixas de transição com o bioma Amazônia (faixa central) e Caatinga (faixa norte) registraram os maiores valores de Tmax8. Os estados que registraram os valores máximos de temperatura em agosto foram: Maranhão; Piauí; Tocantins; noroeste de Goiás, Mato Grosso (centro-norte); centro-sul do Pará e do Amazonas e Rondônia.

Segundo o CPTEC (2018):

“em agosto a temperatura máxima aumenta no norte do Brasil, enquanto que na Região Sul e áreas serranas do Sudeste predominam valores médios inferiores a 22°C. As temperaturas máximas registram valores superiores a 36°C no Piauí, leste do Maranhão, nordeste do Mato Grosso e oeste do Tocantins. Nas Regiões Sudeste e Centro-Oeste ocorre uma diminuição gradativa da influência de massas de ar frio, porém, na Região Sul, as temperaturas mínimas permanecem baixas, com valores que podem variar entre 8°C e 12°C, especialmente nas regiões de Serra”.

Assim, verifica-se que o mês de agosto oferece condições de elevada temperatura,

favorável à reprodução dos vetores da febre amarela, principalmente nas áreas com recorrência de casos (Norte e Centro-Oeste). Porém, com o regime de chuvas ainda reduzido ao extremo norte dos Estados do Amazonas e Roraima, não favorece o acúmulo de água suficiente para a reprodução do vetor.



O Período Pré-sazonal corresponde de meado de Setembro a final de Dezembro. Compreende assim o início da Primavera e início do Verão. As variáveis de temperatura e precipitação relacionadas ao período pré-sazonal da doença são: Precipitação total mensal – Outubro – prec10 (8); Precipitação total mensal – Novembro – prec11 (2) (figuras 4.32 e 4.33). No período pré-sazonal da doença verifica-se que se inicia o período chuvoso no centro-sul da Região Norte e na Região Centro-Oeste. Este período é marcado pelo aumento gradual da atividade convectiva no Brasil Central – área que compreende desde o oeste do Amazonas, passando pelo norte do Mato Grosso, até Minas Gerais.

Segundo o CPTEC (2018):

“os sistemas frontais e a configuração da Zona de Convergência do Atlântico Sul (ZCAS) são os principais sistemas meteorológicos responsáveis pelo aumento da precipitação nestas áreas. Nas porções sul e oeste do Amazonas, sul do Pará e norte do Mato Grosso os totais acumulados de precipitação podem atingir valores entre 500mm e 800mm. No leste da Região Centro-Oeste predominam totais pluviométricos entre 500mm e 700mm. Na Região Sudeste também ocorre o aumento gradual das chuvas, onde os valores de precipitação podem variar entre 300mm e 700mm. Apesar do aumento gradual das chuvas entre Setembro e Dezembro, podem ocorrer baixos índices de umidade relativa do ar que implicam em períodos de estiagem – com destaque para a Região Nordeste, na qual predomina o período de estiagem, com valores acumulados de precipitação mais baixos, entre 25mm e 200mm, no norte do Piauí e na maior parte dos Estados do Ceará, Rio Grande do Norte e Paraíba. Na maior parte da Região Sul, as chuvas estão associadas à passagem de frentes frias e ao início da atuação dos complexos convectivos de mesoescala – nuvens intensas, com grande extensão vertical, que causam trovoadas e rajadas de vento. Nessa Região, os totais variam entre 300mm e 700mm. Especialmente na porção sul do Rio Grande do Sul as frentes avançam mais rapidamente e ocorre uma diminuição das chuvas, onde são esperados acumulados trimestrais entre 200mm e 400mm”.

Nesse contexto, destaca-se que as variáveis Prec10 e Prec11 (figuras 46 e 47) correspondem à Precipitação total mensal nos meses de Outubro e Novembro (parcela da primavera). Os biomas com maiores volumes de chuvas são a Amazônia e Mata Atlântica (porção sul mês outubro e porção central mês de novembro). Os Estados do Amazonas, Pará (porção sul), Tocantins (porção oeste) e Mato Grosso (porção norte) destacam-se com maiores concentrações de chuva nos dois meses. No mês de Outubro destacam-se o Mato Grosso do Sul (porção sudeste), oeste do Paraná e Santa Catarina. Já em Novembro destacam-se as chuvas nos Estados de Goiás e Minas Gerais, principalmente nas porções centrais desses. Verifica-se, assim, que o aumento dos níveis de precipitação no país, principalmente nas regiões Norte, Centro-Oeste e Sudeste podem favorecer o acúmulo de águas em locais propícios à reprodução de vetores transmissores da doença, que acometem inicialmente primatas não humanos, havendo inclusive notificações de epizootias nesse período pré-sazonal.

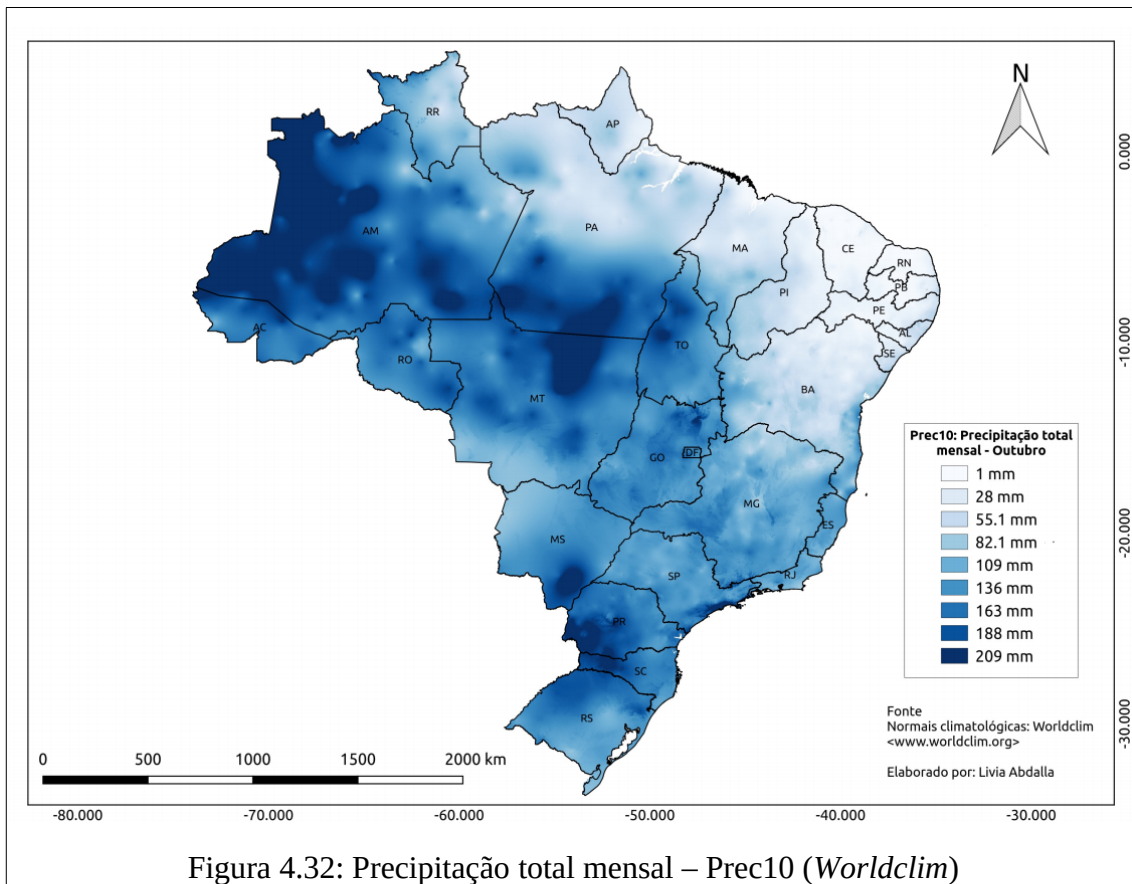


Figura 4.32: Precipitação total mensal – Prec10 (Worldclim)

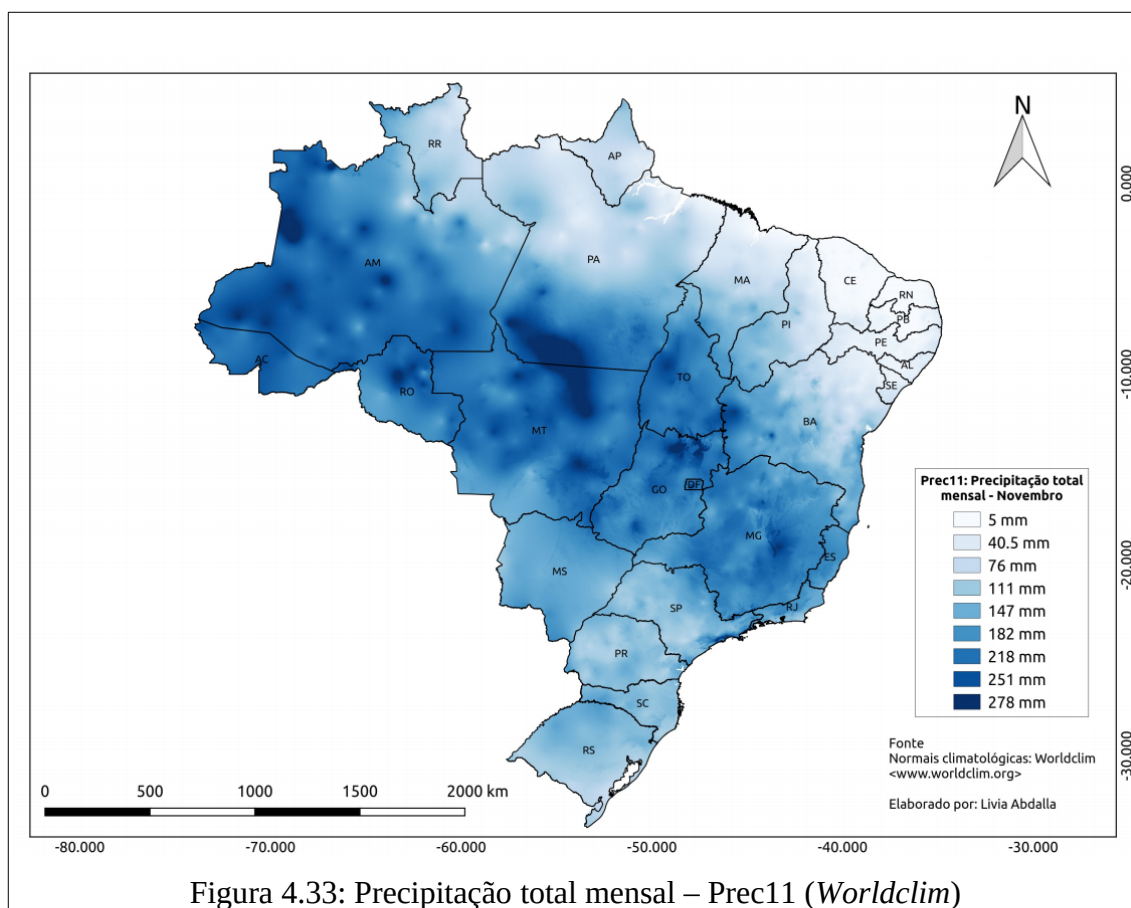


Figura 4.33: Precipitação total mensal – Prec11 (Worldclim)

O Período Sazonal corresponde do final de Dezembro a meado de Maio, que compreende o início do Verão e final do Outono. As variáveis de temperatura e precipitação relacionadas ao período sazonal da doença são: Temperatura média do trimestre mais úmido – bio8 (1); Precipitação do trimestre mais chuvoso – bio16 (2); Precipitação do trimestre mais quente – bio18 (2); Precipitação do mês mais chuvoso – bio13 (1); Precipitação total mensal – Dezembro – prec12 (5); Precipitação total mensal – Fevereiro – prec2 (3). No período sazonal da doença destaca-se que as chuvas são mais acentuadas nos setores central e sul da Região Norte, no norte da Região Centro-Oeste, estendendo-se até o setor central da Região Sudeste.

Segundo o CPTEC (2018):

“as chuvas são decorrentes principalmente da formação de anticiclone que se configura sobre o continente sul-americano (Alta da Bolívia), na alta troposfera, nos meses de verão, mas também

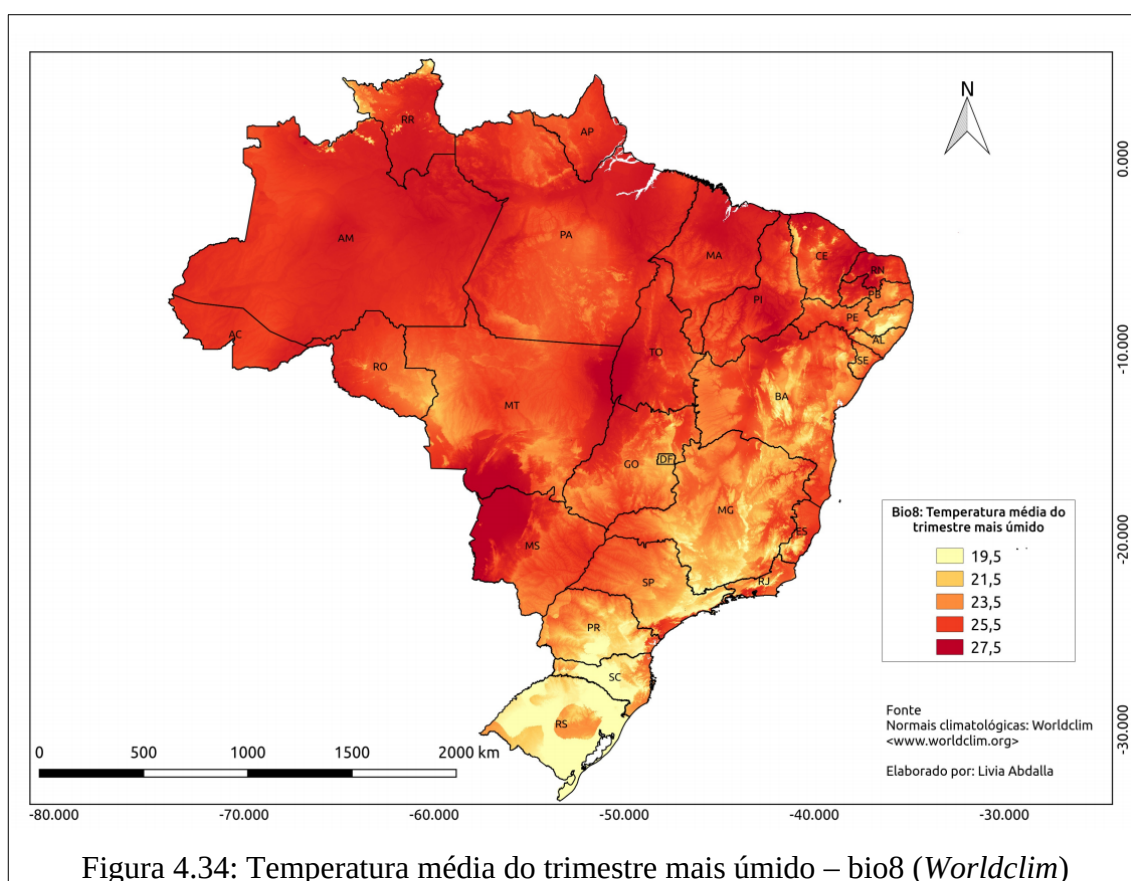
podem estar associadas ao deslocamento dos sistemas frontais que também contribuem para a caracterização de outro sistema conhecido por Zona de Convergência do Atlântico Sul (ZCAS)”.

Nesse período os totais acumulados podem atingir 1.000mm no sudeste do Amazonas e norte do Mato Grosso. No Nordeste, as chuvas aumentam no Piauí, oeste da Bahia e Maranhão, onde os valores podem chegar a 700mm, já na região semiárida e litorânea os totais acumulados podem variar entre 100 e 200mm. Na Região Sul os sistemas frontais avançam de maneira mais rápida e os maiores totais, entre 500mm e 700mm, ocorrem no norte do Paraná. Durante esse período ocorre também a influência da Zona de Convergência Intertropical (ZCIT) sobre o norte do Brasil. Isto provoca chuva por mais dias em toda faixa norte do Brasil e mantém a Região Norte com poucas mudanças. Devido ao posicionamento mais ao sul da ZCIT, o setor norte da Região Nordeste registra um aumento das chuvas, com valores máximos durante o mês de abril. Por outro lado, as Regiões Sudeste e Centro-Oeste apresentam uma redução gradativa da precipitação durante o mês de março. De modo geral, a Região Sul apresenta pouca mudança no Rio Grande do Sul, enquanto que o Paraná e o leste de Santa Catarina evidenciam redução dos totais pluviométricos em comparação com o trimestre anterior (CPTEC, 2018; INMET, 2017).

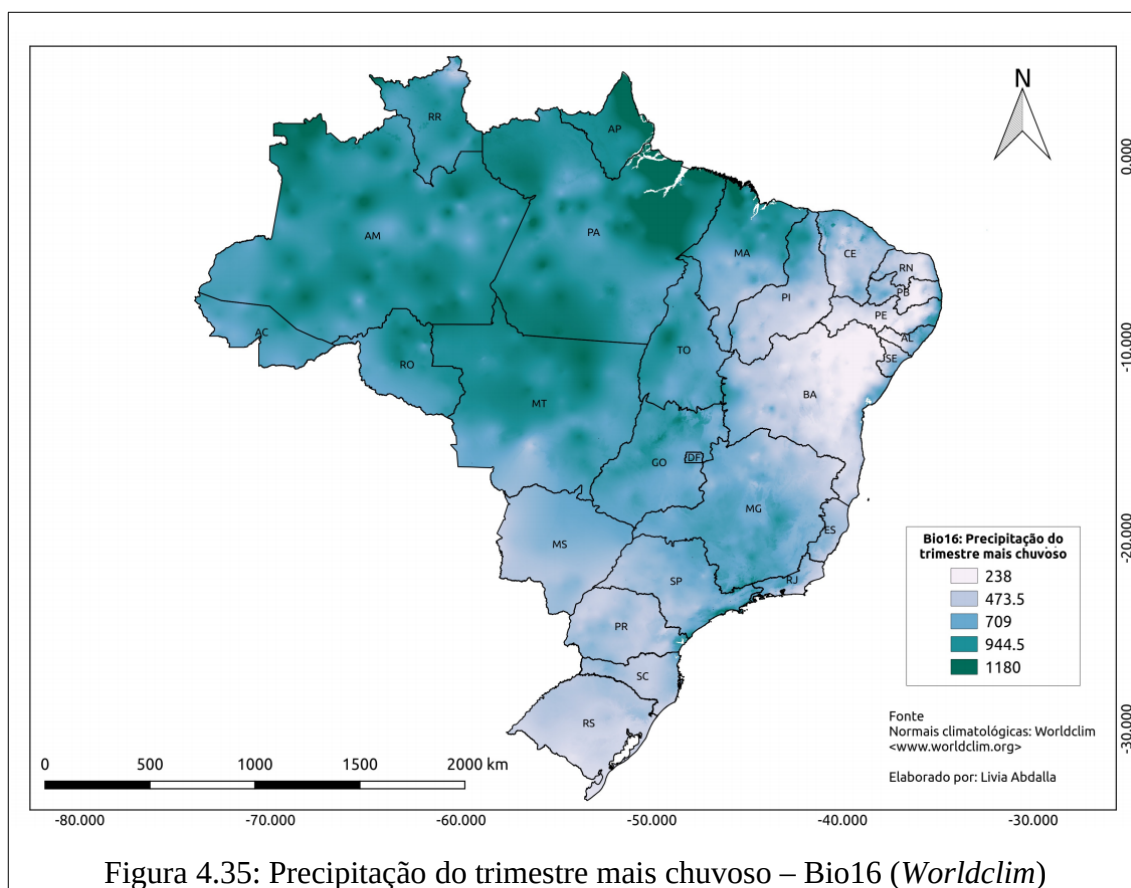
A variável Bio8 (figura 4.34) corresponde à Temperatura média do trimestre mais úmido, ou seja, trata-se do valor médio de temperatura nos três meses com maior acumulado de precipitação. Todos os biomas registraram valores elevados de temperatura nos meses mais chuvosos, somente o Pampa e a porção centro-sul da Mata Atlântica registraram valores menores. Os estados que registraram os maiores valores de temperatura nos meses mais chuvosos (condições ideais para a reprodução de mosquitos transmissores da febre amarela) foram: Amazonas; Pará (centro-norte); tríplice fronteira do Mato Grosso, Tocantins e Goiás; sudoeste do Mato Grosso, noroeste do Mato Grosso do Sul; norte do Maranhão, Ceará e Rio Grande do Norte; centro-norte do Piauí. Segundo o CPTEC (2018), nesse período os valores históricos sobre temperatura máxima variam de 24°C, no leste das Regiões Sul e Sudeste, a 34°C, no Nordeste. Os valores de temperaturas mínimas médias variam entre 12°C, em áreas serranas do Sul e Sudeste, a 24°C, no norte do País.

Nesse período, portanto, verifica-se que em regiões de histórico de manifestação da doença (Amazonas, Pará, Mato Grosso, Tocantins, Goiás, Mato Grosso do Sul, Minas Gerais)

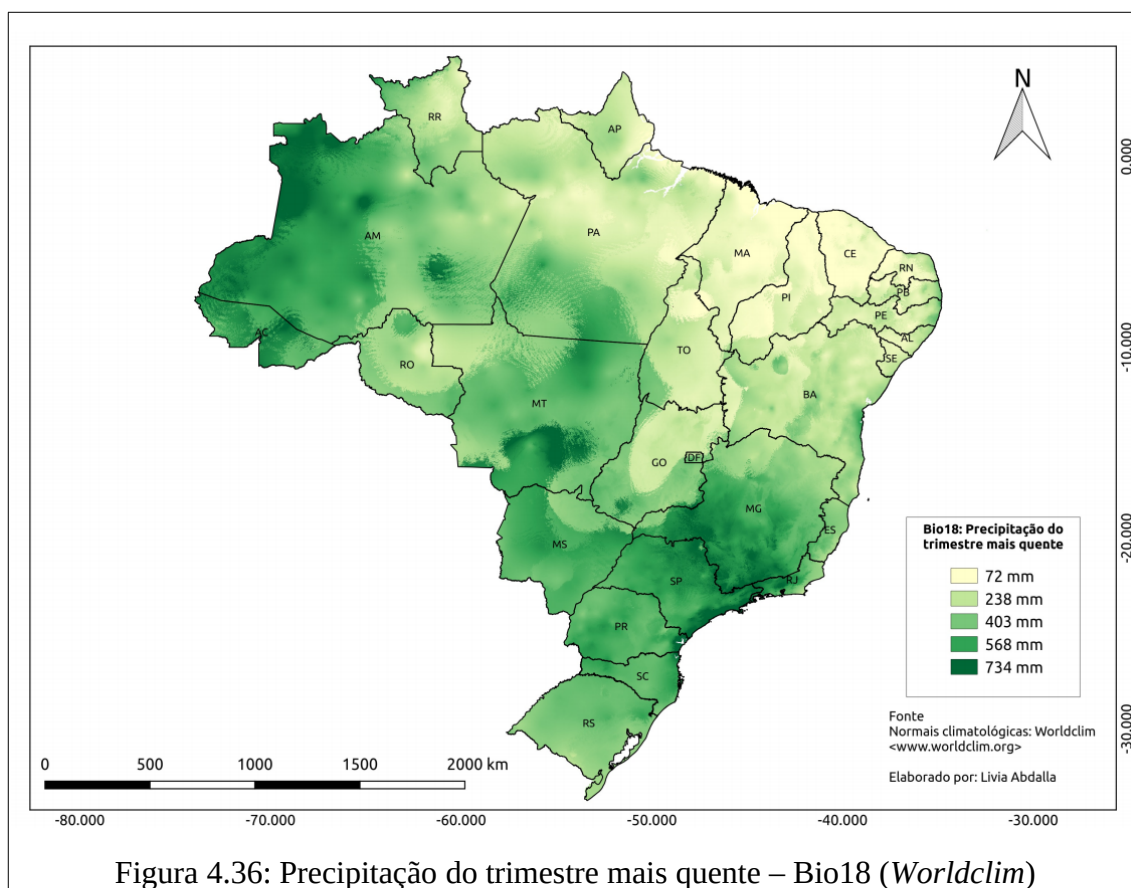
ocorre a combinação de fatores pluviométricos e de temperatura favoráveis (maior índice pluviométrico e registros de temperaturas máximas) à intensa reprodução e infestação dos vetores, que ampliam a circulação de vírus em primatas e, por conseguinte, em humanos. A temperatura elevada favorece a replicação viral nos vetores, o que contribui para altas taxas de carga viral tanto nos mosquitos quanto nos primatas, tornando-os amplificadores (ALMEIDA *et al*, 2016).



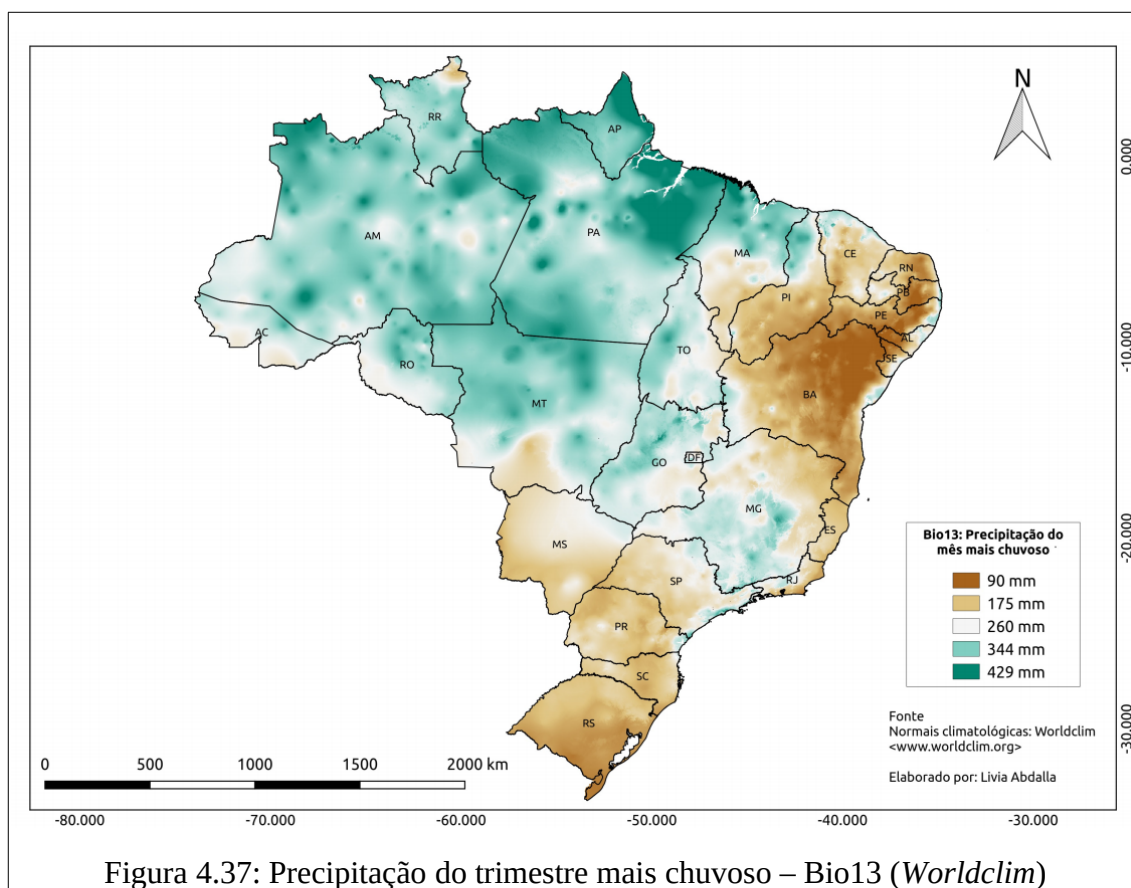
A variável Bio16 (figura 4.35) corresponde à Precipitação do trimestre mais chuvoso (estação Verão), ou seja, o total acumulado de chuvas nos três meses com maior concentração de chuvas. O bioma Amazônia concentra os maiores volumes de chuvas (1.180mm), seguido pelo Cerrado (1.180 a 944mm). Os Estados do Amazonas, Pará e Mato Grosso destacam-se por concentrarem maiores volumes de precipitação no trimestre mais chuvoso.



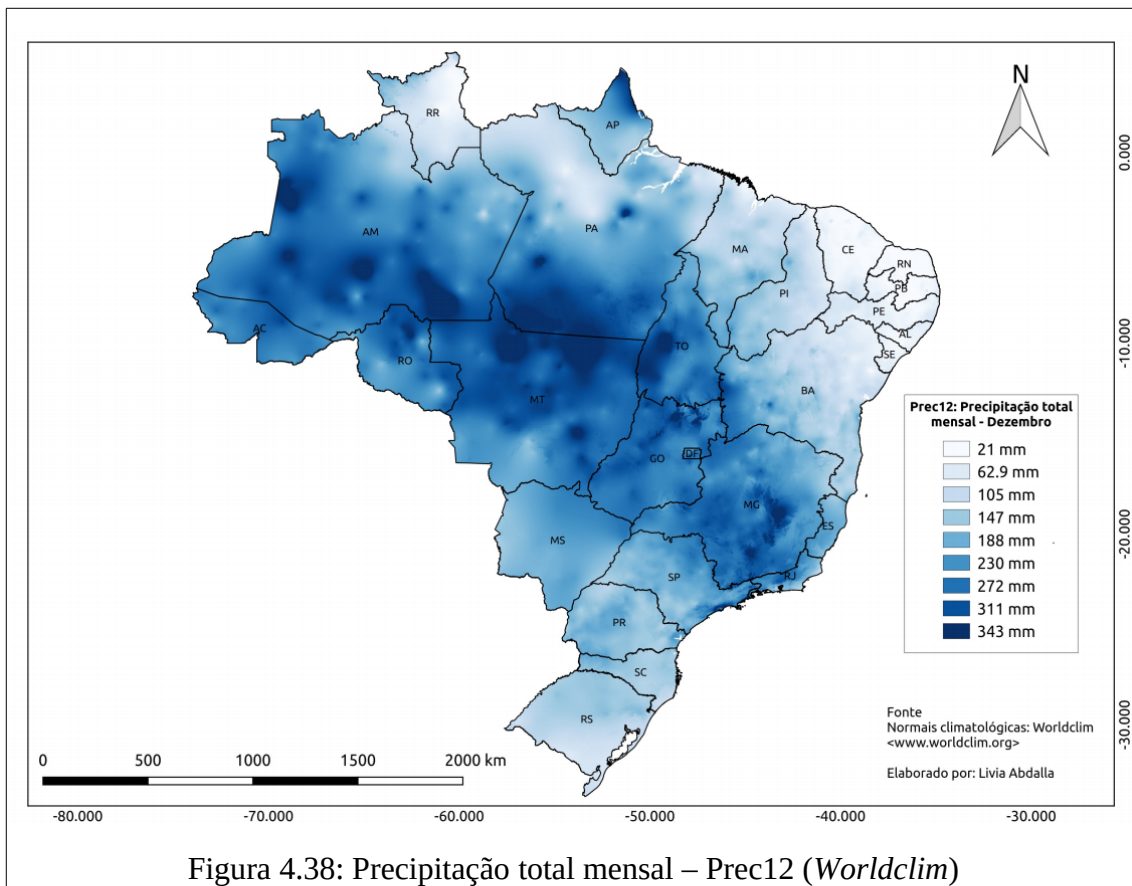
A variável bio18 (figura 4.36) corresponde a Precipitação do trimestre mais quente, condições essas ideais para a reprodução de vetores. Os biomas com maior concentração de volumes de chuvas no verão são a Amazônia (734 a 403 mm), Cerrado (porções oeste e sudeste) e Mata Atlântica (faixa central). Os estados do Acre, Amazonas, Minas Gerais (faixa sul), São Paulo, Mato Grosso e Mato Grosso do Sul destacam-se nesse contexto.

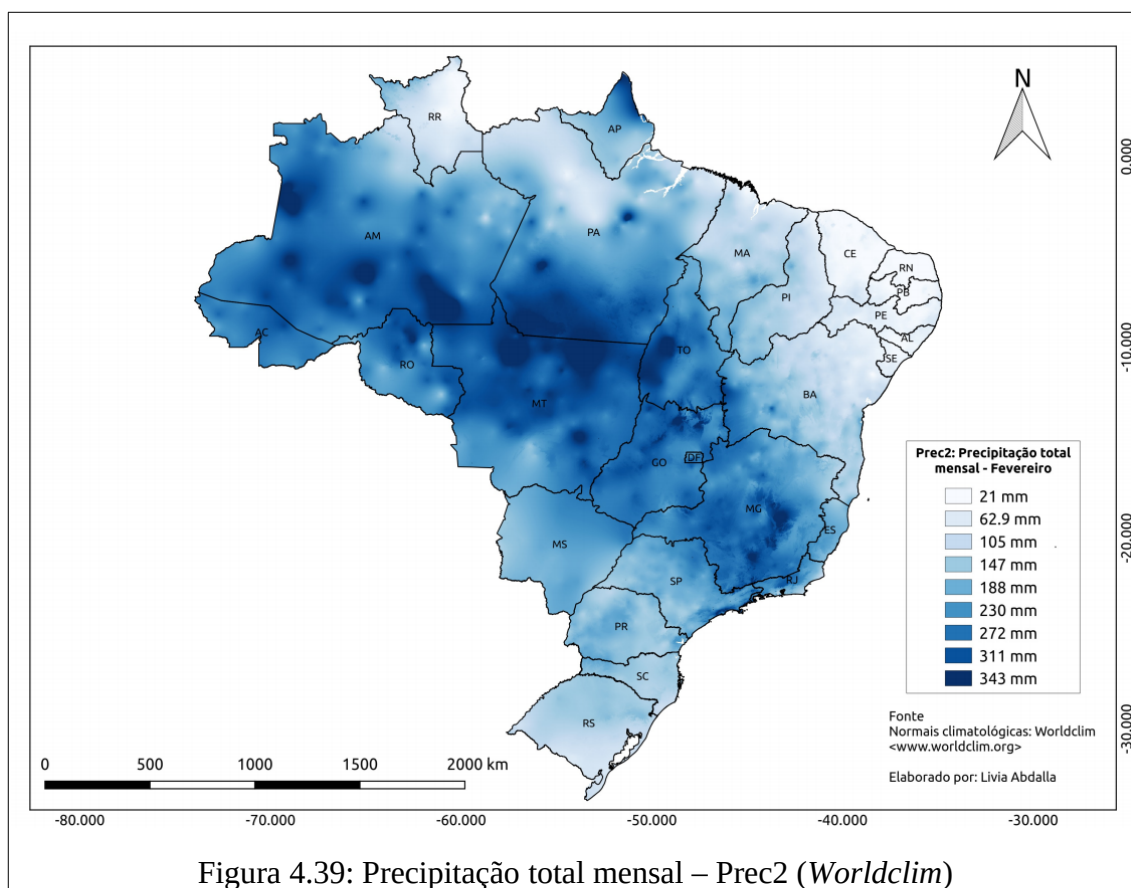


A variável bio13 (figura 4.37) corresponde à Precipitação do mês mais chuvoso, ou seja, trata-se dos maiores acumulados de chuvas do mês com maior precipitação registrada. Os biomas Amazônia, Cerrado e Mata Atlântica (porção central) possuem a maior concentração de chuvas (429 a 260 mm), com destaque para os Estados da Região Norte (Acre, Amazonas, Roraima, Amapá, Pará, Tocantins), Mato Grosso, Goiás, Minas Gerais (centro-sul).



As variáveis Prec12 e Prec2 (figura 4.38 e 4.39) correspondem, respectivamente, a Precipitação total mensal do mês de Dezembro e Fevereiro. Os biomas com maiores concentrações de precipitação nesses meses são Amazônia, Cerrado e Mata Atlântica (343 a 230mm), principalmente nas suas porções centrais. Os Estados do Amazonas, Pará, Mato Grosso, Goiás, Tocantins, Minas Gerais e Rio de Janeiro destacam-se nesse contexto.





Dentre as variáveis de temperatura e precipitação, ainda precisam ser descritas seis, que correspondem a valores médios anuais, ou seja, sem restrições mensais ou trimestrais: Sazonalidade da Temperatura – bio4 (9); Sazonalidade da Precipitação – bio15 (3); Precipitação Anual – bio12 (3); Isotermalidade – bio3 (2); Amplitude térmica anual – bio7 (1); Variação Diurna Média de Temperatura – bio 2 (1).

A variável bio4 (figura 4.40) corresponde a Sazonalidade da Temperatura. Trata-se de um coeficiente de variação que indica o valor da temperatura em razão das quatro estações do ano. Os biomas Pampa e porção Sul da Mata Atlântica registraram maior variação de valores de temperatura, com destaque para os Estados do Rio Grande do Sul, Santa Catarina, Paraná e São Paulo. A menor variação dos valores de temperatura durante as estações do ano na região endêmicas e a maior variação de máximas e mínimas de temperatura (durante as estações) nas áreas extraendêmicas (mas com surtos eventuais ou esporádicos – Rio Grande do Sul, São Paulo) parecem demonstrar uma preferência dos vetores (e consequentemente do vírus) de

habitat nas regiões com menor Sazonalidade da Temperatura.

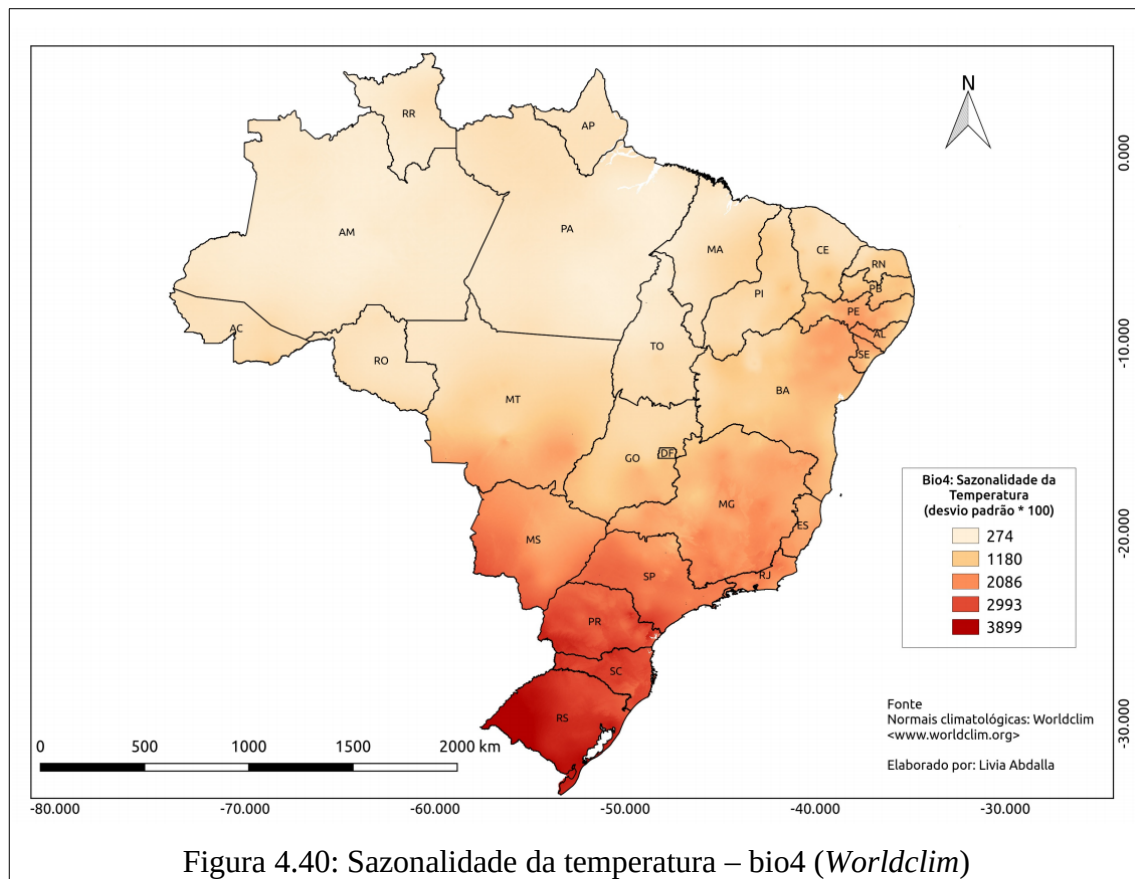


Figura 4.40: Sazonalidade da temperatura – bio4 (Worldclim)

A variável Bio15 (figura 4.41) corresponde à Sazonalidade da Precipitação. Trata-se de um coeficiente de variação que indica o volume de precipitação pluviométrica em razão das quatro estações do ano. Os biomas Caatinga e Cerrado registraram maior variação, principalmente na faixa de transição ou contato desses biomas. Os estados com maior coeficiente de variação foram: Ceará; Piauí; porções centro-oeste dos estados do Rio Grande do Norte, Paraíba, Pernambuco e Bahia; Minas Gerais (centro-norte); sul do Tocantins; e norte de Goiás. Verifica-se que em porções da região endêmica (centro-norte do Amazonas e Pará) ocorre menor variabilidade de precipitação nas estações do ano e, portanto, revelam a preferência dos vetores para sua sobrevivência e manutenção nessas condições. Interessante evidenciar que o período de baixa ocorrência da FA ocorre justamente nessas porções com menor variabilidade de chuvas. Os primatas também possuem preferência, mas com maior

resiliência que os vetores, por locais com frequência de chuvas.

Nowak (1999) descreveu que a riqueza de primatas possui uma associação muito característica com a precipitação da região, já que todos os primatas neotropicais incluem frutos em sua dieta. Alguns gêneros adotam alimentação generalista (*Cebus spp.* e *Sapajus spp.*), enquanto outros adotam dietas especializadas, como frutos imaturos (*Pitheciidae*) e insetos (*Saimiri spp.*), folhas (*Alouatta spp.*), gomas e exsudatos de árvores (*Callitrichidae*). (AYRES, 1989). A procura de alimento relaciona-se com as necessidades nutricionais e morfofisiológicas do animal, assim, cada gênero possui um hábito alimentar que está diretamente relacionado ao ambiente em que vive. Para melhor utilização de fontes alimentares, cada espécie adaptou o número de indivíduos de seus grupos para o máximo aproveitamento de recursos alimentares (AURICCHIO, 1995).

Então, no contexto da Sazonalidade da Precipitação, verifica-se que regiões que possuem menor variabilidade de precipitação durante as estações do ano mantêm condições vitais para sobrevivência dos vetores (e conseqüentemente do vírus) e condições preferenciais para a manutenção de espécies de primatas.

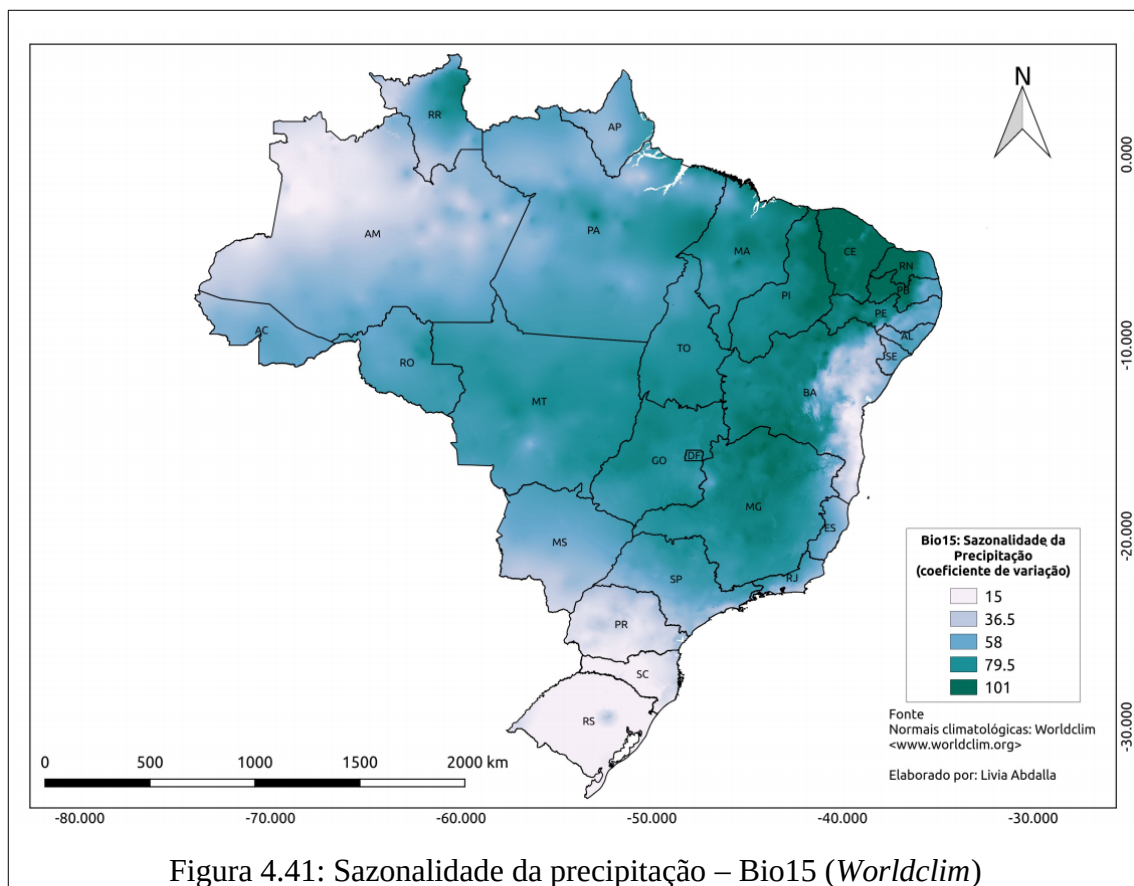
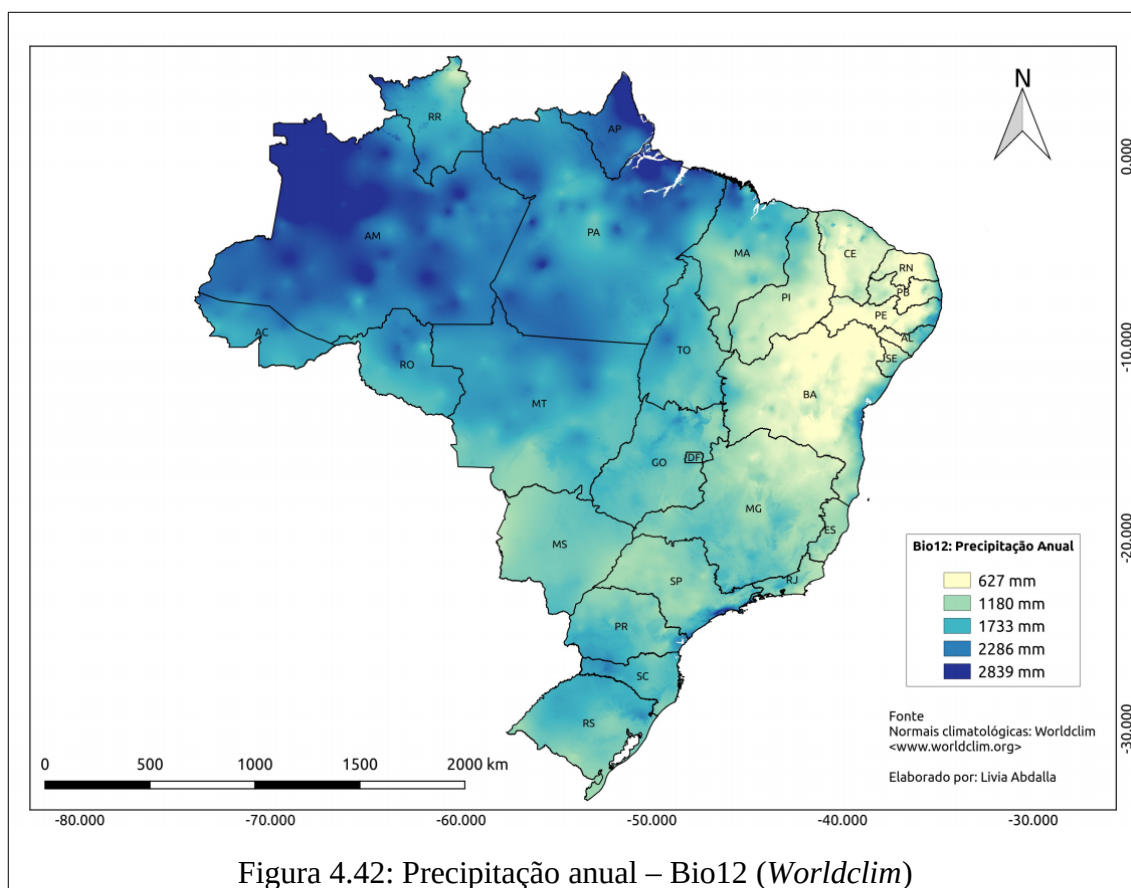


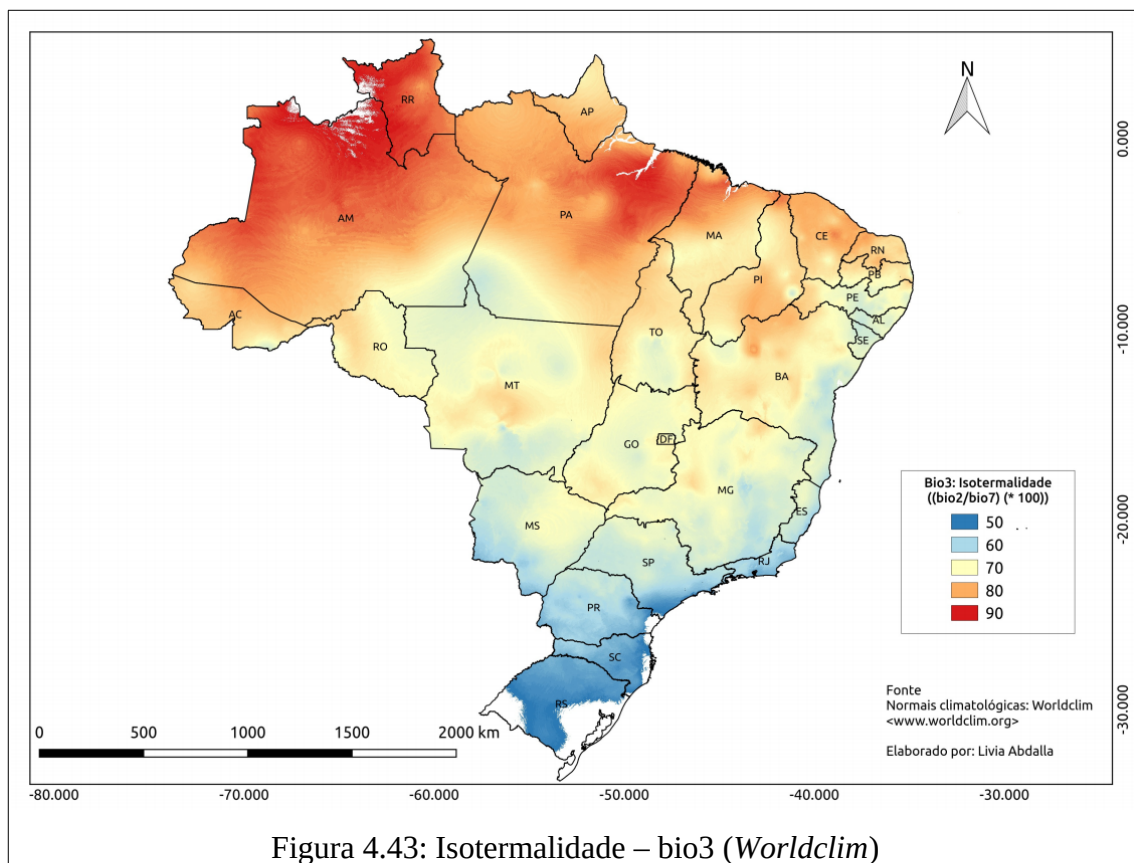
Figura 4.41: Sazonalidade da precipitação – Bio15 (Worldclim)

A variável Bio 12 (figura 4.42) corresponde à Precipitação anual, ou seja, o valor médio de chuvas ao ano. O bioma Amazônia, principalmente nas porções norte e oeste, também concentraram maior total anual de precipitação (entre 2.286 e 2.839mm). Os estados que concentraram maiores volumes de chuva ao ano foram: Amazonas (centro-norte); Pará (nordeste e oeste); Amapá.

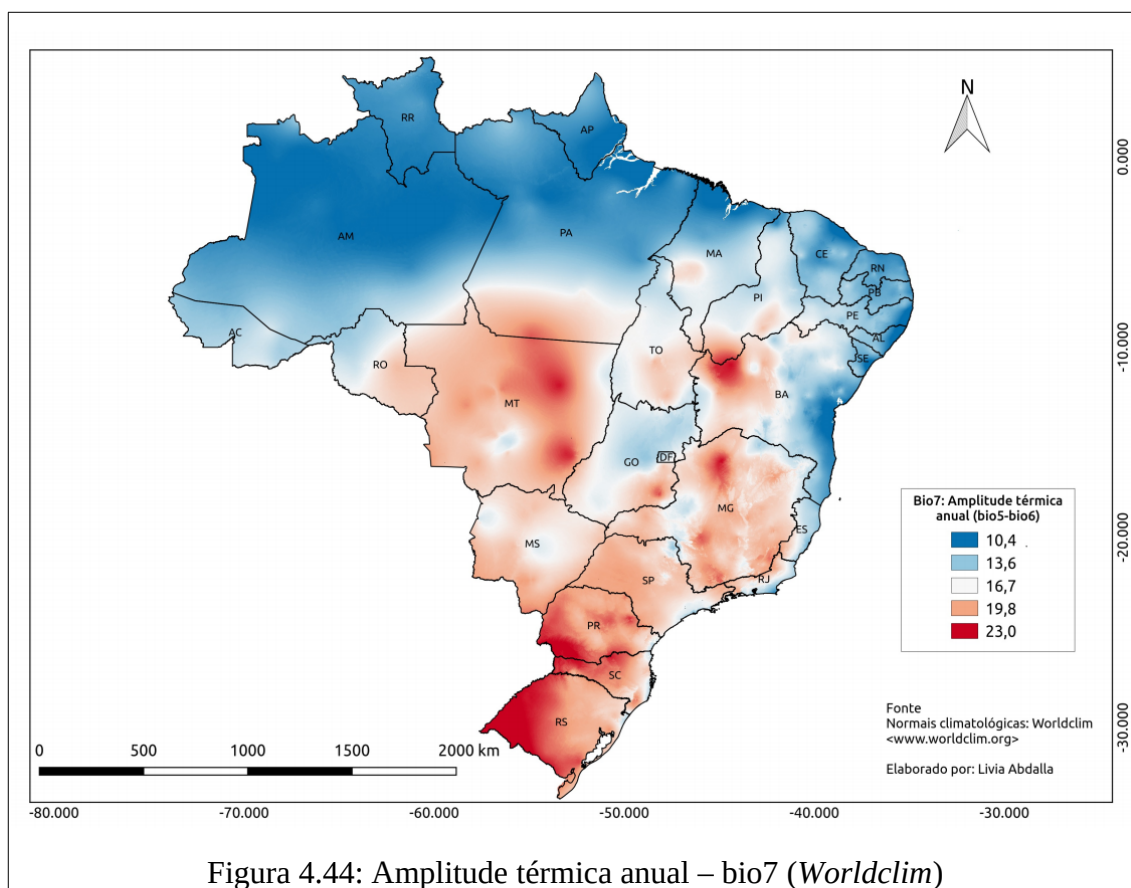
Kay *et al* (1997) indicam que, até 2.500mm de precipitação anual, a riqueza de primatas amplia com o aumento da precipitação e depois desse limite há um decréscimo da riqueza de primatas. Os autores também sugerem que a riqueza decresce após 4.2500mm de precipitação devido à queda da produtividade primária das plantas. A produtividade primária afeta diretamente a capacidade das plantas de produzir itens alimentares, que também é o principal fator que determina a riqueza de primatas neotropicais (STEVENSON 2001; HANYA *et al.* 2011).



A variável Bio3 (figura 4.43) corresponde a Isotermalidade (%). Trata-se de um índice de variabilidade da temperatura contendo a relação entre a faixa média do dia e a faixa média anual. Verifica-se que no bioma Amazônia os valores de isotermalidade possuem valores que variam de 90% a 80% na porção norte e valores entre 80% a 70% nas porções centro-sul. No bioma Caatinga os valores variam entre 90% a 80% em basicamente em todas as porções do bioma. No Cerrado os valores variam entre 80% a 60%. Na Mata Atlântica variam entre 50% e 70%. No Pampa os valores de isoterma não possuem variabilidade significativa, seus valores ficam restritos a 50%. No Pantanal os valores também não possuem variabilidade representativa, sendo basicamente restritos a faixa de 60%. O Amazonas e o Pará destacam-se como os estados que possuem maior variabilidade de valores de isotermalidade.



A variável Bio7 (figura 4.44) é a Amplitude térmica anual, a qual refere-se a diferença entre a temperatura média do mês mais quente e a temperatura média do mês mais frio. O bioma Pampa apresentou maior amplitude térmica anual, seguido pelo bioma Mata Atlântica (porção centro-sul), Cerrado (faixa leste e oeste) e Amazônia (porção sul). Os estados que, em média, registraram maiores variações anuais de temperaturas mínimas e máximas foram: Mato Grosso (centro); Bahia (noroeste); Minas Gerais (centro); centro-oeste do Paraná, Santa Catarina e Rio Grande do Sul.



A variável Bio2 (figura 4.45) corresponde a variação diurna média da temperatura, ou seja, o quanto variaram os valores de temperatura máxima e mínima diariamente, e qual a média dessas oscilações térmicas diárias da temperatura. As faixas de transição entre os biomas Amazônia e Cerrado, e entre os biomas Cerrado e Caatinga concentram os maiores valores de variação térmica diária. Os estados que, em média, registraram maiores variações diárias de temperatura foram: Mato Grosso; Rondônia (Sudeste); Pará (sul); Tocantins (leste); sul do Maranhão e Piauí; centro-oeste da Bahia e Minas Gerais; sudeste de Goiás; oeste de São Paulo e Paraná; e noroeste de Santa Catarina e Rio Grande do Sul.

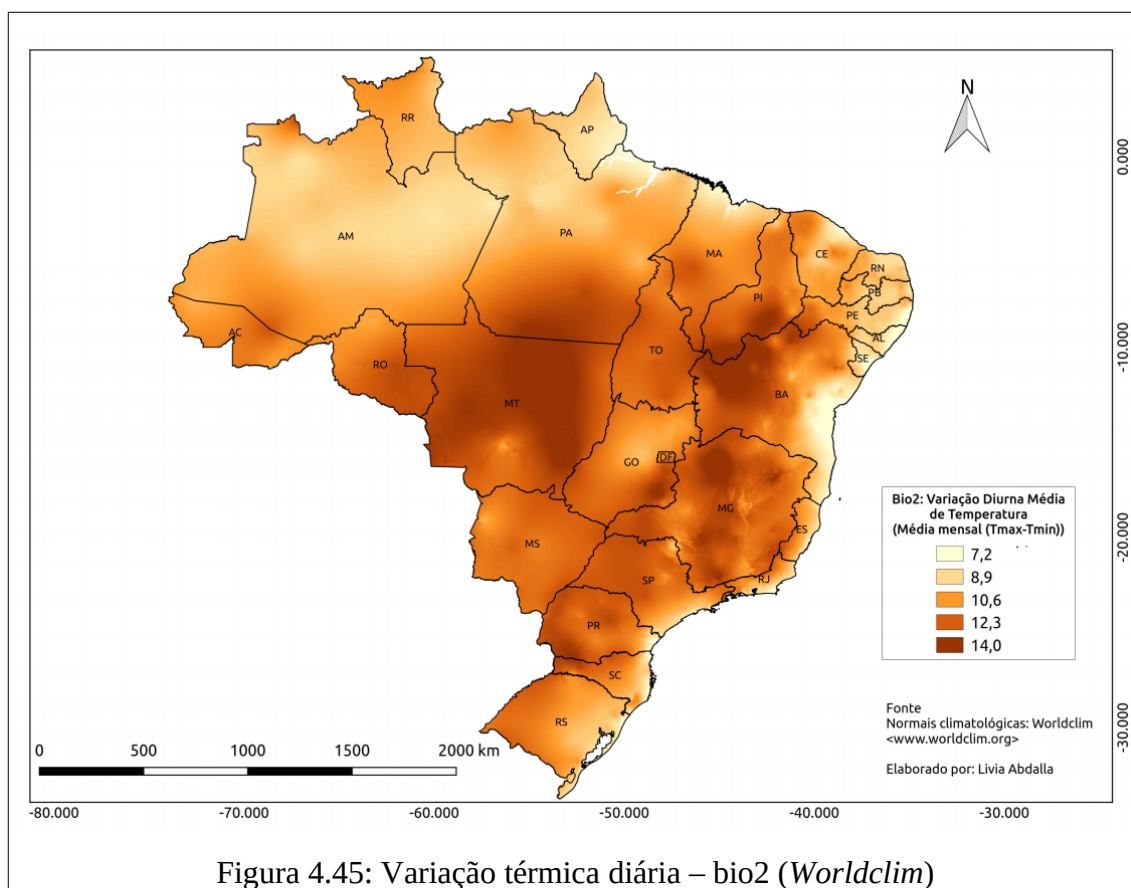


Figura 4.45: Variação térmica diária – bio2 (Worldclim)

Além dessas variáveis do tipo Normais Climatológicas, também foram evidenciadas quatro classes correspondentes à camada temática Climas Zonais e suas Subunidades Regionais (figura 4.46): Tropical Brasil Central Mesotérmico Brando (média entre 10° e 15°C) Úmido com 1 a 3 meses secos (2); Tropical Brasil Central Subquente (média entre 15 e 18°C em pelo menos 1 mês) Superúmido sem seca/subseca (1); Equatorial Quente (média > 18°C em todos os meses) Úmido com 1 a 3 meses secos (1); Temperado Mesotérmico Brando (média entre 10° e 15°C) Superúmido sem seca/subseca (1).

Segundo NIMER (1979):

“o território brasileiro, de proporções continentais, possui grande variabilidade latitudinal (32° S a 5° N) e, portanto, possui ampla diversidade de tipos climáticos, que variam desde climas quentes e secos/úmidos a climas frios e úmidos. Além da variabilidade latitudinal, possui grande variação altimétrica (0m nas planícies litorâneas até 3.014 metros no Pico da Neblina – AM). Além dos efeitos latitudinais e altimétricos, existem as influências da

maritimidade/continentalidade pela presença de extensa massa de água de água do Oceano Atlântico e das massas de ar (Equatorial continental - mEc, Equatorial Atlântica - mEa, Tropical continental - mTc, Tropical atlântica - mTa e a Polar atlântica - mPa). A classificação climática do IBGE considera essa diversidade climática em cinquenta e duas classes. A classificação adotada no mapeamento de Clima do Brasil do IBGE é baseada em 3 sistemas que integram métodos quantitativos e de dinâmica atmosférica. O primeiro sistema, mais abrangente, relativo à gênese climática, fundamentado na climatologia dinâmica e nos padrões de circulação atmosférica, define os 3 climas zonais (Equatorial, Tropical e Temperado) e suas subunidades regionais. O segundo sistema delimita as regiões térmicas (Mesotérmico Mediano e Brando, Subquente e Quente) e fundamenta-se na frequência e médias dos valores extremos mensais. A classificação de regiões quanto aos padrões de umidade e seca mensais (que vão de Superúmido até Semiárido) é resultante do terceiro sistema adotado. Nesta relaciona-se o número de meses secos com tipo de vegetação natural predominante, a fim de mostrar a interação do regime climático com a biogeografia e a ecologia”.

Dentre as classes climáticas, as árvores de decisão evidenciaram quatro zonas climáticas com seus respectivos recortes sub-regionais. O clima Equatorial Quente Úmido domina a região amazônica, caracteriza-se por temperaturas médias (24 a 26°C), baixa amplitude térmica (3°C), regime de chuvas é abundante (mais de 2.500mm/ano), principalmente por influência da massa equatorial continental. O recorte sub-regional deste clima zonal evidenciado foi Equatorial Quente (média > 18°C em todos os meses) Úmido com 1 a 3 meses secos, que corresponde a um arco que transpassa o sul do Acre, Rondônia, norte do Mato Grosso, Pará, nordeste do Amazonas e partes de Roraima e Amapá (IBGE, 2002).

O clima Temperado Mesotérmico Brando (média entre 10° e 15°C) Superúmido sem seca/subseca possui características de invernos frios, com temperatura média do mês mais frio entre 13°C e 15°C e verões quentes, com média do mês mais quente superior a 24°C. As precipitações são regulares durante o ano todo, não apresentando estação seca, com índices pluviométricos anuais entre 1.500mm e 1.600mm. Na região sul predomina este tipo climático (IBGE, 2002). O surto de 2007 e 2009, com grande concentração de epizootias no Rio Grande do Sul, certamente influenciou na escolha deste dado pela árvore de decisão.

O clima Tropical Brasil Central engloba grande parte da área central do país e litoral do meio-norte, este alterna-se entre áreas úmidas e secas. Os recortes sub-regionais evidenciados por essa zona climática foram: Mesotérmico Brando (média entre 10° e 15°C) Úmido com 1 a 3 meses secos e Subquente (média entre 15 e 18°C em pelo menos 1 mês) Superúmido sem seca/subseca. Em ambos os recortes a condição úmido e superúmido destacaram-se. O clima Tropical Brasil Central Subquente (média entre 15 e 18°C em pelo menos 1 mês) Superúmido sem seca/subseca ocorre em porções norte do Paraná, nas porções sul e litoral de São Paulo, e em pequenas porções centrais dos Estados do Rio de Janeiro e Espírito Santo. Já a zona Tropical Brasil Central Mesotérmico Brando (média entre 10° e 15°C) Úmido com 1 a 3 meses secos corresponde a pequenas porções do sul de Minas, do litoral norte de São Paulo, do sudoeste do Espírito Santo e do centro do Rio de Janeiro (IBGE, 2002).

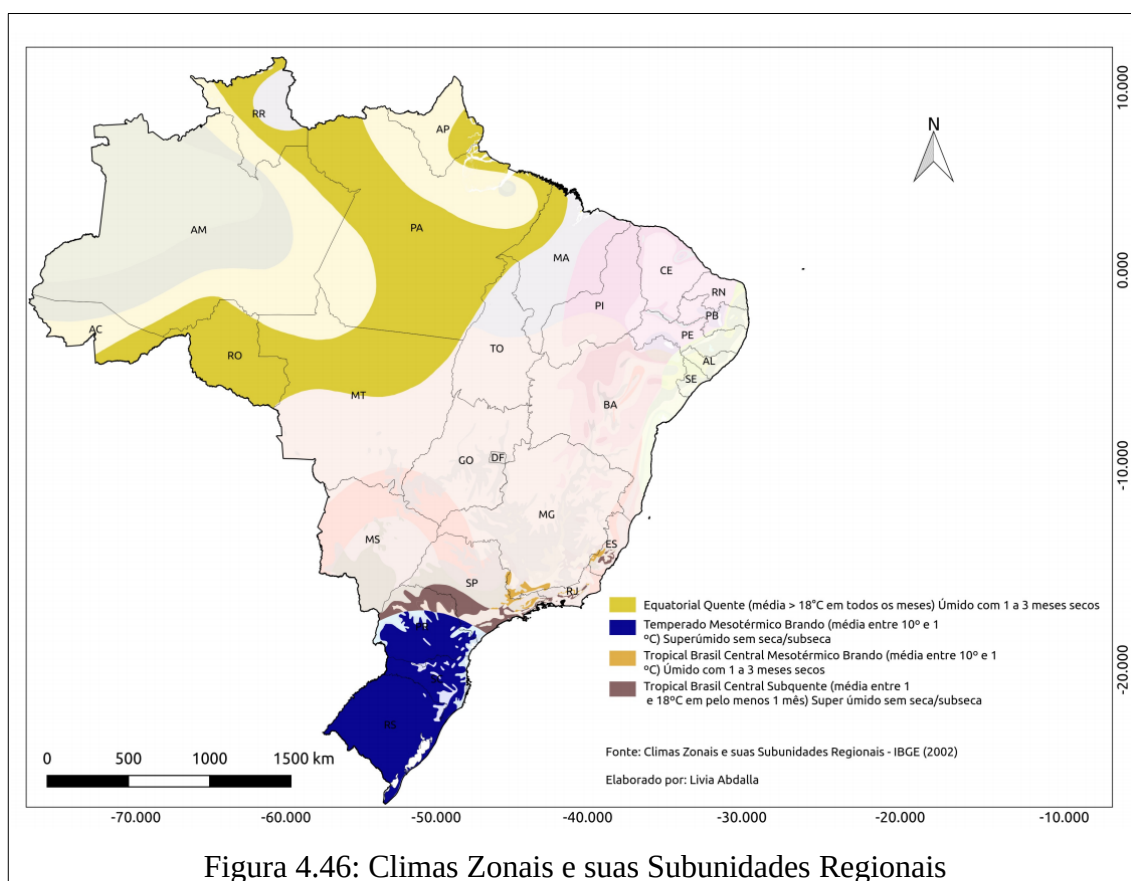


Figura 4.46: Climas Zonais e suas Subunidades Regionais

Nesse subcapítulo já foram indicadas algumas possíveis relações existentes entre a emergência da FA com as variáveis do tipo Normais Climatológicas e com as classes de

Climas Zonais e suas Subunidades Regionais. De modo geral, verificou-se que as condições climáticas possuem convergências diretas com a manifestação da doença, já que as regiões com maior histórico de incidência da doença (Amazonas, Pará, Goiás, Tocantins, Minas Gerais) reúnem, em média, elevados valores de temperaturas e maiores volumes de precipitação.

Fatores climáticos, como temperatura, umidade e duração da estação chuvosa, têm comprovadamente implicações na produção de infecções por FA (VASCONCELOS *et al*, 2001; VASCONCELOS, 2003). Diferentes estudos têm frequentemente citado que fatores meteorológicos como temperatura, chuva e umidade afetam a dinâmica de transmissão de doenças transmitidas por vetores (GUBLER *et al.*, 2001; PATZ e KOVATS, 2002). Na FA há demonstração de que as taxas de transmissão alcançaram níveis muito altos quando esses fatores ambientais se encontram elevados (BARRETT e MONATH, 2003), dado que essas condições propiciam o aumento da população de mosquitos e parecem elevar a frequência de picadas. A longevidade dos mosquitos também é beneficiada pela umidade relativa (VAINIO e CUTTS, 1998) e essa característica favorece a disseminação de um surto, caso não sejam aplicadas medidas de controle, já que uma vez infectado, o mosquito mantém a capacidade de transmitir o vírus até o final de sua vida.

Segundo Almeida *et al* (2016) as chuvas favorecem a reprodução e disseminação dos principais gêneros de mosquitos transmissores da Febre Amarela Silvestre (*Haemagogus* e *Sabethes*) já que esses habitam áreas de mata fechada ou as bordas das matas e depositam seus ovos em cavidades de troncos de árvores e em bromélias, que acumulam água das chuvas.

WHO (2016) indicou vários fatores climatológicos que influenciam nas interações intraespecíficas das espécies e que, conseqüentemente, afetam a transmissão da FA. Alguns estão relacionados ao vírus, como a quantidade de vírus no início da amplificação do ciclo e a virulência; outros dizem respeito ao vetor, como a abundância, longevidade, trofismo, número de repastos sanguíneos/dia, tempo de incubação do vírus no vetor e sua competência vetorial; há também fatores próprios do hospedeiro vertebrado: abundância; taxas de imunidade e suscetibilidade (duração e dimensão da viremia).

Segundo Costa (2005), João Barros Barreto, em publicação de 1949, descreveu que a

emergência da doença tende a ser longa, se iniciar no começo da estação chuvosa, quando há maior proliferação culicidiana; e tenderá a ser curta, se no fim do período de chuvas. Os sintomas iniciais do primeiro caso desta publicação ocorreram em 30 de março e o encontro de macacos mortos na área se deu alguns dias antes, coincidindo com as últimas chuvas da região.

Ainda segundo Costa (2005), também em 1949, Causey e Santos realizaram estudo sobre variações pluviométricas, sua relação com a prevalência de vetores silvestres e a incidência de febre amarela em Passos, Minas Gerais. Descreveram que o pico da pluviosidade foi registrado no mês de dezembro, observando-se maior prevalência de mosquitos *Haemagogus* em janeiro e incidência máxima de casos de febre amarela em fevereiro. Estudo semelhante realizado em 1945 na Colômbia, por Gast-Galvis e Bates, mostrou que o pico das chuvas observado no mês de maio foi seguido por alta prevalência de *Haemagogus* em junho e um mês mais tarde pelo pico de casos de febre amarela.

Monath (1997) relatou que surtos de febre amarela na África foram associados a períodos de chuvas intensas nos países afetados. O padrão observado nesse estudo sugeriu que fatores ecológicos influenciando a atividade viral ocorreram de modo simultâneo em uma extensa região geográfica. Dentre eles, a intensidade de chuva que se antecipou às epidemias na Nigéria em 1978 e Gâmbia em 1986, destacou-se como um dos mais importantes para explicar a amplificação do vírus.

Vasconcelos *et al.* (2000) também relataram que no verão de 1999 e 2000 no Brasil a quantidade de chuvas foi considerada a mais alta para todas as regiões do país, especialmente nas áreas onde houve registro de casos da doença, nas quais as altas taxas de precipitação foram seguidas pelo aumento da população de mosquitos. Segundo os autores durante a epidemia de febre amarela silvestre ocorrida nos estados de Tocantins e Goiás, entre o final de 1999 e os primeiros meses de 2000, estabeleceu-se uma relação de causa e efeito entre a elevada pluviosidade (a mais alta entre todas as regiões geográficas verificadas naquele verão, especialmente nas áreas onde foram notificados casos humanos) e o aumento das populações de mosquitos silvestres transmissores.

Projeções futuras (MEYER *et al.*, 2014; MELO *et al.*, 2016) sugerem que quatro espécies endêmicas de micos-leões (*Leontopithecus spp.*) sofrerão grandes mudanças e/ou reduções de

hábitat nas próximas décadas em função das mudanças climáticas. Similarmente, a espécie muriqui do norte (*Brachyteles hypoxanthus*) deverá sofrer com a redução em mais da metade de sua área atual, com grande declínio das áreas de reservas atualmente protegidas devido às mudanças climáticas. Barrett *et al.* (2013) consideram que as mudanças climáticas intensificarão a exposição dos primatas à parasitas potencialmente nocivos, já que o aumento da temperatura e das chuvas favorecem a ocorrência de períodos mais longos de transmissibilidade do parasita em regiões ricas em primatas.

Diferentes autores indicam que as mudanças climáticas são responsáveis pelas alterações regionais das médias de temperatura e precipitação, influenciando na dinâmica natural de diferentes regiões do globo terrestre e promovendo eventos climáticos extremos, como tempestades tropicais, inundações, ondas de calor, secas prolongadas, nevascas, furacões, tornados, entre outros (WWF, 2015; LEVY e PATZ, 2015; LUBER e LEMERY, 2015; PINKERTON e ROM, 2013). Especialmente nas regiões tropicais o efeito é bastante evidente, considerando as alterações bruscas nos regimes de chuvas, com períodos prolongados de secas, seguidos por chuvas localmente concentradas. Estas alterações climáticas se devem principalmente ao aumento das emissões de gases na atmosfera que causam o efeito estufa, principalmente o dióxido de carbono (CO₂) e, regionalmente, do longo processo de subtração de florestas e exposição do solo ou de cobertura vegetal de baixo porte.

Gouveia *et al.* (2016) destacaram que a fragmentação florestal, resultante das mudanças climáticas, pode limitar a disponibilidade de rotas de dispersão usadas por *Callicebus spp.* nas porções leste do Brasil. Além disso, no futuro, áreas protegidas e parques criados para preservar espécies ameaçadas podem não ser mais adequados, devido à mudanças na cobertura vegetal, e como consequência desse processo, os indivíduos podem migrar para florestas vizinhas e desprotegidas onde eles estarão expostos a caçadores ou residentes locais (ARAÚJO *et al.*, 2004; MALHI *et al.*, 2008; STRUEBIG *et al.*, 2015; WIEDERHOLT e POST, 2010). Por fim, estudos recentes (LAURANCE, 2014; GRAHAM, 2016) destacam ainda que a remoção de florestas para a produção agrícola resulta em dessecação das terras que, quando combinadas com secas e episódios de El Niño, resultam em extensos incêndios florestais, que impactam diretamente as populações de primatas.

4.2.7 Mineração

Foram evidenciados nas árvores de decisão oito tipos de exploração de diferentes recursos minerais no Brasil, com a frequência de indicação em parênteses (figura 4.47): Água mineral (3), Bauxita (2), Ferro (1), Manganês (1), Saibro (1), Caulim (1), Granito (1) e Areia (1). Em todas as 27 Unidades Federativas foram identificadas áreas de exploração de algum desses oito tipos de minérios. Sendo que nas regiões Norte e Sudeste verificaram-se a maior concentração de áreas de extração desses oito minerais listados.

Em 2016 foram exportados pelo Brasil aproximadamente 300 milhões de toneladas de bens minerais (21,6 bilhões de dólares em divisas), o que representa 11,6% das exportações totais do Brasil, e 33% do saldo comercial (IBRAM, 2018). Na indústria mineral brasileira, as substâncias metálicas têm a maior representatividade na economia nacional. Em 2016 essas substâncias responderam por cerca de 77% do valor total da produção mineral brasileira. Dentre as 37 substâncias metálicas, oito destacam-se por corresponderem a 98,6% da produção comercializada da classe: alumínio, cobre, estanho, ferro, manganês, nióbio, níquel e ouro. O valor da produção comercializada dessas oito substâncias totalizou 71,9 bilhões de reais somente em 2016 (DNPM, 2018).

Da mesma maneira que nos subcapítulos anteriores, para cada recurso mineral indicado pelos modelos serão apresentadas suas particularidades no que se refere ao volume de produção, local de exploração e os impactos ambientais associados ao tipo de extração. Ao término desse panorama sobre cada um dos recursos minerais serão discutidas as prováveis relações existentes com os eventos de emergência da febre amarela.

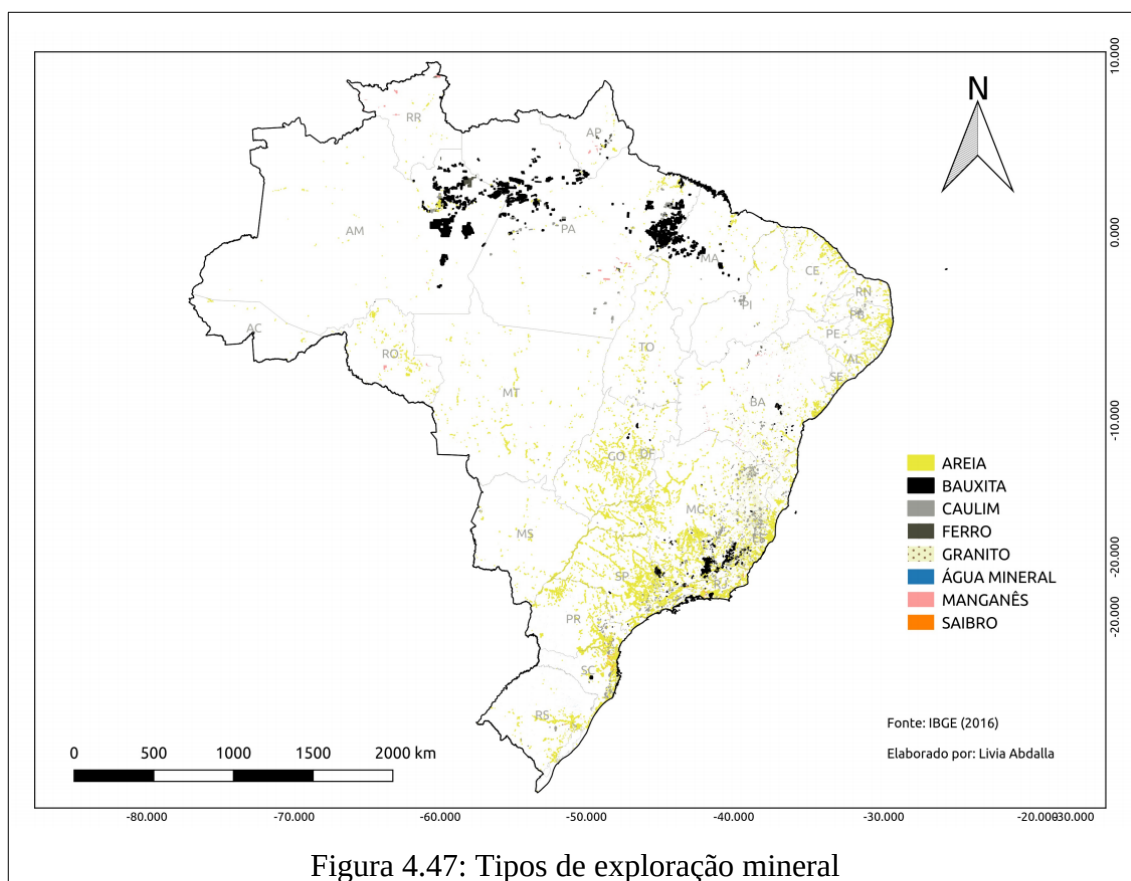


Figura 4.47: Tipos de exploração mineral

O setor de exploração de água mineral no Brasil é disseminado por todo o território nacional, sendo os maiores produtores e consumidores estão concentrados na região sudeste. Em 2009 a produção de água mineral comercializada foi de 7,5 bilhões de litros, sendo 70% para o setor de engarrafamento e 30% para composição de produtos industrializados. Todas as Unidades Federativas possuem participação na produção, em ordem decrescente (DNPM, 2010): SP (25.96%), PE (14.14%), RJ (6.99%), BA (6.50%), MG (5.11%), RS (3.72%), PA (3.53%), PR (3.42%), CE (3.34%), MT (2.94%), AM (2.71%), RN (2.29%), SC (2.09%), AL (1.60%), GO (1.43%), SE (1.37%), ES (1.19%), DF (1.16%), RO (1.11%), PB (1.09%), MA (0.70%), PI (0.68%), AP (0.47%), MS (0.36%), AC (0.29%), TO (0.20%), RR (0.07%), não informado (5.54%). A água mineral ou potável de mesa é obtida diretamente de fontes naturais ou por extração de águas subterrâneas. É fundamental para homologação do local de instalação da jazida a realização de pesquisas geológicas com ênfase nos aspectos hidrogeológicos como clima, vegetação, permeabilidade e sistemas de fraturamento.

Diferentemente dos demais minérios que possuem reservas finitas, a água mineral ou potável de mesa é elemento renovável, se mantidas as condições ambientais e climáticas da região, e respeitada durante o bombeamento, a recarga do aquífero. Para tanto, as indústrias devem realizar o descarte de resíduos (químicos, sólidos e gasosos) de maneira adequada às resoluções do Conselho Nacional do Meio Ambiente (CONAMA nº. 275/2001), evitar a utilização de qualquer equipamento ou instrumento que possa contaminar seu produto e o local de instalação da indústria, bem como implementar projetos de reflorestamento de mata nativa no seu entorno e na zona de recarga do aquífero. Entre as indústrias de mineração, a indústria de água mineral ou potável de mesa é uma das que menos gera resíduos contaminantes. Dessa maneira, por estar associada a ambientes naturais preservados, destaca-se que áreas de extração de água mineral podem estar associadas aos *habitats* de diferentes espécies de primatas.

A Bauxita é o minério a partir do qual se obtém o alumínio, o terceiro elemento em maior abundância na crosta terrestre, depois do oxigênio e silício. Apresenta coloração avermelhada e é encontrada sobretudo em regiões tropicais e subtropicais do planeta, por ação do intemperismo sobre aluminossilicatos. As reservas brasileiras de bauxita estão entre as maiores do mundo e com alta rentabilidade econômica, com concentrações de óxido de alumínio superiores a 40%. A bauxita brasileira participa com cerca de 10% das reservas mundiais, com cerca de 3,6 bilhões de toneladas. A produção nacional é do porte de 25 milhões de toneladas por ano, ocupando o terceiro lugar entre os países produtores, com quase 13% da produção mundial (QUARESMA, 2009a). No ano de 2016 a produção bruta de bauxita no Brasil foi de 51,8 milhões de toneladas, sendo a maior concentração (90,1%) no estado do Pará, com 46,7 milhões de toneladas, seguido por Minas Gerais, com 3,1 milhões, São Paulo, com 1,2 milhões, Goiás, com 736 mil e Santa Catarina, com 4,2 mil toneladas (DNPM, 2018).

Para a recuperação de áreas exploração de bauxita destaca-se a necessidade do manejo e recuperação do solo, tratamento dos resíduos líquidos, bem como o plantio de espécies nativas na área desmatada para a exploração do minério. Para evitar a contaminação do solo e de lençóis freáticos, as áreas destinadas a resíduos de bauxita devem ser impermeabilizadas e a água utilizada no processo de mineração deve ser tratada para ser devolvida ao ambiente

natural. No caso de material sólido as áreas precisam ser reabilitadas e monitoradas na superfície e nos aquíferos.

O minério de ferro destaca-se com a participação mais expressiva dentre as oito substâncias metálicas de maior representatividade na economia nacional, correspondendo a 64,3% (DNPM, 2018). Segundo Quaresma (2009b), o Brasil é o segundo maior produtor de Minério de Ferro no mundo, com média de produção de 390 milhões de toneladas ao ano. Em 2016 o Brasil totalizou 559,5 milhões de toneladas de produção bruta de ferro, sendo os principais estados produtores: Minas Gerais, com 385,5 milhões de toneladas (68,8% da produção); Pará, com 164,5 milhões (29,4%); Mato Grosso do Sul, com 4,8 milhões; São Paulo, 4,5 milhões; Bahia, 70,4 mil; Goiás, com 24 mil; Rio Grande do Norte, com 7 mil; Amazonas, com 4,6 mil; Ceará, com 3,4 mil e Amapá, com 2 mil toneladas (DNPM, 2018). A produção de minério de ferro, tanto nas minas como nas usinas de tratamento de minério, devem estar certificadas com as Licenças de Operações com o objetivo de retenção dos sólidos presentes nos efluentes gerados (lama de beneficiamento), melhorar resultados de reaproveitamento de água e de recuperação de finos de minério.

O minério de manganês é um dos primeiros minerais com produção regular no Brasil, sendo que sua produção data do século XIX. Atualmente a produção média brasileira representa 20% da produção mundial. Característica importante do manganês é que o minério é essencial para a produção do aço. O setor siderúrgico absorve pouco mais de 85% do minério de manganês (forma natural ou transformados em ligas a base de manganês), os 15% restantes sob a forma de dióxido de manganês, 10% são utilizados como componente de pilhas e 5% como insumo da indústria química (adubos, ração) (QUARESMA, 2009c). A produção de manganês no Brasil em 2016 foi de 4,3 milhões de toneladas. As principais reservas nacionais estão localizadas nos estados de Minas Gerais (170 mil), Mato Grosso do Sul (794 mil), Bahia (5 mil), Rondônia (99 mil), Mato Grosso (45,7 mil), Ceará (2,5 mil) e Pará (3,2 milhões). (DNPM, 2018). Assim como as áreas de exploração de bauxita e ferro, as áreas de exploração de manganês também necessitam ser recuperadas, portanto é fundamental realizar o manejo e recuperação do solo, o tratamento dos resíduos sólidos e líquidos, bem como o plantio de espécies nativas na área desmatada.

No Brasil, as reservas de Caulim correspondem a 5,0 bilhões de toneladas (FARIA,

2009). Os estados com maior produção de Caulim são: SC (50.41%), PA (16.51%), MA (10.77%), SP (6.51%), PR (4.64%), MG (3.20%), RS (1.88%), BA (1.47%), GO (1.03%), não informado (3.58%) (DNPM, 2018). Os rejeitos de mineração e resíduos de beneficiamento das indústrias de Caulim, por manusearem produtos argilosos, arenosos e silticos, deve-se atentar para o impacto de qualquer contaminação destes produtos nos sistemas de drenagem, pois estes, em suspensão, podem ser conduzido a grandes distâncias, podendo ser detectados pelo turvamento que causam nas águas. No beneficiamento do caulim, as emissões de gases de efeito estufa correspondem principalmente ao consumo de eletricidade por parte dos equipamentos (motores) usados para o beneficiamento e secagem do minério.

Três tipos de exploração minerais indicados nas árvores de decisão (areia, saibro e granito) pertencem a classe II do Regulamento do Código de Mineração, que correspondem a jazidas de substâncias minerais de emprego imediato na construção civil (ardósias, areias, cascalhos, quartzitos e saibros, quando utilizados "*in natura*" para o preparo de agregados, argamassa ou como pedra de talhe, e que não se destinem, como matéria-prima, à indústria de transformação). Para os minerais de classe II a resolução CONAMA nº 10/90 estabelece critérios específicos para o licenciamento ambiental desses minerais, por considerar um caso de interesse social. Essa mesma resolução deixa a critério do órgão ambiental competente a apresentação ou não de Estudo de Impacto Ambiental (EIA) e de Relatório de Impacto Ambiental (RIMA) por parte do empreendedor, levando em conta a natureza, localização, porte e demais peculiaridades. Contudo, o EIA/RIMA deve ser substituído por outro documento: o Relatório de Controle Ambiental (RCA). Dependendo do caso, o órgão ambiental competente poderá autorizar até a intervenção ou supressão de vegetação em Área de Preservação Permanente (APP). O que representa ameaça aos ecossistemas naturais (como leito de rios), nos quais populações de PNHs habitam ou circulam.

A extração de areia organiza-se em duas principais áreas: fins industriais (tais como fabricação de vidro, moldes para fundição, matéria-prima na indústria cerâmica, entre outros) e construção civil. A areia é extraída de leito de rios, várzeas, depósitos lacustres, mantos de decomposição de rochas, pegmatitos e arenitos decompostos. A distribuição regional é a que segue: SP (23.01%), RS (19.33%), RJ (13.19%), SC (6.61%), MG (4.84%), PR (3.68%), DF (2.68%), CE (1.77%), PE (1.48%), ES (1.33%), GO (1.21%), MT (1.08%), BA (1.06%), MS

(1.02%), MA (0.97%), RO (0.57%), PA (0.49%), AM (0.40%), PB (0.38%), SE (0.33%), PI (0.31%), AP (0.25%), RN (0.23%), AC (0.19%), AL (0.15%), TO (0.14%), RR (0.03%), Mercado Externo (0.25%), não informado (13.02%). A mineração de areia industrial no Brasil concentra-se principalmente nas regiões Sudeste e Sul do País, sendo que a maior demanda por areia industrial encontra-se na região Sudeste do País, principalmente na Grande São Paulo, em função da expressividade da sua zona industrial e na construção civil (COELHO, 2010). A atividade econômica de produção de areia caracteriza-se por grandes volumes produzidos. O transporte responde por cerca de 2/3 do preço final dos produtos, o que impõe a necessidade de produzi-la o mais próximo possível do mercado, que são os aglomerados urbanos.

De modo semelhante a exploração de água mineral, os recursos minerais para a produção de areia são teoricamente abundantes. Entretanto, se não forem adequadamente protegidos poderão ser esgotados pela exploração mineral e urbanização. Portanto, é importante o ordenamento territorial, isto é, zoneamentos que protejam recursos minerais como areia, argila e rocha. Importante destacar que a mineração de areia tem significativa interferência em Áreas de Proteção Permanentes (APP), com remoção da vegetação e alteração de ecossistemas locais. O impacto ambiental promovido varia com o tipo de paisagem e método de extração adotado. Os ambientes mais comuns de exploração do minério são: margens de rios (planícies e terraços aluviais), com a instalação de dutos de bombeamento; restingas e dunas litorâneas; topo de morros e encostas com declividade superior a 45° (arenitos ou quartzitos em platôs com escarpas).

Já o Saibro se origina do intemperismo de rochas graníticas. É bastante utilizado na construção civil e manutenção das redes viárias (OLIVEIRA *et al*, 2009). Sobre as áreas degradadas pela extração de saibro, Silva e Toledo (2010) e Brollo *et al* (2005) descrevem que a extração de Saibro provoca grandes alterações no terreno e entorno, pois para sua obtenção é necessário retirar a cobertura superficial do solo que pode ser muito espessa, podendo chegar a 30 metros ou mais, levando a intensificação de processos erosivos. Como resultado, verifica-se a quase impossibilidade de recuperação vegetal, uma vez que não há mais substrato fértil para o seu desenvolvimento e sim um substrato árido, com alto índice de acidez, devido a característica da rocha. Além disso, a degradação pode encontrar terreno com

alta suscetibilidade a movimentos de massa, ampliando ainda mais a área degradada, especialmente se houver interferência antrópica, resultando em áreas de risco.

Assim como a Areia e o Saibro, a exploração da substância mineral Granito também possui facilidade na obtenção da autorização de extração mineral. Para os minerais de classe II, a questão ambiental é muitas vezes negligenciada. Os impactos causados pela mineração de agregados para construção civil podem ser mais perceptíveis dentro dos centros urbanos, em razão da proximidade das áreas produtoras com estes, porém também representam ameaças aos rios, aos solos, aos fragmentos de matas peri-urbanos em que populações de PNHs podem habitar e/ou transitar.

De modo geral, a mineração é responsável por diferentes impactos socioambientais como: desmatamento; assoreamento de cursos d'água; elevado consumo e contaminação de água nas lavras de mineração; destruição da biodiversidade; apropriação de terras; prejuízos à saúde de populações humanas residentes próximas às áreas de mineração; migração de populações de animais silvestres (ARAÚJO, 2005). Sobre esses impactos, Silva (2007) afirma que eles alteram intensamente a área minerada e as áreas vizinhas, onde são realizados os depósitos de estéril e de rejeito. Além disso, quando se tem a presença de substâncias químicas nocivas na fase de beneficiamento do minério, isto pode significar um problema sério do ponto de vista ambiental. Portanto, a exploração de jazimentos de minério representa uma expressiva ameaça ao equilíbrio de ecossistemas naturais e à produtividade biológica dos corpos d'água, rios, corixos e afluentes, visto o assoreamento provocado pelos rejeitos de minérios e a contaminação por mercúrio. O mesmo autor ainda ressalta que esse processo é ainda mais evidente nas áreas de extração de agregados para a construção civil, pois a produção desses minerais, por fatores mercadológicos, impõe sua atuação próxima dos centros consumidores, caracterizando-se como uma atividade típica das regiões metropolitanas e urbanas. Sobre a mineração em áreas urbanas e periurbanas, o IBAMA (2006) afirma ser um dos fatores responsáveis pela degradação do subsolo. Por esse motivo, pode-se dizer que esse tipo de mineração, se não administrado de maneira correta, também configura um problema de gestão ambiental urbana, uma vez que essas jazidas estão em áreas onde existe certa concentração populacional e que afetam direta ou indiretamente a qualidade ambiental nessas áreas.

Nesse contexto, destaca-se que a fiscalização sobre o cumprimento das medidas de compensação ambiental é responsabilidade do Departamento Nacional de Produção Mineral (DNPM), que deve planejar e fomentar o aproveitamento dos recursos minerais, além de superintender as pesquisas geológicas e minerais. As ações de controle devem seguir hierarquicamente a Constituição Federal de 1988 e as leis estaduais e municipais, bem como as resoluções do CONAMA.

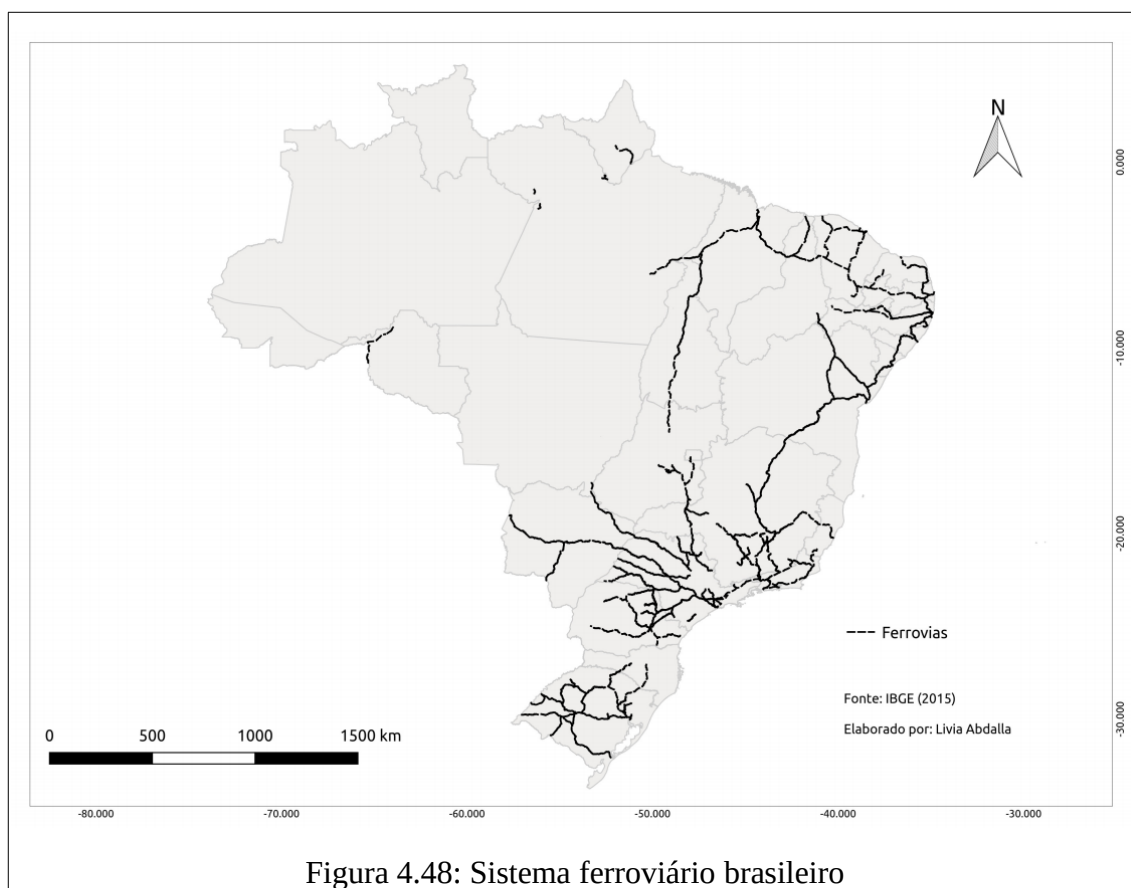
Alvarez-Berríos e Aide (2015) indicam que, entre as ameaças oriundas da exploração mineral aos *habitats* de muitas espécies de animais silvestres (especialmente dos primatas), destacam-se: o desmatamento; a fragmentação ou destruição do *habitat*, o envenenamento e/ou poluição do solo e das águas subterrâneas. O mesmo autor ainda ressalta que no Brasil entre 2001 e 2013, cerca de 1.680 km² de floresta úmida tropical foram perdidos em 1.600 sítios de mineração de ouro, incluindo perdas florestais significativas dentro de 13 áreas protegidas.

Sonter *et al.* (2017) em um estudo mais recente mostrou que, entre 2005 e 2015, a mineração no Brasil aumentou significativamente a perda de florestas na Amazônia (até 70 km além dos limites de locação de minas), causando 11.670 km² de desmatamento (a qual representa 9% de toda a perda florestal da Amazônia durante este período). Por tudo isso, a eliminação de resíduos de mineração é uma ameaça significativa para a biota local, incluindo primatas.

Segundo Garcia *et al* (2017), no Brasil tem-se, além dos impactos ambientais diretos, o risco eminente de rompimentos de 126 barragens de mineração que estão atualmente classificadas sob risco de rompimento. Em um desses casos, o rompimento da barragem poluiu com lama tóxica centenas de quilômetros do Rio Doce, localizado em Mariana – MG, em novembro de 2015, esse rompimento impactou todo o contexto social e ambiental da região, o que inclui as espécies de primatas que habitavam a região. Além disso, a mineração estimula a migração humana, a extração ilegal de madeira e a colonização de áreas florestais, a caça e a construção de estradas e ferrovias (ALAMGIR *et al.*, 2017; BUTT *et al.*, 2013; LAURANCE *et al.*, 2015).

4.2.8 Ferrovias

As redes de ferrovias foram evidenciadas uma vez nas árvores de decisão dos modelos experimentais (figura 4.48). Antes de iniciar a discussão sobre as prováveis relações da doença com esta variável temática, deve-se destacar alguns aspectos sobre o sistema ferroviário brasileiro, o qual conta com 30,6 mil quilômetros de vias ferroviárias funcionais no país, com 3.043 locomotivas e 102.024 vagões em circulação (MTPA, 2018). Os principais produtos transportados são *commodities* para exportação ou atendimento de demandas internas (por exemplo: soja e farelo de soja; indústria cimenteira e de construção civil; minério de ferro; produtos siderúrgicos; ferro gusa; óleo diesel; adubos e fertilizantes; açúcar; carvão mineral; milho; carne; celulose). Tratando-se de um país com dimensões continentais, esses produtos são transportados por longas distâncias entre as regiões onde são produzidos, para polos industriais brasileiros ou para portos de exportação. Segundo a Agência Nacional de Transporte Ferroviário (ANTF, 2018), em 2017 o volume total de cargas transportadas foi de 538,8 milhões de toneladas. Entre os 15 grupos de mercadorias transportadas pelos 12 trechos concedidos à iniciativa privada, a produção agrícola teve a maior alta na comparação com 2016, passando de 10,5 milhões de toneladas para 18,2 milhões.



A crescente demanda global por culturas alimentares e não alimentares (madeira, combustíveis fósseis, minerais, gemas, entre outras), que são produzidas por um pequeno número de nações, resultou em um rápido aumento na produção agrícola, extração de madeira, mineradores itinerantes e extração de petróleo/gás. Isso também levou a uma expansão das redes ferroviárias e rodoviárias e ao desenvolvimento de energia hidrelétrica (ALAMGIR *et al.*, 2017), que promoveu o aumento da perda de florestas, a ocupação ilegal de terras e exploração madeireira, aumentos de mineração itinerante, além do aumento na caça de primatas para comercialização (ESTRADA *et al.*, 2017; LATRUBESSE *et al.*, 2017; PLUMPTRE *et al.*, 2015; SPIRA *et al.*, 2017; TIMPE e KAPLAN, 2017; WINEMILLER *et al.*, 2016).

Dentre as possíveis relações entre os locais de manifestação da febre amarela com as redes ferroviárias, considera-se principalmente a possibilidade de que essas locomotivas, transportando diferentes tipos de *commodities* e matérias-primas, podem estar servindo de

meio de transporte às espécies de vetores transmissores de febre amarela. Isso porque os vagões locomovem-se entre longas distâncias e conectam centros de mineração e siderurgia (ex: Quadrilátero Ferrífero no sul de Minas Gerais), polos industriais e áreas agrícolas do País (especialmente a Região Centro-Oeste) aos mais importantes portos e regiões metropolitanas do país (Santos – SP, Itaqui – MA, Vitória – ES e Rio de Janeiro – RJ). Como exemplo da conexão por meio de ferrovias entre estados endêmicos de febre amarela e estados em que a doença se manifesta de maneira reemergente, Quaresma (2009c) descreve que o escoamento do minério de manganês, das regiões produtoras aos principais clientes domésticos, é feito por ferrovia. Nesse contexto, destacam-se as ferrovias que interligam os produtores de gusa da região norte (PA e MA) e os produtores de ligas de base de manganês (localizados em sua maioria na região sudeste do país – MG, SP, e ES) com uma única unidade de fabricação de liga, no Estado da Bahia.

As regiões nas quais essas atividades (agrícolas, extrativistas, etc) estão estabelecidas reúnem ambientes que foram ecologicamente desestruturados (em razão das práticas de exploração estabelecidas), ao mesmo tempo em que possuem proximidade com fragmentos de matas semipreservados, nos quais espécies de primatas sobrevivem. Nessa dinâmica de desestruturação de ambientes naturais e consequente redução de populações de primatas, as espécies de vetores que se utilizavam dessas espécies para alimentação podem encontrar nos vagões ferroviários um ambiente de fuga e busca de novas espécies de hospedeiros em maior proporção, no caso, humanos. Muitos desses vagões são escuros e fechados, o que permite que muitas espécies de mosquitos possam ser atraídos a adentrar esses vagões e permaneçam neles até o momento de abertura nos locais de destino. Considera-se que esse ambiente é mais favorável aos mosquitos do que transportes como caminhões, por exemplo, que podem ter frestas ou janelas com correntes de ar que expulsam os mosquitos pelo trajeto. Nesse contexto, Benchimol e Silva (2008) destacam que nas pesquisas desenvolvidas por Adolpho Lutz, em 1930, o autor descreve que foi o transporte de mosquitos infectados pela estrada de ferro que determinou os surtos de Rio Claro e Belém do Descavado, em 1889, bem como vários casos isolados entre trabalhadores do correio e da ferrovia, inclusive “numa senhora em Valinhos, que era casada com o chefe da estação e nela morava. Nunca tinha visitado Campinas, enquanto lá havia febre amarela”. Em anos subsequentes a doença, Lutz relatou

que a febre amarela:

“...invadiu o estado de São Paulo seguindo primeiramente a estrada de ferro Paulista. Devastou Limeira, onde eu tinha clinicado durante cinco anos sem ver estegomias, mas poupou Araras e Jundiaí, onde vilas ficam muito distantes das estações. Depois da Paulista foi invadida a Mogiana, e só muito tarde tocou também a vez da Sorocabana. O que impediu uma expansão mais rápida foi, sem dúvida, o fato de que, sem a importação prévia ou simultânea da estegomia, os numerosos casos, que apareceram entre as pessoas fugidas de centros epidêmicos, não podiam propagar a febre amarela.”

Segundo Benchimol e Sá (2005), dois médicos pesquisadores alemães (Otto e Neumann) relataram ainda que a estegomia (nome atribuído ao vetor transmissor da febre amarela no início do século XX) não gostavam de correntes de ar, mas podiam subir a altitudes de várias centenas de metros, não faziam grandes migrações, mas viajavam longas distâncias de trens, bondes, carros e navios, especialmente em cargas de açúcar, melado, frutas e em pacotes úmidos. Otto e Naumann puseram estegomias em malas contendo roupas secas e úmidas e verificaram que as fêmeas mais resistentes sobreviviam até 15 dias. À época em que ocorriam esses surtos (início do século XX), as explicações que se destacavam como as mais inovadoras eram as que relacionavam a capacidade de micro-organismos de viajar por diferentes ambientes e de indivíduo a indivíduo, por intermédio do ar e de outros veículos. Enquanto que a corrente majoritária acreditava que o “germe” sofria alguma transformação no meio exterior antes de se tornar infectante, somente em algumas zonas geográficas e em determinada estação do ano.

4.2.9 Solos

O Brasil possui uma grande diversidade de solos em sua extensão continental, decorrente da ampla diversidade de pedoambientes e de fatores de formação do solo. A ampla peculiaridade observada nos pedoambientes também representa uma condição importante para a avaliação dos potenciais e limitações de cada solo, condicionando a sua sustentabilidade em função dos usos e das práticas de manejo aplicadas (EMBRAPA, 2018). Nas treze classes de solos contidas no Sistema Brasileiro de Classificação de Solos (SiBCS), constata-se a

influência desses fatores por meio da grande variabilidade das características químicas, físicas e morfológicas. No Brasil predominam os Latossolos, Argissolos e Neossolos, que no conjunto se distribuem em aproximadamente 70% do território nacional. As classes Latossolos e Argissolos ocupam aproximadamente 58% da área e são solos profundos, altamente intemperizados, ácidos, de baixa fertilidade natural e, em certos casos, com alta saturação por alumínio. Também ocorrem solos de média a alta fertilidade, em geral pouco profundos em decorrência de seu baixo grau de intemperismo. Estes se enquadram principalmente nas classes dos Neossolos, Luvisolos, Planossolos, Nitossolos, Chernossolos e Cambissolos.

Nas árvores de decisão dos modelos experimentais foram evidenciados três tipos de solos: Latossolo bruno Distrófico (2), Latossolo vermelho Distrófico (1), Neossolo litólico Eutrófico (2) (figura 4.49). Serão inicialmente descritas as principais características pedológicas, distribuição desses solos no território brasileiro, bem como competência para a produção agrícola. Após esta descrição, serão levantadas hipóteses que relacionam a indicação desses solos com os eventos de emergência da doença.

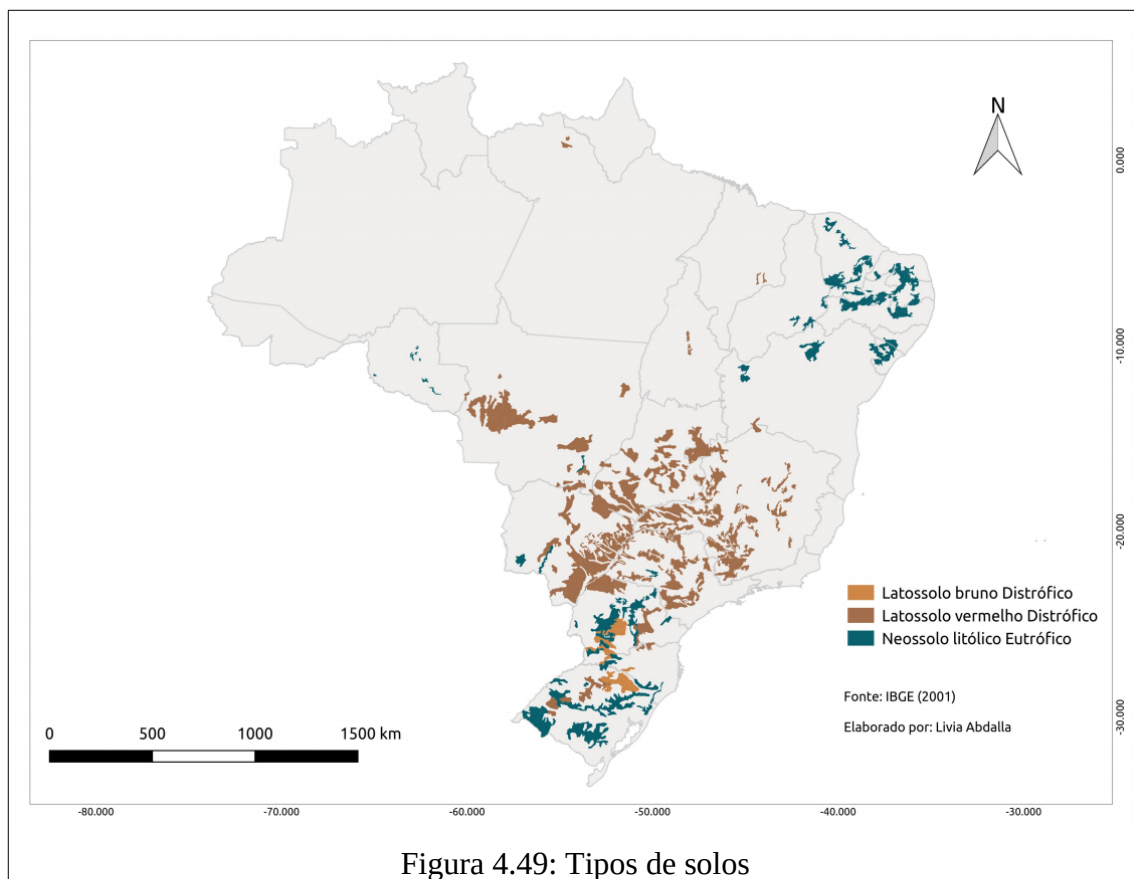


Figura 4.49: Tipos de solos

Segundo Almeida *et al.* (2018)

“os Latossolos Brunos ocorrem principalmente nas áreas elevadas dos planaltos do Sul do Brasil, onde a altitude supera 800m (clima frio e úmido) e com menor expressão no planalto de Poços de Caldas, Minas Gerais, e sua borda em território paulista. No Sul do Brasil, têm sido utilizados para fruticultura”.

Em condições naturais esses solos são profundos ou muito profundos, não apresentando limitação física para o desenvolvimento radicular. No entanto, estes solos possuem baixo potencial nutricional e teores de alumínio em profundidade, o que pode dificultar o enraizamento. Esta condição é agravada ainda pela baixa quantidade de água disponível. Os Latossolos Brunos podem ser classificados no terceiro nível categórico do SiBCS (Acriféricos, Ácricos, Aluminoféricos, Alumínicos, Distroféricos e Distróficos), essas classificações de terceiro nível são relacionadas às características destas classes de solos e às implicações para uso e manejo. As características do tipo Distróficos correspondem a solos de baixa fertilidade (EMBRAPA, 2018).

Latosolos Vermelhos são utilizados em grande parte da produção de grãos do país, pois ocorrem predominantemente em áreas de relevo plano e suave ondulado (em menor expressão, podem ocorrer em áreas de relevo ondulado), tornando propícia a mecanização agrícola. Solos do tipo vermelhos, geralmente possuem grande profundidade, são homogêneos e de boa drenagem e, quase sempre, possuem baixa fertilidade natural, necessitando portanto, de correções químicas para aproveitamento agrícola (EMBRAPA, 2018). Segundo Almeida *et al.* (2018) ocorrem em praticamente todas as regiões do Brasil, mas têm grande expressividade nos chapadões da Região Central (Goiás, Distrito Federal, Mato Grosso, Minas Gerais e outros) e em São Paulo, e com menor expressividade na Região Sul do país. Os latossolos vermelhos evidenciados também são classificados no terceiro nível categórico do SiBCS (Perféricos, Acriféricos, Acrícos, Aluminoféricos, Distroféricos, Distróficos, Eutroféricos e Eutróficos). A nomenclatura do tipo Distróficos também correspondem a solos de baixa fertilidade.

Neossolos Litólicos são geralmente indicados para preservação da flora e fauna, porém em algumas regiões, verifica-se que estes solos são utilizados para produção de café e milho (estados de São Paulo e Minas Gerais); milho, feijão e soja (Santa Catarina); viticultura e pastagem no (Rio Grande do Sul). Esses solos são geralmente associados a relevos mais declivosos, correspondendo a solos rasos, onde geralmente a soma dos horizontes sobre a rocha não ultrapassa 50 cm. As limitações ao uso agrícola estão relacionadas a pequena profundidade, a presença da rocha e aos declives acentuados associados às áreas de ocorrência destes solos, estes fatores portanto, limitam o crescimento radicular, o uso de máquinas e elevam o risco de erosão (EMBRAPA, 2018). Os Neossolos Litólicos evidenciados também são classificados no terceiro nível categórico do SiBCS (Hísticos, Húmicos, Carbonáticos, Chernossólicos, Distro-úmbricos, Distróficos, Eutro-úmbricos, Eutrófico). As características do tipo Eutrófico correspondem a solos de alta fertilidade.

A relação dos casos de febre amarela com os tipos de solos não parece ser direta, mas essa relação pode ser compreendida do ponto de vista que os três tipos de solos possuem relação com áreas destinadas à produção agrícola (neossolo litolítico eutrófico e latossolo vermelho) e com áreas elevadas com clima frio e úmido dos planaltos do Sul do Brasil, onde a altitude supera 800m (latossolo bruno). Conforme foi discutido no subcapítulo 4.1.1.5, a

produção agrícola promove o desmatamento de muitas áreas naturais, que eram ocupadas por populações de primatas (e consequentemente dos vetores) que foram obrigados a migrar de fragmentos, tornando-os mais vulneráveis ao acometimento à doença. Já as regiões de elevadas altitudes, associadas aos solos do tipo latossolo bruno, são áreas ainda preservadas, muitas vezes associadas a áreas de preservação permanente (topo de morros), onde os primatas encontram refúgios para manutenção e sobrevivência de suas populações.

4.2.10 Relevô

Três unidades geomorfológicas foram evidenciadas pelas árvores de decisão: Patamares (2), Serras (2), Planaltos (1) (figura 4.50). Outro dado correspondente ao relevo foi Altitude (1). As unidades geomorfológicas são definidas como um arranjo de formas altimétricas e fisionomicamente semelhantes em seus diversos tipos de modelados (IBGE, 2009). A geomorfogênese e a similitude de formas podem ser explicadas por fatores paleoclimáticos e por condicionantes litológica e estrutural. Cada unidade geomorfológica evidencia seus processos originários, formações superficiais e tipos de modelados diferenciados dos demais. O comportamento da drenagem, seus padrões e anomalias são tomados como referencial à medida que revelam as relações entre os ambientes climáticos atuais ou passados e as condicionantes litológicas ou tectônicas (GUERRA e CUNHA, 1994).

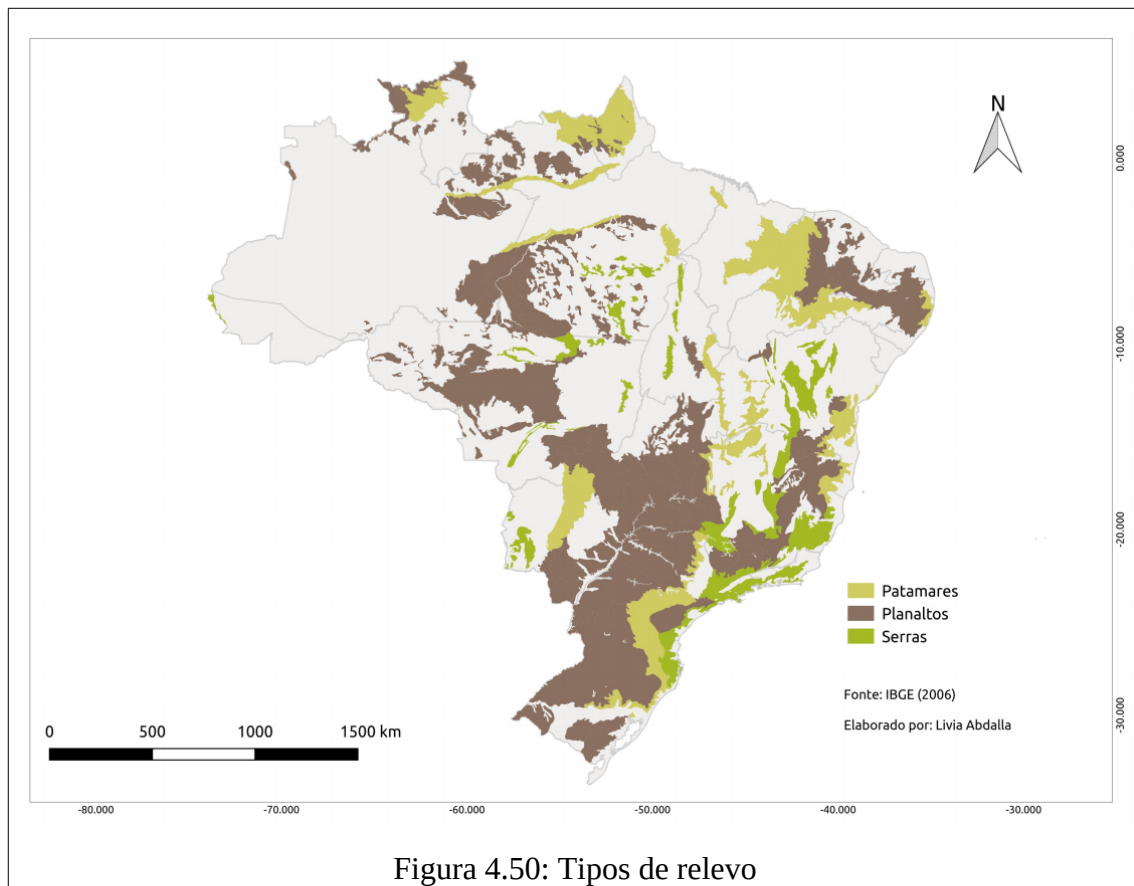


Figura 4.50: Tipos de relevo

Segundo IBGE (2009):

- “Os patamares são relevos planos ou ondulados, elaborados em diferentes classes de rochas, constituindo superfícies intermediárias ou degraus entre áreas de relevos mais elevados e áreas topograficamente mais baixas”;
- “Os planaltos são conjuntos de relevos planos ou dissecados, de altitudes elevadas, limitados, pelo menos em um lado, por superfícies mais baixas, onde os processos de erosão superam os de sedimentação”;
- “As serras constituem relevos acidentados, elaborados em rochas diversas, formando cristas e cumeadas ou as bordas escarpadas de planaltos”.

Essas unidades geomorfológicas estão diretamente relacionadas com clima, temperatura, precipitação, chuvas orográficas, vegetação e logicamente com o *habitat* de espécies de primatas e vetores transmissores do vírus. Todos esses elementos já foram discutidos anteriormente e demonstram relações com a ocorrência da doença. Então, mesmo tratando-se de uma relação indireta com a doença, o modelo foi capaz de identificar que existem uma influência desses padrões de relevo com os locais de ocorrência da doença.

Apesar da relação não parecer direta, algumas hipóteses podem ser levantadas sobre como o relevo pode influenciar no padrão de distribuição e deslocamento das espécies envolvidas. Considera-se que em regiões de planaltos e patamares o deslocamento de espécies de primatas é facilitada, principalmente quando existem corredores ecológicos para sua migração. Já em regiões de serras, com ampla variabilidade altimétrica, o deslocamento de primatas exige maior esforço e, conseqüentemente, maior tempo. Com relação as espécies de vetores, nas regiões de planaltos e patamares eles possuem deslocamento mais lento, já que seu raio de voo é em média de 6 km, o que faz com que o vetor demore mais tempo para se deslocar entre fragmentos de mata em busca de hospedeiros primatas; em ambientes de serras, seu deslocamento vertical ou entre topo de morros é mais facilitado, permitindo que os vetores se alimentem mais rapidamente de outras populações de primatas que habitam diferentes estratos florestais. Conforme descrito por Benchimol e Sá (2005), os médicos Otto e Neumann relataram que o vetor apesar de não gostar de correntes de ar, este podia se deslocar em altitudes de várias centenas de metros.

Essa influência do relevo pode ainda explicar a rapidez de deslocamento de epizootias e maior quantidade de PNHs acometidos nos três surtos de febre amarela. Destaca-se que dispersão de casos no surto mais recente ocorreu em espaço-tempo menor, principalmente nas regiões de Serras de leste de Minas Gerais, centro-oeste do Espírito Santo e centro-sul do Rio de Janeiro. Nos dois surtos anteriores, a dispersão de casos ocorreu de maneira mais lenta e gradual, principalmente entre as regiões de Planaltos de Goiás, Mato Grosso, Mato Grosso do Sul e oeste de Minas Gerais.

4.3 Modelos finais

Conforme descrito anteriormente, o método J48 permite a compreensão e a interpretação de quais atributos são relacionados com a ocorrência ou não ocorrência da doença, enquanto que o método *Random Forest* contribuiu para a obtenção de cenários preditivos mais acurados, sendo possível indicar antecipadamente os novos e prováveis municípios de manifestação da doença. Importante destacar que todas as relações indicadas pelos métodos foram obtidas de forma imparcial, sem quaisquer tendenciosidades (de conhecimento prévio) impostas sobre os classificadores, excetuando-se pela inerente possibilidade de tendência dos

próprios dados utilizados.

4.3.1 Modelos finais – Método J48

A partir da avaliação dos modelos experimentais foram identificados os modelos para cada um dos recortes temporais dos surtos, com os maiores valores de acurácia interna e acurácia preditiva, ou seja, os modelos com maior coerência interna e com maior capacidade de generalização. Dentre os modelos construídos com o método J48, destacam-se: 1.4) Sub-recorte temporal dos dados - Atemporais; 2.0) Todos os atributos; 3.4) Sub-recorte temporal dos dados - Atemporais; 4.5) Sub-recorte temporal dos dados - Sem regularidade temporal. Visto que os modelos produzidos com o método J48 permitem a clara interpretação dos atributos mais representativos e dos valores/estatísticas de separação associados com a manifestação ou não do fenômeno, serão apresentadas e discutidas interpretações factíveis e as principais hipóteses explicativas.

Sobre as árvores de decisão obtidas com o método J48 destaca-se que o critério utilizado para realizar as partições (“Se”) é o da utilidade do atributo para a classificação e o “valor” é o valor desse ganho que melhor divide o conjunto em positivos e negativos. Aplica-se, por este critério, um determinado ganho de informação a cada atributo. O atributo escolhido como atributo teste para o corrente nó é aquele que possui o maior ganho de informação. A partir desta aplicação, inicia-se um novo processo de partição. Nos casos em que a árvore é usada para classificação, os critérios de partição mais conhecidos são baseados na entropia e no índice Gini, já evidenciados anteriormente.

4.3.1.1 Recorte 01 (1999 a 2017)

O modelo construído com todo histórico utilizando o método J48 indicou 16 atributos, sendo oito variáveis relativas à precipitação, três referentes à temperatura, dois tipos de solos, um aspecto geomorfológico, um bioma e um tipo de corpo d'água. Todos os dados evidenciados são do tipo Atemporais, ou seja, variáveis ou dados que não possuem modificações expressivas para a escala de tempo humana. A árvore de decisão produzida contém 16 nós e 7 níveis (figura 4.51), que demonstram quais atributos foram mais

representativos e quais são os respectivos valores normalizados e estatísticas que melhor dividem os municípios em positivos e negativos.

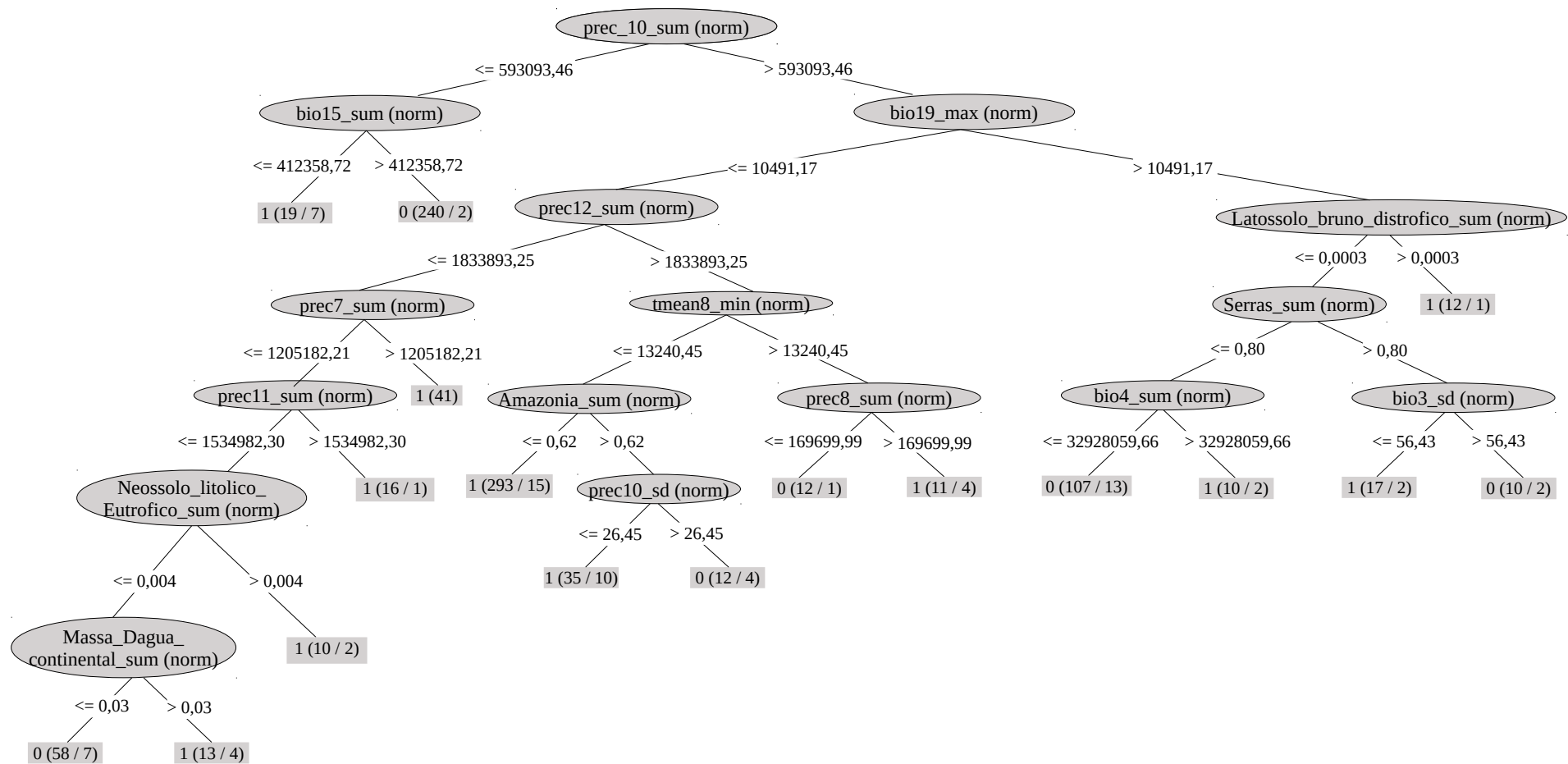


Figura 4.51: Árvore de decisão (método J48) para o recorte 1999 a 2017

Os dados/variáveis evidenciados pela árvore correspondem a 16 atributos, em ordem de indicação na árvore de decisão (de cima para baixo, da esquerda para a direita):

- a) Somatório dos valores de Precipitação total mensal do mês de Outubro (prec10_sum);
- b) Somatório dos valores de Sazonalidade da Precipitação (coeficiente de variação) (bio15_sum);
- c) Valor máximo de Precipitação do trimestre mais frio (bio19_max);
- d) Somatório dos valores de Precipitação total mensal do mês de Dezembro (prec12_sum);
- e) Somatório das áreas de solos do tipo Latossolo_bruno_Distrófico (sum);
- f) Somatório dos valores de Precipitação total mensal do mês de Julho (prec7_sum);
- g) Valor mínimo de Temperatura média mensal do mês de Agosto (tmean8_min);
- h) Somatório das áreas de geomorfologia do tipo Serras_sum;
- i) Somatório dos valores de Precipitação total mensal do mês de Novembro (prec11_sum);
- j) Somatório das áreas do bioma AMAZÔNIA_sum;
- k) Somatório dos valores de Precipitação total mensal do mês de Agosto (prec8_sum);
- l) Somatório dos valores de Sazonalidade da Temperatura(desvio padrão * 100) (bio4_sum);
- m) Desvio padrão dos valores de Isotermalidade ((bio2/bio7) (* 100)) (bio3_sd);
- n) Somatório das áreas de solos do tipo Neossolo litólico Eutrófico (sum);
- o) Desvio padrão dos valores de Precipitação total mensal do mês de Outubro (prec10_sd);
- p) Somatório das áreas de Massa D'agua Continental (sum).

Como os dados utilizados correspondem aos valores normalizados (em razão da área, a fim de evitar vieses sobre a proporção/variabilidade territorial dos municípios), foram então

verificados quais os valores reais desses 16 atributos, para tornar viável a compreensão e discussão sobre os valores de cada um dos nós de separação. Para tanto, foram selecionados exatamente os mesmos atributos (dado/variável e estatística) só que com os valores originais, ou seja, os atributos não normalizados do banco de dados (tanto para a base de treino como para a base de teste). A partir dessa seleta base de dados, contendo apenas 16 atributos, foi novamente aplicado o método J48 (com os mesmos parâmetros de configuração descritos no subcapítulo 3.6), e assim foi obtida uma árvore de decisão, contendo 13 nós e 6 níveis (figura 4.52). Dentre os dezesseis atributos fornecidos ao método, somente oito destes foram utilizados na árvore de decisão: Somatório dos valores de Sazonalidade da Precipitação (coeficiente de variação) (bio15_sum); Valor máximo de Precipitação do trimestre mais frio (bio19_max); Valor mínimo de Temperatura média mensal do mês de Agosto (tmean8_min); Somatório das áreas de geomorfologia do tipo Serras; Somatório dos valores de Precipitação total mensal do mês de Novembro (prec11_sum); Somatório das áreas do bioma AMAZÔNIA_sum; Somatório dos valores de Precipitação total mensal do mês de Agosto (prec8_sum); Somatório das áreas de solos do tipo Neossolo litólico Eutrófico.

A construção de um modelo com menor número de variáveis torna a interpretação mais simples ou realista e, portanto, contribui para a compreensão mais objetiva e precisa sobre os padrões determinantes (no tempo e espaço) para a emergência da doença. A partir da construção desse novo modelo, e árvore decisão correspondente, foi possível recuperar os valores reais de separação somente desses oito atributos. A recuperação dos valores de separação dos atributos não contemplados seria possível com a multiplicação dos valores de área (de cada um dos municípios ajustados a aquela regra) para cada um dos valores de separação de cada um dos dezesseis atributos. Porém, a recuperação exata sobre quais foram os municípios adequados a cada uma das folhas da árvore de decisão não foi possível. Sendo assim, não foi possível obter os valores originais de separação destes atributos evidenciados na árvore de decisão com valores normalizados (prec10_sum; prec12_sum; Latossolo bruno Distrófico (sum); prec7_sum; bio4_sum; bio3_sd; prec10_sd; Massa D'água Continental (sum)). Tendo em vista que a obtenção dos valores reais de separação trata-se de informação básica para orientar com maior precisão a discussão e o aprofundamento do conhecimento

sobre os eventos epidemiológicos da doença, considera-se que esta análise seja satisfatória com os valores reais dos oito atributos contemplados nesta nova árvore de decisão.

Interessante destacar que ainda foram comparadas (a título de complementação aos resultados) as métricas de acurácia entre os modelos 1.4 com dados normalizados, 1.4 com 16 atributos (não normalizados) selecionados e 1.4 com todos os dados não normalizados (tabela 4.10). Na comparação, foram superiores as métricas de acurácia do modelo 1.4 que utiliza dados normalizados. Isto sugere que os atributos normalizados contribuíram para construção de modelos mais acurados. O que reforçou a importância da utilização de variáveis normalizadas. Apesar disso, a variação verificada entre as acurácias dos modelos 1.4 (valores normalizados) e 1.4 (16 atributos selecionados com valores não normalizados) torna válida o procedimento adotado, considerando o ganho de interpretabilidade sobre os valores reais de separação.

Tabela 4.10: Métricas de acurácia com modelos 1.4 com dados normalizados e não normalizados

Métricas de acurácia	Modelo 1.4 (valores normalizados)	Modelo 1.4 com 16 atributos selecionados (valores não normalizados)	Modelo 1.4 (valores não normalizados)
Instâncias classificadas corretamente	85,09%	84,21%	83,33%
Instâncias classificadas incorretamente	14,91%	15,79%	16,67%
Estatística Kappa	0,7018	0,6842	0,6667
Erro absoluto médio	0,1933	0,2373	0,1877
Raiz do erro quadrático médio	0,3504	0,3545	0,3604
Erro absoluto relativo	38,65%	47,47%	37,53%
Raiz do erro quadrático relativo	70,09%	70,89%	72,08%
Taxa TP	0,851	0,842	0,833
Taxa FP	0,149	0,158	0,167
Precision	0,854	0,842	0,834
Recall	0,851	0,842	0,833
F-Measure	0,851	0,842	0,833
Curva ROC	0,907	0,889	0,907

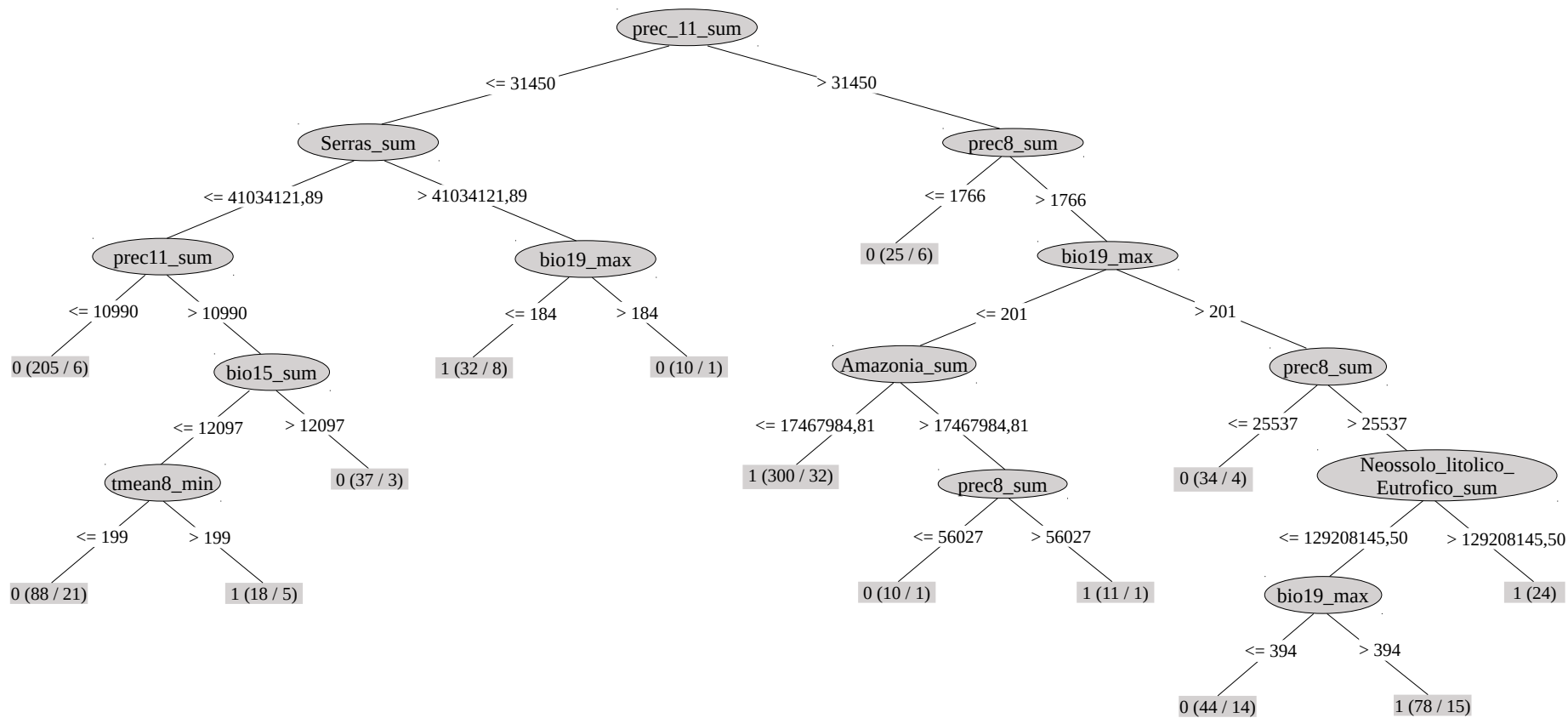


Figura 4.52: Árvore de decisão (método J48) com atributos não normalizados para o recorte 1999 a 2017

A partir da observação da árvore verifica-se que o primeiro nó utiliza o atributo `prec11_sum`, sendo 31.450mm o valor de separação. Deve-se atentar que esse valor extremamente elevado para precipitação não existe na realidade. Este atributo trata-se do somatório dos milímetros de chuva em novembro em cada 1km² pertencente a cada um dos municípios. A partir desse valor abstrato, o método identificou um índice de separação, que não corresponde a um parâmetro de separação “computacional” entre municípios positivos ou negativos, mas realiza a separação entre as duas principais estruturas ou ramificações da árvore.

A partir dessa primeira ramificação foi verificado sobre o somatório dos milímetros de chuva. Se for menor ou igual a 31.450mm, então foi verificada a condição sobre o atributo `Serras_sum`, que corresponde a soma dos valores das áreas de Serras em cada município. O valor de separação (~41km²) também não separou municípios positivos e negativos, mas é parâmetro de separação entre dois outros atributos do terceiro nível de análise da árvore.

No segundo nível de ramificação foi verificado o valor das áreas de Serras. Se for maior a 41km², então foi verificada a condição sobre o atributo `bio19_max` (Valor máximo de Precipitação do trimestre mais frio), que adotou como parâmetro de separação o valor 184mm, sendo que se este for menor ou igual então correspondem a 32 municípios positivos. Senão, verificaram-se 10 municípios negativos. Este atributo retoma a discussão do subcapítulo 4.1.1.6 sobre a variável Bio19, na qual foi verificado que os maiores volumes de precipitação no trimestre mais frio concentram-se na faixa centro-norte do bioma Amazônia (1106 a 835mm). Sobre este atributo e o valor de separação determinado na árvore, compreende-se que o método identificou que entre os positivos para febre amarela 32 municípios atendiam à condição de valor máximo de precipitação menor do que 184mm, enquanto que 8 não atendiam a esta condição. Lembrando que a adequação a esta condição depende da adequação às condições (nós) dos níveis anteriores. Dessa maneira, entende-se que nesse amplo recorte (1999 a 2017) o método reconheceu que o máximo de precipitação no inverno (Junho, Julho e Agosto) foi menor que 184mm para parte dos municípios positivos. Sabendo que de fato, nas regiões Centro-Oeste e Sudeste os volumes de precipitação no inverno são menores (variam entre 24 e 214mm), considera-se pertinente essa condição, já que nessas regiões diversos municípios manifestaram a ocorrência da doença

entre 1999 a 2017. Os 8 municípios positivos que não atendiam a esta condição certamente estão localizados na região norte, já que trata-se da região endêmica mas que concentram, no contexto nacional, o maior volume de chuva no inverno.

Se o valor de `Serras_sum` for menor ou igual a 41 km², então foi verificado atributo `prec11_sum`. Para este foi utilizado como valor de corte 10.990mm (somatório) - se for menor ou igual, então categorizaram-se 205 municípios negativos, senão, segue a análise para o quarto nível. No quarto nível de análise foi verificado o atributo `bio15_sum`, que corresponde ao somatório dos valores da Sazonalidade de Precipitação (coeficiente de variação). O valor de separação deste nó (12.097) também é abstrato, trata-se do somatório dos coeficientes de variação de cada 1 km² pertencente a cada um dos municípios. Se o valor for maior então verificaram-se 37 municípios negativos para febre amarela. Se for menor ou igual, a árvore segue para o quinto nível. No quinto nível foi verificado o atributo `tmean8_min` (Valor mínimo de Temperatura média mensal do mês de Agosto) - se o valor for menor ou igual do que 19,9°C então foram encontrados 88 municípios negativos; se não, foram encontrados 18 municípios positivos para febre amarela.

Retornando à primeira ramificação da árvore de decisão, se o somatório dos milímetros de chuva for maior que 31.450mm, então verificou-se o atributo `Prec8_sum` (Somatório dos valores de Precipitação total mensal do mês de Agosto) com o valor de separação de 1.766mm (soma de valores reais de precipitação por cada 1km² dos municípios); se o valor for menor ou igual, então verificaram-se 25 municípios negativos. Senão, verificou-se o atributo `Bio19_max`. Para esse nó foi utilizado o valor 201mm como parâmetro de separação entre dois novos atributos, pertencentes ao quarto nível de análise. Se o valor de `Bio19_max` for menor ou igual a 201mm, então deve-se verificar o atributo `Amazônia_sum`. O atributo `Amazônia_sum` utiliza como valor de separação 17.467984,81m². Se o município analisado tiver área correspondente ao bioma Amazônia menor ou igual a 17,5 mil km², então correspondem a municípios positivos, nessa regra adequaram-se 300 municípios. Se o valor for superior, então deve-se observar o atributo `prec8_sum` - se o valor de precipitação for menor ou igual a 56.027mm, então os municípios foram negativos, nessa regra adequam-se 10 municípios. Senão, os municípios foram positivos, adequaram-se a essa regra 11 municípios.

Retornando ao terceiro nível de ramificação da árvore de decisão, tem-se que, se o valor do atributo bio19_max for superior a 201 mm, então, foi analisado o atributo prec8_sum. Se o valor for menor ou igual a 25.537 mm, então foram encontrados os municípios negativos para febre amarela (34 municípios adequaram-se a esta regra). Se o valor for maior, então deve-se analisar o tipo de solo Neossolo litólico eutrófico. Se a área de Neossolo for maior a 129.208.145,5 m², então encontraram-se 24 municípios positivos. Se o valor for menor ou igual, então deve-se observar o atributo bio19_max – para essa variável o valor de separação utilizado foi 394mm. Se o valor for menor ou igual, então verificaram-se 44 municípios negativos; senão, verificaram-se 78 municípios positivos.

A amplitude e diversidade de dados evidenciados neste recorte é reflexo da decisão de construção do modelo com todo o histórico de surtos, tornando redundante a tarefa de aprofundar a discussão já realizada nos subcapítulos 4.1.1.1 a 4.1.1.10 e indicação de relações mais específicas desses atributos com a doença do que já foi discutido anteriormente (principalmente por não ser possível a separação sobre o surto ao qual determinada variável foi relacionada). Importante destacar que na maioria das folhas que demonstram quantos municípios adequaram-se a esta regra também foram evidenciados os municípios “exceções”. Assim, foi possível verificar quais atributos e valores foram capazes de ser mais ou menos abrangentes sobre os municípios. De modo geral, os atributos que contém maior nível de profundidade (nível 6) buscaram essas condições de excepcionalidade.

Os municípios com favorabilidade de emergência de febre amarela, indicados pelo modelo do recorte 1999 a 2017 (figura 4.53), evidenciaram a vulnerabilidade para 2.629 municípios, sendo que 220 desses foram verdadeiros positivos, ou seja, tiveram ocorrência real da doença em 2018. O modelo teve 36 falso negativos, ou seja, municípios que tiveram ocorrência real para a doença e o modelo não foi capaz de prever. O modelo indicou que os estados do Amazonas, Acre, Roraima, Rondônia, Mato Grosso, Mato Grosso do Sul, Goiás, Tocantins, Minas Gerais, Espírito Santo, Rio de Janeiro e Rio Grande do Sul possuem quase todos os municípios favoráveis à emergência da febre amarela. Os estados do Pará (porções centro, sul e oeste), Bahia (porções nordeste e oeste), São Paulo (porção centro-norte), Paraná (faixa central), Santa Catarina (porção sul, leste e oeste) tiveram grande ou parcial parte dos seus territórios indicados como áreas favoráveis.

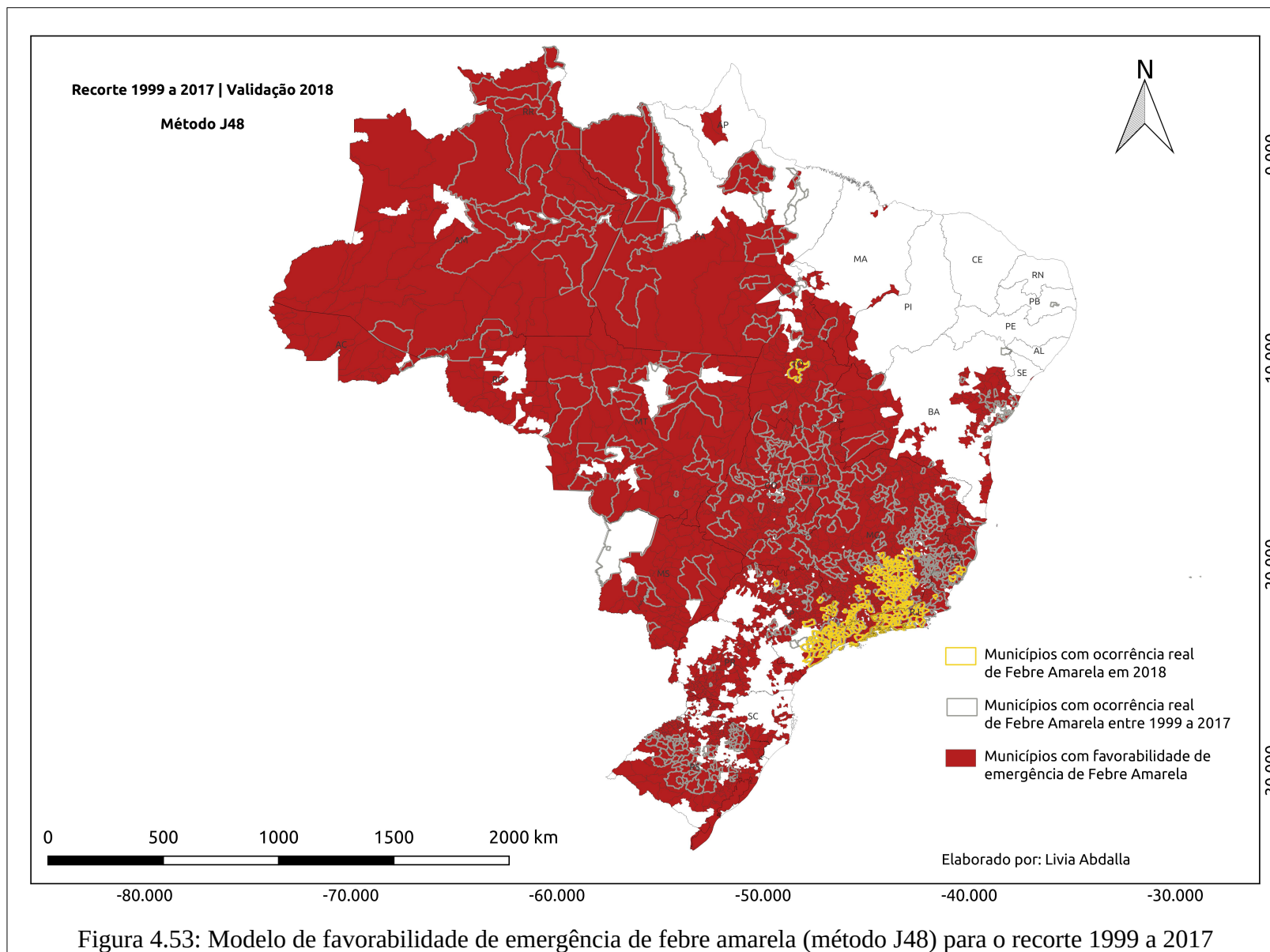


Figura 4.53: Modelo de favorabilidade de emergência de febre amarela (método J48) para o recorte 1999 a 2017

4.3.1.2 Recorte 02 (1999 a 2003)

O modelo construído com o primeiro surto (1999 a 2003) indicou somente duas variáveis: Área total de distribuição do *Callithrix jacchus* e Desvio padrão do valor de Precipitação total acumulada no mês de Agosto (prec 8). Apesar de a base de dados utilizada contemplar dados das diversas temporalidades, os dados selecionados nessa árvore de decisão são do tipo Atemporais (prec8) e Sem regularidade temporal (*Callithrix jacchus*). A árvore de decisão produzida contém 2 nós e 2 níveis (figura 4.54), que demonstram quais atributos foram mais representativos e quais são os respectivos valores normalizados e estatísticas que melhor separam os municípios em positivos e negativos.

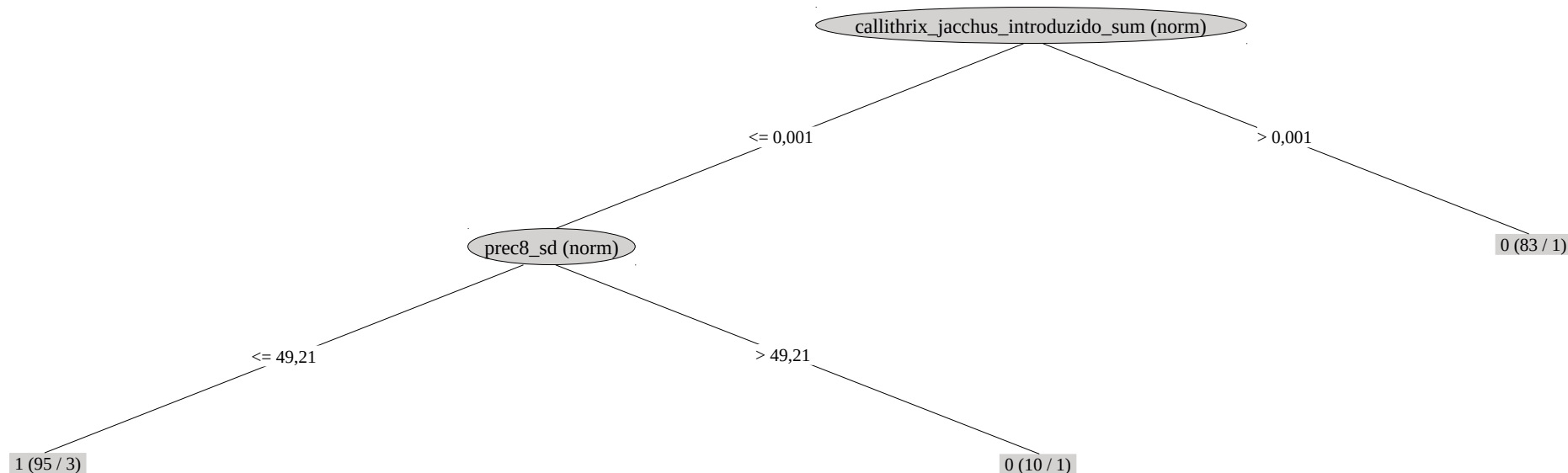


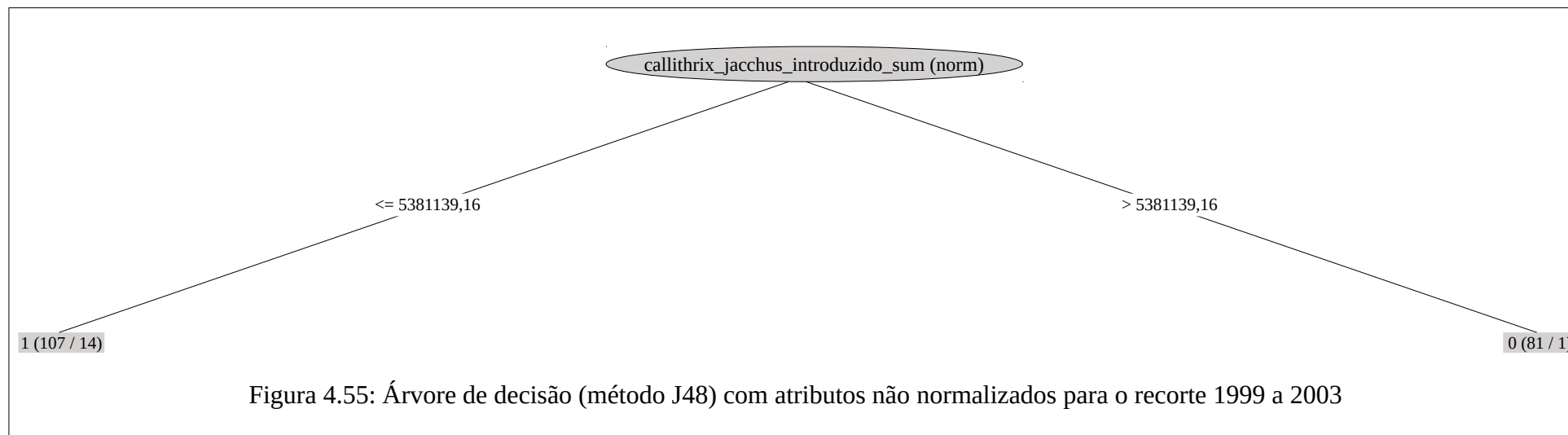
Figura 4.54: Árvore de decisão (método J48) para o recorte 1999 a 2003

Para a compreensão sobre os valores reais de separação dos atributos foram selecionados os mesmos atributos (dado/variável e estatística) somente com valores não normalizados, tanto para a base de treino e teste. A partir dessa base foi aplicado o método J48 (com os mesmos parâmetros de configuração descritos no subcapítulo 3.6), obtendo-se uma árvore de decisão contendo somente o atributo *Callithrix jacchus* (sum) (figura 4.55). Não foi possível recuperar o valor exato do atributo prec8 nesta árvore de decisão.

Destaca-se que, da mesma maneira para o recorte anterior, também foram comparadas as métricas de acurácia entre os modelos 2.0 com dados normalizados, 2.0 com 2 atributos (não normalizados) selecionados e 2.0 com todos os dados não normalizados (tabela 4.11). Na comparação foram superiores as métricas de acurácia do modelo 2.0 que utiliza dados normalizados, indicando que os atributos normalizados contribuíram para construção de modelos mais acurados.

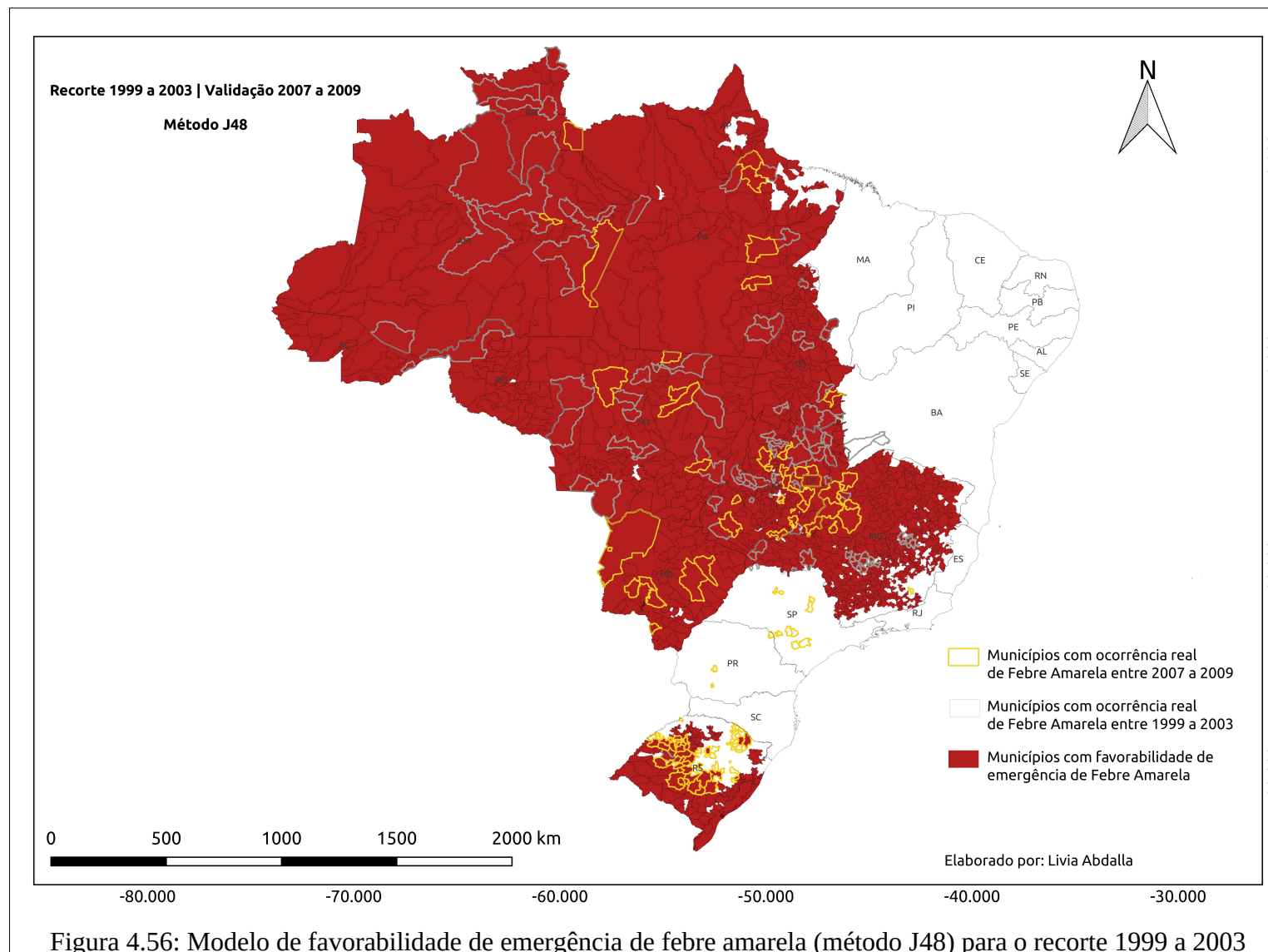
Tabela 4.11: Métricas de acurácia com modelos 2.0 com dados normalizados e não normalizados

Métricas de acurácia	Modelo 2.0 (valores normalizados)	Modelo 2.0 com 2 atributos selecionados (valores não normalizados)	Modelo 2.0 (valores não normalizados)
Instâncias classificadas corretamente	93,48%	86,95%	89,13%
Instâncias classificadas incorretamente	6,52%	13,04%	10,86%
Estatística Kappa	0,8696	0,73	0,78
Erro absoluto médio	0,088	0,18	0,13
Raiz do erro quadrático médio	0,2541	0,34	0,31
Erro absoluto relativo	17,60%	36,67%	26,76%
Raiz do erro quadrático relativo	50,83%	69,39%	62,03%
Taxa TP	0,935	0,87	0,89
Taxa FP	0,065	0,13	0,1
Precision	0,942	0,87	0,91
Recall	0,935	0,87	0,89
F-Measure	0,935	0,87	0,89
Curva ROC	0,929	0,87	0,904



O atributo *Callithrix jacchus* foi evidenciado com valor de separação de 5.381.139,16 m². Se a área de distribuição da espécie for menor ou igual a 5,38 km², então 107 municípios foram positivos para febre amarela; senão, 81 municípios foram negativos. Considerando as regiões de distribuição dessa espécie e os locais de ocorrência da doença, compreende-se que nesse primeiro recorte a ausência (ou presença limitada) da espécie *Callithrix jacchus* em 107 municípios teve maior relação com a ocorrência da doença, o que parece sugerir que outras espécies de PNHs, sem sobreposição de áreas com esta espécie, estiveram mais envolvidas com este surto. Os 81 municípios com maior distribuição dessa espécie não tiveram manifestação da doença. Destaca-se que este primeiro surto ocorreu em um padrão espacial amplo no que se refere aos Estados (12 UFs localizadas em todas as cinco Grandes Regiões – Norte, Centro-Oeste, Nordeste, Sudeste e Sul) e pontual no que se refere aos municípios. Isso pode justificar, portanto, a dificuldade de identificação de uma variável diretamente relacionada ao padrão favorável de manifestação da doença.

O modelo utilizando o recorte 1999 a 2003 indicou favorabilidade para emergência da febre amarela em todos os municípios ou na maior parte dos estados do Amazonas, Roraima, Amapá, Rondônia, Mato Grosso, Mato Grosso do Sul, Goiás, Tocantins, Minas Gerais e Rio Grande Sul (porção sul) (figura 4.56). No total foram indicados 1.451 municípios favoráveis, sendo que 79 destes foram verdadeiros positivos entre 2007 e 2009, enquanto que 54 foram falso negativos, ou seja, foram positivos no período de validação mas o modelo não foi capaz de prever.



4.3.1.3 Recorte 03 (2007 a 2009)

O modelo construído com o segundo surto (2007 a 2009) indicou quatro variáveis: Somatório dos valores de Sazonalidade da Temperatura (desvio padrão * 100) (bio4); Somatório das áreas de solos do tipo Latossolo vermelho Distrófico; Somatório dos valores de Amplitude térmica anual (bio5-bio6) (bio7); Somatório das áreas do bioma Mata Atlântica. Todos os dados evidenciados são do tipo Atemporais, ou seja, variáveis ou dados que não possuem modificações expressivas para a escala de tempo analisada. A árvore de decisão produzida contém 4 nós e 3 níveis (figura 4.57), que demonstram quais atributos foram mais representativos e quais são os respectivos valores normalizados e estatísticas que melhor separam os municípios em positivos e negativos.

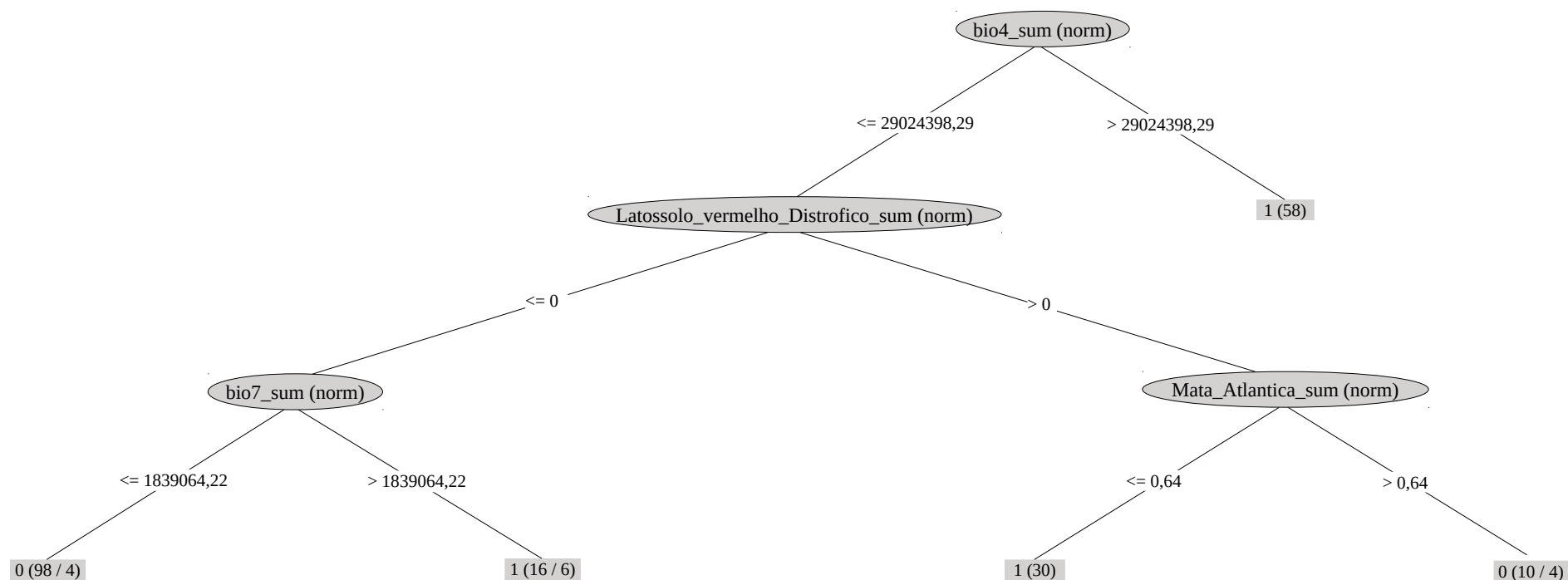


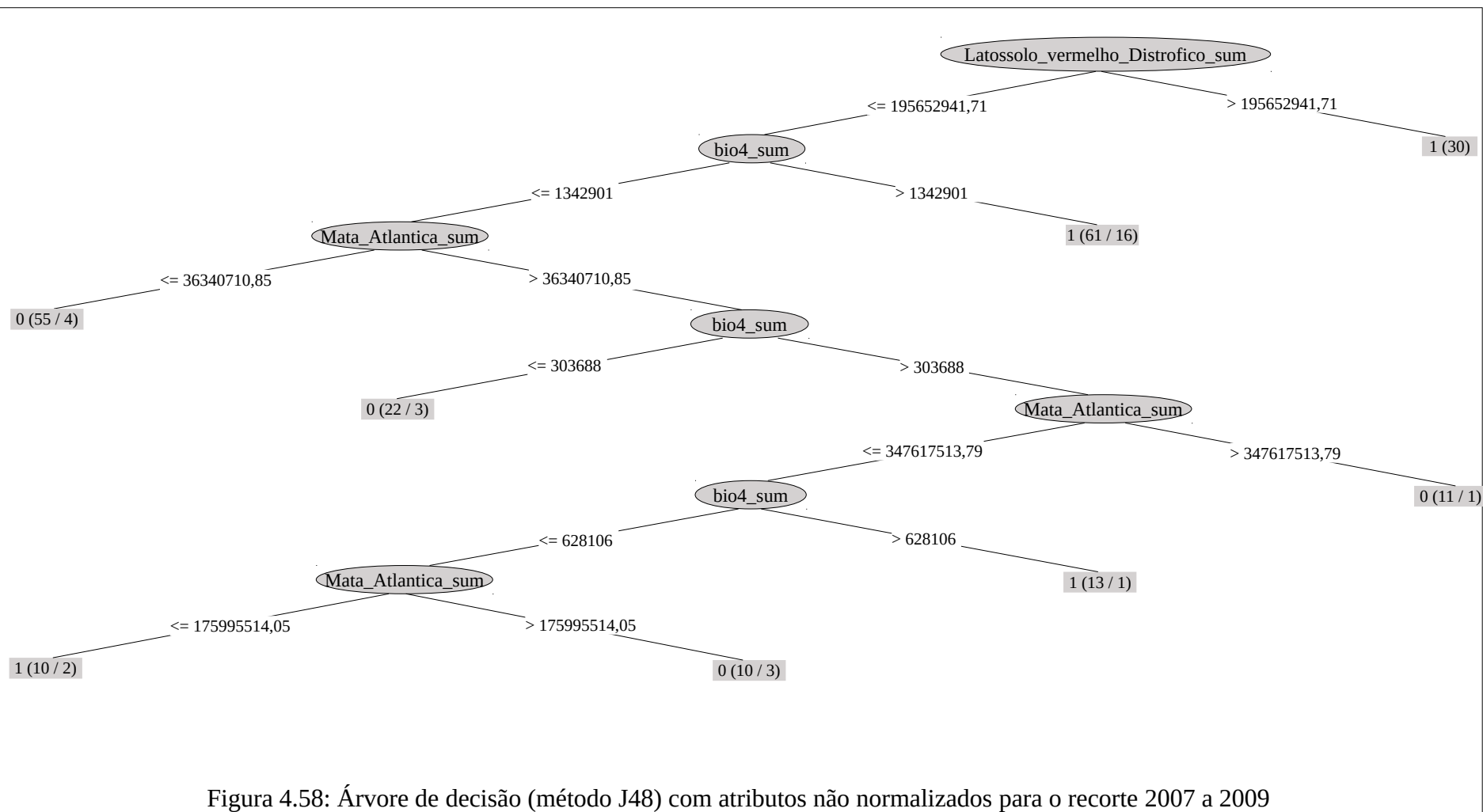
Figura 4.57: Árvore de decisão (método J48) para o recorte 2007 a 2009

Seguindo a mesma lógica de interpretação dos recortes anteriores, foram selecionados os atributos (com valores não normalizados) indicados na figura 4.57 e aplicado sobre esses o método J48 (com parâmetros de configuração idênticos). A árvore de decisão utilizou três dos quatro atributos selecionados (figura 4.58), de modo que não foi possível recuperar o valor exato do atributo Somatório dos valores de Amplitude térmica anual (bio7_sum).

Da mesma maneira que nos dois recortes anteriores, também foram comparadas as métricas de acurácia entre os modelos 3.4 com dados normalizados, 3.4. com 4 atributos selecionados (valores não normalizados) e 3.4 com todos os dados não normalizados (tabela 4.12). Na comparação, as métricas de acurácia foram iguais em sua maioria entre o modelo 3.4 que utiliza dados normalizados e do modelo 3.4 que utiliza dados não normalizados. O fato sugere que, para este recorte, não existiu prejuízo ou ganho na decisão de utilizar dados normalizados ou não normalizados. Portanto, foi utilizada mesma lógica de interpretação sobre valores dos atributos selecionados e não normalizados.

Tabela 4.12: Métricas de acurácia com modelos 3.4 com dados normalizados e não normalizados

Métricas de acurácia	Modelo 3.4 (valores normalizados)	Modelo 3.4 com 4 atributos selecionados (valores não normalizados)	Modelo 3.4 (valores não normalizados)
Instâncias classificadas corretamente	87,04%	81,48%	87,04%
Instâncias classificadas incorretamente	12,96%	18,52%	12,96%
Estatística Kappa	0,7407	0,6296	0,7407
Erro absoluto médio	0,1786	0,2441	0,1438
Raiz do erro quadrático médio	0,3582	0,377	0,3277
Erro absoluto relativo	35,72%	48,81%	28,76%
Raiz do erro quadrático relativo	71,64%	75,40%	65,55%
Taxa TP	0,87	0,815	0,87
Taxa FP	0,13	0,185	0,13
Precision	0,875	0,817	0,871
Recall	0,87	0,815	0,87
F-Measure	0,87	0,815	0,87
Curva ROC	0,871	0,899	0,914



A árvore produzida com atributos selecionados (valores não normalizados) teve como primeiro nó o atributo somatório das áreas de Latossolo vermelho distrófico, sendo o valor de separação igual a 195,65km². Se as áreas correspondentes a esse tipo de solo for maior que este valor, então tiveram 30 municípios positivos para febre amarela. Os municípios pertencentes a esta regra estão localizados nos estados das regiões Centro-Oeste, Sudeste e Sul (TO, GO, DF, MS, MG, SP e PR), regiões estas com intensa transmissão da doença durante os dois primeiros anos (2007-2008) do surto em questão, concentrando 596 epizootias. Se o total de áreas de Latossolo vermelho distrófico for menor ou igual, então foi analisado o atributo Sazonalidade da Temperatura (desvio padrão * 100) (bio4_sum) – se o somatório dos valores do coeficiente de variação for maior que 1.342.901, então 61 municípios foram positivos para febre amarela. Conforme descrito no subcapítulo 4.1.1.6 destaca-se que o atributo Sazonalidade da Temperatura possui os maiores valores de variação no estado do Rio Grande do Sul, justamente o estado com maior concentração de epizootias positivas (950), distribuídas em 103 municípios, para febre amarela no período (principalmente no ano de 2009). Os gêneros mais envolvidos nas epizootias do RS foram o *Alouatta* (942), *Callithrix* (6) e *Cebus* (2).

Se a Sazonalidade da temperatura (bio4_sum) for menor ou igual, então deve ser considerado o atributo Mata Atlântica, sendo que se a área desse bioma for menor ou igual a 36,34 km² então 55 municípios foram negativos. Senão, deve ser novamente considerado o atributo de Sazonalidade da Temperatura – se o valor for menor ou igual a 303.688, então 22 municípios foram positivos. Se for maior, então deve ser considerado novamente a área total do bioma Mata Atlântica – se o valor for maior que 347,61 km², então foram identificados 11 municípios negativos; se for menor ou igual, então verificou-se novamente o atributo Sazonalidade – se o somatório dos valores for maior que 628.106, então foram identificados 13 municípios positivos; se for menor ou igual foi verificado o atributo Mata Atlântica – se o valor total das áreas do bioma Mata Atlântica for maior que 176 km², então foram identificados 10 municípios negativos; se for menor ou igual então foram detectados outros 10 municípios positivos para febre amarela. Destaca-se que, a partir do terceiro nó da árvore, o atributo Mata Atlântica complementa a busca pela separação entre municípios positivos e negativos baseada somente na Sazonalidade da temperatura, principalmente para identificar os

municípios positivos que pertencem aos estados de Minas Gerais e São Paulo, e que ainda não haviam adequado-se as regras dos atributos anteriores.

O modelo produzido para o recorte 2007 a 2009 (figura 4.59) indicou 2.122 municípios favoráveis à emergência da febre amarela, sendo que em 65 municípios ocorreram febre amarela no período de validação (2015 a 2018) (verdadeiros positivos) e em 363 municípios ocorreram febre amarela, mas o modelo não previu como favoráveis (falso negativo). Este é um indicativo sobre como o padrão de manifestação da doença no período de validação (2015 a 2018) foi diferente do surto anterior (2007 a 2009). No modelo de favorabilidade os estados com maior número de municípios favoráveis à doença foram: Mato Grosso, Mato Grosso do Sul, Goiás, Minas Gerais, São Paulo, Paraná, Santa Catarina, Rio Grande do Sul, Rio de Janeiro (faixas norte e oeste), Bahia (porção noroeste). Com exceção dos estados da região Sul (que tiveram manifestação da doença somente no período de 2007 a 2009), o modelo indicou corretamente a tendência de favorabilidade de manifestação da doença nos estados do sudeste (em que de fato houve municípios positivos entre 2015 e 2018). Interessante destacar que, apesar de não existirem municípios positivos (ou existirem poucos registros) nesses estados durante o histórico de ocorrência da doença, o modelo foi capaz de prever a tendência de emergência da febre amarela nessas regiões.

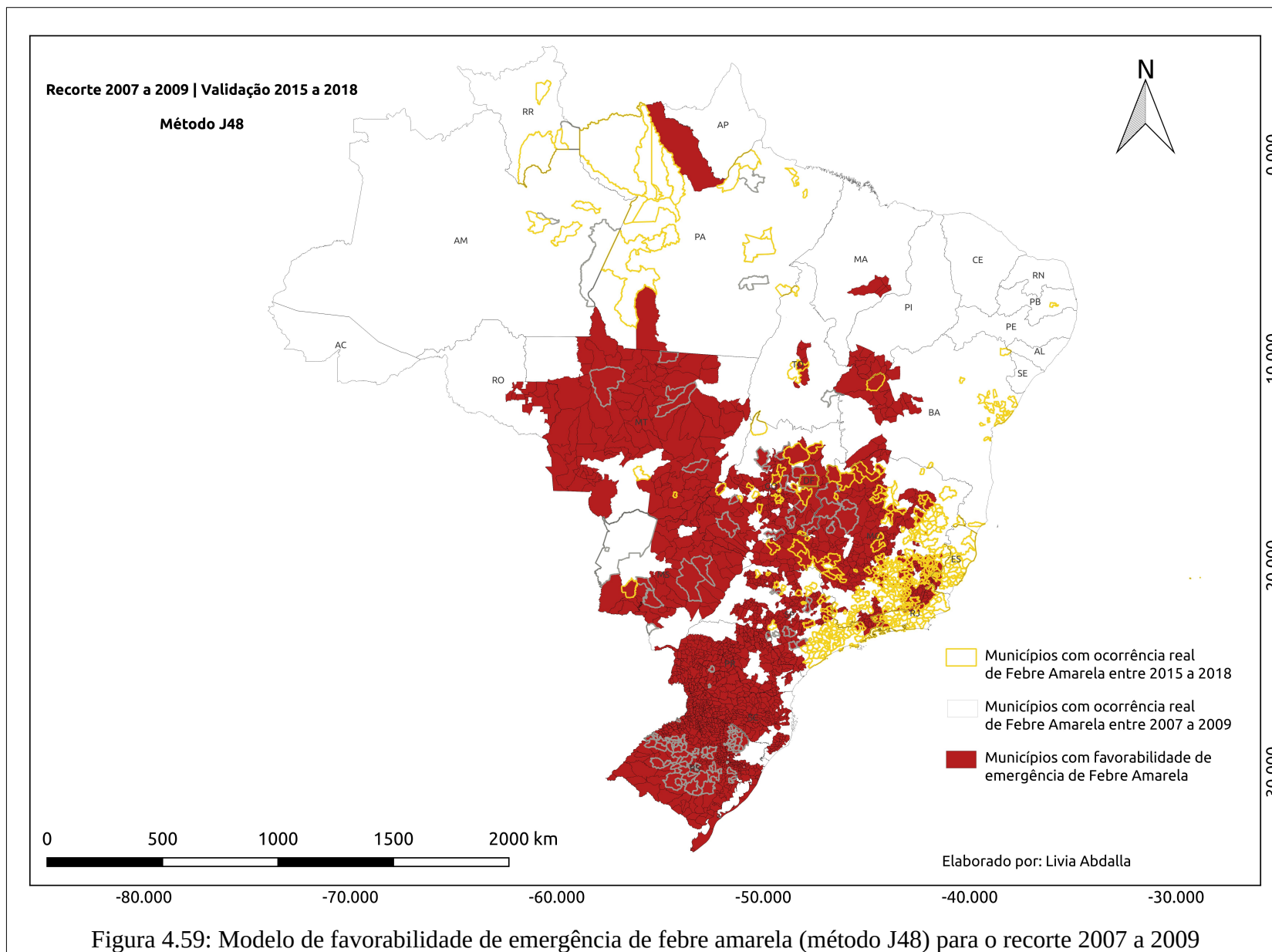


Figura 4.59: Modelo de favorabilidade de emergência de febre amarela (método J48) para o recorte 2007 a 2009

4.3.1.4 Recorte 04 (2015 a 2017)

O modelo construído com o terceiro surto (2015 a 2017) indicou quatro variáveis: Área total de distribuição de *Callithrix penicillata*; Área total de distribuição de *Alouatta macconnelli*; Desvio padrão das áreas de UCs; Área total de distribuição de *Callithrix jacchus*. Todos os dados evidenciados são do tipo Sem regularidade temporal, ou seja, variáveis ou dados que não possuem modificações ou atualizações na escala de tempo analisada. A árvore de decisão produzida contém 4 nós e 3 níveis (figura 5.60), que demonstram quais atributos foram mais representativos e quais são os respectivos valores normalizados e estatísticas que melhor separam os municípios em positivos e negativos.

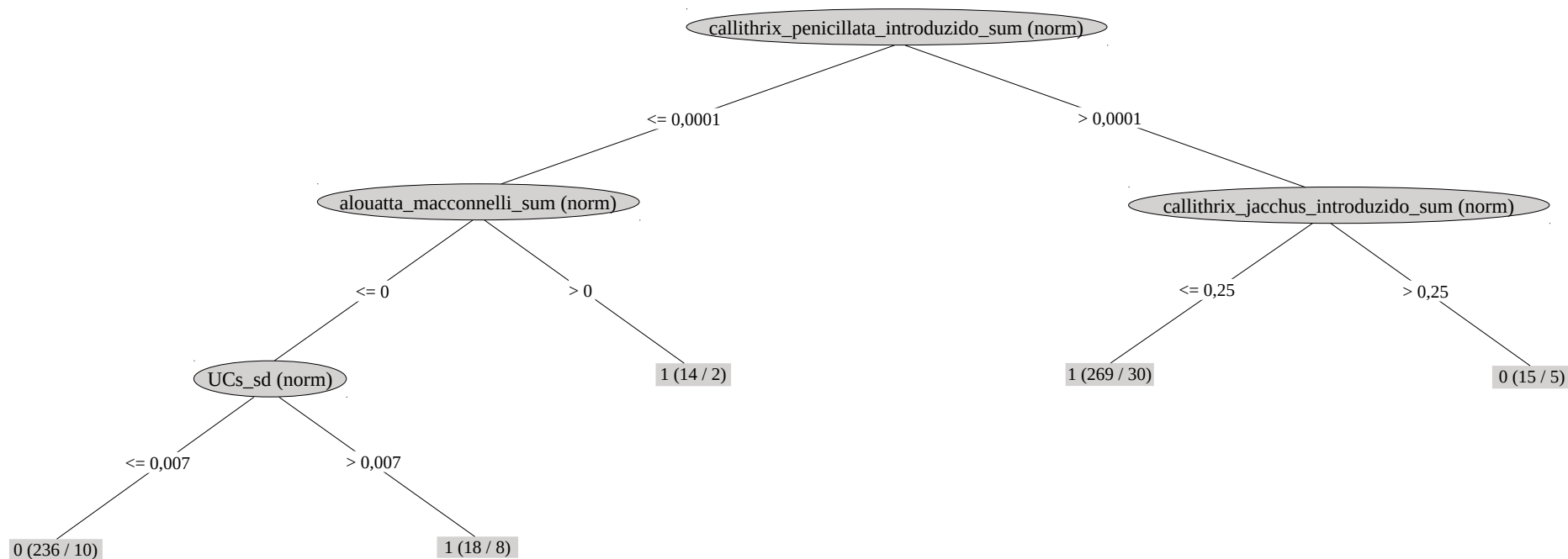


Figura 4.60: Árvore de decisão (método J48) para o recorte 2015 a 2017

Da mesma maneira que nos recortes anteriores, foram selecionados os quatro atributos com valores não normalizados para nova construção do modelo com o método J48. A nova árvore de decisão utilizou os atributos correspondentes às três espécies de primatas, não sendo possível recuperar os valores originais da variabilidade das áreas de Unidades de Conservação (figura 4.61). Igualmente aos três recortes anteriores foram comparadas as métricas de acurácia entre os modelos 4.5 com dados normalizados, 4.5 com 4 atributos selecionados (valores não normalizados) e 4.5 com todos os dados não normalizados (tabela 4.13). Na comparação, as métricas de acurácia foram iguais em sua maioria entre o modelo 4.5 que utiliza dados normalizados e do modelo 4.5 que utiliza dados não normalizados.

Tabela 4.13: Métricas de acurácia com modelos 4.5 com dados normalizados e não normalizados

Métricas de acurácia	Modelo 4.5 (valores normalizados)	Modelo 4.5 com 4 atributos selecionados (valores não normalizados)	Modelo 4.5 (valores não normalizados)
Instâncias classificadas corretamente	86,23%	85,51%	86,23%
Instâncias classificadas incorretamente	13,77%	14,49%	13,77%
Estatística Kappa	0,7246	0,7101	0,7246
Erro absoluto médio	0,19	0,206	0,1868
Raiz do erro quadrático médio	0,3318	0,342	0,342
Erro absoluto relativo	38,00%	41,20%	37,35%
Raiz do erro quadrático relativo	66,36%	68,39%	68,40%
Taxa TP	0,862	0,855	0,862
Taxa FP	0,138	0,145	0,138
Precision	0,862	0,856	0,866
Recall	0,862	0,855	0,862
F-Measure	0,862	0,855	0,862
Curva ROC	0,889	0,873	0,886

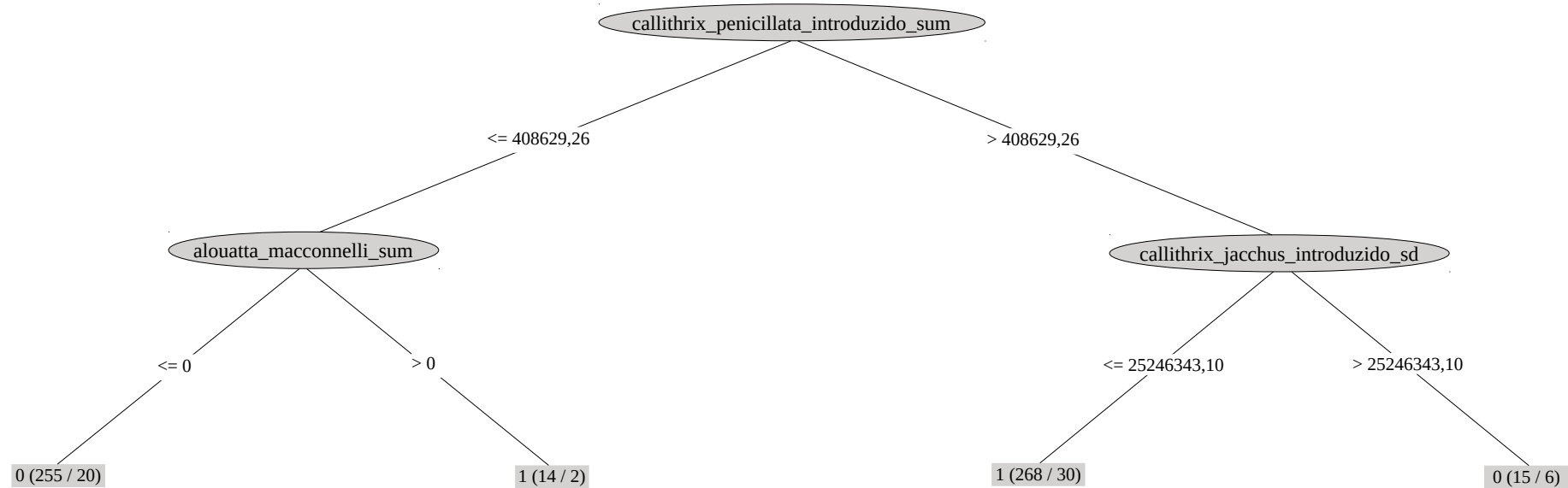


Figura 4.61: Árvore de decisão (método J48) com atributos não normalizados para o recorte 2015 a 2018

O primeiro atributo da árvore corresponde ao somatório das áreas de distribuição da espécie *Callithrix penicillata*. Se a área de ocorrência da espécie for menor ou igual a 0,4 km² (40,86 hectares), então foi verificada a área total de distribuição da espécie *Alouatta macconeli* - se a área for menor ou igual a 0, então 255 municípios foram negativos; se for maior que 0, então 14 municípios foram positivos. Retornando a primeira ramificação da árvore, se a área de *Callithrix penicillata* for maior que 40,86ha, então foi verificada a área total de *Callithrix jacchus* - se a área for menor ou igual a 25,25km², então 268 municípios foram positivos; caso contrário, então foram identificados 15 municípios negativos. Verifica-se que neste terceiro surto a influência dessas espécies de primatas foi evidente, o que parece ser muito coerente, pelo elevado número de epizootias (2.159) ocorrida no surto mais recente.

De modo geral, o terceiro e o recente surto (2015 a 2018) totalizaram 4.194 notificações em 528 municípios, distribuídos em 12 UFs: Amazonas; Pará; Roraima; Tocantins; Mato Grosso do Sul; Mato Grosso; Distrito Federal; Goiás; Bahia; Minas Gerais; Espírito Santo; Rio de Janeiro; São Paulo. Essas espécies de primatas têm ampla disseminação em dez dos treze Estados com casos confirmados de febre amarela, sendo o *Alouatta* extremamente vulnerável à doença e as espécies de *Callithrix* mais resistentes e potenciais amplificadores da doença. A espécie de *Callithrix jacchus* foi introduzida, e é abundante e amplamente distribuída, nos estados da Bahia, Espírito Santo, Rio de Janeiro e São Paulo. A espécie *Callithrix penicillata* é nativa nos estados do Mato Grosso do Sul, Bahia, Minas Gerais, São Paulo, Tocantins, e introduzida (pelo tráfico de animais silvestres) nos estados do Espírito Santo e Rio de Janeiro. Já a espécie *Alouatta macconnelli* é residente e nativa nos estados do Amazonas, Pará e Roraima.

Conforme descrito no subcapítulo 4.1.1.2, estudos indicam que ambas as espécies de *Callithrix* possuem grande capacidade de adaptação (capazes de viver em pequenas porções florestais como parques, jardins urbanos e aldeias rurais, sendo visto habitualmente nas praças e nos quintais das casas em busca de alimento) e de manutenção nas regiões em que foi introduzido, estabelecendo competição com outras espécies de primatas nativos dessas regiões. Ao que tudo indica, o aumento das populações dessas espécies, principalmente em áreas urbanas (em razão da disponibilidade de alimentos e escassez de predadores naturais), possui grande sintonia com a amplitude e velocidade de propagação do surto nos municípios

da região sudeste. Já na região norte, as espécies de *Alouatta macconnelli*, apesar de serem sensíveis e ameaçadas pela doença, demonstram atuar como mantenedoras do ciclo endêmico da doença.

O modelo produzido indicou favorabilidade para todos ou a maioria dos municípios dos estados de Roraima, Amapá, Tocantins, Goiás, Mato Grosso do Sul, São Paulo, Rio de Janeiro, Minas Gerais, Espírito Santo e Bahia (figura 4.62). Indicou favorabilidade também para alguns municípios do Amazonas, Pará, Maranhão (porções norte e sul), Paraná (porção centro-leste), Santa Catarina (porções leste e oeste) e Rio Grande do Sul (porções nordeste e sudoeste). No total, 2.658 municípios foram indicados como favoráveis à emergência da febre amarela, sendo que 237 municípios de fato tiveram ocorrência de febre amarela no período de validação (2018), enquanto que 19 municípios não foram identificados como favoráveis pelo modelo, mas foram positivos em 2018.

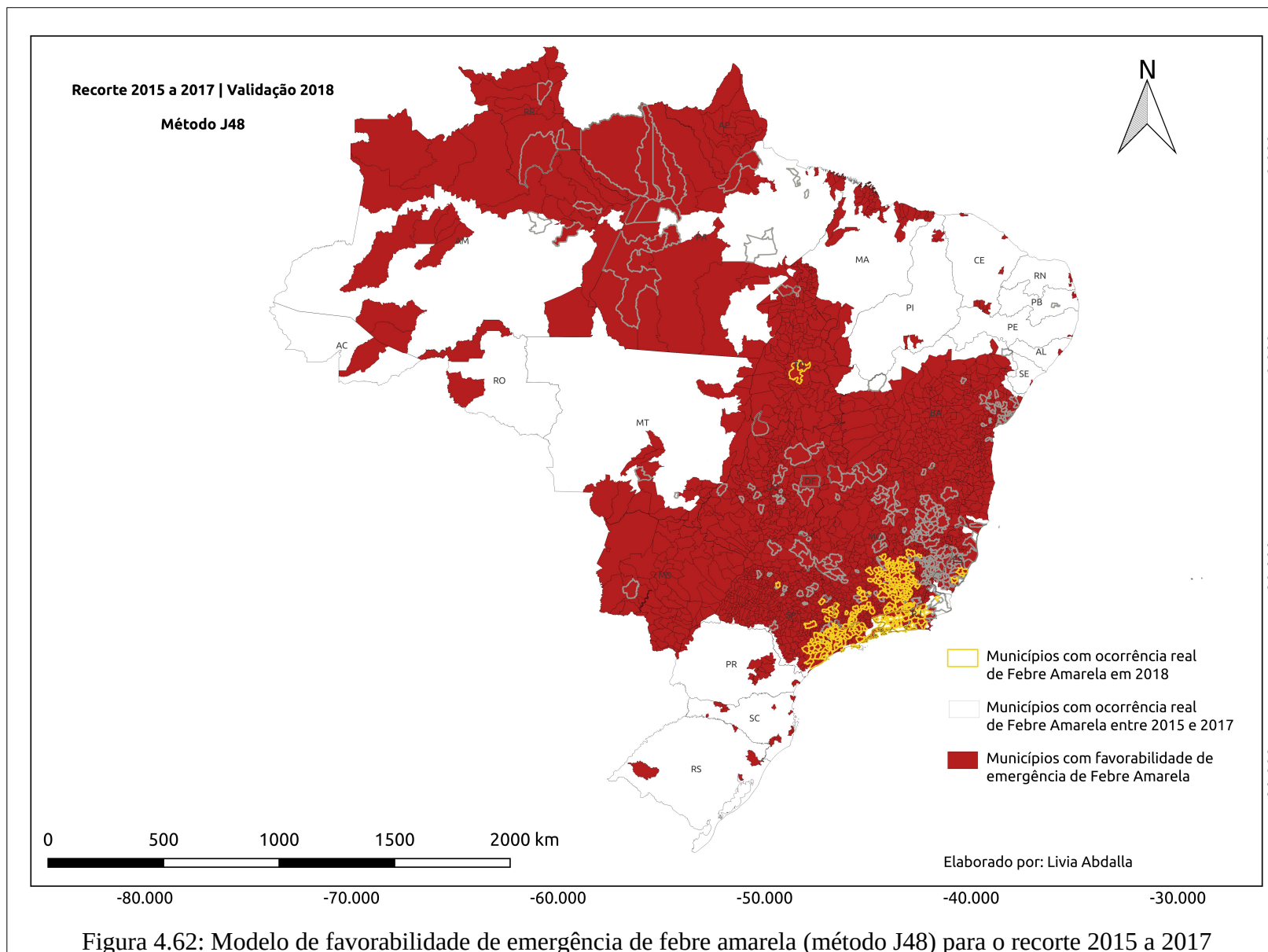


Figura 4.62: Modelo de favorabilidade de emergência de febre amarela (método J48) para o recorte 2015 a 2017

4.3.2 Modelos finais – Método *Random Forest*

A partir da avaliação dos modelos experimentais foram identificados os modelos para cada um dos recortes temporais dos surtos, com os maiores valores de acurácia interna e acurácia preditiva, ou seja, os modelos com maior coerência interna e com maior capacidade de generalização. Dentre os modelos construídos com o método *Random Forest*, destacam-se: 1.9) Atributos selecionados pelos modelos construídos com sub-recorte dos dados temáticos; 2.8) Seleção de atributos pelo método CFS; 3.7) Todos os atributos; 4.8) Seleção de atributos pelo método CFS. As métricas de validação interna e preditiva já foram descritas no subcapítulo 4.1 e definidos os modelos finais. Portanto, a abordagem principal neste momento é apresentar os possíveis cenários de manifestação da doença.

4.3.2.1 Recorte 01 (1999 a 2017)

Para o recorte 1999 a 2017, utilizando o método *Random Forest*, o modelo 1.9 foi identificado como final. O modelo preditivo (figura 4.63) indicou diferentes graus de favorabilidade de emergência de febre amarela para 2.895 municípios, sendo que 2.646 são falsos positivos e 249 verdadeiros positivos. Para cada município a favorabilidade de emergência da doença pode ser maior ou menor, sendo os valores definidos dentro do intervalo que varia entre 0 (não favorável) e 1 (máxima favorabilidade). O modelo preditivo indicou favorabilidade para todos ou parte dos municípios dos Estados do Acre, Rondônia, Amazonas, Roraima, Amapá, Pará, Tocantins, Goiás, Mato Grosso, Mato Grosso do Sul, Bahia (zona da mata), Minas Gerais, Espírito Santo, Rio de Janeiro, São Paulo, Paraná (porção centro-sul), Santa Catarina e Rio Grande do Sul. Destaca-se nesse contexto que os Estados do Espírito Santo, Minas Gerais (porção nordeste), São Paulo (litoral norte) e Rio Grande do Sul (porção sudoeste) foram os que apresentaram municípios com maior favorabilidade de emergência da febre amarela.

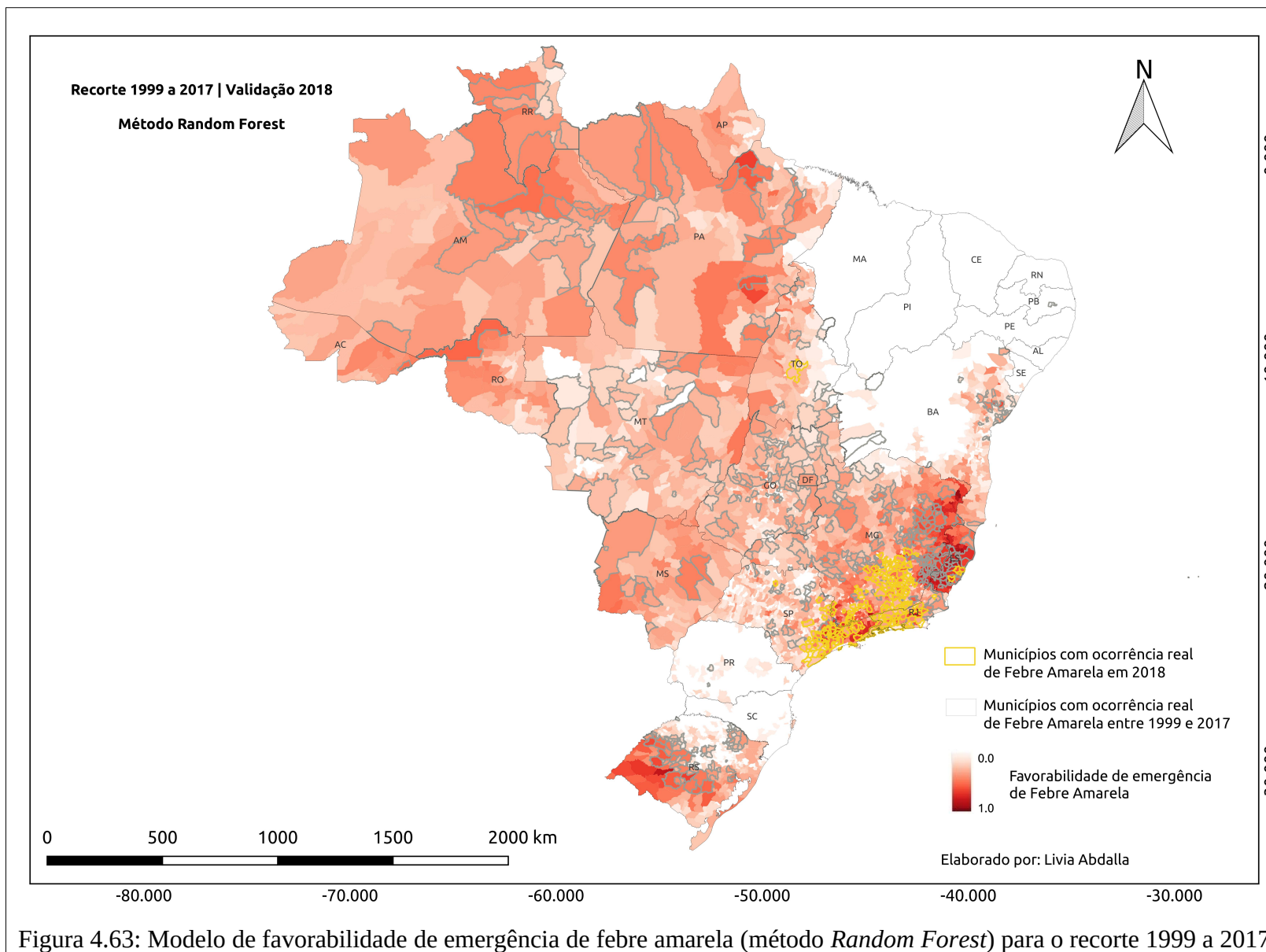


Figura 4.63: Modelo de favorabilidade de emergência de febre amarela (método *Random Forest*) para o recorte 1999 a 2017

4.3.2.2 Recorte 02 (1999 a 2003)

Para o recorte 1999 a 2003 o modelo 2.8 foi selecionado como final. O método *Random Forest* indicou favorabilidade para 2.801 municípios (figura 4.64), sendo que 2.680 falsos positivos e 121 verdadeiros positivos. Para este recorte os Estados com maior número de municípios e com maior grau de favorabilidade foram Acre, Rondônia, Amazonas, Pará, Roraima, Amapá, Tocantins, Goiás, Mato Grosso, Mato Grosso do Sul, Minas Gerais e Rio Grande do Sul. Destaca-se que este modelo preditivo obteve grande assertividade, principalmente, no Estado do Rio Grande do Sul, o qual reuniu o maior número de epizootias no surto ocorrido entre 2007 e 2009.

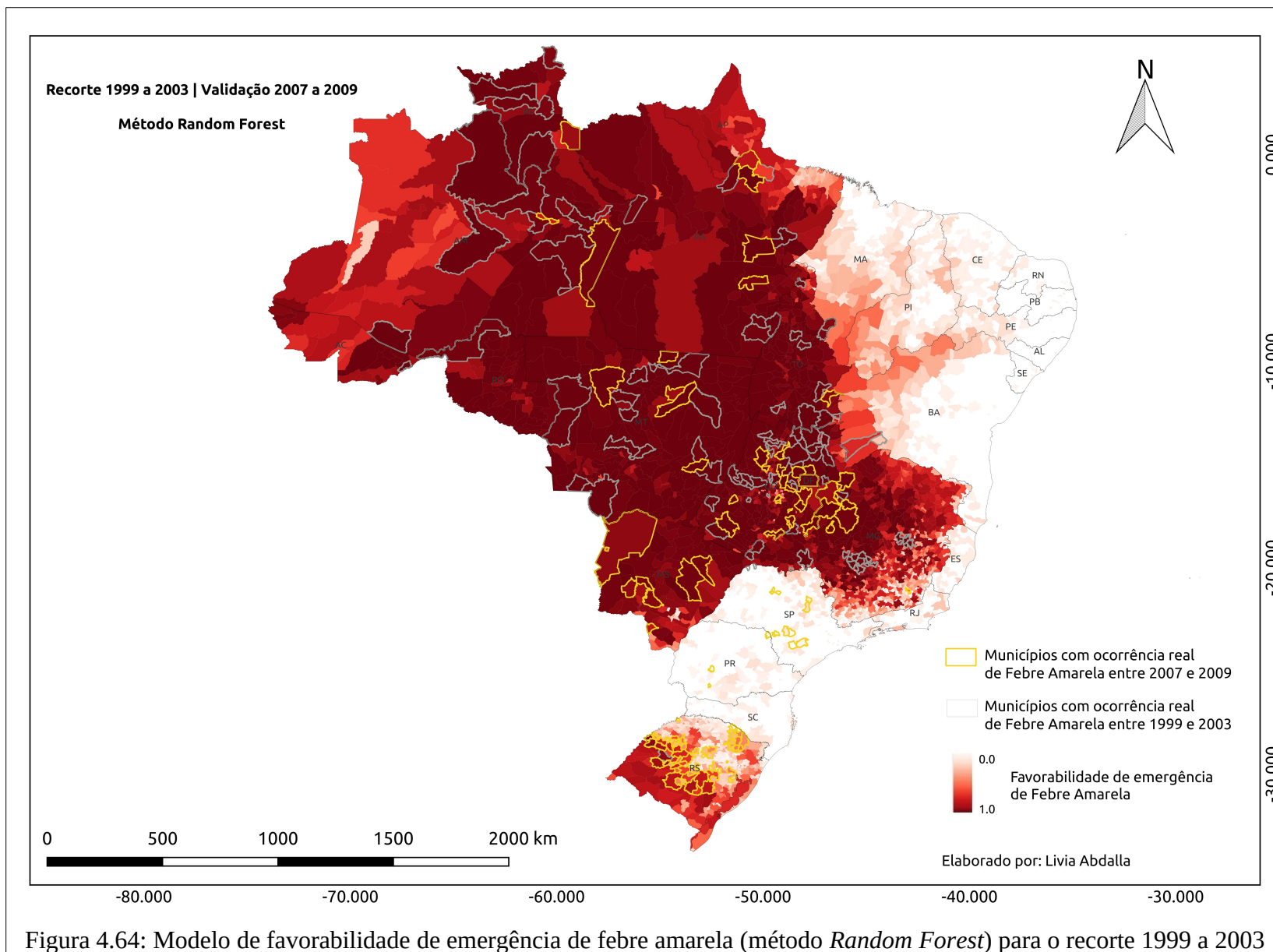


Figura 4.64: Modelo de favorabilidade de emergência de febre amarela (método *Random Forest*) para o recorte 1999 a 2003

4.3.2.3 Recorte 03 (2007 a 2009)

Para o recorte 2007 a 2009 foi selecionado como final o modelo 3.7. O método *Random Forest* indicou favorabilidade de emergência de febre amarela para 2.794 municípios (figura 4.65), sendo 2.623 falsos negativos e 171 verdadeiros positivos. Os estados com maior número de municípios favoráveis foram Acre, Amazonas, Roraima, Amapá, Para, Maranhão (porção centro-sul), Piauí (porção sul), Bahia (porção oeste), Tocantins, Goiás, Mato Grosso, Mato Grosso do Sul, Minas Gerais, São Paulo, Paraná, Santa Catarina e Rio Grande do Sul. Destaca-se que, de modo geral, as regiões Centro-Oeste, Sudeste (MG e SP) e Sul do país foram indicadas com maior número de municípios e com maior grau de favorabilidade, o que concorda, em parte (MS, GO, MG, SP), com as regiões com maior número de ocorrências no surto seguinte, ocorrido entre 2015 e 2018.

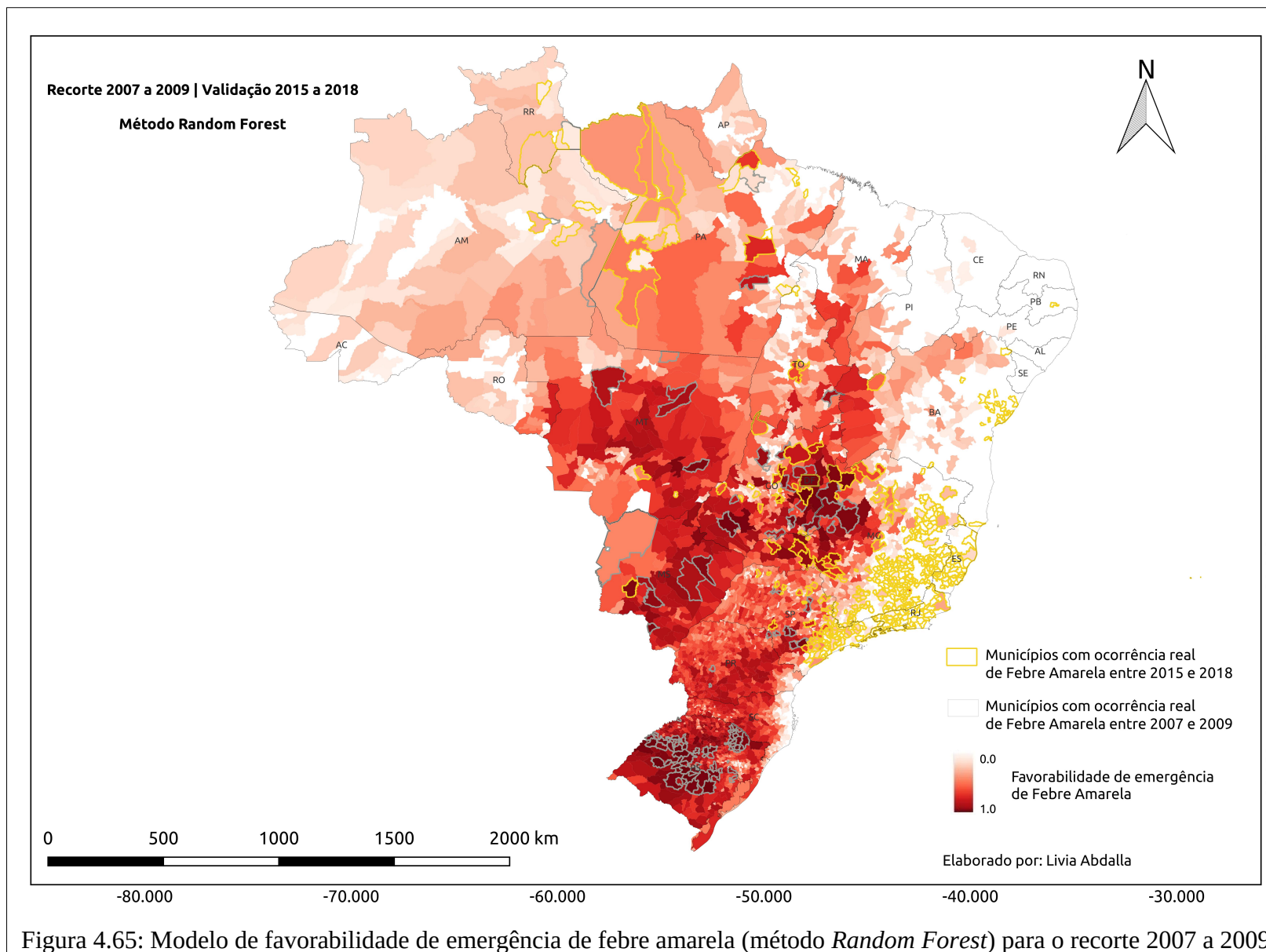
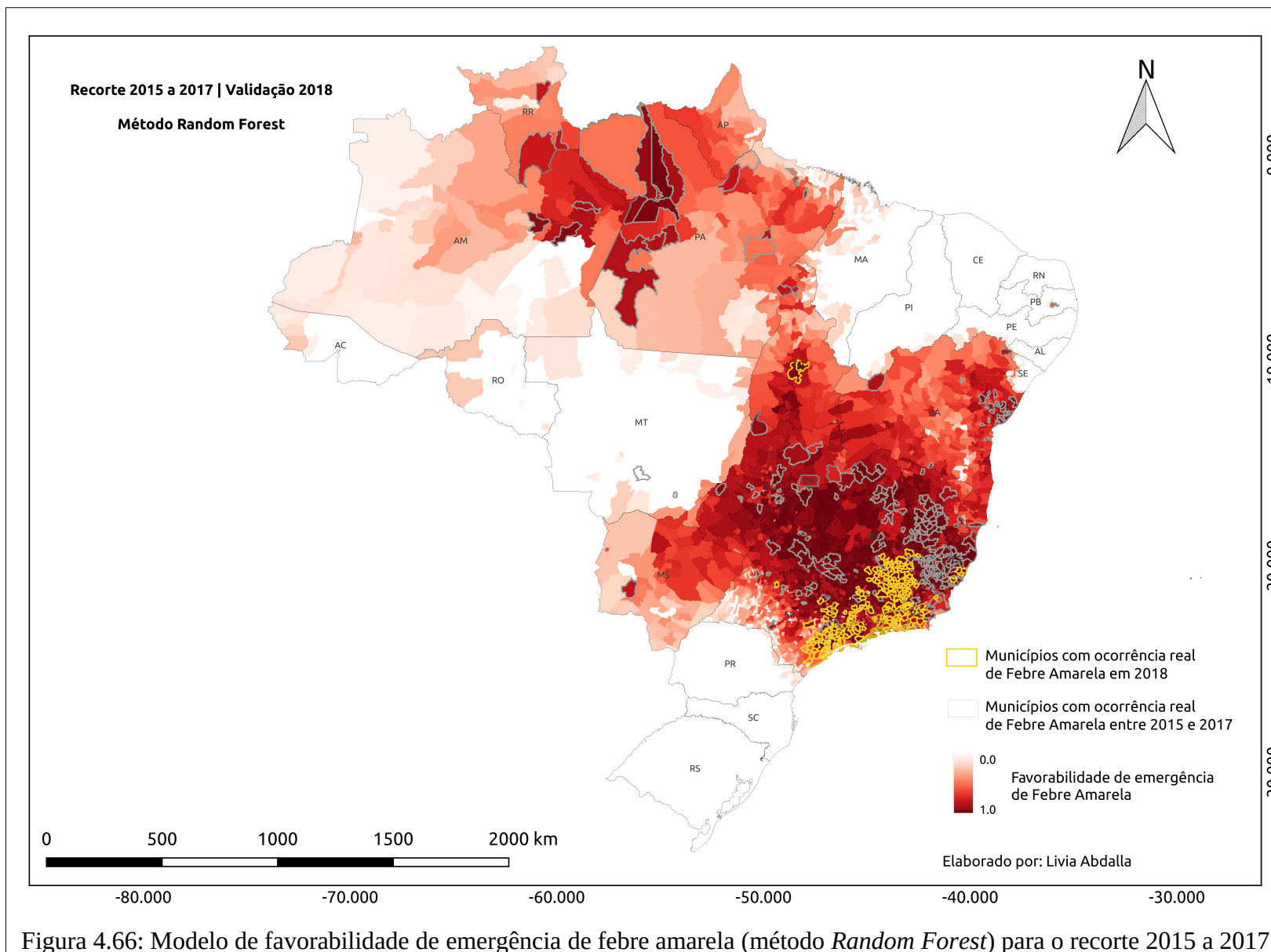


Figura 4.65: Modelo de favorabilidade de emergência de febre amarela (método *Random Forest*) para o recorte 2007 a 2009

4.3.2.4 Recorte 04 (2015 a 2017)

Para o recorte 2015 a 2018 foi selecionado como final o modelo 4.8. O método *Random Forest* indicou favorabilidade para 2.798 municípios (figura 4.66), sendo 2.544 falsos negativos e 254 verdadeiros positivos. Os Estados com maior número de municípios favoráveis foram Amazonas, Para, Roraima, Amapá, Tocantins, Bahia, Goiás, Mato Grosso do Sul, Minas Gerais, Espírito Santo, Rio de Janeiro e São Paulo. Destaca-se que os Estados da Bahia e das regiões Norte (AM, AP e PA), Centro-Oeste (TO, GO, MS) e Sudeste (MG, ES, RJ, SP) foram os que concentraram maior número de municípios com maior grau de favorabilidade, o que demonstra concordância com o padrão espacial de distribuição da doença no surto observado entre 2015 e 2018.



5. CONCLUSÕES

Conforme ressaltado no Capítulo 2, comparativamente a trabalhos recentes de modelagem de distribuição de doenças ou espécies de animais hospedeiros de doenças (SAMY *et al*, 2017; ALLEN *et al*, 2017; ALTAMIRANDA-SAAVEDRA *et al*, 2017; HAMRICK *et al*, 2017; PETERSON e TALUKDAR, 2017; ESCOBAR *et al*, 2016; PETERSON e SAMY, 2016; SAMY *et al*, 2016b; ESCOBAR *et al*, 2015; OLIVEIRA, 2013), destaca-se que esta tese possui contribuições metodológicas originais e robustas, existindo diferenças relevantes, tais como a:

(i) escala municipal de análise dos dados (com maior resolução espacial do que as publicações anteriores a esta pesquisa), com a caracterização dos municípios baseada em ampla descrição estatística, com aproximadamente 22 mil atributos das características socioambientais que compõem os 5.571 municípios brasileiros;

(ii) abordagem espaço-temporal, sendo importante mecanismo de controle e correspondência entre o período de ocorrência e o contexto das condições naturais e antrópicas da paisagem que coexistiram no tempo e espaço para emergência da doença;

(iii) experimentação de dois métodos de aprendizagem de máquina que viabilizaram tanto a interpretabilidade do fenômeno quanto a acurácia preditiva.

Com relação aos resultados destacam-se a:

(i) indicação de 68 variáveis correlacionadas com a emergência da febre amarela;

(ii) construção de 40 modelos experimentais (com diferentes recortes temáticos e/ou temporais) e seleção de 8 modelos finais, com elevada acurácia interna e preditiva, para a indicação de novos municípios favoráveis à ocorrência de febre amarela silvestre.

No que se refere às variáveis correlacionadas com a ocorrência da doença, é fundamental destacar que a sua emergência e/ou reemergência dependem da convergência de um conjunto de condições sociais e ambientais. Dentre as relações evidenciadas no subcapítulo 4.1.1 e que indicaram um possível caminho para compreensão da emergência da febre amarela, destacam-se:

(i) redução de áreas florestais nos biomas Amazônia, Cerrado e Caatinga e em paralelo a recuperação de áreas florestais no bioma Mata Atlântica, o que pode estar relacionado com a dispersão de primatas para os fragmentos de Mata Atlântica, bem como, com a elevada

proporção de primatas acometidos na região Sudeste durante o terceiro surto;

(ii) precarização da gestão e preservação de áreas protegidas, o que aumenta o risco de sobrevivência de diferentes espécies de primatas, o que consequentemente, amplia as condições de acometimento por doenças como a febre amarela;

(iii) desmatamento (principalmente nas regiões Norte e Centro-Oeste) para as diversas espécies de animais silvestres, especialmente muitas espécies de PNHs (*Alouatta caraya*, *Alouatta guariba*, *Callithrix* sp., *Callicebus nigrifrons*, entre outras) tornando-as extremamente vulneráveis e ameaçadas a diversos tipos de riscos diretos e indiretos (estresse, confinamento, escassez de alimentos, competição interespecífica, redução da imunidade ampliando os riscos a doenças infecciosas);

(iv) elevação dos valores de precipitação e temperatura, os quais favorecem a reprodução e disseminação dos principais gêneros de mosquitos transmissores da Febre Amarela Silvestre (*Haemagogus* e *Sabethes*) já que esses habitam áreas de mata fechada ou as bordas das matas e depositam seus ovos em cavidades de troncos de árvores e em bromélias, que acumulam água das chuvas. Especificamente tem-se que a temperatura elevada favorece a replicação viral nos vetores, o que contribui para altas taxas de carga viral tanto nos mosquitos quanto nos primatas, tornando-os amplificadores;

(v) transporte ocasional de espécies de vetores transmissores da FA, por meio de locomotivas ferroviárias;

(vi) variação no padrão do deslocamento de espécies de vetores entre regiões de Serras e regiões de planaltos e patamares. Sabendo que o raio de voo desses vetores é em média de 6 km, considera-se que o mesmo demore mais tempo para se deslocar entre fragmentos de mata de regiões de planaltos e patamares em busca de hospedeiros primatas; enquanto que em ambientes de serras ou mares de morros, seu deslocamento entre topo de morros (com feições florestais) é mais facilitado, permitindo que os vetores se alimentem mais rapidamente de outras populações de primatas que possuem diferentes *habitats*.

Esta pesquisa, portanto, não obteve como conclusão o conhecimento sobre uma única variável temática sintetizadora que evidencie a correlação direta com a ocorrência da doença. O que parece estar em concordância com a realidade, visto o complexo histórico de manifestação espaço temporal da doença. Considera-se ainda que a problemática principal

persiste sobre o desconhecimento, em sua completude, sobre como um conjunto fechado de variáveis que se relacionam para favorecer ou não o aparecimento, ressurgimento da doença. Nesse sentido, essa tese reafirma que não teve o propósito de esgotar ou garantir o conhecimento sobre todas as condições determinantes para emergência da doença; porém, esta foi capaz de indicar um caminho para a compreensão sobre os padrões relacionados com a emergência da doença e prever novos municípios potenciais (ou pelo menos porções territoriais) para emergência de novos casos.

Nesse sentido, o entendimento, a compreensão da emergência da doença, na sua completude, dependerá sempre de um olhar atento e integrado sobre diferentes fatores (ex: supressão de matas que promovem a dispersão, a redução ou fragilização imunológica de população de primatas; os períodos prolongados de chuvas que favorecem a reprodução de mosquitos e replicação do vírus; a oferta e exposição de indivíduos humanos aos vetores transmissores) que ocorrem e convergem no tempo-espço criando oportunidades para surtos epizooticos-epidêmicos da febre amarela.

Além dos fatores indicados por essa pesquisa, outros fatores mais específicos sobre as espécies participantes do ciclo de transmissão podem estar relacionados à dinâmica de emergência da Febre Amarela Silvestre. Dentre elas, sugere-se o aprofundamento na investigação:

- (i) da evolução filogenética do vírus;
- (ii) no papel das espécies hospedeiras, especialmente do gênero *Callithrix*, na manutenção e amplificação viral e sua dinâmica de dispersão entre os fragmentos florestais;
- (iii) da compreensão do efeito da fragmentação florestal e das mudanças climáticas nos *habitats* das espécies integrantes ao ciclo da febre amarela;
- (iv) do impacto da introdução de espécies exóticas invasoras, como é o caso do *Callithrix jacchus*, que por sua capacidade de dispersão e colonização de ambientes antropizados podem ampliar a velocidade e área de circulação do vírus.
- (v) a identificação do papel de cada espécie de primata na manutenção do vírus na natureza, bem como de outras espécies como preguiças, marsupiais, e eventualmente ainda outros mamíferos.

No que se refere à metodologia empregada nesta tese, inicialmente destaca-se que a partir

da observação da diversidade de distribuição espaço-temporal da emergência da febre amarela silvestre no Brasil, conclui-se que foi acertada a escolha pela construção de modelos para 4 períodos distintos. Principalmente porque a dinâmica da doença demonstrou forte mudança no padrão, comportamento e magnitude, dando aspecto bastante particular ao surto mais recente da doença, o qual atingiu regiões em que não haviam manifestações tão significativas há décadas, como nos estados do Espírito Santo e Rio de Janeiro.

Sobre a Modelagem Baseada em Dados, e os métodos de Aprendizagem de Máquina empregados, considera-se que esta foi empregada assertivamente para o alcance dos objetivos pretendidos. Principalmente porque foram capazes de contribuir para o entendimento sobre as condições ambientais relacionadas com a emergência da FAS, e, para a produção de diferentes modelos experimentais e a seleção de modelos preditivos finais com maior acerto sobre municípios positivos reais e menor erro sobre municípios negativos (valores de *Recall* superiores a 0,8 em 3 dos 4 recortes adotados), o que permitiu avançar com maior segurança sobre a indicação de novas porções territoriais candidatas à manifestação da doença. Por tudo isso considera-se que a metodologia de modelagem empregada possui grande potencial para continuidade, havendo margem para obtenção de resultados ainda com melhor desempenho sobre o tema investigado.

Por último, como sugestões para continuidade da pesquisa e/ou para trabalhos futuros destacam-se:

- (i) obtenção de dados georreferenciados de ocorrência da febre amarela e a caracterização socioambiental da área de influência (*buffer*) desses pontos;
- (ii) construção de sub-modelos regionais (ex: área endêmica x não endêmica; biomas);
- (ii) inclusão de outros níveis de informação no processo de aquisição de dados (por exemplo: dados relativos ao monitoramento sistemático de focos de queimadas, desmatamento de florestas ou degradação de pastagens; identificadores sobre o risco geológico para ocorrência de desastres naturais; limites de bacias hidrográficas e suas subdivisões, entre outros);
- (iii) elaboração de outra estratégia de produção de municípios pseudonegativos;
- (iv) atualização da base de dados, e consequente caracterização estatística desses novos dados, principalmente para dados do tipo Anuais;

(v) caracterização da paisagem baseada em dados percentuais de transição entre classes de uso e cobertura vegetal, o que permitirá avançar em análises/comprovações mais aprofundadas sobre as relações existentes entre a perda de ecossistemas naturais e emergência da doença;

(vi) validação dos modelos experimentais baseada em todos os municípios positivos em todos os surtos seguintes ao período de construção do modelo;

(vii) verificar sobreposição de indicação de municípios previstos como positivos em todos os modelos experimentais, e a partir dessa verificação produzir um modelo final único indicando hierarquicamente os municípios mais favoráveis à emergência da FAS;

(viii) utilizar métodos ranqueadores de atributos (ex: *InfoGain*) para verificar e discutir sobre os atributos mais relevantes/impactantes aos modelos;

(ix) aprofundar a investigação sobre os métodos de aprendizagem de máquina, tais como a comparação com métodos generativos, o ajuste de parâmetros dos métodos de AM utilizados e o consequente avanço sobre a seleção de modelos (visto que esta tese não teve como objetivo melhorar as configurações dos métodos);

(x) utilizar outros métodos de validação como *leave-one-out* e/ou *jackknife* (validação cruzada) para avaliação da capacidade de generalização dos modelos;

(xi) aplicar metodologia desenvolvida para a construção de modelos preditivos de emergência de outras zoonoses, utilizando a caracterização socioambiental dos municípios disponibilizada por esta tese.

Por fim, e em conformidade com a discussão apresentada em Chame *et al* (2014), destaca-se que em países megadiversos como Brasil (de dimensões continentais, ecossistemas e realidades socioeconômicas e culturais distintas e serviços de saúde e ambientais limitados) a implementação da vigilância ambiental e epidemiológica para prevenção de surtos é um desafio. Assim, considera-se que este desafio pode ser bem-sucedido com o suporte de pesquisas como esta, que contribuem para identificar/confirmar as condições ambientais favoráveis à ocorrência da doença, prever os caminhos de disseminação da doença, apoiar as tomadas de decisão e políticas de conservação e direcionar os esforços de vigilância e imunização das populações humanas nas regiões mais favoráveis à emergência da FAS.

5.1 Recomendações da pesquisa

Conforme abordado anteriormente, durante o desenvolvimento da pesquisa foram identificadas – e propostas alternativas para compensação (caracterização estatística das feições sociais e ambientais que compõem os municípios brasileiros; produção de pseudonegativos) – as limitações, de cunho operacional, para a utilização dos dados de ocorrência da FAS no Brasil, mais especificamente a falta de georreferenciamento dos registros de casos humanos e epizootias da febre amarela silvestre, o que impôs restrições à modelagem produzida por esta pesquisa.

Assim sendo, e desde já evidenciado que não se trata de ingerência em qualquer instituição, em qualquer instância, seja ela pública ou privada, recomenda-se que, obedecendo a estrutura hierárquica do Sistema Único de Saúde (SUS), possa ser desenvolvido e implementado um sistema georreferenciado, e em tempo real, de monitoramento de epizootias de febre amarela que indique antecipadamente sobre a circulação do agente patogênico em determinadas regiões geográficas do país.

Recomenda-se também a disponibilização dos registros georreferenciados (pretéritos – quando e onde houver – e recentes) de casos humanos e epizootias, já que estes são dados fundamentais para a construção de modelos preditivos mais acurados e precisos, tanto do ponto de vista cartográfico quanto da modelagem baseada em dados, para indicação com maior precisão geodésica possível de porções territoriais (mais refinadas que os limites municipais) mais favoráveis à emergência da doença.

Por fim, sugere-se que o modelo 4.8 produzido por esta tese (utilizando método *Random Forest* sobre o recorte 2015 a 2017) seja utilizado como instrumento de apoio para o desenvolvimento das estratégias de priorização de áreas para a distribuição de vacinas no país, assim como para a ampliação das ações de investigação epidemiológica nessas regiões indicadas com maior favorabilidade à emergência, a fim de prevenir novos surtos da FAS.

REFERÊNCIAS BIBLIOGRÁFICAS

- AB'SÁBER, A. N. **Os domínios de natureza no Brasil: potencialidades paisagísticas**. São Paulo: Ateliê Editorial, 2003, 160 p.
- AB'SÁBER, A. N. **Domínios morfoclimáticos e províncias fitogeográficas do Brasil**. São Paulo: Universidade de São Paulo – USP, Instituto de Geografia, n. 3, p. 45-48, 1967.
- ABERNETHY, K. A.; COAD, L.; TAYLOR, G.; LEE, M. E.; MAISELS, F. Extent and ecological consequences of hunting in Central African rainforests in the twenty-first century. **Philosophical Transactions of the Royal Society B: Biological Sciences** v. 1625, n. 368, p. 1-11, 2013 Disponível em: <<https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.2012.0303>>.
- ABU-MOSTAFA, Y.; MAGDON-ISMAIL, M.; LIN, H. **Learning from data: A short course**. 1 ed. Califórnia: AMLBook.com, 2012, 201 p.
- AGUIAR, L. M.; LUDWIG, G.; ROPER, J. J.; SVOBODA, W. K.; NAVARRO, I. T. e PASSOS, F. C. Howler and capuchin monkey densities in riparian forests on islands and adjacent shores on the Upper Paraná River, southern Brazil. **Neotropical Primates**, v. 18, n. 2, p. 39-43, 2011. Disponível em: <<https://bioone.org/journals/Neotropical-Primates/volume-18/issue-2/044.018.0201/Howler-and-Capuchin-Monkey-Densities-in-Riparian-Forests-on-Islands/10.1896/044.018.0201.full>>.
- ALAMGIR, M.; CAMPBELL, M. J.; SLOAN, S.; GOOSEM, M.; CLEMENTS, G. R.; MAHMOUD, M. I.; LAURANCE, W. F. Economic, socio-political and environmental risks of road development in the tropics. **Current Biology** v. 27, n. 20, p. 1130–1140, 2017. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/29065299>>.
- ALISJAHBANA, A. S.; BUSCH, J. M. Forestry, forest fires, and climate change in Indonesia. **Bulletin of Indonesian Economic Studies** v. 53, n. 2, p. 111–136, 2017. Disponível em: <<https://www.tandfonline.com/doi/pdf/10.1080/00074918.2017.1365404>>.
- ALLEN, T.; MURRAY, K. A.; ZAMBRANA-TORRELIO, C.; MORSE, S. S.; RONDININI, C.; MARCO, M.; BREIT, N.; OLIVAL, K.; DASZAK, P. Global hotspots and correlates of emerging zoonotic diseases. **Nature Communications**. v. 8, n. 1124, p. 1-10. 2017. Disponível em: <https://www.nature.com/articles/s41467-017-00923-8>.
- ALMEIDA, E. P. C.; SANTOS, H. G.; ZARONI, M. J. Agência Embrapa de Informação Tecnológica – Ageitec. **Latossolos Brunos**. Disponível em: <http://www.agencia.cnptia.embrapa.br/gestor/solos_tropicais/árvore/>

CONT000fzyjaywi02wx5ok0q43a0r8ll2dpj.html>. Acesso em: 26 out. 2018.

- ALMEIDA, P. S.; SILVA, J. O.; RAMOS, E. P.; BATISTA, P. M.; FACCENDA, O.; DE PAULA, M. B.; MONTEIRO, A. O.; MUCCI, L. F. Vector aspects in risk areas for silvatic yellow fever in the state of Mato Grosso do Sul, Brazil. **Rev. Patol. Trop**, v. 45, n. 4, p. 398-411. 2016. Disponível em: <<https://revistas.ufg.br/iptsp/article/download/44602/22046>>.
- ALMEIDA, M. A. B.; SANTOS, E.; CARDOSO, J. C.; FONSECA, D. F.; NOLL, C. A.; SILVEIRA, V. R.; MAEDA, A. Y.; SOUZA, R. P.; KANAMURA, C. e BRASIL, R. A. Yellow fever outbreak affecting Alouatta populations in southern Brazil (Rio Grande do Sul State), 2008–2009. **American Journal of Primatology**, v. 74, p. 68-76, 2012.
- ALMEIDA, T. A. e YAMAKAMI, A. Facing the Spammers: A Very Effective Approach to Avoid Junk E-mails. **Expert Systems with Applications**, v. 39, n. 7, p. 6557–6561, 2012. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417411017209>>.
- ALMEIDA, T.; ALMEIDA, J. e YAMAKAMI, A. Spam Filtering: How the Dimensionality Reduction Affects the Accuracy of Naive Bayes Classifiers. **Journal of Internet Services and Applications**, v. 1, n. 3, p. 183–200, 2011. Disponível em: <<https://link.springer.com/article/10.1007/s13174-010-0014-7>>.
- ALPAYDIN, E. **Introduction to machine learning**. 2 ed. Cambridge, Massachusetts: MIT Press, 2009. p. 1-579. Disponível em: <[https://kkpatel7.files.wordpress.com/2015/04/alppaydin_machinelearning_2010.pdf](http://kkpatel7.files.wordpress.com/2015/04/alppaydin_machinelearning_2010.pdf)>.
- ALTAMIRANDA-SAAVEDRA, M.; ARBOLEDA, S.; PARRA, J. L.; PETERSON, A.T.; CORREA, M. M. Potential distribution of mosquito vector species in a primary malaria endemic region of Colombia. **PloS ONE** v. 12, n.6, p. 1-14, 2017. Disponível em: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0179093>>
- ALVAREZ-BERRÍOS, N. L.; AIDE, T. M. Global demand for gold is another threat for tropical forests. **Environmental Research Letters** v. 10, n. 1, p. 1-11, 2015. Disponível em: <https://www.researchgate.net/publication/272180225_Global_demand_for_gold_is_another_threat_to_tropical_forests>.
- ALVES, S. L. **Efeitos do tipo de floresta e da estrutura de habitat em assembleias de primatas no sudoeste da Amazônia**. 2013. 56 p. Dissertação (Mestrado em Zoologia) - Museu Paraense Emílio Goeldi, Universidade Federal do Pará - UFPA. Pará, 2013. Disponível em: <<http://repositorio.ufpa.br/jspui/handle/2011/6911>>.
- ALVES, S. L. Ampliação da distribuição geográfica de *Alouatta caraya* (Humboldt, 1812) e simpatria com *Alouatta seniculus* (Linnaeus, 1766) no Estado de Rondônia, Brasil. In: I

- Congresso Sul-Americano de Mastozoologia, 2006, Gramado (RS). **Anais**. Gramado: Sociedade Brasileira de Mastozoologia, 2006, p. 111.
- ANGELOV, P. e GUTHKE, R. A genetic-algorithm-based approach to optimization of bioprocesses described by fuzzy rules. **Bioprocess Engineering**, v. 16, n. 5, p. 299–303, 1997. Disponível em: <<https://link.springer.com/content/pdf/10.1007%2Fs004490050326.pdf>>.
- ANTF – Associação Nacional dos Transportes Ferroviários. **O setor ferroviário de carga brasileiro - 2018**. Disponível em: <<https://www.antf.org.br/informacoes-gerais/>>. Acesso em: 26 out. 2018.
- ARAÚJO, M. B.; CABEZA, M.; THUILLER, W.; HANNAH, L.; WILLIAMS, P. H. Would climate change drive species out of reserves? An assessment of existing reserve-selection methods. **Global Change Biology** v. 10, n. 9, p. 1618–1626, 2004. Disponível em: <<https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1365-2486.2004.00828.x>>.
- ARAÚJO, F. A. A.; RAMOS, D. G.; SANTOS, A. L.; PASSOS, P. H. O.; ELKHOURY, A. N. S. M.; COSTA, Z. G. A.; LEAL, S. G.; ROMANO, A. P. M. Epizootias em primatas não humanos durante reemergência do vírus da febre amarela no Brasil, 2007 a 2009. **Epidemiol. Serv. Saúde, Brasília**, v. 20, n. 4, p. 527-536. 2011. Disponível em <http://scielo.iec.gov.br/scielo.php?script=sci_arttext&pid=S1679-49742011000400012&lng=pt&rm=iso>.
- ARAÚJO, G. H. S; RIBEIRO, A. J; GUERRA, A. J. T. **Gestão Ambiental de áreas degradadas**. Rio de Janeiro: Bertrand Brasil, 2005.
- AUGUSTO, D. **Programação Genética Multi-Populacional e Co-evolucionária para Classificação de dados**. 2009. 165 p. Tese (Doutorado em Engenharia Civil) – Programa de Pós-graduação em Engenharia Civil - COPPE, Universidade Federal do Rio de Janeiro – UFRJ. Rio de Janeiro, 2009. Disponível em: <<https://tesesufrrj.wordpress.com/2010/09/16/programacao-genetica-multi-populacional-e-co-evolucionaria-para-classificacao-de-dados/>>.
- AURICCHIO, P. **Primatas do Brasil**. São Paulo: Terra Brasilis Editora, 168p. 1995. Disponível em: <https://www.researchgate.net/publication/266615951_Primatas_do_Brasil>.
- AYRES, J. M.; MARTINS, E. **Biologia e ecologia de primatas não-humanos na Amazônia: avaliação e perspectivas**. Belém: Museu Paraense Emílio Goeldi, 1989, p. 241-52.
- BABUSKA, R. **Fuzzy Modeling for Control**. Boston: Kluwer Academic Publishers, 1998.
- BALDI, P. e BRUNAK, S. **Bioinformatics: The Machine Learning Approach**. 2 ed. MIT

Press. 2001.

BANCO MUNDIAL. **Country and Economies 2018**. Disponível em: <http://data.worldbank.org/country>. Acesso em: 28 out. 2018.

BANCO MUNDIAL. **The Cost of Fire: An Economic Analysis of Indonesia's 2015 Fire Crisis**. Washington - D.C.: World Bank, 2016.

BARCELLOS, C.; RAMALHO, W. M.; GRACIE, R. Georreferenciamento de dados de saúde na escala submunicipal: algumas experiências no Brasil. **Epidemiol. Serv. Saúde**, v. 17, n. 1, p. 59-70, 2008. Disponível em: http://repositorio.unb.br/bitstream/10482/11481/1/ARTIGO_GeorreferenciamentoDadosSaude.PDF>.

BARCELLOS C, RAMALHO WM. Situação atual do geoprocessamento e da análise de dados espaciais em saúde no Brasil. **Revista de Informática Pública**; v. 4, n. 2, p.221-230, 2002. Disponível em: https://www.arca.fiocruz.br/bitstream/icict/768/2/BARCELLOS_Situacao%20atual%20do%20geoprocessamento_2002.pdf>.

BARRETT, M. A.; BROWN, J. L.; JUNGE, R. E.; YODER, A. D. Climate change, predictive modeling and lemur health: assessing impacts of changing climate on health and conservation in Madagascar. **Biological Conservation** v. 157, p. 409–422, 2013. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0006320712003989>>.

BARRETT, A. D. T. e MONATH, T. P. Epidemiology and ecology of yellow fever vírus. **Adv vírus Res** v. 61, p. 291–315, 2003.

BARTON, R. R. e MECKESHEIMER, M. Metamodel-based simulation optimization. **Handbooks in Operations Research and Management Science**, v. 13, p. 535–574, 2006. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0927050706130182>>.

BEJARANO, J. F. R. Complejos patógenos de los vírus transmitidos por artrópodos. **Rev Sanit Milit Argent**; v. 70, n. 1, p. 51-81, 1971.

BENAYAS, J. M. R.; NEWTON, A. C.; DIAZ, A.; BULLOCK, J. M. Enhancement of biodiversity and ecosystem services by ecological restoration: A meta-analysis. **Science**, v. 325, n. 5944, p. 1121-1124, 2009. Disponível em: http://www.lerf.eco.br/img/publicacoes/2009_1811%20Enhancement%20of%20Biodiversity%20and%20Ecosystem%20Services%20by%20Ecological%20Restoration.pdf>.

- BENCHIMOL, J. L.; SILVA, A. F. C. Ferrovias, doenças e medicina tropical no Brasil da Primeira República. **Hist. cienc. saude-Manguinhos**, v. 15, n. 3, p. 719-762, 2008. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-59702008000300009&lng=en&enrm=iso>.
- BENCHIMOL, J. L. e SÁ, M. R. Insetos, humanos e doenças: Adolpho Lutz e a medicina tropical. In: Benchimol, J. L. e Sá, M. R. (Orgs.), Adolpho Lutz, Obra Completa, volume II, livro 1: **Febre amarela, malária e protozoologia**. Yellow Fever, Malaria e Protozoology. Rio de Janeiro: Ed. Fiocruz, 2005, p.43-244. Disponível em: <<http://books.scielo.org/id/pmpkw/pdf/benchimol-9788575414071-05.pdf>>.
- BENCHIMOL, J. L. (coord). **Febre amarela: a doença e a vacina, uma história inacabada**. Rio de Janeiro: Editora FIOCRUZ, 2001. 470 p. Disponível em: <<http://books.scielo.org/id/4nktq/pdf/benchimol-9788575413951.pdf>>.
- BICCA-MARQUES, J. C.; CALEGARO-MARQUES, C.; RYLANDS, A. B.; STRIER, K. B.; MITTERMEIER, R. A.; ALMEIDA, M. A. B.; CASTRO, P. H. G.; CHAVES, Ó. M.; FERRAZ, L. P.; FORTES, V. B.; HIRANO, Z. M. B.; JERUSALINSKY, L.; KOWALEWSKI, M.; MARTINS, W. P.; MELO, F. R.; MENDES, S. L.; NEVES, L. G.; PASSOS, F. C.; PORT-CARVALHO, M.; RIBEIRO, S.; ROMANO, A. P. M.; RUIZ-MIRANDA, C. R.; SANTOS, E. O.; SOUZA, J. C.; TEIXEIRA, D. S. Yellow fever threatens Atlantic Forest primates. **Science Advances** v. 3, n. 1, 2017. Disponível em: <https://www.researchgate.net/publication/317162898_Yellow_fever_threatens_Atlantic_Forest_primates>.
- BICCA-MARQUES, J. C. How do howler monkeys cope with habitat fragmentation? In: Marsh, L. K. (ed.) **Primates in Fragments: Ecology and Conservation**. New York: Kluwer Academic/Plenum Publishers, 2003, p. 283-303. Disponível em: <<https://link.springer.com/content/pdf/10.1007%2F978-1-4757-3770-7.pdf>>.
- BIEHN, N. Wired. **The missing V's in big data: viability and value - 2013**. Disponível em: <<https://www.wired.com/insights/2013/05/the-missing-vs-in-big-data-viability-and-value>>. Acesso em: 25 set. 2017.
- BISHOP, C. M. Introduction. **Pattern Recognition and Machine Learning**. Springer, 2006, p. 1-66. Disponível em: <<http://users.isr.ist.utl.pt/~wurmd/Livros/school/Bishop%20-%20Pattern%20Recognition%20And%20Machine%20Learning%20-%20Springer%20%202006.pdf>>.
- BOHRER, C. B. A. Mata Atlântica e formações associadas. In: Mozart, V.; Serra; M. T. F. (Orgs.). **Guia de História Natural do Rio de Janeiro**. Rio de Janeiro: Cidade Viva, 2012, p. 138-157.
- BONVICINO, C. R. Ecologia e comportamento de *Alouatta belzebul* (Primates: Cebidae) na Mata Atlântica. **Revista Nordestina de Biologia**, v. 6, n. 2, p. 149-179, 1989. Disponível

- em: <<http://www.periodicos.ufpb.br/ojs/index.php/revnebio/article/view/16763/10222>>.
- BOUBLI, J. P.; DI FIORE, A. e MITTERMEIER, R. A. **Alouatta macconnelli**. IUCN Red List of Threatened Species, versão 2011.2, 2008. Disponível em www.iucnredlist.org. Acesso em: 23 out. 2018.
- BORGES, C. C. H. **Algoritmos genéticos para otimização em dinâmica de estruturas**. 1999, 216 p. Tese (Doutorado em Engenharia), Programa de Engenharia Civil – COPPE, Universidade Federal do Rio de Janeiro - UFRJ, Rio de Janeiro, 1999. Disponível em: <<file:///home/livia/Downloads/Carlos%20Cristiano%20Hasenclever%20Borges.pdf>>.
- BRANCALION, P. H. S.; GARCIA, L. C.; LOYOLA, R., RODRIGUES, R. R.; PILLAR, V. D.; LEWINSOHN, T. M. A critical analysis of the Native Vegetation Protection Law of Brazil (2012): updates and ongoing initiatives. **Nat Conserv**, v. 14, p. 1–15, 2016. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1679007316300020>>.
- BRASIL / MS - MINISTÉRIO DA SAÚDE / SVS - SECRETARIA DE VIGILÂNCIA EM SAÚDE. Departamento de Vigilância das Doenças Transmissíveis. **Guia de vigilância de epizootias em primatas não humanos e entomologia aplicada à vigilância da febre amarela** – 2. ed. – Brasília: Ministério da Saúde, 2017, 100 p. Disponível em: <http://bvsmms.saude.gov.br/bvs/publicacoes/guia_vigilancia_epizootias_primatas_entomologia.pdf>.
- BRASIL / MS - MINISTÉRIO DA SAÚDE / SVS - SECRETARIA DE VIGILÂNCIA EM SAÚDE. Departamento de Vigilância das Doenças Transmissíveis. **Manual de vigilância epidemiológica de febre amarela**. Brasília: Ministério da Saúde, 2004, 69 p. Disponível em: <http://bvsmms.saude.gov.br/bvs/publicacoes/manual_vigilancia_epid_febre_amarela.pdf>.
- BRAVO, S. P. e SALLENAVE, A. Foraging behavior and activity patterns of *Alouatta caraya* in the northeastern argentinean flooded Forest. **International Journal of Primatology**, v. 24, n. 4, p. 825-846, 2003. Disponível em: <<https://link.springer.com/article/10.1023/A:1024680806342>>.
- BREIMAN, L.; GHAMRANI, Z. **Consistency for a simple model of random forests**. California: Statistical Department, University of California at Berkeley, 2004 (Relatório Técnico n° 670). Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=D8389C3D89DD8ACE5F4EDCA02024258C?doi=10.1.1.618.90&rep=rep1&type=pdf>>.
- BREIMAN, L. Random Forests. **Machine Learning**. v. 45, p. 5-32, 2001. Disponível em: <<https://link.springer.com/article/10.1023/A:1010933404324>>.
- BREIMAN, L.; FRIEDMAN, J.; OLSHEN, R. e STONE, C. **Classification and Regression**

Trees. Wadsworth Intl, 1984.

BRODIE, J. F.; HELMY O. E.; BROCKELMAN W. Y.; MARON J. L. Bushmeat poaching reduces the seed dispersal and population growth rate of a mammal-dispersed tree. **Ecological Applications** v. 19, n. 4, p. 854–863, 2009. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/19544729>>.

BROLLO, M. J.; FERREIRA, C. J.; SILVA, P. C. F.; TOMINAGA, L. K.; VEDOVELLO, R.; GUEDES, A. C. M.; LOPES, E. A.; OLIVEIRA, A. R.; UMMUS, M. E.; GRIPPS, J. C. Caracterização das áreas degradadas por mineração de saibro em Ubatuba: condicionantes para sua recuperação ambiental. In: 11º Congresso Brasileiro de Geologia e Engenharia Ambiental, 2005, Florianópolis – SC. **Anais...** Disponível em: <https://www.researchgate.net/publication/316856443_Caracterizacao_das_areas_degradadas_por_mineracao_de_saibro_em_Ubatuba_condicionantes_para_a_sua_recuperacao_ambiental>.

BUTCHART, S. H. M.; CLARKE, M.; SMITH, R. J.; SYKES, R. E.; SCHARLEMANN, J.P.W.; HARFOOT, M.; BUCHANAN, G.M.; ANGULO, A.; BALMFORD, A.; BERTZKY, B.; BROOKS, T. M.; CARPENTER, K. E.; COMEROS-RAYNAL, M. T.; CORNELL, J.; FICETOLA G. F.; FISHPOOL, L. D. C.; FULLER, R. A.; GELDMANN, J.; HARWELL, H.; HILTON-TAYLOR, C.; HOFFMANN, M.; JOOLIA, A.; JOPPA, L.; KINGSTON, N.; MAY, I.; MILAM, A.; POLIDORO, B.; RALPH, G.; RICHMAN, N.; ONDININI, C.; SEGAN, D. B.; SKOLNIK, B.; SPALDING, M. D.; STUART, S. N.; SYMES, A.; TAYLOR, J.; VISCONTI, P.; WATSON, J. E. M.; WOOD, L.; BURGESS, N. D. Shortfalls and solutions for meeting national and global conservation area targets. **Conservation Letters** v. 8, n. 5, p. 329–337, 2015. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1111/conl.12158>>.

BUTT, N.; BEYER, H. L.; BENNETT, J. R.; BIGGS, D.; MAGGINI, R.; MILLS, M.; RENWICK, A. R.; SEABROOKE, L. M.; POSSINGHAM, H. P. Biodiversity risks from fossil fuel extraction. **Science** v. 342, n. 6157, p. 425–426, 2013. Disponível em: <<http://science.sciencemag.org/content/342/6157/425>>.

CÂMARA G, MONTEIRO AMV. Geocomputation techniques for spatial analysis: are they relevant to health data? **Cadernos de Saúde Pública** v. 17, n. 5, p.1059-1081, 2001. Disponível em: <<http://www.scielo.br/pdf/csp/v17n5/6309.pdf>>.

CAMARGO, E. P. Doenças tropicais. **Estudos av.** v. 22, n. 64, p. 95-110, 2008. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142008000300007&lng=en&rm=iso>.

CAMILO, C. O.; SILVA, J. C. **Mineração de dados: Conceitos, Tarefas, Métodos e Ferramentas.** Goiás: Instituto de Informática, Universidade Federal de Goiás – UFG, 2009 (Relatório Técnico). Disponível em: <http://www.portal.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_001-09.pdf>.

- CARUANA, R., LOU, Y., GEHRKE, J., KOCH, P., STURM, M., ELHADAD, N. Intelligible Models for HealthCare: Predicting Pneumonia Risk and Hospital 30-day Readmission. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, Sydney: NSW. *Anais...* 2015, p. 10-13,. Disponível em: <http://people.dbmi.columbia.edu/noemie/papers/15kdd.pdf>.
- CARVALHO, R. P. **Análise dos instrumentos legais de proteção ambiental para exploração da silvicultura de espécies exóticas do gênero *pinus* e *eucalyptus* no estado do Rio Grande do Sul (Brasil) e na argentina sob a óptica do direito comparado**. 2018. 142 p. Dissertação (Mestrado em Direito) - Universidade de Caxias do Sul – UCS, Caxias do Sul, Rio Grande do Sul, 2018. Disponível em: <<https://repositorio.ucs.br/xmlui/bitstream/handle/11338/3814/Dissertacao%20Rodrigo%20Pinto%20Carvalho.pdf?sequence=1&isAllowed=y>>.
- CARVALHO, A. S. **Mamíferos de médio e grande porte na Floresta Nacional de Carajás, Pará: riqueza, abundância e efeitos da fitofisionomia e do impacto da mineração**. 2010. 134p. Dissertação (Mestrado em Ecologia e Evolução) - Universidade do Estado do Rio de Janeiro - UERJ, Rio de Janeiro. 2010. Disponível em: <http://bdtd.ibict.br/vufind/Record/UERJ_c74f5ff35396b79d244ecd4b5e3e724a>.
- CASAGRANDE, R. A. PANNUTI, C. S.; KANAMURA, C.; FREIRE, W. S.; GRESPAN, A.; MATUSHIMA, E. R. Infecção fatal por Herpesvírus simplex Tipo 1 (HHV-1) em saguis de cativeiro (*Callithrix jacchus* and *Callithrix penicillata*) no Brasil: caracterização clínica e patológica. **Pesq. Vet. Bras** (online). v. 34, n. 11, p. 1109-1114, 2014. Disponível em: <http://www.scielo.br/scielo.php?pid=S0100-736X2014001100013&script=sci_abstract&lng=pt>.
- CASTRO, L. N.; FERRARI, D. M. **Introdução à mineração de dados**. Conceitos básicos, algoritmos e aplicações. São Paulo: Editora Saraiva, 2016, 351 p.
- CAUGHLIN, T. T.; FERGUSON, J. M.; LICHSTEIN, J. W.; ZUIDEMA, P. A.; BUNYAVEJCHEWIN, S.; LEVEY, D. J. Loss of animal seed dispersal increases extinction risk in a tropical tree species due to pervasive negative density dependence across life stages. **Proceedings of the Royal Society B: Biological Sciences**. v. 282, n. 1798, 2015. Disponível em: <<https://royalsocietypublishing.org/doi/full/10.1098/rspb.2014.2095>>.
- CEBALLOS, G.; EHRLICH, P. R. Mammal population losses and the extinction crisis. **Science** v. 296, n. 5569, p. 904–907, 2002. Disponível em: <<http://science.sciencemag.org/content/296/5569/904>>.
- CERQUEIRA, F. R.; FERREIRA, T. G.; DE PAIVA OLIVEIRA, A.; AUGUSTO, D. A.; KREMPSE, E.; CORRÊA BARBOSA, H. J.; DO CARMO CASTRO FRANCESCHINI, S.; DE FEITAS, B. A. C.; GOMES, A. P.; SIQUEIRA-BATISTA, R.

- NICeSim: An open-source simulator based on machine learning techniques to support medical research on prenatal and perinatal care decision making. **Artificial Intelligence in Medicine - (Print) JCR**, v. 62, p. 193-201, 2014. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0933365714001043?via%3Dihub#bib0120>>.
- CHAME, M.; BARBOSA, H. J. C.; GADELHA, L. M. R.; AUGUSTO, D. A.; KREMPSE, E.; ABDALLA, L. SISS-Geo: Leveraging Citizen Science to Monitor Wildlife Health Risks in Brazil. **Journal of Healthcare Informatics Research**. p. 1-27, 2019. Disponível em: <<http://www.springer.com/home?SGWID=0-0-1003-0-0&aqId=3793896&download=1&checkval=89aae73ee2ff8ba4f4f91779a9e5ea79>>.
- CHAME, M.; BARBOSA, H. J. C.; GADELHA, L. M. R.; AUGUSTO, D. A.; KREMPSE, E.; ABDALLA, L. Sistema de Informação em Saúde Silvestre - SISS-Geo. In: Salgado, A.C.; Motta, C.L.R.; Santoro, F.M. (orgs). **Grandes Desafios da Computação no Brasil – Relatos do 3º Seminário**. Rio de Janeiro: SBC. p. 72-87, 2014. Disponível em: < <http://www.sbc.org.br/documentos-da-sbc/send/141-grandes-desafios/802-grandesdesafiosdacomputaonobrasil>>.
- CHAME, M.; LABARTHE, N. **Saúde Silvestre e Humana: Experiências e perspectivas**. Rio de Janeiro: Fundação Oswaldo Cruz-Fiocruz, Rio de Janeiro, 2013. 105p.
- CHIARELLO, A. G. Effects of fragmentation of the Atlantic forest on mammal communities in south-eastern Brazil. **Biological Conservation**, v. 89, p. 71-82, 1999. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S000632079800130X>>.
- CHIAVEGATTO FILHO, A. D. P. Uso de big data em saúde no Brasil: perspectivas para um futuro próximo. **Epidemiologia e Serviços de Saúde**, v. 24, n.2, p.325-332, 2015. Disponível em: http://www.scielo.br/scielo.php?pid=S2237-96222015000200325&script=sci_abstract&lng=pt.
- CHIU, S. L. Fuzzy model identification based on cluster estimation. **Journal of Intelligent and Fuzzy Systems**, v. 2, n. 3, 1994. Disponível em: <https://www.researchgate.net/publication/233932671_Fuzzy_Model_Identification_Base_d_on_Cluster_Estimation>.
- COELHO, J. M. **Areia industrial** (Relatório Técnico 44 - 2010). Disponível em: <http://www.mme.gov.br/documents/1138775/1256650/P28_RT44_Perfil_Areia_Industrial.pdf/bbfaa3ec-8809-436f-8c20-a9bfd70edd14>.
- CONAWAY, D. **The data science venn diagram**. Disponível em: <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>. Acesso em: 22 out. 2018.
- COSTA, Z. G. A.; ROMANO, A. P. M.; ELKHOURY, A. N. M.; FLANNERY, B. Evolução

- histórica da vigilância epidemiológica e do controle da febre amarela no Brasil. **Rev. Pan-Amaz Saúde**. v. 2, n. 1, p. 11-26, 2011. Disponível em: <<http://scielo.iec.gov.br/pdf/rpas/v2n1/v2n1a02.pdf>>.
- COSTA, Z. G. A. C. **Estudo das características epidemiológicas da febre amarela no Brasil, nas áreas fora da Amazônia Legal, no período de 1999 a 2003**. 2005, 164 P. Dissertação (Mestrado em Saúde Pública) - Escola Nacional de Saúde Pública – ENSP, Fiocruz, Rio de Janeiro, 2005. Disponível em: <<https://www.arca.fiocruz.br/handle/icict/13327>>.
- COURA, J. R. **Dinâmica das doenças infecciosas e parasitárias**. 2. ed. Rio de Janeiro: Guanabara Koogan, 2013.
- COUTO-LIMA, D.; MADEC, Y.; BERSOT, M. I.; CAMPOS, S. S.; MOTTA, M. A.; SANTOS, F. B. D.; VAZEILLE, M.; VASCONCELLOS, P. F. D. C.; LOURENÇO-DE-OLIVEIRA, R.; FAILLOUX, A. B. Potential risk of re-emergence of urban transmission of yellow fever vírus in Brazil facilitated by competent Aedes populations. **Scientific Report**. v. 4848, n. 7, 2017 Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/28687779>>
- COVER, T. e HART, P. Nearest Neighbor Pattern Classification. **IEEE Transactions on Information Theory**, v. 13, n. 1, p. 21-27, 1967. Disponível em: <<https://ieeexplore.ieee.org/document/1053964?denied=>>>.
- CPTEC. **Climatologias de Precipitação e Temperatura no Brasil - 2018**. Disponível em: <<http://infoclima1.cptec.inpe.br/~rinfo/infoclima/climatologia.shtml>>. Acesso em: 21 out. 2018.
- CRIST, E.; MORA, C.; ENGELMAN, R. The interaction of human population, food production, and biodiversity protection. **Science** v. 6334, n. 356, p. 260–264, 2017. Disponível em: <<http://science.sciencemag.org/content/356/6335/260>>
- CRISTIANINI, N.; SCHÖLKOPF, B. Support Vector Machines and Kernel Methods. The New Generation of Learning Machines. **AI Magazine**. v. 23, n. 3, 2002. Disponível em: <http://homepages.math.uic.edu/~lreyzin/f14_mcs548/cristianini02.pdf>.
- DATASUS - Departamento de Informática do SUS. **Informações de Saúde, Epidemiológicas e Morbidade: banco de dados** - 2019. Disponível em: <<http://www2.datasus.gov.br/DATASUS/index.php?%20area=0203>>.
- DAVENPORT, T., FAKIH, S., KIMITI, S., KLEINE, L., FOLEY, L., DE LUCA, D. Zanzibar’s endemic red colobus Piliocolobus kirkii: first systematic and total assessment of population, demography and distribution. **Oryx**. v. 53, n. 1, p. 36-44, 2017. Disponível em: <<https://www.cambridge.org/core/journals/oryx/article/zanzibars-endemic-red>>.

[colobus-piliocolobus-kirkii-first-systematic-and-total-assessment-of-population-demography-and-distribution/F084062464F9C99208E5F7E8CB155880](https://doi.org/10.1007/s10764-007-9173-1)>.

DECANINI, D. P.; MACEDO, R. H. Sociality in *Callithrix penicillata*: I. Intragroup male profile. **International Journal of Primatology**, v. 29, p. 433-447, 2008. Disponível em: <<https://link.springer.com/article/10.1007/s10764-007-9173-1>>.

DECOSTE, D.; SCHOELKOPF, B. Training Invariant Support Vector Machines. **Machine Learning**, v. 46, n. 1-2-3, p. 161-190. 2002. Disponível em: <<https://link.springer.com/article/10.1023/A:1012454411458>>.

DEFLER, T. R. **Primates de Colombia**. Bogotá, D.C: Conservation International, 2003. 543p.

DIGNUM, S.; POLI, R. Generalisation of the limiting distribution of program sizes in tree-based genetic programming and analysis of its effects on bloat. In: GECCO '07: Proceedings of the 9th annual conference on Genetic and evolutionary computation, 2007, London: ACM Press, **Anais...** 2007, v. 2, p. 1588–1595,. Disponível em: <https://www.researchgate.net/publication/220742724_Generalisation_of_the_limiting_distribution_of_program_sizes_in_tree-based_genetic_programming_and_analysis_of_its_effects_on_bloat>.

DNPM - Departamento Nacional de Produção Mineral. PINHEIRO, W. F.; FERREIRA FILHO, O. B.; NEVES, C. A. R. (coord. Geral) **Anuário Mineral Brasileiro: Principais Substâncias Metálicas**. Brasília: DNPM, versão 1, 2018. 33 p. Disponível em: <http://www.anm.gov.br/dnpm/publicacoes/serie-estatisticas-e-economia-mineral/anuario-mineral/anuario-mineral-brasileiro/amb_metalicos2017>

DNPM – Departamento Nacional de Produção Mineral. **Anuário Mineral Brasileiro**. Brasília: DNPM, 2010, 871 p. Disponível em: <<http://www.anm.gov.br/dnpm/publicacoes/serie-estatisticas-e-economia-mineral/anuario-mineral/anuario-mineral-brasileiro/anuario-mineral-brasileiro-2010>>

DOMINGOS, P. **The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World**. New York: Basic Books, 2015.

DOSEN, J.; RABOY, B. E.; FORTIB, M. J. Restoration strategies to improve connectivity for golden-headed lion tamarins (*Leontopithecus chrysomelas*) in the Bahian Atlantic Forest, Brazil. **International Journal of Primatology** v. 38, n. 5, p. 962–983, 2017. Disponível em: <<https://link.springer.com/article/10.1007/s10764-017-9991-8>>.

EFRON, B. Bootstrap methods: Another look at the jackknife. **The Annals of Statistics**. v. 7, n. 1, p. 1-26, 1979. Disponível em: <http://jet1.uni-freiburg.de/studenten_seminar/stud_sem_SS_09/EfronBootstrap.pdf>.

- EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária. SANTOS, H. G.; [JACOMINE, P. K. T.](#); ANJOS, L. H. C.; OLIVEIRA, V. A.; LUMBRERAS, J. F.; COELHO, M. R.; ALMEIDA, J. A.; ARAUJO FILHO, J. C.; OLIVEIRA, J. B.; CUNHA, T. J. F (Orgs). **Sistema brasileiro de classificação dos solos**. 1º ed., 2018. Disponível em: <<https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1094003/sistema-brasileiro-de-classificacao-de-solos>>.
- ENGEL, A. R.; VASCONCELOS, P. F.; MCARTHUR, M. A.; BARRETT, A. D. Characterization of a viscerotropic yellow fever vaccine variant from a patient in Brazil. **Vaccine**. v. 24, n. 15, 2006. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/16464518>>.
- ESCOBAR, L. E.; QIAO, H.; PETERSON, A. T. Forecasting Chikungunya spread in the America via data-data empirical approaches. **Parasites and Vectors**. v. 112, n. 9, 2016. Disponível em: <https://parasitesandvectors.biomedcentral.com/articles/10.1186/s13071-016-1403-y>.
- ESCOBAR, L. E.; PETERSON, A. T.; PAPES, M.; FAVI, M.; YUNG, V.; RESTIF, O.; MEDINA-VOGEL, G. Ecological approaches in veterinary epidemiology: mapping the risk of bat-borne rabies using vegetation indices and night-time light satellite imagery. **Veterinary Research**, v. 46, n. 1, p. 2-10, 2015. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4558958/>>.
- ESPÍNDOLA, R. P. **Sistema Inteligente para Classificação de Dados**. 2004, 123 p. Tese (Doutorado em Engenharia Civil), Pós-graduação em Engenharia Civil - COPPE / Universidade Federal do Rio de Janeiro - UFRJ, Rio de Janeiro, 2004. Disponível em: <<http://www.coc.ufrj.br/pt/teses-de-doutorado/148-2004/984-rogerio-pinto-espindola>>.
- ESTRADA, A.; GARBER, P. A.; MITTERMEIER, R. A.; WICH, S.; GOUVEIA, S.; DOBROVOLSKI, R.; NEKARIS, K. A. I.; NIJMAN, V.; RYLANDS, A. B.; MAISELS, F.; WILLIAMSON, E. A.; BICCA-MARQUES, J.; FUENTES, A.; JERUSALINSKY, L.; JOHNSON, S.; MELO, F. R.; OLIVEIRA, L.; SCHWITZER, C.; ROOS, C.; CHEYNE, S. M.; KIERULFF, M. C. M.; RAHARIVOLOLONA, B.; TALEBI, M.; RATSIMBAZAFY, J.; SUPRIATNA, J.; BOONRATANA, R.; WEDANA, M.; SETIAWAN, A. Primates in peril: the significance of Brazil, Madagascar, Indonesia and the Democratic Republic of the Congo for global primate conservation. **PeerJ** v. 6, 2018. Disponível em: <<https://peerj.com/articles/4869/>>.
- ESTRADA, A.; GARBER, P. A.; RYLANDS, A. B.; ROOS, C.; FERNANDEZ-DUQUE, E.; DI FIORE, A.; NEKARIS, K. A. I.; NIJMAN, V.; HEYMANN, E. W.; LAMBERT, J. E.; ROVERO, F.; BARELLI, C.; SETCHELL, J. M.; GILLESPIE, T. R.; MITTERMEIER, R. A.; ARREGOITIA, L. V.; GUINEA, M.; GOUVEIA, S.; DOBROVOLSKI, R.; SHANEE, S.; SHANEE, N.; BOYLE, S. A.; FUENTES, A.; MACKINNON, K. C.; AMATO, K. R.; MEYER, A. L. S.; WICH, S.; SUSSMAN, R. W.; PAN, R.; KONE, I.; LI, B. Impending extinction crisis of the world's primates: why primates matter. **Science**

- Advances** v. 3, 2017. Disponível em: <<http://advances.sciencemag.org/content/3/1/e1600946>>.
- ESTRADA, A. Socioeconomic contexts of primate conservation: population, poverty, global economic demands and sustainable land use. **American Journal of Primatology** v. 1, n. 75, p. 30–45, 2013. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1002/ajp.22080>>
- ESTRADA-PEÑA, A.; OSTFELD, R. S.; PETERSON, A. T; POULIN, R.; FUENTE, J. Effects of Environmental change on zoonotic disease risk: an ecological primer. **Trends in Parasitology**, v. 30, n. 4, p. 205-214, 2014. Disponível em: <[http://www.cell.com/trends/parasitology/abstract/S1471-4922\(14\)00032-4](http://www.cell.com/trends/parasitology/abstract/S1471-4922(14)00032-4)>.
- FAWCETT, T. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, 2006. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S016786550500303X?via%3Dihub>>.
- FAO – **Estatísticas 2018**. Disponível em: <<http://www.fao.org/faostat/en/>>
- FARIA, J. O. G. **Outras rochas e minerais industriais. Relatório técnico 39. Perfil do Caulim** - 2009. Disponível em: <http://www.mme.gov.br/documents/1138775/1256650/P28_RT39_Perfil_do_Caulim.pdf/c122682a-3c84-40e8-b7f4-e18894f3d42c>. Acesso em: 15 set. 2018.
- FARIA, N. R.; KRAEMER, M. U. G.; HILL, S. C.; GOES DE JESUS, J.; DE AGUIAR, R. S.; IANI, F. C. M.; XAVIER, J.; QUICK, J.; DU PLESSIS, L; DELLICOUR, S.; THÉZÉ, J.; CARVALHO, R. D. O.; BAELE, G.; WU, C. H.; SILVEIRA, P. P.; ARRUDA, M. B.; PEREIRA, M. A.; PEREIRA, G. C.; LOURENÇO, J.; OBOLSKI, U.; ABADE, L.; VASYLYEVA, T. I.; GIOVANETTI, M.; YI, D.; WEISS, D.J.; WINT, G. R. W.; SHEARER, F. M.; FUNK, S.; NIKOLAI, B.; ADELINO, T. E. R.; OLIVEIRA, M. A. A.; SILVA, M. V. F.; SACCHETTO, L.; FIGUEIREDO, P. O.; REZENDE, I. M.; MELLO, E. M.; SAID, R. F. C.; SANTOS, D. A.; FERRAZ, M. L.; BRITO, M. G.; SANTANA, L. F.; MENEZES, M. T.; BRINDEIRO, R. M.; TANURI, A.; DOS SANTOS, F. C. P.; CUNHA, M. S.; NOGUEIRA, J. S.; ROCCO, I. M.; DA COSTA, A. C.; KOMNINAKIS, S. C. V.; AZEVEDO, V.; CHIEPPE, A. O.; ARAUJO, E. S. M.; MENDONÇA, M. C. L.; DOS SANTOS, C. C.; DOS SANTOS, C. D.; MARES-GUIA, A. M.; NOGUEIRA, R. M. R.; SEQUEIRA, P. C.; ABREU, R. G.; GARCIA, M. H. O.; ALVES, R. V.; ABREU, A. L.; OKUMOTO, O.; KROON, E. G.; DE ALBUQUERQUE, C. F. C.; LEWANDOWSKI, K.; PULLAN, S. T.; CARROLL, M.; SABINO, E. C.; SOUZA, R. P.; SUCHARD, M. A.; LEMEY, P.; TRINDADE, G. S.; DRUMOND, B. P.; FILIPPIS, A. M. B.; LOMAN, N. J.; CAUCHEMEZ, S.; ALCANTARA, L. C. J.; PYBUS, O. G. Genomic and epidemiological monitoring of yellow fever vírus transmission potential. **BioRxiv**. 2018, p. 1-35. Disponível em: <<https://www.biorxiv.org/content/biorxiv/early/2018/04/16/299842.full.pdf>>.

- FEELDERS, A. Data mining in economic science. In: Dealing with the Data Flood 166-175, Holanda, Haia: STT/Beweton. **Anais...** 2002.
- FERNANDEZ-DUQUE, E.; CORLEY, M. K. e SPENCE-AIZENBERG, A. Family Aotinae (nigth monkeys). Pp. 414-431. In: Mittermeier, R.A.; Rylands, A.B. e Wilson, D.E. (eds.). **Handbook of the Mammals of the World**. Barcelona: Lynx Edicions, 2013, 951p.
- FERRARI, S. F.; IWANAGA, S.; RAVETTA, A. L.; FREITAS, F. C.; SOUSA, B. A. R.; SOUSA, L. L.; COSTA, C. G. e COUTINHO, P. E. G. Dynamic of primate communities along the Santarém-cuiabá highway in south-Central Brazilian Amazonia. p. 123-124. In: Marsh, L.K. **Primates in Fragments - Ecology and Conservation**. Kluwer Academic/Plenum Publishers 2003, 404p. Disponível em: <https://link.springer.com/chapter/10.1007/978-1-4757-3770-7_9>.
- FIALHO, M. S.; PRINTES, R. C.; ALMEIDA, M. A. B.; LAROQUE, P. O.; SANTOS, E.; JERUSALINSKY, L. Avaliação do impacto da epizootia de Febre Amarela sobre as populações de primatas não humanos nas unidades de conservação do Rio Grande do Sul, Brasil. **Biotemas** v. 25, n. 3, p. 217-225, 2012. Disponível em: <<https://periodicos.ufsc.br/index.php/biotemas/article/view/2175-7925.2012v25n3p217>>.
- FIELDING, A. H. e BELL, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. **Environmental Conservation** v. 24, p. 38-49, 1997. Disponível em: <http://www.dpi.inpe.br/referata/arq/14.Camilo/fielding_Bell_1997.pdf>.
- FICK, S. E.; HIJMANS, R. J. Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. **International Journal of Climatology**. v. 37, p. 4302-4315, 2017. Disponível em: <<https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/joc.5086>>.
- FINZER, W. The data science education dilemma. **Technology Innovations in Statistics Education**, v. 7, n. 2, p. 1-10, 2013. Disponível em: <<http://escholarship.org/uc/item/7gv0q9dc>>.
- FIORAVANTI, C. **The monkey alarm**, v. 3, 2018. Revista Pesquisa FAPESP. Disponível em: <https://issuu.com/pesquisafapesp/docs/ingles_dez2018_completo>.
- FOGUESATTO, K. **Redução de bandos de Alouatta caraya (Primates, Atelidae) no município de Tupanciretã-RS, após surto de febre amarela silvestre**. 2011. 29 p. Monografia (Graduação em Ciências Biológicas), Universidade de Cruz Alta – UNICRUZ, Rio Grande do Sul, 2011.
- FONSECA, G. A. B; LACHER JÚNIOR, T. E. Exudate-feeding by *Callithrix jacchus penicillata* in semideciduous woodland (Cerradão) in Central Brazil. **Primates**, n. 25, v. 4, p. 441-449, 1984. Disponível em:

<<https://link.springer.com/article/10.1007/BF02381667>>.

FONSECA, L. G. **Algoritmos Genéticos Assistidos por Metamodelos Baseados em Similaridade**. 2009, 229 p. Tese (Doutorado em Modelagem Computacional), Programa de pós-graduação em Modelagem Computacional, Laboratório Nacional de Computação Científica - LNCC, Petrópolis, 2009. Disponível em: <http://bdtd.ibict.br/vufind/Record/LNCC_d9f879a5c9c54e29738b1535926fd0b3>.

FORATTINI, O. P. **Epidemiologia, Ecologia e Sociedade**. 2ª ed. São Paulo: Artes Médicas 1999. 209p.

FRANK, E.; HALL, M. A.; WITTEN, I. H. The WEKA Workbench. **Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"**, Morgan Kaufmann, 4º ed, 2016.

FREITAS, D. S. e BICCA-MARQUES, J. C. The impact of a yellow fever outbreak on *Alouatta caraya* in a fragmented landscape in southern Brazil. **American Journal of Primatology**, v. 75, n. 1, p. 41-41, 2013. Disponível em: <https://www.researchgate.net/publication/293639382_THE_IMPACT_OF_A_YELLOW_FEVER_OUTBREAK_ON_ALOUATTA_CARAYA_IN_A_FRAGMENTED_LANDSCAPE_IN_SOUTHERN_BRAZIL>.

FREITAS, D. S.; BICCA-MARQUES, J. C. Evaluating the impact of an outbreak of yellow fever on the black-and-gold howler monkey in southern Brazil. **Oryx** v. 45, p. 16–17, 2011. Disponível em: <https://www.researchgate.net/publication/301358906_Evaluating_the_impact_of_an_outbreak_of_yellow_fever_on_the_black-and-gold_howler_monkey_in_southern_Brazil>.

FREUND, Y.; SCHAPIRE, R. E. Experiments with a new boosting algorithm. Machine Learning: Proceedings of the Thirteenth International Conference, 1996, Itália. **Anais...** Disponível : <<https://cseweb.ucsd.edu/~yfreund/papers/boostingexperiments.pdf>>.

FUNDAÇÃO SOS MATA ATLÂNTICA / INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - INPE. **Atlas da evolução dos remanescentes florestais e ecossistemas associados no domínio da Mata Atlântica no período 1995-2000**. São Paulo: INPE. 2002, 1-45 p. Disponível em: <<http://mtc-m12.sid.inpe.br/col/sid.inpe.br/jeferson/2003/06.02.07.45/doc/RelatorioAtlas.pdf>>.

FUNK, C.; PETERSON, P.; LANDSFELD, M.; PEDREROS, D.; VERDIN, J.; ROWLAND, J.; ROMERO, B.; HUSAK, G.; MICHAELSEN, J.; e VERDIN A. **A Quasi-global Precipitation Time Series for Drought Monitoring**, USA: US Geological Survey, 2014. Disponível em: <<https://pubs.usgs.gov/ds/832/pdf/ds832.pdf>>.

GAMA, J. **Árvores de Decisão**. Porto: Universidade do Porto, 2004. Disponível em

http://www.liaad.up.pt/~jgama/Aulas_ECD/arv.pdf.

GARCIA, L. C.; RIBEIRO, D. B.; OLIVEIRA-ROQUE, F.; OCHOA-QUINTERO, J. M.; LAURANCE, W. F. Brazil's worst mining disaster: corporations must be compelled to pay the actual environmental costs. **Ecological Applications** v. 27, n. 1, p. 5–9, 2017. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/27770593>>.

GFW - Global Forest Watch. **Global Forest Watch - 2018**. Disponível em: <<http://www.globalforestwatch.org/country/COD>>. Acesso em: 12 nov. 2018.

GNANAMBAL, S.; THANGARAJ, M.; MEENATCHI, V. T.; GAYATHRI, V. Classification Algorithms with Attribute Selection: an evaluation study using WEKA. **Int. J. Advanced Networking and Applications**. v. 9, n. 6, p. 3640-3644, 2018. Disponível em: <<http://oaji.net/articles/2017/2698-1528114152.pdf>>.

GOLDBERG, D. **Genetic Algorithms in Search, Optimization and Machine Learning**. Massachusetts - USA: Addison-Wesley Publishing Co., 1989.

GOMES, A. C.; SILVA, N. N.; BERNAL, R. T. I.; SOUZA, A. Estimação da infestação predial por *Aedes aegypti* (Díptera: Culicidae) por meio da armadilha Adultrap. **Epidemiol e Serviços Saúde**. v. 17, n. 4, p. 293–300, 2008. Disponível em: <https://bdpi.usp.br/bitstream/handle/BDPI/13400/art_GOMES_Estimacao_da_infestacao_predial_por_Aedes_aegypti_2008.pdf?sequence=1>.

GOMES, L. M.; REIS, R. B.; CRUZ, C. B. M. Análise da cobertura florestal da Mata Atlântica por município no Estado do Rio de Janeiro. In: XIV Simpósio Brasileiro de Sensoriamento Remoto, 2009, Natal, **Anais...** São José dos Campos: INPE, 2009, p. 3849-3857. Disponível em: <<http://marte.sid.inpe.br/col/dpi.inpe.br/sbsr@80/2008/11.18.00.28/doc/3849-3857.pdf?languagebutton=pt-BR>>.

GÓMEZ, M. M.; ABREU, F. V. S.; SANTOS, A. A. C.; MELLO, I. S.; SANTOS, M. P.; RIBEIRO, I. P.; FERREIRA-DE-BRITO, A. F.; MIRANDA, R. M.; CASTRO, M. G.; RIBEIRO, M. S.; LATERRIÈRE JUNIOR, R. C.; AGUIAR, S. F.; MEIRA, G. L. S.; ANTUNES, D.; TORRES, P. H. M.; MIR, D.; VICENTE, A. C. P.; GUIMARÃES, A. C. R.; CAFFARENA, E. R.; BELLO, G.; LOURENÇO-DE-OLIVEIRA, R.; BONALDO, M. C. Genomic and structural features of the yellow fever virus from the 2016–2017 Brazilian outbreak. **Journal of General Virology**. n. 99; p. 536–548, 2018. Disponível em: <http://www.microbiologyresearch.org/docserver/fulltext/jgv/99/4/536_vir001033.pdf?expires=1543945053&eid=ideacname=capesmemberchecksum=B439E4F0213241A7414802EF6E9BA2A8>

GOUVEIA, S. F., SOUZA-ALVES, J. P., SOUZA, B. B., BELTRÃO-MENDES, R., JERUSALINSKY, L., FERRARI, S. F. Functional planning units for the management of an endangered Brazilian titi monkey. **American Journal of Primatology** v. 79, n. 5,

2017. Disponível em: <https://www.researchgate.net/publication/312021750_Functional_planning_units_for_the_management_of_an_endangered_Brazilian_titi_monkey>.
- GOUVEIA, S. F.; SOUZA-ALVES, J. P.; RATTIS, L.; DOBROVOLSKI, R.; JERUSALINSKY, L.; BELTRÃO-MENDES, R.; FERRARI, S. F. Climate and land use changes will degrade the configuration of the landscape for titi monkeys in eastern Brazil. **Global Change Biology** v. 22, p. 6, p. 2003–2012, 2016 Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/26663738>>.
- GRAHAM, T. L.; MATTHEWS, H. D.; TURNER, S. E. A global-scale evaluation of primate exposure and vulnerability to climate change. **International Journal of Primatology** v. 37, n. 2, p. 158–174, 2016. Disponível em: <<https://link.springer.com/article/10.1007/s10764-016-9890-4>>.
- ROBEL, J.; GRAUMANN, A. **NOAA KLM User's Guide with NOAA-N,-P Supplement**. 2014. Disponível em: <<https://www1.ncdc.noaa.gov/pub/data/satellite/publications/podguides/N-15%20thru%20N-19/pdf/0.0%20NOAA%20KLM%20Users%20Guide.pdf>>.
- GRENFELL, B. T. e DOBSON, A. P. **Ecology of Infectious Diseases in Natural Populations**. Cambridge: University Press. Publications of the Newton Institute, 1995.
- GUBLER, D. J.; REITER, P.; EBI, K. I.; YAP, W.; NASCI, R.; PATZ, J. A. Climate variability and change in the United States: Potential impacts on vector- and rodent-borne diseases. **Environ. Health Perspect.** v. 109, p. 223–233, 2001. Disponível em: <<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1240669/>>.
- GUERRA, A. J. T; DA CUNHA, S. B. (Orgs); **Geomorfologia uma utilização de bases e conceitos**, Rio de Janeiro: Bertrand Brasil, p. 93 – 199, 1994.
- HALL, A. M. **Correlation-based Feature Selection for Machine Learning**. 1999, 198 p. Tese (Doutorado em Filosofia) - Universidade de Waikato, Nova Zelândia. 1999,. Disponível em: <<https://www.cs.waikato.ac.nz/~mhall/thesis.pdf>>.
- HALL, M; FRANK, E. Combining naive bayes and decision tables. In: 2008 FLAIRS Conference - AAAI, 2008, Florida – USA. **Anais...** Disponível em: <<https://www.cs.waikato.ac.nz/~eibe/pubs/HallAndFrankFLAIRS08.pdf>>.
- HALL, M. A., HOLMES, G. Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. **IEEE Transactions on Knowledge and Data Engineering**. v. 15, n. 3, 2003. Disponível em: <<https://ieeexplore.ieee.org/abstract/document/1245283>>.
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I.

- H. The weka data mining software: An update. **ACM SIGKDD Explor Newslett**, n. 11, v. 1, pp. 10-18, 2009. Disponível em: <<https://dl.acm.org/citation.cfm?id=1656278>>.
- HAMMANN, E.; LAGUARDIA, J. Reflexões sobre a vigilância epidemiológica: mais além da notificação compulsória. **Inf Epidemiol SUS**. v. 9, n. 3, p. 211-19, 2000. Disponível em: <<http://scielo.iec.gov.br/pdf/iesus/v9n3/v9n3a06.pdf>>.
- HAMRICK, P. N.; ALDIGHERI, S.; MACHADO, G.; LEONEL, D. G.; VILCA, L. M.; URIONA, S. Geographic patterns and environmental factors associated with human yellow fever presence in the Americas. **PloS Negl Trop Dis**. v. 11, n. 9, p. 1-27. 2017. Disponível em: <https://doi.org/10.1371/journal.pntd.0005897>.
- HAN, J.; KAMBER, M. **Data Mining: Concepts and Techniques** (The Morgan Kaufmann Series in Data Management Systems). 2 ed. San Francisco-USA: Morgan Kaufmann, 2006.
- HANSKI, I.; ZURITA, G. A.; BELLOCQ, M. I.; RYBICKI, J. Species–fragmented area relationship. **Proceedings of the National Academy of Sciences of the United States of America** v. 110, n. 31, p. 12715–12720, 2013. Disponível em: <<https://www.pnas.org/content/110/31/12715>>.
- HANYA, G.; STEVENSON, P.; VAN NOORDWIJK, M.; TE WONG, S.; KANAMORI, T.; KUZE, N.; AIBA, S.I.; CHAPMAN, C. A. Seasonality in fruit availability affects frugivorous primate biomass and species richness. **Ecography** v. 34, p. 1009-1017, 2011. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1111/j.1600-0587.2010.06775.x>>.
- HATT J. M., GREEST P., POSTHAUS H. e BOSSART W. Serologic survey in a colony of captive common marmosets (*Callithrix jacchus*) after infection with Herpes simplex type1-like virus. **J. Zoo Wildl. Med.** v. 35, p. 387-390, 2004. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/15526895>>.
- HERNÁNDEZ-CAMACHO, J. e COOPER, R. W. The nonhuman primates of Colombia. Pp. 35-69. In: Thorington Jr., R.W. e Heltne, P.G. (eds.). **Neotropical Primates: Field Studies and Conservation**. Washington: National Academy of Sciences, 1976, 135p. Disponível em: <<https://www.nap.edu/catalog/18666/neotropical-primates-field-studies-and-conservation-proceedings-of-a-symposium>>.
- HEY, T.; TANSLEY, S.; TOLLE, K. M. **The fourth paradigm: data-intensive scientific discovery**, v. 1. Washington - USA: Microsoft research Redmond, 2009. Disponível em: <<https://www.immagic.com/eLibrary/ARCHIVES/EBOOKS/M091000H.pdf>>.
- HIJMANS, R. J.; CAMERON, S. E.; PARRA, J. L.; JONES, P. G.; JARVIS, A. Very high resolution interpolated climate surfaces for global land areas. **International Journal of Climatology**, v. 25, p. 1965–1978, 2005. Disponível em:

<<https://rnmets.onlinelibrary.wiley.com/doi/10.1002/joc.1276>>.

HILL, A. **The Encyclopedia of Operations Management: A Field Manual and Glossary of Operations Management Terms and Concepts**. FT Press., 2012.

HIRSCH, A. L. G.; DIAS, L. O.; MARTINS, R. F.; CAMPOS, E. C. LANDAU, N. A. T. R. BDGEOPRIM – Database of geo-referenced localities of neotropical primates. **Neotropical Primates**, v. 10, p. 79-84, 2002. Disponível em: <https://www.researchgate.net/publication/280311930_BDGEOPRIM_Database_of_geo-referenced_localities_of_Neotropical_primates>.

HIRZEL, A. H. e GUIAN, A. Which is the optimal sampling strategy for habitat suitability modelling? **Ecological Modelling** v. 157, p. 331-341, 2002. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S030438000200203X>>.

HOFFMANN, C.; ZIMMERMAN, F.; BIEK, R.; KUEHL, H.; NOWAK, K.; MUNDY, R.; AGBOR, A.; ANGEDAKIN, S.; ARANDJELOVIC, M.; BLANKENBURG, A.; BRAZOLLA, G.; COROGENES, K.; COUACY-HYMAN, E.; DESCHNER, T.; DIEGUEZ, P.; DIERKS, K.; DUÛX, A.; DUPKE, S.; ESHUIS, H.; FORMENTY, P.; YUH, Y.G.; GOEDMAKERS, A.; GOGARTEN, J. F.; GRANJON, A. C.; MCGRAW, S.; GRUNOW, R.; HART, J.; JONES, S.; JUNKER, J.; KIANG, J.; LANGERGRABER, K.; LAPUENTE, J.; LEE, K.; LEENDERTZ, S. A.; LÉGUILLON, F.; LEINERT, V.; LÖHRICH, T.; MARROCOLI, S.; MÄTZ-RENSING, K.; MEIER, A.; MERKEL, K.; METZGER, S.; MURAI, M.; NIEDORF, S.; NYS, H.; SACHSE, A.; SCHIJNDEL, J.; THIESEN, U.; TON, E.; WU, D.; WIELER, L. H.; BOESCH, C.; KLEE, S. R.; WITTIG, R. M.; CALVIGNAC-SPENCER, S.; LEENDERTZ, F. H. Persistent anthrax as a major driver of wildlife mortality in a tropical rainforest. **Nature** v. 548, n. 7665, p. 82–86, 2017. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S030438000200203X>>.

HONG Y. S.; LEE, H.; TAHK, M.-J. Acceleration of the convergence speed of evolutionary algorithms using multi-layer neural networks. **Engineering Optimization**, v. 35, n. 1, p. 91–102, 2003. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/0305215031000069672>>.

HONGYU, K.; SANDANIELO, V. L. M.; JUNIOR, G. J. O. Análise de Componentes Principais: resumo teórico, aplicação e interpretação. **E&S - Engineering and Science**. v.5, n. 1, 2016. Disponível em: <<http://periodicoscientificos.ufmt.br/ojs/index.php/eng/article/view/3398>>.

HYLANDER, K.; EHRLÉN, J. The mechanisms causing extinction debts. **Trends in Ecology and Evolution** v. 28, n. 6, p. 341–346, 2013. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S016953471300027X>>.

IBAMA - Instituto Brasileiro de Meio Ambiente e dos Recursos Naturais Renováveis. **O**

estado dos solos. 1º ed., 2006. Disponível em: <<http://www2.ibama.gov.br/~geobr/Livro/cap2/subsolos.pdf>>. Acesso em: 28 set. 2018.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Projeção da população, 2018.** Projeção da população do Brasil e das Unidades da Federação. Disponível em: <<https://www.ibge.gov.br/apps/populacao/projecao/index.html>>. Acesso em: 16 set. 2018.

IBGE - Instituto Brasileiro de Geografia e Estatística. **Base Cartográfica Contínua do Brasil – 1:1000000 - 2016.** 5ª versão digital. Disponível em: <ftp://geofp.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas/bcimftp://geofp.ibge.gov.br/cartas_e_mapas/bases_cartograficas_continuas/bcim/versao2016>. Acesso em: 16 out. 2018.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Atlas do censo demográfico 2010.** Rio de Janeiro: IBGE, 2013, 160 p. Disponível em: <<https://censo2010.ibge.gov.br/apps/atlas/>>

IBGE – Instituto Brasileiro de Geografia e Estatística. **CENSO DEMOGRÁFICO 2010. Características da população e dos domicílios: resultados do universo.** Rio de Janeiro, IBGE, 2011. 270 p. Disponível em: <https://ww2.ibge.gov.br/home/estatistica/populacao/censo2010/caracteristicas_da_populacao/default_caracteristicas_da_populacao.shtm>

IBGE – Instituto Brasileiro de Geografia e Estatística. **Atlas nacional do Brasil Milton Santos.** Rio de Janeiro: IBGE, 2010. 307 p. Disponível em: <<https://biblioteca.ibge.gov.br/index.php/biblioteca-catalogo?view=detalheseid=247603>>.

IBGE – Instituto Brasileiro de Geografia e Estatística / Coordenação de Recursos Naturais e Estudos Ambientais. **Manual técnico de geomorfologia.** 2. ed., Rio de Janeiro : IBGE, 2009. 182 p. Disponível em: <<https://biblioteca.ibge.gov.br/visualizacao/livros/liv66620.pdf>>.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Mapa de Biomas do Brasil, primeira aproximação.** Rio de Janeiro: IBGE, 2004. Disponível em: <www.ibge.gov.br>.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Mapa de Vegetação do Brasil.** Rio de Janeiro: IBGE, 2004. Disponível em: <www.ibge.gov.br>.

IBGE – Instituto Brasileiro de Geografia e Estatística. **Mapa de Clima do Brasil.** Rio de Janeiro, IBGE, 2002. Escala 1:5.000.000. Disponível em: <<https://www.ibge.gov.br/geociencias/cartas-e-mapas/informacoes-ambientais/15817-clima.html?=&t=o-que-e>>

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA/Departamento de Cartografia. **Noções básicas de cartografia.** Rio de Janeiro: IBGE, 1999. 130p.

- Disponível em: <https://biblioteca.ibge.gov.br/visualizacao/livros/liv8595_v1.pdf>.
- IBRAM – INSTITUTO BRASILEIRO DE MINERAÇÃO. **O portal da mineração**. 2018. Disponível em: <<http://portaldamineracao.com.br/quem-somos/>>. Acesso em: 13 set. 2018.
- ICMBIO. Fauna brasileira, 2014. **Espécies ameaçadas – Lista 2014 - Mamíferos**. Disponível em: <<http://www.icmbio.gov.br/portal/fauna-brasileira/lista-de-especies>>
- INEA – INSTITUTO ESTADUAL DO AMBIENTE. Fauna e recursos faunísticos do estado do Rio de Janeiro – Considerações. **Revista INEANA**. v.1, n.1, 2012. Rio de Janeiro: INEA, 2012. Disponível em: <http://www.inea.rj.gov.br/cs/groups/public/documents/document/zwff/mde0/~edisp/inea_014687.pdf>.
- INMET – Instituto Nacional de Meteorologia. Banco de Dados Meteorológicos para Ensino e Pesquisa. Anomalias de temperaturas médias - 2017. Disponível em: <<http://www.inmet.gov.br>>. Acesso em: 13 set. 2018.
- ISMAIL-ZADEH, A.; KOROTKII, A.; TSEPELEV, I. **Data-Driven Numerical Modelling in Geodynamics: Methods and Applications**. Springer Briefs in Earth Sciences, 2016. Disponível em: <<https://www.springer.com/kr/book/9783319278001>>.
- IUCN. **Red List of Threatened Species, Versão 2018.2**. Disponível em: <<https://www.iucnredlist.org/>>. Acesso em: 22 out. 2018.
- IWANAGA, S. e FERRARI, S. F. Geographic distribution of red howlers (*Alouatta seniculus*) in southwestern Brazilian Amazonia, with notes on *Alouatta caraya*. **International Journal of Primatology**, v. 23, n. 6, p. 1245-1256, 2002. Disponível em: <<https://link.springer.com/article/10.1023/A:1021174803915>>.
- IWASHITA, F. **Sensibilidade de modelos de distribuição de espécies à qualidade do posicionamento de dados de coleta**. 2007, 103 p. Dissertação (Mestrado em Sensoriamento Remoto). Pós-graduação em Sensoriamento Remoto, Instituto Nacional de Pesquisas Espaciais - INPE, São José dos Campos, 2007. Disponível em: <<http://mtc-m16b.sid.inpe.br/col/sid.inpe.br/mtc-m17@80/2007/06.13.12.04/doc/publicacao.pdf>>
- JANG, J. A. Adaptive-Network-Based Fuzzy Inference System. **IEEE Transactions on Systems Man and Cybernetics**, v. 23, n. 3, 1993. Disponível em: <<https://ieeexplore.ieee.org/document/256541>>.
- JANG, J.; SUN, E. e MIZUTANI, C. **Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence**. Prentice-Hall, 1 ed., 1997.
- JOACHIMS, T. Training Linear SVMs in Linear Time. In: Proceedings of the ACM

- SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006. Philadelphia, USA. **Anais...** 2006. p. 217-226. Disponível em: <https://www.cs.cornell.edu/people/tj/publications/joachims_06a.pdf>.
- JONES, K.; PATEL, N. G.; LEVY, M. A.; STOREYGARD, A.; BALK, D.; GITLEMAN, J. L.; DASZAK, P. Global trends in emerging infectious diseases. **Nature**, v. 451, p. 990-993, 2008. Disponível em: <<http://www.nature.com/nature/journal/v451/n7181/full/nature06536.html>>.
- JOPPA, L. N.; PFAFF, A. High and far: biases in the location of protected areas. **Plos One** v. 4, n. 12, 2009. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0008273>>.
- KALAMANDEEN, M.; GLOOR, E.; MITCHARD, E.; QUINCEY, D.; ZIV, G.; SPRACKLEN, D.; SPRACKLEN, B.; ADAMI, M.; ARAGÃO, L. E. O. C.; GALBRAITH, D. Pervasive rise of small-scale deforestation in Amazonia. **Scientific Reports** v. 8, n. 1, p. 1-10, 2018. Disponível em: <<https://www.nature.com/articles/s41598-018-19358-2>>.
- KAREGOWDA, A.; JAYARAM, M. A; MANJUNATH, A. S. Feature Subset Selection Problem using Wrapper Approach in Supervised Learning. **International Journal of Computer Applications**. v. 1, p. 169-295, 2010. Disponível em: <https://www.researchgate.net/publication/43656124_Feature_Subset_Selection_Problem_using_Wrapper_Approach_in_Supervised_Learning/download>.
- KARSTENSEN, J.; PETERS, G. P.; ANDREW, R. M. Attribution of CO2 emissions from Brazilian deforestation to consumers between 1990 and 2010. **Environmental Research Letters** v. 8, n 2, 2013. Disponível em: <<https://iopscience.iop.org/article/10.1088/1748-9326/8/2/024005>>.
- KASABOV, N. K. **Foundations of Neural Networks, Fuzzy Systems, and Knowledge Engineering**. Cambridge, MA, USA: MIT Press, 1996.
- KAY, R. F.; MADDEN, R. H.; VAN SCHAIK, C.; HIGDON, D. Primate species richness is determined by plant productivity: Implications for conservation. **Proceedings of the National Academy of Sciences USA** v. 94, p. 13023-13027, 1997. Disponível em: <<https://www.pnas.org/content/94/24/13023>>.
- KEESING, F.; HOLT, R. D; OSTFELD, R. S. Effects of species diversity on disease risk. **Ecology Letters**, v. 9, p. 485-495, 2006. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/16623733>>.
- KESSLER, R. C.; ROSE, S.; KOENEN, K. C.; KARAM, E. G.; STANG, P. E.; STEIN, D. J.; HEERINGA, S. G.; HILL, E. D.; LIBERZON, I.; MCLAUGHLIN, K. A.; MCLEAN, S. A.; PENNELL, B. E.; PETUKHOVA, M.; ROSELLINI, A. J.; RUSCIO, A. M.;

- SHAHLY, V.; SHALEV, A. Y.; SILOVE, D.; ZASLAVSKY, A. M.; ANGERMEYER, M. C.; BROMET, E. J.; DE ALMEIDA, J. M. C.; DE GIROLAMO, G.; DE JONGE, P.; DEMYTTENAERE, K.; FLORESCU, S. E.; GUREJE, O.; HARO, J. M.; HINKOV, H.; KAWAKAMI, N.; KOVACS-MASFETY, V.; LEE, S.; MEDINA-MORA, M. E.; MURPHY, S. D.; NAVARRO-MATEU, F.; PIAZZA, M.; POSADA-VILLA, J.; SCOTT, K.; TORRES, Y.; CARMEN VIANA, M. How well can post-traumatic stress disorder be predicted from pre-trauma risk factors? **World Mental Health Surveys**. v.13, p. 265–274, 2014. Disponível em: <https://www.researchgate.net/publication/266553987_How_well_can_post-traumatic_stress_disorder_be_predicted_from_pretrauma_risk_factors_An_exploratory_study_in_the_WHO_World_Mental_Health_Surveys_World_Psychiatry_13_265-274>.
- KHU, S. T.; SAVIC, D.; KAPELAN, Z. Evolutionary-based Meta-modelling: The relevance of using approximate models in hidroinformatics. **Practical Hydroinformatics**. Berlim: Springer-Verlag, 2008, 275 p. Disponível em: <https://www.researchgate.net/publication/226796771_Evolutionary-based_Meta-modelling_The_Relevance_of_Using_Approximate_Models_in_Hydroinformatics>.
- KINZEY, W. G. The titi monkeys, genus *Callicebus*. p. 241-276. In: COIMBRA-FILHO, A. F. e MITTERMEIER, R. A. (eds). **Ecology and Behavior of Neotropical Primates**. Academia Brasileira de Ciências. 1981, v. 1, 496p.
- KOHAVI, R.; QUINLAN, J. R. **Data mining tasks and methods: Classification: decision-tree discovery, Handbook of data mining and knowledge Discovery**. New York : Oxford University Press, 2002.
- KOZA, J. R. **Genetic programming: On the programming of computers by natural selection**. Cambridge: MIT Press, 1992. Disponível em: <<https://doc.lagout.org/science/Artificial%20Intelligence/Evolutionary%20computation/Genetic%20programming%20Complex%20adaptive%20systems%20-%20Koza%20J.R..pdf>>.
- KREMPSE, E. **Evolução diferencial para a otimização restrita**. 2009, 115p. Dissertação (Mestrado em Modelagem Computacional), Programa de Pós-graduação em Modelagem Computacional, Laboratório Nacional de Computação Científica – LNCC, Petrópolis, 2009. Disponível em: <http://bdtd.ibict.br/vufind/Record/LNCC_f2e28bf13638b10054df5bd0ea73ef3a/Details>.
- KREMPSE, E. **Uso de metamodelos na Evolução Diferencial para problemas envolvendo simulações de alto custo computacional**. 2014. 211 p. Tese (Doutorado em Modelagem Computacional) – Programa de Pós-graduação em Modelagem Computacional, Laboratório Nacional de Computação Científica – LNCC. Petrópolis, 2014. Disponível em: <http://bdtd.ibict.br/vufind/Record/LNCC_8dace5e664d31d1ff6062e8bf8d883bb>.

- KREMPSER, E.; BARBOSA, H. J. C. Um estudo da evolução diferencial na otimização estrutural. In: Congresso Nacional de Matemática Aplicada e Computacional, 2009, Cuiabá – MT, **Anais...** 2009.
- KRITICOS, D. J.; WEBBER, B. L.; LERICHE, A.; OTA, N.; MACADAM, I.; BATHOLS, J.; SCOTT, J.K. CliMond: global high resolution historical and future scenario climate surfaces for bioclimatic modelling. **Methods in Ecology and Evolution**. v. 3, p. 53-64, 2012. Disponível em: <<https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/j.2041-210X.2011.00134.x>>.
- KUTZ, J. N. **Data-driven Modeling e Scientific Computation**. Methods for Complex Systems and Big Data. Inglaterra: Oxford University Press, 2013.
- LACERDA A. Em busca de um tempo perdido: fontes visuais para a história da vacina contra a febre amarela. In: Benchimol, J. (ed). **Febre amarela: a doença e a vacina, uma história inacabada**. Rio de Janeiro: FIOCRUZ, 2001. 470 p.
- LACHER JÚNIOR, T. E.; FONSECA, G. A. B.; ALVES J. R. C.; MAGALHÃES-CASTRO, B. Parasitism of trees by marmosets in a central Brazilian gallery forest. **Biotropica**, n. 16, v. 3, p. 202-209, 1984. Disponível em: <https://www.jstor.org/stable/2388053?seq=4#metadata_info_tab_contents>.
- LANDIS, J. R.; KOCH, G. G. The measurement of observer agreement for categorical data. **Biometrics**. v. 33, n. 1, p. 159-174, 1977. Disponível em: <https://www.dentalage.co.uk/wp-content/uploads/2014/09/landis_jr_koch_gg_1977_kappa_and_observer_agreement.pdf>.
- LANEY, D. 3D Data Management: Controlling Data Volume, Velocity and Variety. Application Delivery Strategies, v. 949, p. 1-4. 2001. Disponível em: <<https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>>.
- LAROSE, D. T. **Discovering Knowledge in Data: An Introduction to Data Mining**. John Wiley and Sons, Inc, 2005.
- LATRUBESSE, E. M.; ARIMA, E. Y.; DUNNE, T.; PARK, E.; BAKER, V. R.; D'HORTA, F. M.; WIGHT, C.; WITTMANN, F.; ZUANON, J.; BAKER, P. A.; RIBAS, C. C.; NORGAARD, R. B.; FILIZOLA, N.; ANSAR, A.; FLYVBJERG, B.; STEVAUX, J. C. Damming the rivers of the Amazon basin. **Nature** v. 546, n. 7658, p. 363–369, 2017. Disponível em: <<https://www.nature.com/articles/nature22333>>.
- LAURANCE, W. F.; SLOAN, S.; WENG, L.; SAYER, J. A. Estimating the environmental costs of África's massive “development corridors”. **Current Biology** v. 25, n. 24, p.

- 3202–3208, 2015. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/26628009>>.
- LAURANCE, W. F.; SAYER, J.; CASSMAN, K. G. Agricultural expansion and its impacts on tropical nature. **Trends in Ecology e Evolution** v. 29, n. 2, p. 107–116, 2014. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0169534713002929>>.
- LAZARO-PEREA, C.; SNOWDON, C. T.; ARRUDA, M. F. Scent-marking behavior in wild groups of common marmosets (*Callithrix jacchus*). **Behavioral Ecology and Sociobiology, Göttingen**, v. 46, p. 313–324, 1999. Disponível em: <<https://link.springer.com/article/10.1007/s002650050625>>.
- LÁZARO, N. I. L. **Cuidado parental e relações sociais entre fêmeas adultas e seus filhotes de bugios-pretos-e-dourados (*Alouatta caraya*; Primates, Atelidae) em fragmentos urbanos de Cerrado-Pantanal em Aquidauana, Mato Grosso do Sul.** 2012, 63p. Dissertação (Mestrado em Biologia Animal). Universidade Federal de Mato Grosso do Sul - UFMS, Campo Grande, 2012. Disponível em: <<http://repositorio.cbc.ufms.br:8080/jspui/handle/123456789/1996?mode=full>>.
- LEE, H. D. **Seleção de atributos importantes para a extração de conhecimento de base de dados.** 2005, 182 p. Tese (Doutorado), Instituto de Ciências Matemáticas e de Computação – ICMS, Universidade de São Paulo - USP. São Carlos, 2005. Disponível em: <http://www.teses.usp.br/teses/disponiveis/55/55134/tde-22022006-172219/publico/tese_huei.pdf>.
- LEFAUX, B.; DUPREZ, R.; TANGUY, M.; LONGEART, L.; GESSAIN, A.; BOULANGER E. Letter to the editor. **Vet. Pathol.** v. 41, p.302–303, 2004.
- LE MOS, A.; MAIA, R.; INÁCIO, M.; CAMINHAS, W. Uma metodologia para geração automática de redes neurais nebulosas a partir de árvores de decisão. In: IX Congresso Brasileiro de Redes Neurais/Inteligência Computacional, 2009, Ouro Preto. **Anais...** Ouro Preto, 2009.. Disponível em: <http://abricom.org.br/eventos/cbrn_2009/181_CBRN2009/>.
- LEVY, B.; PATZ, J. **Climate Change and Public Health.** Inglaterra: Oxford University Press, 2015. 368 p.
- LEWIS, S. L.; EDWARDS, D. P.; GALBRAITH, D. Increasing human dominance of tropical forests. **Science** v. 349, n. 6250, p. 827–832, 2015. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/26293955>>.
- LI, B.; CHEN, Y. W.; CHEN, Q. The nearest neighbor algorithm of local probability centers. **IEEE Trans. Syst. Man Cybernet.** v.38, n. 1, p.141–154. 2008. Disponível em: <<https://ieeexplore.ieee.org/document/4389962>>.

- LIMA, M. I. C. **Projeto RADAM: Uma Saga Amazônica**. Paka-Tatu, Belém, Pará, Brazil. 132 pp, 2008.
- LIU, Y.; SUN, F. A fast differential evolution algorithm using k-nearest neighbour predictor. **Expert Syst Appl** v. 38, n. 4, p.4254–4258, 2011. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0957417410010493>>.
- LODOLI, B. C.; MAGNANI, H. M.; PAIVA, R. P. **Ferramenta da geração de redes neurais para classificação de criptogramas**. 2016, 60 p. Monografia (Graduação em Engenharia de Computação) - Instituto Militar de Engenharia – IME. Rio de Janeiro, 2016. Disponível em: <<http://www.comp.ime.eb.br/graduacao/pfc/repositorio-pfc/2016/2016-Lodoli.pdf>>.
- LOUKIDES, M. **What is data science?** Sebastopol, CA: O'Reilly. 2011.
- LOVEJOY, T. E. Global change and epidemiology: nasty synergies. In: **Emerging víruses** Morse, S. E. (ed.), New York/Oxford: Oxford University Press, 1993, p. 281-268.
- LUBER, G.; LEMERY, J. **Global Climate Change and Human Health: From Science to Practice**. San Francisco. Jossey-Bass, 2015. 672p.
- LUDWIG, G. **Área de vida e uso do espaço por Alouatta caraya (Humboldt, 1812) em ilha e continente na região de Porto Rico, Paraná**. 2006, 88 p. Dissertação (Mestrado em Zoologia) - Universidade Federal do Paraná - UFPR. Paraná, 2006. Disponível em: <https://www.researchgate.net/publication/26977247_Area_de_vida_e_uso_do_espaco_p_or_Alouatta_caraya_Humboldt_1812_em_ilha_e_continente_do_Alto_Rio_Paraná>.
- MALHI, Y.; ROBERTS, J. T.; BETTS, R. A.; KILLEEN, T. J.; LI, W.; NOBRE, C. A. Climate change, deforestation, and the fate of the Amazon. **Science** v. 319, n. 5860, p. 169–172, 2008. Disponível em: <<http://science.sciencemag.org/content/319/5860/169>>.
- MALHOTRA, R.; NARINDER, S.; SINGH, Y. Genetic Algorithms: Concepts, Design for Optimization of Process Controllers. **Computer and Information Science**. v. 4, n. 2, 2011. Disponível em: <<http://www.ccsenet.org/journal/index.php/cis/article/view/9804>>.
- MAN, K. F.; TANG, K. S.; KWONG, S. **Genetic Algorithms**. Springer, 2001, 344p.
- MAPBIOMAS. 2018. **Coleção 2.3 da Série Anual de Mapas de Cobertura e Uso do Solo do Brasil**. Disponível em: <mapbiomas.org>. Acesso em: 12 set. 2018.
- MARSLAND, S. **Machine learning: an algorithmic perspective**. Boca Raton: CRC Press, 2009.
- MASCIA, M. B.; PAILLER, S.; KRITHIVASAN, R.; ROSHCHANKA, V.; BURNS, D.; MLOTHA, M. J.; MURRAY, D. R.; PENG, N. Protected area downgrading, downsizing,

- and degazettement (PADDD) in África, Asia, and Latin America and the Caribbean, 1900–2010. **Biological Conservation** v. 169, p. 355–361, 2014. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S000632071300400X>>.
- MATZ-RESING K.; JENTSCH, K. D.; RENSING, S.; LANGENHUYZEN S.; VERSCHOOR, E.; NIPHUIS H.; KAUP F. J. Fatal *Herpes simplex* infection in a group of common marmosets (*Callithrix jacchus*). **Vet. Pathol.** v. 40, p. 405-411, 2003. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/12824512>>
- McCARTHY, J. **What is artificial intelligence?** Standard University, 2007. Disponível em: <<http://www.formal.stanford.edu/jmc/whatisai.html>>.
- MEDRONHO, R. A.; BLOCH, K. V.; LUIZ, R. R.; WERNECK, G. L. **Epidemiologia**. 2 ed. São Paulo: Atheneu. 2008. 790 p.
- MELO, F. R. **Primatas e áreas prioritárias para a conservação da biodiversidade no vale do rio Jequitinhonha, Minas Gerais**. 2004, 154p. Tese (Doutorado em Ecologia, Conservação e Manejo de Vida Silvestre) - Universidade Federal de Minas Gerais – UFMG. Belo Horizonte, 2004. Disponível em: http://pos.icb.ufmg.br/pgecologia/teses/T22_Fabiano_Rodrigues_de_Melo.pdf>.
- MELO, F. R.; TERRIBILE, L. C.; OLIVEIRA, G.; LIMA-RIBEIRO, M. S.; STRIER, K. B. The efficiency of Atlantic forest conservation areas to preserve the Northern Muriquis (*Brachyteles hypoxanthus*) in a changing climate. In: XXVIth Congress of the International Primatological Society and 39th meeting of the American Society of Primatologists, 2016, Chicago, IL, USA. **Anais...** ASP/IPS, 2016, Chicago Scientific Program. Chicago, USA: IPS and ASP, 2016. v. 1, 400 p.
- MENDES, S. L. Estudo ecológico de *Alouatta fusca* (Primates, Cebidae) na estação ecológica de Caratinga, MG. **Revista Nordestina de Biologia**, v. 6, n. 2, p. 71-104, 1989. Disponível em: <<http://www.periodicos.ufpb.br/ojs/index.php/revnebio/article/view/16760>>.
- MENDES, S. L.; RYLANDS, A.B.; KEIRULFF, M.C.M. e DE OLIVEIRA, M.M. **Alouatta guariba - 2008**. In: IUCN Red List of Threatened Species, versão 2010.2. Disponível em www.iucnredlist.org. Acesso em: 14 set. 2018.
- MEYER, A. L. S.; PIE, M. R.; PASSOS, F. C. Assessing the exposure of lion tamarins (*Leontopithecus* spp.) to future climate change. **American Journal of Primatology** v. 76, n. 6, p. 551–562, 2014. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/24346860>>.
- MERSCHMANN, L. H. C. PCC142/BCC444 – **Mineração de dados, avaliação de classificadores**. 2018. Departamento de Computação – Universidade Federal de Ouro Preto. 24 slides. Disponível em:

<http://www.decom.ufop.br/luiz/site_media/uploads/arquivos/bcc444_pcc142/slides09_md.pdf>.

MITCHELL, T. M. **Machine learning**. Nova Iorque: McGraw-Hill, 1997.

MMA – Ministério do Meio Ambiente. **Cadastro Nacional de Unidades de Conservação, 2018.** Disponível em: <http://www.mma.gov.br/images/arquivo/80229/CNUC_JUL18%20-%20C_Bio.pdf>. Acesso em: 14 set. 2018.

MMA – Ministério do Meio Ambiente. **Sistema Nacional de Unidades de Conservação da natureza, 2014.** Disponível em: <http://www.mma.gov.br/estruturas/240/Publicacao/240_Publicacao05072011052536.pdf> Acesso em: 14 set. 2018.

MONATH T. P.; VASCONCELOS, P. F. C. Yellow fever. **J Clin Virol** v. 64, p. 160-173, 2015. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/25453327>>.

MONATH, T. P.; CETRON, M. S.; TEUWEN, D. E. Yellow fever vaccine. In: Plotkin SA, Orenstein WA, Offit PA, editors. **Vaccines**, 5° ed. Philadelphia: **Elsevier**, 2008, p. 959-1055.

MONATH, T. P. Epidemiology of yellow fever: current status and speculations on future trends. In: Saluzzo JF, Dodet B, editors. **Factors in the emergence of arbovirus diseases**. Paris: Elsevier, 1997, p.143-156.

MONDET, B. Considérations sur l'épidémiologie de la fièvre jaune au Brésil. **Bull Soc Pathol Exot.**, v. 94, p. 260-267, 2001. Disponível em: <http://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers12-10/010027819.pdf>.

MONICO, J. F. G.; PÓZ, A. P. D.; GALO, M.; DOS SANTOS, M. C.; DE OLIVEIRA, L. C. Acurácia e Precisão: Revendo os conceitos de forma acurada. **Bol. Ciência Geod.**, v. 15, n.3, p. 469-483, 2009. Disponível em: <<https://revistas.ufpr.br/bcg/article/view/15513/10363>>.

MOURA, A. C. Primate group Size and abundance in the caatinga dry forest, northeastern Brazil. **International Journal Primatology**, v. 28, p. 1279-1297, 2007. Disponível em: <https://www.researchgate.net/publication/225243969_Primate_Group_Size_and_Abundance_in_the_Caatinga_Dry_Forest_Northeastern_Brazil>.

MTPA – Ministério dos Transportes, Portos e Aviação Civil. **Síntese – Setor Ferroviário. 2018.** Disponível em: <<http://www.transportes.gov.br/grandes-números/88-dados-de-transportes/5380-sintese-ferroviario.html>> . Acesso em: 14 set. 2018.

MUHLE, C. B. **Estratégias adaptativas de um grupo de bugios-pretos, Alouatta caraya**

MUÑOZ, M. E. S.; GIOVANNI, R.; SIQUEIRA, M.F.; SUTTON, T.; BREWER, P.; PEREIRA, R. S.; CANHOS, D. A. L.; CANHOS, V. P. "openModeller: a generic approach to species' potential distribution modelling". **GeoInformatica**, v. 15, n. 1, p. 111-135, 2011. Disponível em: <<https://link.springer.com/article/10.1007%2Fs10707-009-0090-7>>.

NEWMARK, W. D.; JENKINS, C. N.; PIMM, S. L.; MCNEALLY, P. B.; HALLEY, J. M. Targeted habitat restoration can reduce extinction rates in fragmented forests. **Proceedings of the National Academy of Sciences of the United States of America** v. 114, n. 36, p. 9635–9640, 2017. Disponível em: <<https://www.pnas.org/content/114/36/9635>>.

NIMER, E. Um modelo metodológico da classificação de climas. **Revista Brasileira de Geografia - IBGE**, Rio de Janeiro, n. 4, v. 41, p. 59-89, 1979.

NOBLE, W. S. What is support vector machine? **Nature Biotechnology**, n. 24, pp. 1565-1567, 2006. Disponível em: <<https://www.nature.com/articles/nbt1206-1565>>.

NOWAK, R. M. **Primates of the world**. Baltimore: John Hopkins University Press, 1999, 224p.

NUÑEZ-ITURRI, G.; OLSSON, O.; HOWE, H. F. Hunting reduces recruitment of primate-dispersed trees in Amazonian Peru. **Biological Conservation** v. 141, n. 6, p. 1536–1546, 2008. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0006320708001250>>.

NUNN, C. L.; ALTIZER, S. **Infectious Diseases in Primates: Behavior, Ecology and Evolution**. Oxford: Oxford University Press, 2006.

NUNN, C. L.; GILLESPIE, T. R. Infectious disease and primate conservation. In: Wich SA, Marshall AJ, eds. **An Introduction to Primate Conservation**. Oxford: Oxford University Press, 2016, p. 157–173. Disponível em: <<https://people.nunn-lab.org/charlie/PDFs/123-Nunn-Gillespie-2016.pdf>>.

OLIVEIRA-FILHO, A. T.; RATTER, J. A. A study of the origin of central Brazilian forests by the analysis of plant species distribution patterns. **Edinburgh Journal of Botany**, v. 52, n. 2, p. 141-194, 1995. Disponível em: <https://www.researchgate.net/publication/231823072_A_study_of_the_origin_of_central_Brazilian_forests_by_the_analysis_of_plant_species_distribution_patterns>.

- OLIVEIRA, D. G. R. **Primatas do Cerrado: Conservação, biogeografia e mudanças climáticas**. 2015, 242 p. Tese (Doutorado em Ecologia) - Programa de Pós-Graduação em Ecologia. Universidade de Brasília - UNB. Brasília, 2015. Disponível em: <<http://repositorio.unb.br/handle/10482/19434>>.
- OLIVEIRA, L. C.; GRELLE, C. E. V. Introduced primate species of an Atlantic Forest region in Brazil: present and future implications for the native fauna. **Tropical Conservation Science**, v. 5, n. 1, p. 112–120, 2012. Disponível em: <<https://journals.sagepub.com/doi/abs/10.1177/194008291200500110>>.
- OLIVEIRA, R. A.; REGO, W. A.; FERREIRA, S. R. M.; MOTA, J. M. F. Manifestações patológicas em revestimentos de argamassas com saibro. X Congresso Latinoamericano de Patología y XIII Congreso de Calidad em la construcción – CONPAT, Valparaíso – Chile, 2009. **Anais...** 2009.
- OLIVEIRA, S. V. **Análise preditiva da distribuição geográfica de hantavírus no Brasil**. 2013. 129 p. Dissertação (Mestrado em Medicina Tropical) Universidade de Brasília, Brasília – UNB. Brasília, 2013. Disponível em: http://repositorio.unb.br/bitstream/10482/14159/1/2013_StefanVilgesOliveira.pdf
- OSTFELD, R. S.; KEESING, F. e EVINER, V. T. (Eds). **Infectious disease ecology – Effects of ecosystems on disease and disease on ecosystems**. Princeton Univ. Press. 506p, 2008.
- OSTFELD, R. S. e KEESING, F. The function of biodiversity in the ecology of vector-borne zoonotic diseases. **Canadian Journal of Zoology** v. 78, p. 2061–2078, 2000. Disponível em: <https://www.caryinstitute.org/sites/default/files/public/reprints/Ostfeld_and_Keesing_2000_Can_J_Zool_78_2061-2078.pdf>.
- PACHECO L. F.; SIMONETTI, J. A. Genetic structure of a mimosoid tree deprived of its seed disperser, the spider monkey. **Conservation Biology** v. 14, n. 6, p. 1766–1775, 2000. Disponível em: <https://www.researchgate.net/publication/227680568_Genetic_Structure_of_a_Mimosoid_Tree_Deprived_of_Its_Seed_Disperser_the_Spider_Monkey>.
- PATZ, J. A.; KOVATS, R. S. Hotspots in climate change and human health. **BMJ** v. 9, n. 25, p. 1094–8, 2002. Disponível em: <<https://www.bmj.com/content/325/7372/1094.1.long>>.
- PERES, C. A. The fire next time. The potential for a catastrophic blaze threatens the Amazon. **Time Magazine**, p. 1-48, 2001.
- PERES, C. A.; EMILIO, T. J.; SCHIETTI, S.; DESMOULIÈRE, J. M.; LEVI, T. Dispersal limitation induces long-term biomass collapse in overhunted Amazonian forests. **Proceedings of the National Academy of Sciences of the United States of America** v. 113, n. 4, p. 892–897, 2016. Disponível em: <<https://www.pnas.org/content/113/4/892>>.

- PETERSON, A. T.; SAMY, A. Geographic potential of disease caused by Ebola and Marburg víruses in África. **Acta Tropica**. v. 162, p. 114-124, 2016. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0001706X16303795>.
- PETERSON, A. T.; SOBERÓN, J.; PEARSON, R.; ANDERSON, R.; MARTINÉS-MEYER, E.; NAKAMURA, M.; ARAÚJO, M. B. **Ecological Niches and Geographic Distributions**. New Jersey: Princeton University Press, 2011, p. 32-34. Disponível em: <https://www.researchgate.net/profile/Andrew_Peterson10/Publication/230709994_Ecological_Niches_and_Geographic_Distributions/links/5860be4408ae8fce4903bc00.pdf>.
- PETERSON, A. T.; TALUKDAR, G. Preliminary risk maps for transmission of kyanasanur forest disease in Southern India. **Indian Journal of Public Health**. v. 61, p. 47-50, 2017. Disponível em: <<http://www.ijph.in/article.asp?issn=0019-557X;year=2017;volume=61;issue=1;spage=47;epage=50;aualast=Peterson>>.
- PHILLIPS, S. J.; ANDERSON, R. P. e SCHAPIRE, R. E. Maximum entropy modeling of species geographic distributions. **Ecological Modelling** v. 190, p. 231-259, 2006. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S030438000500267X>>.
- PINA, A. L. C. B.; SOUZA, L. L. S.; FERRARI, S. F. Spacing patterns of *Alouatta belzebul* groups. In: Lisboa, P.L.B. (ed.). **Caxiuanã: Populações Tradicionais, Meio Físico e Diversidade Biológica**. Museu Paraense Emílio Goeldi, 2002, p. 685-695.
- PINKERTON, K. E.; ROM, W. N. **Global Climate Change and Public Health**. New York: Springer Science e Business Media, 2013. 406p. ISBN 9781461484172.
- PLUMPTRE, A. J.; NIXON, S.; CRITCHLOW, R.; VIEILLEDENT, G.; NISHULI, R.; KIRKBY, A.; WILLIAMSON, E. A.; HALL, J. S.; KUJIRAKWINJA, D. **Status of Grauer's Gorilla and Chimpanzees in Eastern Democratic Republic of Congo: Historical and Current Distribution and Abundance**. New York: Wildlife Conservation Society, Fauna e Flora International and Institut Congolais pour la Conservation de la Nature, 2015, p. 1-25. Disponível em: <http://fscdn.wcs.org/2016/04/04/inbumeq9_Status_of_Grauers_gorilla_and_eastern_chimpanzee_Report_Final.pdf>.
- PORTO, M. Marcos técnicos e legais para a descentralização do controle de endemias. In: Saúde MdSFNd, editor. **Descentralização do controle de endemias**. Brasília: Ministério da Saúde, 1994. p. 35- 9.
- POTAPOV, P.; HANSEN, M. C.; YAROSHENKO, A.; MINNEMEYER, S.; THIES, C.; ESIPOVA, E.; LAESTADIUS, L.; SMITH, W.; TURUBANOVA, S.; ZHURAVLEVA, I.; KOMAROVA, A. The last frontiers of wilderness: tracking loss of intact forest landscapes from 2000 to 2013. **Science Advances**. v. 3, n. 1, 2017. Disponível em:

<<http://advances.sciencemag.org/content/3/1/e1600821>>.

POULIN, R.; FORBES, M. Meta-analysis and research on host-parasite interactions: past and future. **Evol. Ecol.**, v. 26, p. 1169-1185, 2012. Disponível em: <<https://link.springer.com/article/10.1007/s10682-011-9544-0>>.

POWERS, D. M. W. **Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation**. Austrália: School of Informatics and Engineering, Flinders University, 2011. 24 p. (Relatório Técnico - SIE-07-001). Disponível em: <<https://csem.flinders.edu.au/research/techreps/SIE07001.pdf>>.

PRATES, H. M. **Vivendo no limite? Ecologia e comportamento de um grupo de bugios-pretos (*Alouatta caraya*) habitante de um pomar em Alegrete, RS**. 2007, 95 p. Dissertação (Mestrado em Zoologia). Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 2007. 81p. Disponível em: <<http://tede2.pucrs.br/tede2/handle/tede/298>>.

PRINTES, R. C.; JERUSALINSKY, L.; SOUSA, M. C.; RODRIGUES, L. R. R.; HIRSCH, A. Zoogeography, genetic variation and conservation of the *Callicebus personatus* group. In: Veiga, L.M.; Barnett, A.; Ferrari, S.F. e Norconk, M. (eds.). **Evolutionary biology and conservation of titis, sakis, and uacaris**. Cambridge University Press, 2013, p. 43-50. Disponível em: <https://www.researchgate.net/publication/280780056_Zoogeography_Genetic_Variation_and_Conservation_of_the_Callicebus_personatus_Group>.

QGIS Development Team. **Geographic Information System User Guide**. Open Source Geospatial Foundation Project. 2019. QGIS versão 2.18 Documentação eletrônica: <https://docs.qgis.org/2.18/pt_PT/docs/user_manual/index.html>.

QIAO, H.; PETERSON, A. T.; CAMPBELL, L. P.; SOBERÓN, J.; JI, L.; ESCOBAR, L. E. NicheA: Creating virtual species and ecological niches in multivariate environmental scenarios. **Ecography**. n. 39, p. 805-813, 2016. Disponível em: <<https://onlinelibrary.wiley.com/doi/pdf/10.1111/ecog.01961>>.

QUARESMA, L. F. Minério de Alumínio. Relatório Técnico 22. **Perfil da Mineração de Bauxita**. 2009a. Disponível em: <http://www.mme.gov.br/documents/1138775/1256650/P11_RT22_Perfil_da_Minerao_de_Bauxita.pdf/1713eb90-cbf9-42e5-a502-18abf47d9a1f>.

QUARESMA, L.F. Minério de Ferro. Relatório Técnico 18. **Perfil da Mineração de Ferro**. 2009b. Disponível em: <http://www.mme.gov.br/documents/1138775/1256650/P09_RT18_Perfil_da_Minerao_de_Ferro.pdf/e9ef3b57-cdbd-4166-b39d-f0f26f28e3ca>.

QUARESMA, L.F. Minério de Manganês. Relatório Técnico 19. **Perfil da mineração de**

- manganês. 2009c. Disponível em: <http://www.mme.gov.br/documents/1138775/1256650/P11_RT19_Perfil_da_mineraxo_de_manganxs.pdf/190c1abf-3223-4c19-9dfb-b267d06742c6>
- QUEIROZ, G. R.; FERREIRA, K. R. **Tutorial sobre Bancos de Dados Geográficos**. São José dos Campos: Instituto Nacional de Pesquisas Espaciais - INPE, 2006. Disponível em: <http://www.dpi.inpe.br/TutorialBdGeo_GeoBrasil2006.pdf>.
- QUINLAN, J. R. Improved use of continuous attributes in c4.5. **Journal of Artificial Intelligence Research**, v. 4, p. 77-90, 1996.
- QUINLAN, J. R. **C4.5: Programs for Machine Learning**. San Mateo, California: Morgan Kaufmann, 1993.
- R Development Core Team (2019). **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: <<http://www.R-project.org/>>.
- RAO, S. S. **Engineering optimization**. New York: Wiley, 1996. Disponível em: <<https://onlinelibrary.wiley.com/doi/book/10.1002/9780470549124>>.
- RASHEED, K. Informed operators: Speeding up genetic-algorithm-based design optimization using reduced models. In: Proceedings of genetic and Evolutionary Computation Conference, 2000, Las Vegas. **Anais...** Las Vegas: Morgan Kaufmann, 2000, p. 628–635. Disponível em: <https://www.researchgate.net/publication/2631230_Informed_operators_Speeding_up_genetic-algorithm-based_design_optimization_using_reduced_models>.
- REDFORD, K. H.; FONSECA, G. A. B. The role of Gallery Forests in the Zoogeography of the Cerrado's non-volant mammalian fauna. **Biotropica** v. 18, p. 126-135, 1986. Disponível em: <https://www.jstor.org/stable/2388755?seq=1#metadata_info_tab_contents>.
- REIS, N. R.; PERACCHI, A. L.; ANDRADE, F. R. **Primatas brasileiros**. Universidade Estadual de Londrina - UEL. Relatório Técnico. Londrina, 2008.
- RENTAS – Rede Nacional Contra o Tráfico de Animais Silvestres. **1º Relatório Nacional sobre o Tráfico de Fauna Silvestre**. 2004. Rio de Janeiro. Disponível em: <http://www.rentas.org.br/wp-content/uploads/2014/02/REL_RENTAS_pt_final.pdf>
- REY, L. **Parasitologia**. 4ª edição. Rio de Janeiro: Editora Guanabara Koogan, 2008.
- RIBEIRO, M. C.; METZGER, J. P.; MARTENSEN, A. C.; PONZONI, F. J.; HIROTA, M. M. The Brazilian Atlantic Forest: how much is left, and how is the remaining forest distributed? Implications for conservation. **Biol Conserv** v. 142, p. 1141–1153, 2009.

- Disponível em: <https://www.sciencedirect.com/science/article/pii/S0006320709000974>.
- RÍMOLI, J.; NANTES, R. DOS S. e LÁZARO-JÚNIOR, A. Diet and activity patterns of black howler monkeys *Alouatta caraya* (Humboldt, 1812, Primates, Atelidae) in ecotone Cerrado-Pantanal in left bank of Aquidauana river, Mato Grosso do Sul, Brazil. **Oecologia Australis**, v. 16, n. 4, p. 836-851, 2012. Disponível em: <https://revistas.ufrrj.br/index.php/oa/article/view/8247/0>.
- ROGERS, S.; GIROLAMI, M. **A first course in machine learning**. Boca Raton: CRC Press, 2011, 305 p. Disponível em: https://www.researchgate.net/publication/261804084_A_First_Course_in_Machine_Learning_by_Simon_Rogers_and_Mark_Girolami.
- ROMANO, A. P. M. R. **A vigilância de primatas não humanos e a febre amarela silvestre no Brasil**. 2012. 123 p. Tese (Doutorado em Ciência Animal) – Escola de Veterinária e Zootecnia, Universidade Federal de Goiás – UFG. Goiânia, 2012. Disponível em: <https://repositorio.bc.ufg.br/tede/handle/tede/4303>.
- ROMANO, A. P. M.; RAMOS, D. G.; ARAÚJO, F. A. A.; SIQUEIRA, G. A. M.; RIBEIRO, M. P. D.; LEAL, S. G.; ELKHOURY, A. N. M. S.. Febre amarela no Brasil: recomendações para a vigilância, prevenção e controle. **Epidemiol. Serv. Saúde** [online]. v. 20, n. 1, p. 101-106, 2011. Disponível em: http://scielo.iec.gov.br/scielo.php?script=sci_arttext&id=S1679-49742011000100011&lng=pt.
- ROSE, S. Mortality Risk Score Predictions in an Elderly Population Using Machine Learning. **American Journal of Epidemiology**. v. 177, n. 5, p. 443-452, 2013. Disponível em: <https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kws241>.
- ROTHMAN, K.; GREENLAND, S.; LASH, T. **Epidemiologia moderna**. 3 ed. São Paulo: Art Med, 2011, 888p.
- ROVERO, F.; MTUI, A.; KITEGILE, A.; JACOB, P.; ARALDI, A.; TENAN, S. Primates decline rapidly in unprotected forests: evidence from a monitoring program with data constraints. **PLOS ONE**. v. 10, n. 2, 2015. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118330>.
- RUMELHART, D.; HILTON, G.; WILLIAMS, R. Learning internal representations by error propagation. In: Rumelhart, D.; McClelland, J. (eds) **Parallel Distributed Processing: Explorations in the microstructure of cognition**. v. 1. Foundations. Cambridge, MA : MIT Press, p. 318-363, 1986. Disponível em: https://web.stanford.edu/class/psych209a/ReadingsByDate/02_06/PDPVolIChapter8.pdf.
- RUSSEL, S. J.; NORVIG, P. **Artificial intelligence: a modern approach**. 3 ed. Nova Jérsei:

- Prentice Hall, 2009. Disponível em: <<https://www.cin.ufpe.br/~tfl2/artificial-intelligence-modern-approach.9780131038059.25368.pdf>>.
- RYLANDS, A. B.; SCHNEIDER, H.; LANGGUTH, A.; MITTERMEIER, R. A.; GROVES, C. P.; RODRÍGUEZ-LUNA, E. An assessment of the diversity of New World primates. **Neotrop. Primates**. v. 8, p. 261–93, 2000.
- RYLANDS, A. B. Exudate-eating and tree-gouging by marmosets (Callitrichidae, Primates). In: Chadwick, A.C. e Sutton, S.L. (eds.). **Tropical Rain Forest: The Leeds Symposium**. Leeds, Inglaterra: Leeds Philosophical and Literary Society. 1984, p. 155-168. Disponível em: <https://www.researchgate.net/publication/274389488_Exudate-eating_and_tree-gouging_by_marmosets_Callitrichidae_Primates>.
- RYLANDS, A. B.; FARIA, D. S. Habitats, feeding ecology and range size in the genus *Callithrix*. In: Rylands, A.B. (eds.). **Marmosets and Tamarins: Systematics, Behaviour, and Ecology**. Oxford - Inglaterra: Oxford University Press,, 1993, p. 262-272.
- RYLANDS, A. B.; MENDES, S. L. *Callithrix penicillata* - 2008.. In: IUCN Red List of Threatened Species, Versão 2018.2. Disponível em www.iucnredlist.org.
- RYLANDS, A. B.; MITTERMEIER, R. A.; OLIVEIRA, M. M.; KIERULFF, M. C. M. *Callithrix jacchus* - 2008. In: IUCN Red List of Threatened Species, Versão 2018.2. Disponível em www.iucnredlist.org.
- RYLANDS, A. B.; SPIRONELO, W. R.; TORNISIELO, V. L.; LEMOS DE SÁ, R. M.; KIERULFF, M. C. M. e SANTOS, I. B. Primates of the Rio Jequitinhonha valley, Minas Gerais, Brazil. **Primate Conservation**, n. 9, p. 100-109, 1988. Disponível em: <https://www.researchgate.net/publication/330261745_Primates_of_the_Valleys_of_the_Rios_Jequitinhonha_and_Mucuri_Brazil_Occurrence_and_Distribution>.
- SABROZA, P. C.; TOLEDO, L. M.; OSANAI, C. H. Organização do espaço e os processos endêmico-epidêmicos. In: LEAL, M. C.; SABROZA, P. C.; RODRIGUEZ, R.H.; BUSS P.M. (org.). **Saúde, ambiente e desenvolvimento** – v. 2. São Paulo: Hucitec, 1992, p.56-77.
- SAMY, A. M.; ANNAJAR, B. B.; DOKHAN, M. R.; BOUSSAA, S.; PETERSON, A. T. Coarse-resolution Ecology of Etiological Agent, Vector, and Reservoirs of Zoonotic Cutaneous Leishmaniasis in Libya. **PLoS Negl Trop Dis** v. 10, n. 2, 2016b. Disponível em: <<http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0004381>>.
- SAMY, A. M.; THOMAS, S. M.; WAHED, A. A. E.; COHOON, K. P.; PETERSON, A. T. Mapping the global geographic potential of Zika vírus spread. **Mem. Inst. Oswaldo Cruz** v. 111, n. 9, p. 559-560, 2016a. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0074-02762016000900559&lng=en>.

- SAMY, A. M.; PETERSON, A. T.; HALL, M. Phylogeography of Rift Valley Fever vírus in África and the Arabian Peninsula. **PLoS Negl Trop Dis** v. 11, n. 1, 2017. Disponível em: <<http://journals.plos.org/plosntds/article?id=10.1371/journal.pntd.0005226>>.
- SCHWARZKOPF, L.; RYLANDS, A. B. Primate species richness in relation to habitat structure in Amazonian rainforest fragments. **Biol Conserv** v. 48, p. 1-12, 1989. Disponível em: <<https://www.sciencedirect.com/science/article/pii/0006320789900554>>.
- SEABRA, H. F.; ENCINAS, J. M.; FELFILI, J. M. Análise estrutural da mata ciliar do Córrego Capetinga – DF, habitat de *Callithrix jacchus penicillata*. **Pesquisa Agropecuária Brasileira**, n. 26, v. 1, p. 11-17, 1991. Disponível em: <<https://www.embrapa.br/busca-de-Publicacoes/-/Publicacao/105725/análise-estrutural-da-mata-ciliar-do-corrego-capetinga---df-habitat-de-callithrix-jacchus-penicillata-l>>.
- SHMUELI, G. To Explain or to Predict? **Institute of Mathematical Statistics in Statistical Science**, v. 25, n. 3, pp. 289-310, 2010. Disponível em: <<https://www.stat.berkeley.edu/~aldous/157/Papers/shmueli.pdf>>.
- SILVA, J. M. C. D.; BATES, J. M. Biogeographic patterns and conservation in the South American Cerrado: A tropical savanna hotspot. **BioScience** v. 52, p. 225-234, 2002. Disponível em: <<https://academic.oup.com/bioscience/article/52/3/225/312213>>.
- SILVA, J. P. S. Impactos ambientais causados por mineração. **Revista Espaço da Sophia**, n. 8, 2007. Disponível em: <<http://www.registro.unesp.br/sites/museu/basededados/arquivos/00000429.pdf>>.
- SILVA, A. M; NUNES, V.; LOPES, J. Culicídeos associados a entrenós de bambu e bromélias, com ênfase em *Aedes (Stegomyia) albopictus* (Diptera, Culicidae) na Mata Atlântica, Paraná, Brasil. **Iheringia, Sér. Zool.**, v.94, n.1, p.63-66, 2004.
- SILVA, K. M.; TOLEDO, C. C. **Diagnóstico de área degradada por extração de saibro**. Vianna Sapiens, Juiz de Fora. (Relatório Técnico v.1, edição especial), 2010. Disponível em: <<http://www.viannasapiens.com.br/revista/article/view/34/24>>.
- SMOLA, A. J.; BARTLETT, P. J.; SCHÖLKOPF, B.; SCHUURMANS, D. (eds). **Advances in Large Margin Classifiers**. Londres – Inglaterra: MIT Press., 2000, p. 1-422. Disponível em: <https://www.eecs.yorku.ca/course_archive/2005-06/F/6002B/Readings/lmc-book.pdf>.
- SOLOMATINE, D. P.; SEE, L. M.; ABRAHART, R. J. Data-driven Modeling: Concepts Approaches and Experiences. In: Abrahart, R. J.; SOLOMATINE, D.; SEE, L. M. (eds) **Practical hydroinformatics: computations intelligence and technological developments in water applications**. Berlim: Springer, 2008, p. 17-30. Disponível em:

<https://www.researchgate.net/publication/321614268_Practical_Hydroinformatics_Computational_Intelligence_and_Technological_Developments_in_Water_Applications>.

SOLOMATINE, D. P. e TORRES, L. A. Neural network approximation of hidrodynamic model in optimizing reservoir operation. In: Proc. 2nd Int. Conf. on Hydroinformatics, Balkema, Rotterdam. **Anais...** p. 201-206, 1996. Disponível em: <https://www.researchgate.net/publication/222108332_Neural_network_approximation_of_a_Hydrodynamic_Model_in_Optimizing_Reservoir_Operation>.

SOLOMATINE, D. P.; OSTFELD, A. Data-driven modeling: some past experiences and new approaches. **Journal of Hydroinformatics**. p. 3-21, 2008. Disponível em: <http://jh.iwaponline.com/content/ppiwajhydro/10/1/3.full.pdf>.

SONG, Y.; HUANG, J.; ZHOU, D.; ZHA, H.; GILES, C. L. (2007) Iknn: Informative k-nearest neighbor pattern classification. In KOK J.N., KORONACKI J., LOPEZ DE MANTARAS R., MATWIN S., MLADENIČ D., SKOWRON A. (eds) Knowledge Discovery in Databases: PKDD 2007. PKDD 2007. Lecture Notes in Computer Science, vol 4702. Springer, Berlin, Heidelberg. **Anais...** Disponível em: <https://link.springer.com/chapter/10.1007/978-3-540-74976-9_25>.

SONTER, L. J.; HERRERA, D.; BARRETT, D. J.; GALFORD, G. L.; MORAN, C. J.; SOARES-FILHO, B. S. Mining drives extensive deforestation in the Brazilian Amazon. **Nature Communications** v. 8, n. 1, p. 1013, 2017. Disponível em: <<https://www.nature.com/articles/s41467-017-00557-w>>.

SPIRA, C.; KIRKBY, A.; KUJIRAKWINJA, D.; PLUMPTRE, A. J. The socio-economics of artisanal mining and bushmeat hunting around protected areas: Kahuzi–Biega National Park and Itombwe Nature Reserve, eastern Democratic Republic of Congo. **Oryx**. v. 53, n. 1, p. 136-144, 2017. Disponível em: <<https://www.cambridge.org/core/journals/oryx/article/socioeconomics-of-artisanal-mining-and-bushmeat-hunting-around-protected-areas-kahuzibiega-national-park-and-itombwe-nature-reserve-eastern-democratic-republic-of-congo/E311BA3CE96B578F773F29142E6A70F2>>.

SPRACKLEN, B. D.; KALAMANDEEN, M.; GALBRAITH, D.; GLOOR, E.; SPRACKLEN, D. V. A global analysis of deforestation in moist tropical forest protected areas. **PLOS ONE** v. 10, n. 12, 2015. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0143886>>.

STAVIS, V. K. **Área de vida, alimentação e comportamento social de uma população de *Alouatta caraya* em fragmento de Cerrado**. 2013, 71 p. Dissertação (Mestrado em Biologia Animal) - Universidade Federal de Mato Grosso do Sul – UFMS. Mato Grosso do Sul, 2013. Disponível em: <<http://repositorio.cbc.ufms.br:8080/jspui/bitstream/123456789/1986/1/Vanessa%20Katherinne%20Stavis.pdf>>.

- STEPHENS, P. R.; ALTIZER, S.; SMITH, K. F.; AGUIRRE, A. A.; BROWN, J. H.; BUDISCHAK, S. A.; BYERS, J. E.; DALLAS, T. A.; DAVIES, T. J.; DRAKE, J. M.; EZENWA, V. O.; FARRELL, M.; GITTLEMAN, J. L.; HAN, B. A.; HUANG, S.; HUTCHINSON, R. A.; JOHNSON, P.; NUNN, C. L.; ONSTAD, D.; PARK, A.; VAZQUEZ-PROKOPEC, G. M.; SCHMIDT, J. P.; POULIN, R. The macroecology of infectious diseases: a new perspective on global-scale drivers of pathogen distribution and impacts. **Ecology Letters**, p. 1-13, 2016. Disponível em: <<https://www.researchgate.net/publication/304609813>>.
- STEVENSON, P. The relationship between fruit production and primate abundance in Neotropical communities. **Biological Journal of the Linnean Society** v. 72, p. 161- 178, 2001. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0024406600904978>>.
- STEVENSON, P. R.; ALDANA, A. M. Potential effects of ateline extinction and forest fragmentation on plant diversity and composition in the western Orinoco Basin, Colombia. **International Journal of Primatology** v. 29, n. 2, p. 365–377, 2008. Disponível em: <<https://link.springer.com/article/10.1007/s10764-007-9177-x>>.
- STOCKWELL, D. R. B. Genetic algorithms II. In: Fielding, A. H. (ed). **Machine learning methods for ecologic applications**. Boston: Kluwer Academic Publishers. 1999, p. 123-144. Disponível em: <https://link.springer.com/chapter/10.1007/978-1-4615-5289-5_5>.
- STRANO, J.; DOOLEY, J. R.; ISHAK, K. G. **Manual sobre la fiebre amarilla y su diagnóstico diferencial histopatológico**. Washington: OPS, 1975. p. 1-53.
- STRASSBURG, B. B. N.; BROOKS, T.; FELTRAN-BARBIERI, R.; IRIBARREM, A.; CROUZEILLES, R.; LOYOLA, R.; LATAWIEC, A. E.; OLIVEIRA FILHO, F. J. B.; SCARAMUZZA, C. A. D. E. M.; SCARANO, F. R.; SOARES-FILHO, B.; BALMFORD, A. Moment of truth for the Cerrado hotspot. **Nature Ecology e Evolution** v. 1, 2017. Disponível em: <https://www.researchgate.net/publication/315549179_Moment_of_truth_for_the_Cerrado_hotspot>.
- STRIER, K. B.; POSSAMAI, C. B.; TABACOW, F. P.; PISSINATTI, A.; LANNA, A. M.; RODRIGUES DE MELO, F.; MOREIRA, L.; TLEBI, M.; BREVES, P.; MENDES, S. L.; JERUSALINSKY, L. Demographic monitoring of wild miqui populations: criteria for defining priority areas and monitoring intensity. **PLOS ONE** v. 12, n. 12, 2017. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0188922>>.
- STRUEBIG, M. J.; FISCHER, M.; GAVEAU, D. L.; MEIJAARD, E.; WICH, S. A.; GONNER, C.; SYKES, R.; WILTING, A.; KRAMER-SCHADT, S. Anticipated climate and land-cover changes reveal refuge areas for Borneo's orang-utans. **Global Change**

- Biology** v. 21, n. 8, p. 2891–2904, 2015. Disponível em: <<https://onlinelibrary.wiley.com/doi/full/10.1111/gcb.12814>>.
- SKABA D.A., CARVALHO MS, BARCELLOS C, MARTINS PC, TERRON SL. Geoprocessamento dos dados da saúde: o tratamento dos endereços. **Cadernos de Saúde Pública**. v. 20, n. 6, p. 1753-1756, 2004. Disponível em: <<http://www.scielo.br/pdf/csp/v20n6/37.pdf>>.
- SUTTON, R. S.; BARTO, A. **Reinforcement Learning: An Introduction**. 2 ed. A Bradford Book. Londres – Inglaterra: MIT Press, 2012, 1-334 p. Disponível em: <http://people.inf.elte.hu/lorincz/Files/RL_2006/SuttonBook.pdf>.
- TAN, P. N.; STEINBACH, M.; KUMAR, V. **Introduction to data mining**. Massachusettes, USA: Pearson Addison-Wesley, 2006, p. 1-169. Disponível em: <<https://www-users.cs.umn.edu/~kumar001/dmbook/sol.pdf>>.
- TANG, L.; ZHENG A. Y.; XIE, X.; YUAN, J.; YU, X.; HAN, J. **Retrieving k-Nearest Neighboring Trajectories by a Set of Point Locations**. Berlim: Springer, 2011, p. 223–241. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-642-22922-0_14>.
- TAUBERT, F.; FISCHER, R.; GROENEVELD, J.; LEHMANN, S.; MÜLLER, M. S.; RÖDIG, E.; EIGAN, T.; HUTH, A. Global patterns of tropical forest fragmentation. **Nature** v. 554, n. 7693, p. 519–522, 2018. Disponível em: <<https://www.nature.com/articles/nature25508>>.
- TEIXEIRA, L. A. Da Transmissão hídrica a culicidiana: a febre amarela na sociedade de medicina e cirurgia de São Paulo. **Rev Bras História**. v. 21, n. 41, p. 217-242, 2001. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-01882001000200012>.
- TERZIAN, A. C. B.; ZINI, N.; SACCHETTO, L.; ROCHA, R. F.; PARRA, M. C. P.; DEL SARTO, J. L.; DIAS, A. C. F.; COUTINHO, F.; RAYRA, J.; SILVA, R. A.; COSTA, V. V.; FERNANDES, N. C. C. A.; RÉSSIO, R.; DÍAZ-DELGADO, J.; GUERRA, J.; CUNHA, M. S.; CATÃO-DIAS, J. L.; BITTAR, C.; REIS, A. F. N.; SANTOS, I. N. P.; FERREIRA, A. C. M.; CRUZ, L. E. A. A.; RAHAL, P.; ULLMANN, L.; MALOSSI, C.; ARAÚJO JR, J. P.; WIDEN, S.; REZENDE, I. M.; MELLO, E.; PACCA, C. C.; KROON, E. G.; TRINDADE, G.; DRUMOND, B.; CHIARAVALLLOTI-NETO, F.; VASILAKIS, N.; TEIXEIRA, M. M.; NOGUEIRA, M. L. Evidence of natural Zika vírus infection in neotropical non-human primates in Brazil. **Scientific Reports**. v. 8; p. 1-15, 2018. Disponível em: <<https://www.nature.com/articles/s41598-018-34423-6.pdf>>
- TIMPE, K.; KAPLAN, D. The changing hydrology of a dammed Amazon. **Science Advances** v. 3, n. 11, 2017. Disponível em: <<http://advances.sciencemag.org/content/3/11/e1700611>>.

- TING, K. M. Confusion Matrix. In: Sammut C., Webb G.I. (eds) **Encyclopedia of Machine Learning**. Boston, MA: Springer, 2011. Disponível em: <<https://doi.org/10.1007/978-0-387-30164-8>>.
- TUCKER, C.; PINZON, J.; BROWN, M.; SLAYBACK, D.; PAK, E.; MAHONEY, R.; VERMOTE, E.; EL SALEOUS, N. An extended AVHRR 8-km NDVI dataset compatible with MODIS and SPOT vegetation NDVI data. **Int. J. Remote Sens.** v. 26, p. 4485–4498, 2005. Disponível em: <<https://www.tandfonline.com/doi/abs/10.1080/01431160500168686>>.
- VAINIO, J.; CUTTS, F. (1998). **Yellow Fever**. New York: Thomas Nelson e Sons, 1948.
- VASCONCELOS, P. F.; LUNA, E. J.; GALLER, R.; SILVA, L. J.; COIMBRA, T. L.; BARROS, V. L.; MONATH, T. P.; RODIGUES, S. G.; LAVAL, C.; COSTA, Z. G.; VILELA, M. F.; SANTOS, C. L.; PAPAIOORDANOU, P. M.; ALVES, V. A.; ANDRADE, L. D.; SATO, H. K.; ROSA, E. S.; FROGUAS, G. B.; LACAVA, E.; ALMEIDA, L. M.; CRUZ, A. C.; ROCCO, I. M.; SANTOS, R. T.; OLIVA, O. F. Serious adverse events associated with yellow fever 17DD vaccine in Brazil: a report of two cases. **Lancet**. n. 358, v. 9276, p. 91-97, 2001. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/11463409>>.
- VASCONCELOS, P. F. C. Yellow fever. In: Marcondes C. B. (ed). **Arthropod Borne Diseases**. New York: Springer, 2017, p. 101–113.
- VASCONCELOS, P. F. C. Febre amarela. **Rev Soc Bras Med Trop**. v. 36, n. 2, p. 275-293, 2003. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0037-86822003000200012>.
- VASCONCELOS, P. F. C. Febre amarela: reflexões sobre a doença, as perspectivas para o século XXI e o risco da reurbanização. **Rev. Bras. Epidemiologia**. v. 5, n. 3, p. 244-258, 2002. ISSN 1415-790X. Disponível em: <<http://dx.doi.org/10.1590/S1415-790X2002000300004>>. Disponível em: <<http://www.scielo.br/pdf/rbepid/v5n3/04.pdf>>.
- VASCONCELOS, P. F. C.; COSTA, Z. G.; TRAVASSOS DA ROSA, E. S.; LUNA, E.; RODIGUES, S. G.; BARROS, V. L. R. S.; DIAS, J. P.; MONTEIRO, H. A. O.; OLIVA, O. F. P.; VASCONCELOS, H. B.; BARBOSA DA SILVA, J.; CRUZ, A. C. R.; MARTINS, E. C.; TRAVASSOS DA ROSA, J. F. S. Epidemic of Jungle Yellow Fever in Brazil, 2000: Implications of Climatic Alterations in Disease Spread. **Journal of Medical Virology**, v. 65, p. 598-604, 2001. Disponível em: <<https://www.ncbi.nlm.nih.gov/pubmed/11596099>>.
- VEIGA, J. B.; FORTES, V. B.; BICCA-MARQUES, J. C. Population viability analysis of a

- brown howler monkey metapopulation after an outbreak of yellow fever in southern Brazil. In: Abstracts of the XXV Congress of the International Primatological Society, Hanoi, 30, 2014. *Anais...* Disponível em: <https://www.researchgate.net/publication/326723770_Population_Viability_Analysis_of_howler_monkeys_Alouatta_guariba_clamitans_Cabrera_1940_in_a_fragmented_landscape_in_Southeast_Brazil>.
- VEIGA, L. M. e FERRARI, S. F. 2008. **Callicebus moloch**. In: IUCN Red List of Threatened Species, Versão 2011.2. Disponível em: www.iucnredlist.org.
- VEIGA, L. M.; KIERULFF, C. e OLIVEIRA, M. M. 2008. **Alouatta belzebul**. IUCN Red List of Threatened Species, Versão 2011.2. Disponível em www.iucnredlist.org. Acesso em 22 set. 2018.
- VEIGA, L. M.; KIERULFF, C. M.; OLIVEIRA, M. M. e MENDES, S. L. 2008. **Callicebus nigrifrons**. In: IUCN Red List of Threatened Species, Versão 2011.2. www.iucnredlist.org.
- VENTER, O.; FULLER, R. A.; SEGAN, D. B.; CARWARDINE, J.; BROOKS, T.; BUTCHART, S. H. M.; DI MARCO, M.; IWAMURA, T.; JOSEPH, L.; O'GRADY, D.; POSSINGHAM, H. P.; RONDININI, C.; SMITH, R. J.; VENTER, M.; WATSON, J. E. M. 2014. Targeting global protected area expansion for imperiled biodiversity. **PLOS Biology** v. 12, n. 6, 2014. Disponível em: <<https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001891>>.
- VERIKAS, A.; VAICIUKYNAS, E.; & GELZINIS, A.; PARKER, J.; OLSSON, M. C. Electromyographic Patterns during Golf Swing: Activation Sequence Profiling and Prediction of Shot Effectiveness. **Sensors**. v. 16, n. 592, p. 1-26. Disponível em: <https://www.researchgate.net/publication/301638643_Electromyographic_Patterns_during_Golf_Swing_Activation_Sequence_Profiling_and_Prediction_of_Shot_Effectiveness/link/5720b9ca08ae26e7210068/download>.
- WAEBER, P. O.; WILMÉ, L.; MERCIER, J. R.; CAMARA, C.; LOWRY, P. P. How effective have thirty years of internationally driven conservation and development efforts been in Madagascar? **PLOS ONE** v. 11, n. 8, 2016. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0161115>>.
- WANG, L. e MENDEL, J.. Generating Fuzzy Rules by Learning from Examples. **IEEE Transactions on Systems Man and Cybernetics**, v. 22, n. 6, p. 1414–1427, 1992. Disponível em: <<https://ieeexplore.ieee.org/document/199466>>.
- WANG, Y.; SHI, Y.; YUE, B.; TENG, H. An efficient differential evolution algorithm with approximate fitness functions using neural networks. In: Proceedings of the 2010 international conference on Artificial intelligence and computational intelligence: Part II, AICI'10. *Anais...* Berlim, Heidelberg: Springer-Verlag, p. 334–341, 2010.. Disponível

- em: <https://link.springer.com/chapter/10.1007/978-3-642-16527-6_42>.
- WARD, S.; BARKER, A. **Undefined by data: a survey of big data definitions**. p. 1-2, 2013. Disponível em: <<https://arxiv.org/abs/1309.5821>>.
- WARNER, M. D. Assessing habitat utilization by Neotropical primates: a new approach. **Primates** v. 43, p. 59-71, 2002. Disponível em: <https://www.researchgate.net/publication/11283450_Assessing_habitat_utilization_by_neotropical_primates_A_new_approach>.
- WERBOS, P. J. **The Roots of Backpropagation**. New York: John Wiley e Sons. 1994, p. 1-319.
- WEBB, G. I. Model Evaluation. In: Sammut C., Webb G.I. (Eds) **Encyclopedia of Machine Learning**. Boston, MA: Springer, 2011. Disponível em: <<https://doi.org/10.1007/978-0-387-30164-8>>.
- WHO – World Health Organization. **Updates on yellow fever vaccination recommendations for international travellers related to the current situation in Brazil**. 2017. Disponível em: <<https://www.who.int/ith/updates/20180116/en/>>. Acesso em: 29 set. 2018.
- WHO – World Health Organization. **Febre amarela**. 2016. Disponível em: <<http://www.who.int/mediacentre/factsheets/fs100/pt/>>. Acesso em: 29 set. 2018.
- WIEDERHOLT, R.; POST, E. Tropical warming and the dynamics of endangered primates. **Biology Letters** v. 6, n. 2, p. 257–260, 2010. Disponível em: <<https://royalsocietypublishing.org/doi/full/10.1098/rsbl.2009.0710>>.
- WILLMOTT, C. J., & MATSUURA, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. **Climate Research**, v. 30, n. 1, p. 79–82, 2005. Disponível em: <<https://www.int-res.com/articles/cr2005/30/c030p079.pdf>>.
- WINEMILLER, K.; MCINTYRE, P.; CASTELLO, L.; FLUET-CHOUINARD, E.; GIARRIZZO, T.; NAM, S.; BAIRD, I.; DARWALL, W.; LUJAN, N.; HARRISON, I.; STIASSNY, M. L.; SILVANO, R. A.; FITZGERALD, D. B.; PELICICE, F. M.; AGOSTINHO, A. A.; GOMES, L. C.; ALBERT, J. S.; BARAN, E.; PETRERE, M. J. R.; ZARFL, C.; MULLIGAN, M.; SULLIVAN, J. P.; ARANTES, C. C.; SOUSA, L. M.; KONING, A. A.; HOEINGHAUS, D. J.; SABAJ, M.; LUNDBERG, J. G.; ARMBRUSTER, J.; THIEME, M. L.; PETRY, P.; ZUANON, J.; TORRENTE, V. G.; SNOEKS, J.; RAINBOTH, W.; PAVANELLI, C. S.; AKAMA, A.; VAN SOESBERGEN, A.; SAENZ, L. Balancing hydropower and biodiversity in the Amazon, Congo, and Mekong. **Science** v. 351, n. 6269, p. 128–129, 2016. Disponível em:

<<http://science.sciencemag.org/content/351/6269/128>>.

WOODALL, J. Biodiversidade e Saúde: a necessidade absoluta de monitoramento. In: Chame, M., Labarthe, N. (orgs) **Caderno de Texto Workshop Estado da Arte da Saúde Silvestre no Brasil**. 2009. Disponível em: <http://www.biodiversidade.ciss.fiocruz.br/sites/www.biodiversidade.ciss.fiocruz.br/files/Caderno_Texto_Base_Workshop_Saude_Silvestre_2009.pdf>.

WU, X.; KUMAR, V.; QUINLAN, J. R.; GHOSH, J.; YANG, Q.; MOTODA, H.; MCLACHLAN, N. A.; LIU, B.; YU, P. S.; ZHOU, Z. H.; STEINBACH, M.; HAND, D. J.; e STEINBERG, D. Top 10 algorithms in data mining. **Knowledge and Information Systems**, v. 14, n. 1, p. 1–37, 2007. Disponível em: <<http://www.cs.uvm.edu/~icdm/algorithms/10Algorithms-08.pdf>>.

WUETHRICH, B. Biodiversity: reconstructing Brazil's Atlantic Rainforest. **Science**, Washington, v. 315, n. 5815, p. 1070-1072, 2007. Disponível em: <<http://science.sciencemag.org/content/315/5815/1070/tab-article-info>>.

WWF – World Wide Fund for Nature. **Impact of Climate Change on Species (publicado em 12 nov. 2015)**. Suíça: WWF, 2015. 21p. ISBN: 978-2-940529-25-4. Disponível em: <<http://wwf.panda.org/?256190/impact-of-climate-change-on-species>>.

WWF – World Wide Fund for Nature. **Relatório Planeta Vivo - 2018**. Disponível em: <<https://promo.wwf.org.br/planetavivo2018>>. Acesso em 20 out. 2018.

XAVIER, S. D. C.; ROQUE, A. L. R.; LIMA, V. S.; MONTEIRO, K. J. L., OTAVIANO, J. C. R. Lower Richness of small wild mammals species and Chagas disease risk. **PLoS Neglected Tropical Diseases** v. 6, n. 5, p. 1-11, 2012. Disponível em: <<http://www.plosntds.org/article/info%3Adoi%2F10.1371%2Fjournal.pntd.0001647>>.

YAGER, R. e FILEV, D.. Approximate Clustering via The Mountain Method. **IEEE Transactions on Systems Man and Cybernetics**, v. 24, n. 8, p. 1279–1284, 1994. Disponível em: <<https://ieeexplore.ieee.org/document/299710>>.

YANG, H. M. **Epidemiologia matemática**. São Paulo: Unicamp. 2001. 1-239p.

ZHANG, S.; LI, X.; ZONG, M.; ZHU, X; CHENG, D. Learning k for KNN Classification. **ACM Transactions on Intelligent Systems and Technology**. v. 8, n. 3, p. 1-19. 2017. Disponível em: <http://delivery.acm.org/10.1145/3000000/2990508/a43-zhang.pdf?ip=177.195.34.241eid=2990508eacc=OAekey=4D4702B0C3E38B35%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35%2EC2C253FD690A2A4FeCFID=791633153eCFTOKEN=81874470e_acm=1501426785_eae1af5f9930bb60f1cbadb0b600f6ec>.