

**Leandro Dias Gomes de Carvalho**

**Exposição ambiental e a internação por  
asma nos municípios da Amazônia e  
Pantanal**

Niterói - RJ, Brasil

23 de setembro de 2021

**Leandro Dias Gomes de Carvalho**

**Exposição ambiental e a internação  
por asma nos municípios da  
Amazônia e Pantanal**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em  
Estatística pela Universidade Federal Fluminense.

Orientador(a): Profa. Dra. Ludmilla da Silva Viana Jacobson

Co-Orientador(a): Profa. Dra. Sandra de Souza Hacon

Niterói - RJ, Brasil

23 de setembro de 2021

**Leandro Dias Gomes de Carvalho**

**Exposição ambiental e a internação por asma  
nos municípios da Amazônia e Pantanal**

Monografia de Projeto Final de Graduação sob o título “*Exposição ambiental e a internação por asma nos municípios da Amazônia e Pantanal*”, defendida por Leandro Dias Gomes de Carvalho e aprovada em 23 de setembro de 2021, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

---

**Profa. Dra. Ludmilla da Silva Viana Jacobson**  
Departamento de Estatística – UFF

---

**Profa. Dra. Sandra de Souza Hacon**  
Escola Nacional de Saúde Pública – ENSP

---

**Prof. Dr. Jony Arrais Pinto Junior**  
Departamento de Estatística – UFF

---

**Profa. Dra. Márcia Marques de Carvalho**  
Departamento de Estatística – UFF

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

C331e Carvalho, Leandro Dias Gomes de  
Exposição ambiental e a internação por asma nos  
municípios da Amazônia e Pantanal / Leandro Dias Gomes de  
Carvalho ; Ludmilla da Silva Viana Jacobson, orientadora ;  
Sandra de Souza Hacon, coorientadora. Niterói, 2021.  
112 f. : il.

Trabalho de Conclusão de Curso (Graduação em  
Estatística)-Universidade Federal Fluminense, Instituto de  
Matemática e Estatística, Niterói, 2021.

1. Internação por asma. 2. Exposição ambiental. 3.  
Séries Temporais. 4. Modelos Lineares Generalizados. 5.  
Produção intelectual. I. Jacobson, Ludmilla da Silva Viana,  
orientadora. II. Hacon, Sandra de Souza, coorientadora. III.  
Universidade Federal Fluminense. Instituto de Matemática e  
Estatística. IV. Título.

CDD -

# Resumo

No Brasil, grande parte das queimadas é realizada pelo homem por diversas razões como o desmatamento, as disputas de terras e os protestos. As queimadas antropogênicas realizadas de forma descontrolada ou as naturais, provocadas, por exemplo, por falta de chuvas, muitas das vezes empobrecem o solo e destroem a fauna e flora local. Como consequência, podem levar à extinção e morte de diversos animais e plantas nativas, assim como à emissão de poluentes no ar e à alteração da temperatura do local. No âmbito municipal e regional a fumaça das queimadas pode provocar diversos problemas de saúde, como a asma. O objetivo principal deste trabalho foi avaliar a relação entre a internação por asma em crianças e adolescentes (com 14 anos ou menos) e à exposição às variáveis ambientais Focos de Queimadas, Temperatura e Material Particulado Fino ( $PM_{2.5}$ ) nos biomas Amazônia e Pantanal. Foram utilizados os dados fornecidos pelo Instituto Nacional de Pesquisas Espaciais (INPE) sobre o número de Focos de Queimadas (variável proxy que mede a intensidade das queimadas), utilizando o satélite de referência *AQUA M-T*. As informações sobre as variáveis meteorológicas da Temperatura mínima, máxima e média mensal foram obtidas através do European Centre for Medium-Range Weather Forecasts (ECMWF). Os dados sobre as concentrações de Material Particulado Fino ( $PM_{2.5}$ ) mensal foram obtidos do Sistema de Informações Ambientais Integrado a Saúde (SISAM). E os dados sobre o número de internações hospitalares por asma foram obtidos do Sistema de Informações Hospitalares do SUS (SIH/SUS) por meio do Departamento de Informática do Sistema Único de Saúde (DATASUS). O período de análise do estudo se refere a Jan./2010 até Dez./2020. Na análise dos dados, foram estimados Modelos Lineares Generalizados através da distribuição de Quasi-Poisson e Modelos de Séries Temporais para avaliar a tendência das variáveis, a partir do SARIMA. Os resultados mostraram que para o Pantanal o número de Focos de Queimadas observados em 2020 foi muito superior ao esperado pelo comportamento da série temporal. Por outro lado, para ambos os biomas o número observado de internações por asma foi muito inferior ao que seria esperado de acordo com o modelo de séries temporais. Quanto aos resultados dos Modelos Lineares Generalizados, a temperatura indicou ser um fator de risco para as internações por asma em crianças e adolescentes menores de 14 anos nos dois biomas, quando o modelo foi ajustado por  $PM_{2.5}$  e Focos de Queimadas, de modo que para o bioma Pantanal, o aumento em um grau na temperatura média mensal aumenta o risco de internações por asma em aproximadamente 20%. Enquanto que para o bioma Amazônia, os resultados sugerem que o aumento em um grau na temperatura mínima mensal aumenta o risco de internações por asma em aproximadamente 18%.

Palavras-chave: Queimadas. Poluentes ambientais. Asma. Modelos. Séries Temporais.

# Agradecimentos

Aos meus pais e meus irmãos que desde sempre me apoiaram e me fizeram acreditar de que tudo daria certo.

A professora e orientadora, Ludmilla da Silva Viana Jacobson por todo o ensinamento e apoio ao longo deste trabalho.

Aos professores do departamento de Estatística da UFF, em especial a Ana Maria Lima de Farias por todos os ensinamentos e conversas, ao Hugo Henrique Kegler dos Santos por uma ótima didática e pelas portas abertas, a Jessica Quintanilha Kubrusly pela ótima didática e conhecimento, ao Jony Arrais Pinto Junior por ter aceitado fazer parte da Banca Examinadora e por todos os ensinamentos e conversas formais e informais de forma a deixar mais dinâmica as matérias, a Márcia Marques de Carvalho por também ter aceitado fazer parte da Banca Examinadora e por ter me ensinado e ajudado desde o primeiro período até agora, a Núbia Karla de Oliveira Almeida pela ótima didática e por seus ensinamentos e ao Rafael Santos Erbisti que ajudou enormemente no entendimento de Modelagem nestes tempos de Educação à Distância.

A todos os professores em especial para o qual eu tive a honra de ser aluno.

A Sandra de Souza Hacon, doutora, professora e orientadora, a qual tive a honra de conhecer e realizar projeto.

Aos meus cinco amigos em especial que conheci na faculdade e estão comigo desde então, Matheus Alves, Thiago Augusto, Letícia Felix, Beatriz Loureira e João Pedro Cigliato e a todos os outros que fizeram parte desta minha caminhada.

Ao meu melhor amigo Matheus Gomes Marinho que sempre me apoiou e me deu forças para continuar seguindo.

Em especial a todos que fizeram parte deste trabalho e desta jornada. Obrigado.

# Sumário

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 16
1.1	Organização . . . . .	p. 18
<b>2</b>	<b>Revisão Bibliográfica</b>	p. 20
<b>3</b>	<b>Materiais e Métodos</b>	p. 25
3.1	Área de Estudo . . . . .	p. 25
3.2	Banco de Dados . . . . .	p. 26
3.2.1	Internações por asma . . . . .	p. 26
3.2.2	Variáveis Meteorológicas . . . . .	p. 26
3.2.3	Focos de Queimadas . . . . .	p. 27
3.2.4	Material Particulado . . . . .	p. 27
3.3	Métodos . . . . .	p. 28
3.3.1	Séries Temporais . . . . .	p. 28
3.3.1.1	Estacionariedade . . . . .	p. 29
3.3.1.2	Tendência . . . . .	p. 29
3.3.1.3	Sazonalidade . . . . .	p. 30
3.3.1.4	Função de Autocorrelação . . . . .	p. 31
3.3.1.5	Função de Autocorrelação Parcial . . . . .	p. 31

3.3.1.6	Correlograma . . . . .	p. 32
3.3.1.7	Teste de Estacionariedade . . . . .	p. 32
3.3.1.8	Modelos Box-Jenkins . . . . .	p. 33
3.3.1.9	Tipos de Modelos . . . . .	p. 34
3.3.1.10	Análise do Modelo . . . . .	p. 36
3.3.1.11	Previsão com modelo SARIMA . . . . .	p. 37
3.3.2	Modelo de Regressão Linear Generalizado de Poisson . . . . .	p. 37
3.3.2.1	Estimação . . . . .	p. 39
3.3.2.2	Adequabilidade do modelo . . . . .	p. 40
3.3.2.3	Análise dos Resíduos . . . . .	p. 41
3.3.2.4	Cr�terio de Informa�o de Akaike (AIC) . . . . .	p. 43
3.3.2.5	Coeficiente de Determina�o . . . . .	p. 43
3.3.2.6	Multicolinearidade . . . . .	p. 44
3.3.2.7	Superdispers�o . . . . .	p. 45
3.3.2.8	Natural Cubic Splines . . . . .	p. 46
<b>4</b>	<b>An�lise dos Resultados</b> . . . . .	<b>p. 47</b>
4.1	An�lise Descritiva . . . . .	p. 47
4.1.1	Pantanal . . . . .	p. 48
4.1.2	Amaz�nia . . . . .	p. 52
4.2	S�ries Temporais . . . . .	p. 56
4.2.1	Pantanal . . . . .	p. 56
4.2.2	Amaz�nia . . . . .	p. 65
4.3	Modelos Lineares Generalizados . . . . .	p. 75
4.3.1	Pantanal . . . . .	p. 78
4.3.2	Amaz�nia . . . . .	p. 82

<b>5 Conclusões</b>	p. 87
<b>Referências</b>	p. 90
<b>Apêndice 1 – Gráficos Análises Descritivas</b>	p. 94
<b>Apêndice 2 – Gráficos Séries Temporais Pantanal</b>	p. 99
<b>Apêndice 3 – Gráficos Séries Temporais Amazônia</b>	p. 103
<b>Apêndice 4 – Gráficos Modelagem</b>	p. 108
<b>Apêndice 5 – Variável de Temperatura</b>	p. 110

# Lista de Figuras

1	Diagrama de Fluxo PRISMA . . . . .	p. 22
2	Exemplo de um gráfico de Correlograma . . . . .	p. 32
3	Exemplo do uso de envelope simulado . . . . .	p. 42
4	Exemplo de homocedasticidade aceita . . . . .	p. 42
5	Histograma e Densidade das Interações por Asma e Focos de Queimadas nos municípios que abrangem o bioma do Pantanal de jan/2010 até dez/2019 . . . . .	p. 49
6	Histograma e Densidade do Material Particulado Fino médio e Temperatura Média nos municípios que abrangem o bioma do Pantanal de jan/2010 até dez/2019 . . . . .	p. 49
7	Boxplot das Interações por Asma, Focos de Queimadas, PM2.5 médio e Temperatura média no bioma do Pantanal agrupados por mês de jan/2010 até dez/2019 . . . . .	p. 50
8	Série Temporal das Interações por Asma e Focos de Queimadas no bioma do Pantanal de jan/2010 até dez/2019 . . . . .	p. 51
9	Série Temporal do Material Particulado Fino médio e Temperatura média no bioma Pantanal de jan/2010 até dez/2019 . . . . .	p. 52
10	Histograma e Densidade das Interações por Asma e Focos de Queimadas nos municípios que abrangem o bioma da Amazônia de jan/2010 até dez/2019 . . . . .	p. 53
11	Histograma e Densidade do Material Particulado Fino médio e Temperatura Média nos municípios que abrangem o bioma da Amazônia de jan/2010 até dez/2019 . . . . .	p. 53

12	Boxplot das Internações por Asma, Focos de Queimadas, PM <sub>2.5</sub> médio e Temperatura média no bioma da Amazônia agrupados por mês de jan/2010 até dez/2019 . . . . .	p. 54
13	Série Temporal das Internações por Asma e Focos de Queimadas no bioma da Amazônia de jan/2010 até dez/2019 . . . . .	p. 55
14	Série Temporal do Material Particulado Fino médio e Temperatura média no bioma da Amazônia de jan/2010 até dez/2019 . . . . .	p. 56
15	Série Temporal das Internações por Asmas em crianças e adolescentes de 0 a 14 anos no bioma do Pantanal de jan/2010 até dez/2019 . . . . .	p. 57
16	Decomposição aditiva da série temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma do Pantanal de jan/2010 até dez/2019 . . . . .	p. 57
17	Valores reais <i>versus</i> valores ajustados para a série temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma Pantanal para o período de 2010 a 2020 . . . . .	p. 60
18	Previsão de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma do Pantanal para o ano de 2020 . . . . .	p. 61
19	Valores reais <i>versus</i> valores ajustados para a série temporal de Focos de Queimadas no bioma Pantanal para o período de 2010 a 2020 . . . . .	p. 62
20	Previsão dos Focos de Queimadas no bioma do Pantanal para o ano de 2020 . . . . .	p. 63
21	Valores reais <i>versus</i> valores ajustados para a série temporal de PM <sub>2.5</sub> médio no bioma Pantanal para o período de 2010 a 2019 . . . . .	p. 64
22	Previsão do Material Particulado Fino médio no bioma do Pantanal para o ano de 2020 . . . . .	p. 65
23	Série Temporal das Internações por Asmas em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia de jan/2010 até dez/2019 . . . . .	p. 66
24	Decomposição aditiva da série temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia de jan/2010 até dez/2019 . . . . .	p. 67

25	Valores reais <i>versus</i> valores ajustados para a série temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma Amazônia para o período de 2010 a 2020 . . . . .	p. 70
26	Previsão de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia para o ano de 2020 . . . . .	p. 71
27	Valores reais <i>versus</i> valores ajustados para a série temporal de Focos de Queimadas no bioma Amazônia para o período de 2010 a 2020 . . . . .	p. 72
28	Previsão dos Focos de Queimadas no bioma da Amazônia para o ano de 2020 . . . . .	p. 73
29	Valores reais <i>versus</i> valores ajustados para a série temporal de $PM_{2.5}$ médio no bioma Amazônia para o período de 2010 a 2019 . . . . .	p. 74
30	Previsão do Material Particulado Fino médio no bioma da Amazônia para o ano de 2020 . . . . .	p. 75
31	Uso do envelope simulado no modelo 3 para o bioma do Pantanal . . . . .	p. 80
32	Homocedasticidade dos dados do modelo 3 para o bioma do Pantanal . . . . .	p. 81
33	Avaliação de pontos influentes do modelo 3 para o bioma do Pantanal . . . . .	p. 81
34	Uso do envelope simulado no modelo 12 para o bioma da Amazônia . . . . .	p. 84
35	Homocedasticidade dos dados do modelo 12 para o bioma da Amazônia . . . . .	p. 85
36	Avaliação de pontos influentes do modelo 12 para o bioma da Amazônia . . . . .	p. 85
37	Boxplot das variáveis de $pm2.5min$ , $pm2.5max$ e $pm2.5p90$ no bioma do Pantanal de jan/2010 até dez/2019 . . . . .	p. 94
38	Boxplot das variáveis de $tempmin$ , $tempmax$ , $tempp90$ e $tempamplitude$ no bioma do Pantanal de jan/2010 até dez/2019 . . . . .	p. 95
39	Série Temporal das variáveis de $pm2.5min$ , $pm2.5max$ e $pm2.5p90$ no bioma do Pantanal de jan/2010 até dez/2019 . . . . .	p. 95
40	Série Temporal das variáveis de $tempmin$ , $tempmax$ , $tempp90$ e $tempamplitude$ no bioma do Pantanal de jan/2010 até dez/2019 . . . . .	p. 96
41	Boxplot das variáveis de $pm2.5min$ , $pm2.5max$ e $pm2.5p90$ no bioma da Amazônia de jan/2010 até dez/2019 . . . . .	p. 96

42	Boxplot das variáveis de <i>tempmin</i> , <i>tempmax</i> , <i>tempp90</i> e <i>tempamplitude</i> no bioma da Amazônia de jan/2010 até dez/2019 . . . . .	p. 97
43	Série Temporal das variáveis de <i>pm2.5min</i> , <i>pm2.5max</i> e <i>pm2.5p90</i> no bioma da Amazônia de jan/2010 até dez/2019 . . . . .	p. 97
44	Série Temporal das variáveis de <i>tempmin</i> , <i>tempmax</i> , <i>tempp90</i> e <i>tempamplitude</i> no bioma da Amazônia de jan/2010 até dez/2019 . . . . .	p. 98
45	Autocorrelação e Autocorrelação Parcial da série temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma do Pantanal de jan/2010 até dez/2019 . . . . .	p. 99
46	Série Temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma do Pantanal sem tendência . . . . .	p. 100
47	Autocorrelação e Autocorrelação Parcial da série temporal de Internações por Asma no bioma do Pantanal sem tendência . . . . .	p. 100
48	Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Internações por Asma SARIMA(3,1,1)(0,0,1)[12] para o bioma Pantanal . . . . .	p. 101
49	Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Internações por Asma SARIMA(1,1,1)(0,0,1)[12] para o bioma Pantanal . . . . .	p. 101
50	Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Focos de Queimadas SARIMA(1,0,0)(1,1,1)[12] para o bioma Pantanal . . . . .	p. 102
51	Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Material Particulado Fino médio SARIMA(0,0,1)(2,0,0)[12] para o bioma Pantanal . . . . .	p. 102
52	Autocorrelação e Autocorrelação Parcial da série temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia de jan/2010 até dez/2019 . . . . .	p. 103
53	Série Temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia sem tendência . . . . .	p. 104
54	Autocorrelação e Autocorrelação Parcial da série temporal de Internações por Asma no bioma da Amazônia sem tendência . . . . .	p. 104
55	Série Temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia sem tendência e sazonalidade . . . . .	p. 105

56	Autocorrelação da Série Temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia sem tendência e sazonalidade . . . . .	p. 105
57	Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Internações por Asma SARIMA(1,1,1)(1,1,1)[12] para o bioma da Amazônia	p. 106
58	Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Internações por Asma SARIMA(1,0,0)(0,1,1)[12] para o bioma da Amazônia	p. 106
59	Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Focos de Queimadas SARIMA(1,0,0)(2,1,0)[12] para o bioma da Amazônia . .	p. 107
60	Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Material Particulado Fino médio SARIMA(1,0,0)(1,1,0)[12] para o bioma da Amazônia . . . . .	p. 107
61	Autocorrelação dos resíduos do modelo 3 do bioma Pantanal . . . . .	p. 108
62	Autocorrelação Parcial dos resíduos do modelo 3 do bioma Pantanal . .	p. 108
63	Autocorrelação dos resíduos do modelo 12 do bioma Amazônia . . . . .	p. 109
64	Autocorrelação Parcial dos resíduos do modelo 12 do bioma Amazônia	p. 109
65	Overview e Download data ERA5 Land site ECMWF . . . . .	p. 110
66	Exemplo de uma requisição realizada pela API do site ECMWF . . . . .	p. 112

# Lista de Tabelas

1	Estudos sobre queimadas e seus poluentes e a relação com a asma no Brasil, com foco nos municípios do bioma Pantanal e Amazônia entre os anos de 2010 a 2020. . . . .	p. 23
2	Variáveis de estudo . . . . .	p. 48
3	Resumo das variáveis de estudo do bioma Pantanal de jan/2010 até dez/2019 . . . . .	p. 48
4	Resumo das variáveis de estudo do bioma Amazônia de jan/2010 até dez/2019 . . . . .	p. 52
5	Resumo da análise dos modelos de Série Temporal para o bioma do Pantanal . . . . .	p. 59
6	Resumo da análise dos modelos de Série Temporal para o bioma da Amazônia . . . . .	p. 68
7	Resumo da análise do modelo com somente uma variável de exposição - com a variável Estações . . . . .	p. 76
8	Resumo da análise do modelo com somente uma variável de exposição - sem a variável Estações . . . . .	p. 77
9	Resumo da análise dos melhores modelos completos para o bioma do Pantanal . . . . .	p. 79
10	Diagnóstico da Multicolinearidade do modelo 3 por meio do VIF para o bioma do Pantanal . . . . .	p. 80
11	Resumo da análise do melhor modelo para o bioma do Pantanal . . . . .	p. 82
12	Resumo da análise dos melhores modelos completos para o bioma da Amazônia . . . . .	p. 83
13	Diagnóstico da Multicolinearidade do modelo 12 por meio do VIF para o bioma da Amazônia . . . . .	p. 84

14	Resumo da análise do melhor modelo para o bioma da Amazônia . . . .	p.86
----	---	------

# 1 Introdução

A asma é uma doença inflamatória obstrutiva crônica que se caracteriza através da hiperresponsividade das vias aéreas e pela limitação do fluxo aéreo. Essa doença pode ser reversível por meio de tratamentos ou espontaneamente, manifestando-se a partir de episódios recorrentes de falta de ar, aperto no peito, tosse, sibilância ou desconforto torácico. As crises asmáticas podem ocorrer em consequência de mudanças bruscas na temperatura, poluentes ambientais, infecções respiratórias, ácaros, fungos, entre outros fatores (SALDANHA; BOTELHO, 2008).

Alguns estudos foram realizados na região dos biomas da Amazônia e do Cerrado sobre a prevalência, sazonalidade e variáveis associadas à ocorrência de asma na população. Na região da Amazônia brasileira, um estudo concluiu que “as hospitalizações por asma em crianças são mais frequentes nos meses chuvosos, com maior magnitude nos estados que compõem o arco do desmatamento da Amazônia brasileira, especialmente Rondônia e Maranhão” (SILVA et al., 2009). Quanto aos idosos, a sazonalidade das internações por asma também é significativa, no entanto o período seco é o que apresenta as maiores taxas de internação neste grupo, principalmente em Rondônia e Mato Grosso (RODRIGUES et al., 2010). Um estudo transversal realizado no município de Tangará da Serra, no Estado do Mato Grosso, observou que a prevalência de asma entre os escolares (crianças com 6 e 7 anos de idade) foi de 25,2%, enquanto que entre os adolescentes com idade de 13 e 14 anos foi de 15,9% (ROSA et al., 2009). Já foi demonstrado que existe associação entre as queimadas da floresta Amazônica e do Cerrado, mensuradas pelos focos de calor ou pelas concentrações de poluentes atmosféricos, com a ocorrência de desfechos respiratórios, como a internação e atendimentos ambulatoriais por asma ((SALDANHA; BOTELHO, 2008); (SILVA et al., 2013); (RIBEIRO; ASSUNÇÃO, 2002); (HORN, 2016); (ARBEX et al., 2007); (SILVA et al., 2016); (GOUVEIA et al., 2006); (MENDES et al., 2017); (ALVES, 2020); (SILVA et al., 2009); (SANTOS, 2018); (NASCIMENTO; MEDEIROS, 2012)).

As queimadas que ocorrem frequentemente tanto no bioma Amazônia quanto no bioma Pantanal, de forma natural ou por meio de ações humanas, liberam uma enorme

quantidade de monóxido e dióxido de carbono, metano, material particulado, entre outros. Os poluentes liberados nesses incêndios apresentam efeitos diretos na saúde humana, especialmente para o sistema respiratório (NASCIMENTO; MEDEIROS, 2012), onde as partículas respiráveis ( $PM_{10}$  e  $PM_{2.5}$ , material particulado igual ou abaixo de 10 micra e igual ou abaixo de 2.5 micra respectivamente) são facilmente inaladas e penetram as vias respiratórias ((SALDANHA; BOTELHO, 2008); (LEMES; REBOITA; CAPUCIN, 2020)). A fumaça proveniente das queimadas é também formada por material particulado fino ( $PM_{2.5}$ ), composto tóxico e multielementar gerado pela queima de biomassa. Segundo a Organização Mundial da Saúde (OMS), o  $PM_{2.5}$  provoca diversos efeitos na saúde da população, afetando principalmente o sistema respiratório e cardiovascular. De acordo com a OMS, recomenda-se o limite de  $25\mu g/m^3$  para a média diária (24h) de  $PM_{2.5}$ , para efeitos na saúde humana.

No bioma Amazônia, as queimadas estão inseridas no processo produtivo do local, impulsionando o processo migratório iniciado na década de 1970 (BECKER, 2005). Além disso, a prática das queimadas para uso do solo e coleta de madeira é comum e intensificada todo ano durante a estação da seca (SILVA et al., 2009), (MOUTINHO et al., 2020), (ARTAXO et al., 2005), onde nas propriedades privadas a queima deste material só é permitida se autorizada pelo órgão governamental capacitado, mas grande parte tem sido feita de forma ilegal. Entretanto, para o ano de 2020 em específico, uma parte do que foi derrubado em 2019 não queimou devido às ações de combate realizadas pelos estados e pelo governo federal (MOUTINHO et al., 2020) e com isso essa parte foi somada ao que foi derrubado e queimado em 2020.

No âmbito do bioma Pantanal, a cobertura vegetal predominante é o cerrado (MORAES et al., 2017) e as queimadas já ocorrem na atividade pecuária, durante a época da estiagem, praticadas com o objetivo de promover a renovação dos alimentos aos rebanhos (COUTINHO; CORNÉLIO, 2010). De 2002 até 2015, 74,7% do total de focos de queimadas do bioma Pantanal ocorreram nas áreas de cerrado (MORAES et al., 2017). Desse modo, a pecuária no bioma Pantanal não apresenta tratos culturais específicos, acarretando a degradação do solo, principalmente incêndios, queimadas e desmatamentos para instalação de pastagens (HARRIS et al., 2005).

Com isso, as concentrações de poluentes atmosféricos provenientes da queima de biomassa colaboram para o aumento da morbidade respiratória, especialmente na estação da seca, nos municípios que pertencem ao arco do desmatamento, na região da Amazônia, onde, de acordo com Reddington e colaboradores (2015), as queimadas explicam 80% do

aumento do material particulado fino na atmosfera.

Além desses achados, existem evidências de que esses poluentes emitidos pelas queimadas exercem efeitos na temperatura dos locais, no qual o impacto ambiental das mesmas provoca um elevado grau de evapotranspiração, que por sua vez reduz o teor de umidade da vegetação durante o período seco ((KAUFMAN et al., 1998); (BOND; WOODWARD; MIDGLEY, 2005)). De acordo com Aragão e colaboradores (2020), a temperatura do Oceano Atlântico em 2020 encontrava-se acima da média histórica e, essa característica tende a causar secas no sudoeste da Amazônia, principalmente no estado do Acre e em regiões adjacentes, nas quais podem intensificar os incêndios nas áreas desmatadas, aumentando portanto, a quantidade de poluentes na atmosfera, como o  $PM_{2.5}$ .

As projeções dos cenários de aquecimento evidenciam que as temperaturas extremas serão cada vez mais frequentes nas regiões Norte, Nordeste e Centro-Oeste e, assim, este fenômeno pode intensificar as queimadas e emissão de poluentes (ARAGÃO; JUNIOR; ANDERSON, 2020).

Deste modo, o **objetivo principal** deste trabalho é avaliar a relação entre a internação por asma em crianças e adolescentes (com 14 anos ou menos) e a exposição aos Focos de Queimadas, Temperatura e Material Particulado 2.5 ( $PM_{2.5}$ ) nos biomas Amazônia e Pantanal. Os **objetivos específicos** são:

- (i) Estimar um Modelo Linear Generalizado para avaliar a relação entre a internação por asma e variáveis ambientais em cada bioma;
- (ii) Avaliar a tendência da internação por asma em cada bioma e estimar um modelo preditivo de Séries Temporais;
- (iii) Comparar os resultados observados da internação por asma com os previstos para o ano de 2020;
- (iv) Avaliar a diferença entre os resultados observados da internação por asma com os previstos para o ano de 2020.

## 1.1 Organização

Este projeto final de curso está organizado em capítulos. No Capítulo 2 será apresentado uma revisão da literatura através de uma metodologia sistemática. No Capítulo 3 serão apresentados os materiais e métodos estudados e aplicados neste trabalho. São eles: Modelo Linear Generalizado e Modelo de Séries Temporais para avaliar a tendência

das variáveis e todos os testes relacionados. No Capítulo 4 será realizada uma descrição dos bancos de dados utilizados e os resultados obtidos para os dois métodos de modelos preditivos estudados. No Capítulo 5 serão feitas as conclusões finais. O Apêndice 2 apresenta os gráficos de séries temporais do bioma Pantanal. Já o Apêndice 3 apresenta os gráficos de séries temporais do bioma Amazônia. No Apêndice 4 são apresentados gráficos referentes aos modelos estimados para os dois biomas. Por fim, no Apêndice 5 é apresentado o passo a passo para o download dos dados de temperatura.

## 2 Revisão Bibliográfica

Foi realizada uma revisão da literatura por meio de metodologia sistemática nas bases Pubmed ([www.pubmed.com.br](http://www.pubmed.com.br)) e Scielo ([www.scielo.org](http://www.scielo.org)).

Os artigos pesquisados tiveram restrição de idioma para Português e Inglês. Não houve limite para o ano de publicação do artigo e a pesquisa bibliográfica foi realizada utilizando o método de MeSH Terms nas duas bases a partir das palavras chaves “Asthma AND Brazil AND Temperature”; “Biomass Burning AND Asthma AND Brazil”; “Forest AND Asthma AND Brazil”; “Asthma AND Brazil AND Climate Change”; “Asthma AND Brazil AND Air Pollution” e “Asthma AND Brazil AND Particulate Matter”.

Os critérios de elegibilidade foram:

- Desfecho em asma;
- População de estudo sendo crianças e adolescentes com idade igual ou inferior a 14 anos;
- Região de estudo sendo o bioma da Amazônia, Pantanal e Cerrado.

Nas duas bases de dados foram identificados 289 artigos, dos quais 247 foram observados na base da Pubmed e 42 na base da Scielo. Após ter o título dos 289 artigos, foram verificados, em um primeiro momento, aqueles que estavam duplicados, totalizando 120. Na segunda fase, dos 169 artigos restantes, 167 foram removidos por não compactuar com pelo menos um dos critérios de elegibilidade a partir da leitura do resumo de cada um deles.

Assim, na terceira fase, os 2 artigos que sobraram foram lidos na íntegra e foi realizada uma busca em suas referências para verificar se haveria mais algum que estivesse dentro dos critérios de elegibilidade e que não haviam aparecidos na busca. Dessa forma, foi acrescentado mais 1 artigo.

Por fim, também foi utilizado um método livre nas bases Pubmed ([www.pubmed.com.br](http://www.pubmed.com.br)), Scielo ([www.scielo.org](http://www.scielo.org)) e Google Scholar ([www.scholar.google.com.br](http://www.scholar.google.com.br)) para busca de mais artigos que compreendessem os critérios de elegibilidade atribuídos através das combinações de palavras “Asma”, “Queimadas”, “Amazônia”, “Pantanal”, “Doenças Respiratórias”, “Amazônia Asma”, “Pantanal Asma”, “Queimadas Asma”, retornando assim 4 artigos que se encaixavam nesses critérios, sendo eles os 4 últimos da Tabela 1 e totalizando 7 artigos estudados.

A Figura 1 abaixo apresenta, através do método PRISMA analisado a partir de (MOHER et al., 2010) e pelo sítio PRISMA, um fluxograma mostrando a seleção dos artigos para revisão sistemática, enquanto a Tabela 1 abaixo apresenta um resumo dos artigos finais.

Dessa forma, ao analisar a Tabela 1, observa-se que, dos 7 artigos apresentados, 2 tratam sobre o material particulado e sua forte correlação com os desfechos das doenças apresentadas nesta tabela. Além disso, dos 7 artigos, 1 desses utilizam modelos aditivos generalizados a partir da distribuição estatística de Poisson para analisar essa correlação e verificar o incremento dado as variáveis explicativas de cada um destes modelos.

Além disso, dos 7 artigos analisados nesta revisão, 4 deles utilizam as variáveis meteorológicas como tipo de exposição para o desfecho da asma. Sendo assim, nota-se que o material particulado fino e as variáveis meteorológicas (Temperatura, Precipitação e umidade relativa do ar) podem ser melhores indicadores para a análise do desfecho de saúde deste trabalho ao se comparar com os focos de queimadas, que se encontram em apenas 1 dos artigos da tabela.

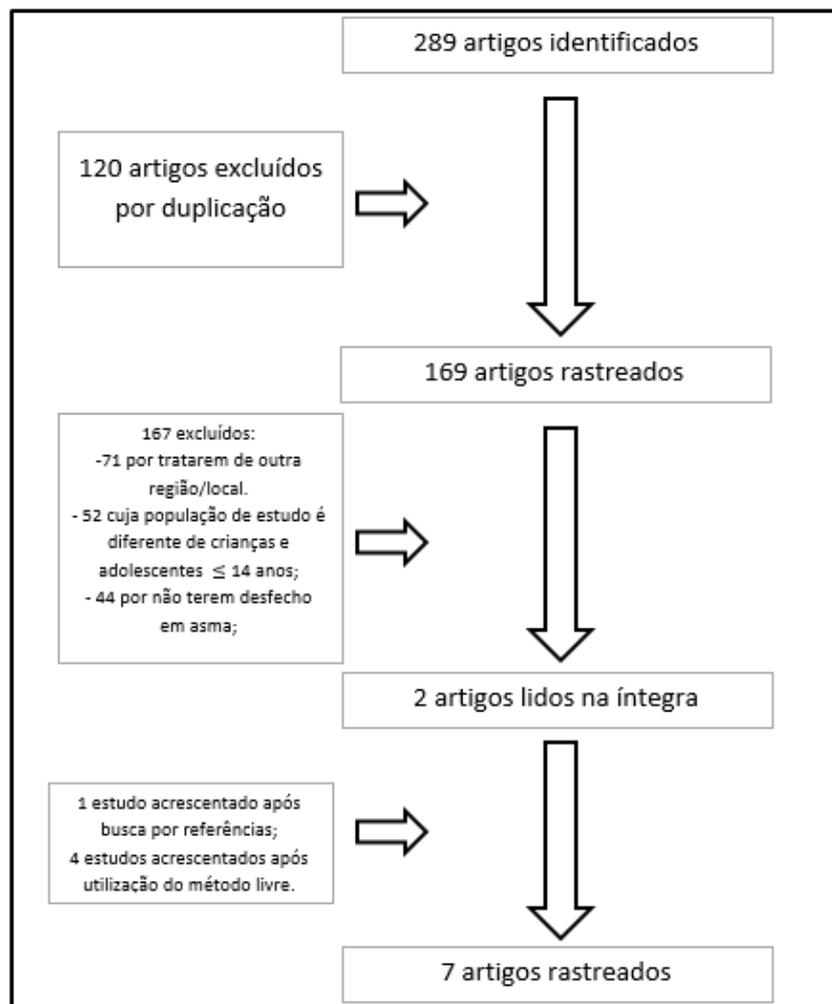


Figura 1: Diagrama de Fluxo PRISMA

Tabela 1: Estudos sobre queimadas e seus poluentes e a relação com a asma no Brasil, com foco nos municípios do bioma Pantanal e Amazônia entre os anos de 2010 a 2020.

<b>Autores</b>	<b>Ano de publicação</b>	<b>Região (país)</b>	<b>População de Estudo</b>	<b>Período do Estudo</b>	<b>Desenho e Tamanho da Amostra</b>	<b>Desfecho</b>	<b>Tipo de Exposição</b>	<b>Método Estatístico</b>	<b>Variáveis de Ajuste</b>
Dias et al.	2020	Belo Horizonte, Minas Gerais, Brasil.	Crianças de 0-14 anos	2002-2012	Estudo observacional descritivo	Autorizações de internações hospitalares (AIHs) por asma.	Temperatura, umidade e precipitação	Correlação de Pearson; Modelo de Regressão Generalizado; Sazonalidade.	Faixa etária, taxas de hospitalização por asma de 0-14 anos e de 0-4 anos, de bronquiolite de 0-4 anos e variáveis meteorológicas.
Rosa et al.	2009	Tangará da Serra, Mato Grosso, Brasil.	Crianças de 6-7 anos e adolescentes de 13-14 anos	Março - abril de 2007	Estudo transversal e descritivo	Prevalência de asma e sintomas de acordo com o gênero e grupo etário.	-	Teste Qui-Quadrado.	Indicador da presença de asma, local de residência sexo, tipo de escola.
Oliveira et al.	2012	Tangará da Serra, Mato Grosso, Brasil.	Crianças de 6-14 anos	Janeiro - dezembro de 2008	Estudo de painel ( $n = 234$ )	Estimação do risco toxicológico em asmáticos.	$PM_{2.5}$ proveniente de queimadas, temperatura, umidade relativa do ar e precipitação.	Simulação de Monte Carlo; Teste T-Student; ANOVA; Correlação de Spearman.	Idade, gênero, indicador de asma, peso e indicador de período de chuva.
Saldanha et al.	2005	Cuiabá, Mato Grosso, Brasil.	Crianças $\geq$ 5 anos	Janeiro - dezembro de 1999	Estudo ecológico	Prontuários de atendimentos em asmáticos.	Variações climáticas.	Teste Qui-Quadrado.	Sexo, faixa etária da idade, diagnóstico, número de atendimentos e indicadora de necessidade de internação.

Autores	Ano de publicação	Região (país)	População de Estudo	Período do Estudo	Desenho e Tamanho da Amostra	Desfecho	Tipo de Exposição	Método Estatístico	Variáveis de Ajuste
Saldanha et al.	2008	Cuiabá, Mato Grosso, Brasil.	Crianças de 0-5 anos	Janeiro - dezembro de 1999	Estudo ecológico	Atendimentos ambulatorial e hospitalar das crianças com asma e outros diagnósticos compatíveis com a asma	Focos de queimadas	Teste Qui-Quadrado; Teste de Kruskal-Wallis; Teste de Fisher-Snedecor.	Prontuários médicos de crianças < de 5 anos agrupadas por mês.
Silva et al.	2009	Estados do bioma Amazônia, Brasil.	Crianças de 0-14 anos	2001-2007	Estudo descritivo	Interações hospitalares por asma.	Temperatura	Indicador de Kernel; Sazonalidade; Distribuição espacial.	Taxas de hospitalização / internação por asma anuais e mensais e proporção das internações por asma agrupadas por período de seca ou chuvoso.
Silva et al.	2016	Rio Branco, Acre Brasil.	Escolares de 6-14 anos	Agosto - outubro de 2009	Estudo de risco	Estimação do risco toxicológico pelas variáveis de exposição nos escolares asmáticos e não asmáticos.	$PM_{2.5}$ e $O_3$	Simulação de Monte Carlo; Modelos probabilísticos; Testes de Kolmogorov-Smirnov; ANOVA.	Sexo, idade, altura, peso, IMC, indicador de presença da asma.

## 3 Materiais e Métodos

### 3.1 Área de Estudo

A área de estudo é composta pelos municípios que abrangem os biomas Amazônia e Pantanal, no período de 2010 a 2020, compreendendo, ao todo, 691 municípios, conforme descrição obtida no Instituto Brasileiro de Geografia e Estatística (IBGE) para o ano de 2019.

O bioma Amazônia ocupa cerca de 49%, 4.196.943 km<sup>2</sup> do território brasileiro abrangendo os Estados do Acre, Amapá, Amazonas, Pará, Rondônia, Roraima e parte dos Estados do Mato Grosso (54%), Maranhão (34%) e Tocantins (9%). Possui a maior floresta tropical do mundo, equivalente a um terço das reservas de florestas tropicais úmidas, que representa cerca de 90% do bioma (MIRANDA; MARTINHO; CARVALHO, 2020), que abrigam a maior quantidade de espécies da flora e da fauna, 1,5 milhão de espécies vegetais catalogadas, 3 mil espécies de peixes, 950 espécies de pássaros e diversas espécies de insetos, répteis e mamíferos, conforme descrição retirada do sítio do IBGE.

O bioma Pantanal, situado na região Centro-Oeste, ocupa aproximadamente 2%, cerca de 150.355 km<sup>2</sup>, do território nacional e abrange parte dos Estados do Mato Grosso e Mato Grosso do Sul conforme descrição retirada do sítio do IBGE. A vegetação predominante é o cerrado e este bioma é reconhecido como a maior planície de inundação contínua do Planeta Terra, o que estabelece o principal fator para a sua formação e diferenciação em relação aos demais biomas (MORAES et al., 2017). Além disso, é o bioma mais preservado, embora a criação de gado, introduzida nessa área desde 1740, seja a atividade econômica principal do bioma (ABDON et al., 2007).

## 3.2 Banco de Dados

Para este trabalho, foi realizado um estudo ecológico descritivo e de séries temporais. Este tipo de estudo analisa uma população ou grupo de indivíduos de um determinado lugar, neste caso dois biomas, com o intuito de avaliar como as situações ambientais e sociais interferem na saúde dos mesmos (CARVALHO; ROCHA, 2008).

Durante todo o trabalho, todas as análises foram realizadas por meio do *software* estatístico R (R Core Team, 2014) na versão 4.0.3 e utilizando a IDE RStudio na versão 1.4.1. Os pacotes que contém as funções utilizadas neste trabalho através do *software* R foram: *readr*, *readxl*, *tidyverse*, *abjutils*, *stringi*, *stringr*, *ggplot2*, *gridExtra*, *splines*, *hnp*, *car*, *rsq*, *TSPred*, *keyring*, *ecmwfr*. Por fim, na análise dos dados foi adotado um nível de significância de 5%.

### 3.2.1 Internações por asma

Os dados sobre internação por asma nos municípios que abrangem os biomas Amazônia e Pantanal foram obtidos através do Departamento de Informática do Sistema Único de Saúde (DATASUS) por local de residência e agrupados por municípios, disponibilizados pelo Ministério da Saúde por meio do Sistema de Informações Hospitalares do SUS (SIH/SUS). Para este estudo, foram utilizadas como unidade de observação o ano e mês de internação, incluídas como variável de desfecho todas as internações com diagnóstico principal de asma a partir da Classificação Internacional de Doenças (CID-10) – **J45** – em crianças e adolescentes de 0 a 14 anos, residentes nos municípios que abrangem os biomas Amazônia e Pantanal e que tenham sido internados em hospitais públicos e/ou conveniados com o Sistema Único de Saúde.

Os dados foram obtidos para todos os meses dos 10 anos de estudo em análise e os mesmos foram filtrados para conterem apenas os municípios reconhecidos pelo sítio do IBGE.

### 3.2.2 Variáveis Meteorológicas

Os dados meteorológicos da temperatura mínima ( $T_{min_{tb}}$ ), média ( $T_{med_{tb}}$ ) e máxima ( $T_{max_{tb}}$ ) mensal, onde  $t = 1, 2, \dots, n$ , onde  $n$  é o total de meses ao longo dos anos de 2010 à 2019 (120) e  $b \in (\text{Amazônia}, \text{Pantanal})$ , foram obtidos através do *European Centre for Medium-Range Weather Forecasts* (ECMWF), onde foram selecionados os dados de

temperatura acima de 2 metros do solo, para toda a região do Brasil através dos dados diários, de modo que estes dados são disponibilizados a cada hora. Com isso, foi feito o *download* destes dados para todos os horários, dias, meses e anos de janeiro de 2010 até dezembro de 2019. Devido a sua complexibilidade e grande taxa de uso computacional, estes dados tiveram que ser tratados e baixados em intervalos de 2 em 2 meses para cada um dos biomas em estudo neste trabalho, para no final serem agrupados e criadas as devidas variáveis de temperatura média, máxima, mínima, percentil 90 e a amplitude da mesma. Dessa forma, um guia que demonstra como os dados foram buscados e baixados se encontra no Apêndice 5.

### 3.2.3 Focos de Queimadas

Os dados diários relacionados aos focos de queimadas  $Focos_{tb}$ , onde  $t = 1, 2, \dots, n$ , onde  $n$  é o total de meses ao longo dos anos de 2010 à 2019 (120) e  $b \in (\text{Amazônia, Pantanal})$ , foram obtidos através do Instituto Nacional de Pesquisas Espaciais (INPE), de forma que foi selecionado o continente América do Sul, país Brasil, todos os estados do mesmo, utilizando o satélite de referência *AQUA M-T* e sendo selecionado todos os biomas para todos os meses (120 meses) para cada ano de estudo. Após o *download* dos arquivos, os dados foram agrupados em um único banco de dados e filtrados para conter apenas os municípios relacionados aos biomas estudados neste presente trabalho de acordo com o sítio do IBGE. Por fim, agrupou-se estes dados para seus respectivos meses para ficar de acordo com as outras variáveis de pesquisa.

### 3.2.4 Material Particulado

Os dados diários de material particulado fino  $PM_{2.5_{tb}}$ , onde  $t = 1, 2, \dots, n$ , onde  $n$  é o total de meses ao longo dos anos de 2010 à 2019 (120) e  $b \in (\text{Amazônia, Pantanal})$ , foram obtidos através do Sistema de Informações Ambientais Integrado a Saúde (SISAM). O SISAM disponibiliza as concentrações do poluente medidos em 4 momentos diferentes do dia. Para este trabalho foi calculado a média aritmética diária dessas quatro medidas. Assim, cada município de estudo pertencente a cada bioma estudado neste trabalho tem uma média diária de concentração de  $PM_{2.5}$ , para o período de 2010 a 2019. Os dados em seguida foram agregados por mês por meio de média aritmética. Por fim, um único banco de dados foi construído com todas as variáveis de pesquisa, por mês e ano.

### 3.3 Métodos

Neste estudo, serão estimados Modelos de Regressão Linear de Poisson, onde as internações por asma em cada bioma de estudo será a variável resposta  $Y_{tb}$ , onde  $Y$  são as internações por asma (variável resposta),  $t = 1, 2, 3, \dots, n$ , onde  $n$  é o total de meses ao longo dos anos de 2010 à 2019 (120) e  $b \in (\text{Amazônia}, \text{Pantanal})$ . Neste tipo de modelagem, a relação entre a variável explicativa e a variável resposta não necessita ser de forma linear. Além disso, também serão estimados Modelos de Séries Temporais, onde serão avaliados a tendência e sazonalidade da variável de desfecho, internação por asma nos municípios de cada um dos biomas em estudo  $Y_{tb}$  e das variáveis ambientais de maneira agregada por bioma. As subseções abaixo descrevem a teoria referente.

#### 3.3.1 Séries Temporais

Uma série temporal é um conjunto de observações analisados ao longo de um tempo  $t$ , que pode ser dia, semana, mês, ano, etc. Estas observações  $t$  são dependentes ao longo do tempo, dado que eventos passados podem influenciar em eventos futuros. Então deseja-se analisar e modelar o comportamento dessa dependência. Além disso, a ordem dos dados neste caso deve ser respeitada.

Em séries temporais, a análise gráfica ao longo do tempo é de suma importância, onde pode revelar diversos padrões de comportamento. Tendências de decrescimento ou crescimento, padrões periódicos, entre outros, são identificados com base nas observações dessas análises.

Sendo assim, algumas características presentes na série temporal  $Y_{tb}$ , onde  $Y$  são as internações por asma (variável resposta),  $t = 1, 2, 3, \dots, n$ , onde  $n$  é o total de meses ao longo dos anos de 2010 à 2019 (120) e  $b \in (\text{Amazônia}, \text{Pantanal})$ , podem ser concebidas levando em conta o modelo de decomposição aditiva, dado da forma:

$$Y_{tb} = T_{tb} + S_{tb} + \epsilon_{tb}, \quad (3.1)$$

onde  $T_{tb}$  representa a tendência da série  $Y_{tb}$ ,  $S_{tb}$  representa a sazonalidade,  $\epsilon_{tb}$  é a componente que representa o erro aleatório,  $t$  representa o tempo da série que abrange os meses dos anos de 2010 até 2020 e  $b$  são os biomas em estudo, neste caso Amazônia e Pantanal. Dessa forma, a tendência é definida como o padrão de crescimento ou decrescimento das internações por asma em um período de tempo  $t$ . Já a sazonalidade é definida como uma

repetição que ocorre em um determinado período idêntico de tempo  $t$ .

### 3.3.1.1 Estacionariedade

Para uma melhor compreensão de um processo estacionário, deve-se reconhecer o que é um processo estocástico. De acordo com Alves (1997), um processo estocástico pode ser descrito como uma sequência de variáveis aleatórias indexadas a uma variável (geralmente a variável tempo), sendo representado por  $Y = \{Y(t), t \in T\}$ , de forma que para  $t \in T$ ,  $Y(t)$  é uma variável aleatória.

Dessa forma, como neste trabalho serão utilizados modelos para representar séries temporais, há a necessidade de reconhecer sobre processo estacionário.

Sendo assim, segundo Morettin (2006) uma série temporal é dita estacionária se ela avança ao longo do tempo aleatoriamente ao redor de uma média constante, retratando alguma forma de equilíbrio estável.

Caso a série temporal em estudo nesse trabalho não seja estacionária, é necessário transformar os dados, de forma que a transformação mais comum, segundo Morettin (2006), consiste em tomar diferenças sucessivas da série original de forma a se obter uma série estacionária, onde a primeira diferença de  $Y(t)$  é dada pela expressão abaixo:

$$\Delta Y(t) = Y(t) - Y(t - 1). \quad (3.2)$$

A segunda diferença é

$$\Delta^2 Y(t) = \Delta[\Delta Y(t)] = \Delta[Y(t) - Y(t - 1)] = Y(t) - 2Y(t - 1) + Y(t - 2). \quad (3.3)$$

E de modo geral, a  $j$ -ésima diferença de  $Y(t)$  é da forma:

$$\Delta^j Y(t) = \Delta[\Delta^{j-1} Y(t)]. \quad (3.4)$$

### 3.3.1.2 Tendência

A tendência, neste trabalho indicado por  $T_{tb}$ , representa o padrão estável ao longo do tempo, livre dos ruídos, que podem ser sazonais ou aleatórios. A tendência pode ser descrita por uma função qualquer sendo a mais simples a função linear ou outra suave.

Supondo que a componente sazonal não esteja presente nesta série temporal em estudo, o modelo de 3.1 pode ser reescrito da forma:

$$Y_{tb} = T_{tb} + \epsilon_{tb}. \quad (3.5)$$

Dessa forma, há vários métodos para estimar a tendência. Os mais utilizados são:

- Suavizar ou filtrar os valores da série ao redor de um ponto e estimar a tendência naquele ponto. A média móvel é dado como um exemplo de suavização ou filtro.
- Ajustar uma função do tempo  $t$ , como uma função linear, ou polinomial, ou exponencial.
- Suavizar os valores da série por meio de sucessivos ajustes de retas de mínimos quadrados ponderados.

Depois de estimar a tendência, pode-se eliminar este componente da série original. No caso da tendência linear estimada por  $\hat{T}_{tb}$  a série sem tendência é estimada da forma:

$$Z_{tb} = Y_{tb} - \hat{T}_{tb}.$$

Além disso, uma outra forma de eliminar a tendência, neste caso a forma utilizada neste trabalho, é através da primeira diferença da série ( $\Delta Y_{tb}$ ) dado por:

$$\Delta Y_{tb} = Y_{tb} - Y_{t-1,b}.$$

### 3.3.1.3 Sazonalidade

A sazonalidade representa os elementos mediadores sobre a tendência, ora elevando seu valor, ora reduzindo-lhe. Consideram-se como sazonais os fenômenos que ocorrem regularmente na série em um mesmo período de tempo  $t$ . Essa sazonalidade são oscilações regulares em torno da linha de tendência.

Existem dois tipos de sazonalidade. São eles:

- Determinística

Quando subentende-se um padrão sazonal regular e estável no tempo, sendo possível, desta forma, prever o comportamento periódico completamente a partir dos dados passados.

- Estocástica

Quando a componente sazonal varia com o tempo, ou seja, para séries cuja sazonalidade é estocástica, onde o método das médias móveis é o método mais apropriado.

Neste trabalho, iremos tratar de sazonalidade determinística, visto que o padrão dos dados é regular e estável no tempo.

### 3.3.1.4 Função de Autocorrelação

A função de autocorrelação (FAC ou ACF) é aplicada neste trabalho e se refere como um procedimento para reconhecer o processo estocástico que gera a série temporal. Esta função é dada por:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \tau \in Y. \quad (3.6)$$

A FAC  $\rho_\tau$  satisfaz as propriedades abaixo:

1.  $\rho_0 = 1$ ;
2.  $\rho_{-\tau} = \rho_\tau$ ;
3.  $|\rho_\tau| \leq 1, \forall \tau \in \mathbb{Y}$ .

### 3.3.1.5 Função de Autocorrelação Parcial

A função de autocorrelação parcial (FACP ou PACF) é um outro método aplicado neste trabalho que funciona como identificação do processo estocástico que mede a correlação entre  $Y_{tb}$  e  $Y_{tb-\tau}$  depois de removido o efeito das observações  $Y_{tb-1}, Y_{tb-2}, \dots, Y_{tb-\tau+1}$ , denotado por  $\Phi_{gg}$ , ou seja,

$$\Phi_{gg} = Cov(Y_{tb}, Y_{tb-\tau} | Y_{tb-1}, Y_{tb-2}, \dots, Y_{tb-\tau+1}). \quad (3.7)$$

### 3.3.1.6 Correlograma

O correlograma é o gráfico composto por valores de autocorrelação (ACF) ou autocorrelação parcial (PACF) em diversas defasagens. Por meio deste gráfico, é possível analisar a aleatoriedade da série assim como identificar possíveis tendências e sazonalidades e identificar também os parâmetros dos modelos de séries temporais, conforme descrito na subseção 3.3.1.9.

Na Figura 2, o eixo horizontal indica a defasagem e o eixo vertical indica a autocorrelação. A linha tracejada em azul indica onde a autocorrelação é significativamente diferente de zero. Ou seja, na figura abaixo, o coeficiente de autocorrelação na primeira defasagem é baixo, próximo de 0,05 e permanece próximo de zero ao longo das defasagens, característico de uma série estacionária, indicando que a série é aleatória.

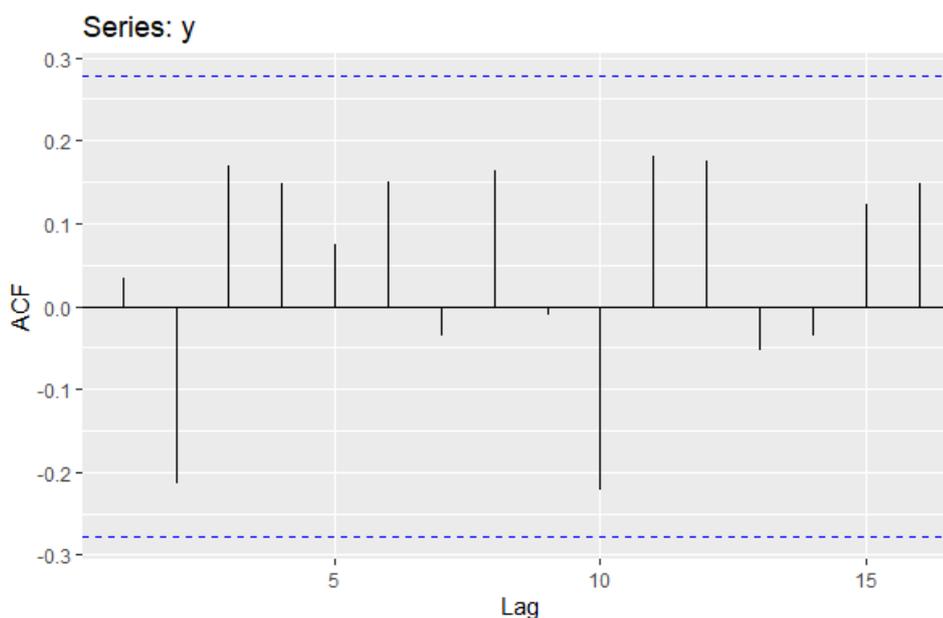


Figura 2: Exemplo de um gráfico de Correlograma

### 3.3.1.7 Teste de Estacionariedade

Para analisar a suposição de estacionariedade da série, foi utilizado neste trabalho o teste de Dickey-Fuller (1979). Neste teste, o interesse é verificar a existência de raízes unitárias na série, o que implica na existência ou não de estacionariedade. Para analisar o que é uma raiz unitária seja o modelo abaixo:

$$Y_{tb} = \rho \cdot Y_{tb-1} + \epsilon_{tb}, \quad (3.8)$$

onde  $\rho$  é coeficiente de auto-correlação que varia de -1 a 1 e  $\epsilon_{tb}$  segue as hipóteses do termo aleatório de média zero, variância  $\sigma^2$  e não autocorrelacionado.

Se na Equação 3.8  $\rho = 1$ , tem-se que  $Y_{tb}$  depende do  $Y_{tb-1}$ , indicando uma alta correlação entre  $Y_{tb}$  e  $Y_{tb-1}$  e então  $Y_{tb}$  é dito como um processo não estacionário. Se for verificado que de fato  $\rho = 1$ , dizemos que a variável  $Y_{tb}$  tem uma raiz unitária. Sendo assim, caso  $\rho = 1$  em 3.8, temos:

$$\Delta Y_{tb} = \epsilon_{tb}, \quad (3.9)$$

indicando então que a série  $\Delta Y_{tb}$  é um ruído branco (erro aleatório) e então é estacionária. Ou seja, a série original não é estacionária porém a série de primeira diferença é. Então, para o teste de Dickey-Fuller tem-se o modelo abaixo em que  $\epsilon_{tb}$  possui média zero e variância  $\sigma^2$ , da forma:

$$\Delta Y_{tb} = \beta_0 + \beta_1 \cdot Y_{tb-1} + \epsilon_{tb},$$

onde  $\beta_1 = \rho_1 - 1$ . E então as seguintes hipóteses devem ser testadas:

$$\begin{cases} H_0 : \beta_1 = 0 \text{ Existência da raiz unitária} \\ H_1 : \beta_1 < 0 \text{ Não existência da raiz unitária} \end{cases}$$

Sob a hipótese nula de que  $\beta_1 = 0$ , o valor estimado  $t$  do coeficiente  $Y_{tb-1}$  segue a estatística  $\tau$  abaixo:

$$\tau = \frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}},$$

onde os valores da estatística  $\tau$  foram tabelados segundo Dickey e Fuller (1979) através de simulações de Monte Carlo e variam em casos de presença somente do intercepto ( $\beta_0$ ), presença somente de tendência e presença de ambos.

Caso não haja estacionariedade, deve-se realizar sucessivas diferenças na série de forma a se obter uma série estacionária.

### 3.3.1.8 Modelos Box-Jenkins

A metodologia de Box e Jenkins (2015) foi utilizada neste trabalho na análise de modelos paramétricos e serviu para regular modelos autorregressivos integrados de médias

móveis (ARIMA)(p,d,q). O método consiste de três etapas, partindo de uma série temporal estacionária:

- Identificação

Este primeiro passo consiste em descobrir os valores apropriados de  $p$ ,  $d$  e  $q$  com base no correlograma, correlograma parcial e outros critérios;

- Checagem do diagnóstico

Este segundo passo consiste em estimar e testar os parâmetros dos melhores ou melhor modelo(s) identificado(s) através de algoritmos computacionais para obter os coeficientes que representam o melhor modelo identificado.

- Adequação do modelo escolhido e previsão

Este passo serve para avaliar a conformidade do modelo, verificando qual o melhor modelo através do AIC, descrito na subseção 3.3.1.10, depois checando se os resíduos assemelham-se ao erro aleatório (ruído branco), possuindo média zero, variância constante e sendo independentes e identicamente distribuídos (i.i.d.). Esta última verificação é feita através do teste de Ljung-Box (1978) e para a normalidade, deve-se ser aplicado o teste de Shapiro-Wilk (1965) nos resíduos da série. Além disso, este último passo de adequação do modelo escolhido e previsão também consiste em prever os valores futuros da série.

Caso o modelo não seja adequado, o passo 1 deve ser repetido e buscar um modelo que melhor adeque os dados.

### 3.3.1.9 Tipos de Modelos

Os modelos *AR*, *MA* e *ARMA*, são modelos que descrevem as séries temporais estacionárias. Porém, diversas séries não são estacionárias, mas são integradas ou possuem uma raiz unitária. Ou seja, se a série  $Y_{tb}$  for integrada de ordem 1, sua primeira diferença torna a série estacionária e nesse caso,  $I(1)$ . Sendo assim, se  $Y_{tb}$  precisar ser diferenciada duas vezes para se tornar estacionária, então  $Y_{tb}$  é  $I(2)$ . Em geral, se a série for  $I(d)$ , depois de diferenciá-la  $d$  vezes, obtém-se uma série estacionária.

Portanto, se for necessário diferenciar uma série temporal  $d$  vezes para torná-la estacionária e aplicar a ela um modelo  $ARMA(p, q)$ , dizemos que a série temporal original é  $ARIMA(p, d, q)$ , ou seja, é uma série temporal auto-regressiva integrada de média móvel

em que  $p$  indica o número de termos auto-regressivos,  $d$  o número de vezes em que a série tem de ser diferenciada para se tornar estacionária e  $q$  o número de termos de média móvel.

Sendo assim, quando os fenômenos que ocorrem durante o tempo se repetem a cada período exato de tempo, pode-se dizer que esta série é sazonal. Se por acaso o período em questão for mensal no decorrer do ano, haverão 12 fatores sazonais, de forma que cada fator seja relacionado a cada mês. Por conta disto, para este trabalho será utilizado o modelo *SARIMA*, que são os Modelos Sazonais Autorregressivos, Integrados e de Médias Móveis, e que, de acordo com Espinosa (2010), este modelo se mostra o mais adequado para realizar tais previsões, pois foram observados efeitos sazonais ao longo dos anos e meses considerados neste estudo. Além de ajustar os dados com relação a tendência e estacionariedade, como o *ARIMA*, este modelo também irá levar em conta a sazonalidade. Com isso, o modelo  $SARIMA(p, d, q)(P, D, Q)_\alpha$ , de forma que os parâmetros  $(p, d, q)$  são a parte não sazonal e os parâmetros  $(P, D, Q)$  são a parte sazonal do modelo, pode ser escrito como:

$$\phi(B)\Phi(B^\alpha)(1 - B^\alpha)^D(1 - B)^d Y_{tb} = \theta(B)\Theta(B^\alpha)\epsilon_{tb}, \quad (3.10)$$

onde

- $B$  é o operador de retardo dado da forma  $B^m Y_{tb} = Y_{tb-m}$ ;
- $\phi(B) = (1 - \phi_1 B^\alpha - \dots - \phi_p B^p)$  é o polinômio auto-regressivo não sazonal ou estacionário de ordem  $p$ ;
- $\Phi(B^\alpha) = (1 - \Phi_1 B^\alpha - \dots - \Phi_p B^{p\alpha})$  é o polinômio auto-regressivo sazonal de ordem  $P$  e periodicidade  $\alpha$ ;
- $(1 - B^\alpha)^D$  é o operador de diferença sazonal em que  $D$  indica o número de diferenças sazonais e periodicidade  $\alpha$ ;
- $(1 - B)^d$  é o polinômio de integração não sazonal de ordem  $d$ ;
- $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$  é o polinômio sazonal de médias móveis de ordem  $q$ ;
- $\Theta(B^\alpha) = (1 - \Theta_1 B^\alpha - \dots - \Theta_q B^{q\alpha})$  é o polinômio sazonal de médias móveis de ordem  $Q$ ;

Para encontrar esses parâmetros, deve-se seguir os seguintes passos:

1. Analisar se a série é estacionária. Caso seja,  $\mathbf{d} = 0$ . Caso não seja, deve-se diferenciar a séries a quantidade de vezes em que for necessária, de forma a obter a estacionariedade, tendo assim o parâmetro  $\mathbf{d}$ .
2. Neste passo deve-se identificar os parâmetro  $\mathbf{p}$  e  $\mathbf{q}$ , que são reconhecidos ao analisar os gráficos da FACP (PACF) e da FAC (ACF), respectivamente.
3. Encontrar o parâmetro  $\alpha$  reconhecido a partir da periodicidade da série.
4. Para este passo, deve-se analisar se a série possui sazonalidade. Caso não possua,  $\mathbf{D} = 0$ . Caso possua, também deve-se diferenciar a série a quantidade de vezes em que for necessária até não haver mais sazonalidade. Porém, a série deve-se ser diferenciada em relação ao instante  $t - \alpha$ . Feito isso, temos o parâmetro  $\mathbf{D}$ .
5. Após a remoção da sazonalidade, deve-se identificar os parâmetros  $\mathbf{P}$  e  $\mathbf{Q}$ . Para isso, deve-se novamente analisar os gráficos da FACP (PACF) e FAC (ACF), respectivamente.

### 3.3.1.10 Análise do Modelo

Conforme dito na subseção 3.3.1.8 e utilizado neste trabalho, ao obter os parâmetros dos modelos em análise pode-se comparar os mesmos através do critério de informação de Akaike, ou AIC, de modo que conforme (XAVIER, 2016), este critério minimiza a divergência de Kullback-Leibler (K-L). A estimativa produzida pelo AIC para um determinado modelo é dada por

$$AIC = -2L + 2K,$$

onde  $L$  é a log-verossimilhança máxima e  $K$  é o número de parâmetros do modelo. Com isso, o modelo com menor valor de AIC é considerado o que possui melhor ajuste.

Após identificar e estimar o modelo que melhor se ajusta às observações da série temporal, deve-se analisar se ele representa adequadamente os dados ou não. A verificação pode ser realizada através da análise dos resíduos.

Sendo assim, primeiro deve-se analisar o correlograma, a função de autocorrelação e o gráfico dos resíduos, que são dados da forma:

$$r_{tb} = \frac{e_{tb}}{\gamma_e(o)}, \quad (3.11)$$

onde  $\gamma_e(o)$  é a variância do resíduo  $e_{tb}$ ,  $t = 1, \dots, T$  e  $b \in (\text{Amazônia, Pantanal})$ .

Dessa forma, também conforme descrito na subseção 3.3.1.8, o teste de Shapiro Wilk (1965) utilizado neste trabalho possui as seguintes hipóteses:

$$\begin{cases} H_0 : \text{Os resíduos seguem uma distribuição normal} \\ H_1 : \text{Os resíduos não seguem uma distribuição normal} \end{cases}$$

Por fim, também foi testado neste trabalho se os resíduos são i.i.d., conforme dito na subseção 3.3.1.8. Para isso, o teste proposto por Ljung e Box (1978), que testa se a FAC (ACF) das defasagens  $1, \dots, H$  são iguais a zero através das hipóteses a seguir:

$$\begin{cases} H_0 : \rho_1 = \rho_2 = \dots = \rho_H = 0 \text{ os resíduos são i.i.d} \\ H_1 : \text{Pelo menos um } \rho_k \neq 0, k = 1, \dots, H \text{ os resíduos não são i.i.d} \end{cases}$$

### 3.3.1.11 Previsão com modelo SARIMA

Para fazer a previsão de um determinado valor  $Y_{tb+h}$ ,  $h \geq 1$  neste trabalho, partindo do pressuposto que todos os valores até o instante  $t$  são conhecidos, ou seja,  $Y_{tb-2}, Y_{tb-1}, Y_{tb}$ , considerando  $t$  como a origem das previsões, temos que, o modelo *SARIMA* obtido em 3.10 pode ter os valores previstos para o instante  $Y_{tb+h}$  da forma:

$$Y_{tb+h} = \Phi_1 Y_{tb+h-1} + \Phi_2 Y_{tb+h-2} + \dots + \Phi_p Y_{tb+h-p} + \epsilon_{tb+h} - \theta_1 \epsilon_{tb+h-1} - \theta_2 \epsilon_{tb+h-2} - \dots - \theta_q \epsilon_{tb+h-q},$$

em que, quanto maior o horizonte de previsão, maior o erro previsto.

## 3.3.2 Modelo de Regressão Linear Generalizado de Poisson

De acordo com Nelder (1972), os Modelos Lineares Generalizados (MLG) são aqueles que se estendem dos modelos lineares de forma a acomodar distribuições de respostas não normais. Neste tipo de modelagem, a variável resposta  $Y$  pode ser categórica ao invés de contínua e sua relação com as variáveis explicativas não precisa ser linear.

Sendo assim, os MLG's são caracterizados por três componentes. São eles:

### 1. Componente Linear

O MLG é composto por um conjunto de  $N$  variáveis aleatórias independentes e identicamente distribuídas condicionalmente ao parâmetro  $\beta$ , da forma  $Y_1, \dots, Y_N$ ,

cada uma com uma distribuição de probabilidade pertencente a família exponencial. Dessa forma, uma distribuição de probabilidade pertence a esta família quando sua função de probabilidade pode ser escrita da forma:

$$f(\mathbf{y}; \theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)], \quad (3.12)$$

onde  $a(y)$ ,  $b(\theta)$ ,  $c(\theta)$  e  $d(y)$  são funções conhecidas.

Dessa forma, neste trabalho, a variável  $Y_{tb}$  é a variável resposta e trata do número de internações por asma em cada um dos biomas em estudo, onde  $b \in (\text{Amazônia}, \text{Pantanal})$ , e como é um dado de contagem em um determinado tempo  $t$ , onde  $t = 1, 2, \dots, n$ , de modo que  $n$  é o total de mês ano (120), esta variável  $Y_{tb}$  possui distribuição de Poisson com parâmetro  $\lambda$ , onde sua função de probabilidade pertence a família exponencial dado que ela pode ser escrita da forma:

$$\begin{aligned} f(\mathbf{y}; \lambda) &= \frac{\lambda^y e^{-\lambda}}{y!}, \quad y = 0, 1, 2, 3, \dots \\ &= \exp[y \log(\lambda) - \lambda - \log(y!)], \end{aligned}$$

onde as funções são:

$$\begin{aligned} a(y) &= y; \\ b(\lambda) &= \log(\lambda); \\ c(\lambda) &= \lambda; \\ d(y) &= \log(y!). \end{aligned}$$

## 2. Componente Sistemático

O componente sistemático é composto pelo preditor linear  $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ , onde  $\mathbf{x}_i^T$  é o vetor de variáveis explicativas referentes a  $i$ -ésima unidade, com dimensão  $1 \times P$ , enquanto  $\boldsymbol{\beta}$  é o vetor de parâmetros desconhecidos do modelo referente, com dimensão  $P \times 1$ .

## 3. Função de Ligação

É uma função monótona e diferenciável denotada por  $g$  de forma que relaciona a média da variável resposta com o preditor linear do modelo da forma:

$$g(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta}; i = 1, 2, \dots, n, \quad (3.13)$$

em que

$$\mu_i = E(Y_{ib})$$

Se a função de ligação for escolhida de forma que  $g(\mu_i) = b(\theta_i)$ , concluí-se que o preditor linear irá modelar diretamente o parâmetro natural, chamada de função de ligação canônica. Para a distribuição de Poisson, sua ligação canônica é dada da forma:

$$\eta = \log(\lambda), \quad (3.14)$$

chamada de função de ligação canônica logarítmica, de forma que esta função aceite a restrição  $\lambda > 0$ .

Sendo assim, a partir da função de ligação, obtém-se que  $\mu_i = \exp(x_i^T \beta)$ . E com isso, o Risco Relativo associado a variável explicativa  $x_k$  é dado da forma  $\exp(\beta_k)$ , onde resultará em um efeito multiplicativo  $\exp(\beta_k)$  na taxa  $\mu$ , onde o resultado deverá ser multiplicado por 100 para obter os valores em porcentagem, de modo a saber quantos % a variável explicativa afeta na variável resposta.

Além disso, a partir do Risco Relativo obtido, pode-se calcular o Intervalo de Confiança para o  $\beta_{ib}$ , onde neste trabalho foi feito o IC de 95% para  $\beta_{ib}$  através da fórmula:

$$IC(RR) = \exp IC(\beta_{ib}) = \beta_{ib} \pm z * st(\beta_{ib}), \quad (3.15)$$

onde  $\beta_{ib}$  é o valor da estimativa da variável explicativa  $i$  do bioma  $b$  encontrada no modelo,  $z$  é o valor tabelado da distribuição normal padrão (neste caso é 1,96) e  $st(\beta_{ib})$  é o Erro Padrão estimado da variável explicativa  $i$  do bioma  $b$  do modelo.

### 3.3.2.1 Estimação

Para obter as estimativas pontuais e intervalares dos parâmetros de um Modelo Linear Generalizado, é utilizado o método de máxima verossimilhança. Este método tem por objetivo estimar os  $\beta$ 's que estão relacionados aos  $Y_{bi}$ 's de forma que para cada  $Y_{bi}$ , a função de log-verossimilhança é dada por:

$$l_i = l(\theta_i; \mathbf{y}_{bi}) = y_{bi}c(\theta_i) + d(\theta_i) + f(y_{bi}), \quad (3.16)$$

onde  $c$ ,  $d$  e  $f$  são funções conhecidas e

$$\begin{aligned} E(Y_i) &= \mu_i = -\frac{d'(\theta_i)}{c'(\theta_i)}; \\ \text{Var}(Y_i) &= \frac{[c''(\theta_i)d'(\theta_i) - d''(\theta_i)c'(\theta_i)]}{[c'(\theta_i)]^3}; \\ g(\mu_i) &= \mathbf{x}_i^T \boldsymbol{\beta} = \eta_i. \end{aligned}$$

E então a função de log-verossimilhança para todos os  $Y'_{bi}$ s é da forma:

$$\ln L(\boldsymbol{\theta}; \mathbf{y}) = \sum_{i=1}^N li = \sum y_{bi}c(\theta_i) + \sum d(\theta_i) + \sum f(y_{bi}). \quad (3.17)$$

E, como o modelo do estudo se trata de uma distribuição de Poisson, a função de log-verossimilhança é dada por:

$$\ln L(\boldsymbol{\lambda}; \mathbf{y}) = \prod_{i=1}^N \frac{e^{-\lambda_i} \lambda_i^{y_{bi}}}{y_{bi}!} = \sum_{i=1}^N [\lambda_i + y_{bi} \ln(\lambda_i) - \ln(y_{bi}!)]. \quad (3.18)$$

Para achar o estimador de máxima verossimilhança para  $\boldsymbol{\beta}$ , deve-se derivar a equação de log-verossimilhança em relação a cada  $\beta_k$  e igualar a zero.

$$\frac{\partial}{\partial \beta} \ln L(\boldsymbol{\beta}; \mathbf{y}) = 0.$$

No entanto, como não há uma solução analítica para o ponto de máximo desta equação acima, foi utilizado neste trabalho um método iterativo para encontrar uma aproximação para a estimativa de máxima verossimilhança de  $\boldsymbol{\beta}$  através da utilização da matriz de informação de Fisher, conforme escrito por (DOBSON; BARNETT, 2018).

### 3.3.2.2 Adequabilidade do modelo

Para avaliar a adequabilidade de um modelo, de acordo com Dobson (2018), pode-se comparar o mesmo com um modelo geral com o número máximo de parâmetros que podem ser estimados. Este modelo deve possuir as seguintes características:

- Possuir a mesma distribuição que a do modelo de interesse;
- Ter a mesma função de ligação que o modelo de interesse;

- O número total de parâmetros neste modelo deve ser igual ao total de observações  $N$ .

Dadas as  $N$  observações, pode-se ajustar um modelo que contenha até  $N$  parâmetros. Sendo assim, o modelo mais simples, conhecido como **modelo nulo**, tem apenas um parâmetro, de forma que represente uma média comum para os  $Y$ 's. Entretanto, no outro extremo, o **modelo completo** terá  $N$  parâmetros, um para cada observação.

Dessa forma, para o **modelo completo**, algumas observações podem ter o mesmo componente linear ou padrão de covariável e assim, o número máximo de parâmetros que podem ser estimados no **modelo saturado (modelo completo)** é igual ao número de combinações lineares potencialmente diferentes, podendo ser menor que  $N$ .

### 3.3.2.3 Análise dos Resíduos

Na modelagem, a análise dos resíduos facilita a investigação dos aspectos específicos do modelo. Sendo assim, esta análise abrange técnicas para investigação da adequabilidade de um modelo de regressão a partir dos resíduos. Dessa forma, os pressupostos dos modelos que foram validados e analisados neste trabalho para que os resultados da análise e do modelo sejam confiáveis foram os de **normalidade** e **homocedasticidade**.

- Gráfico normal de probabilidade

De acordo com Fernandes (2019), o gráfico de probabilidade é uma técnica gráfica utilizada para realizar comparações de duas distribuições de probabilidade. Para isso, o gráfico quantil (QQplot) se torna bastante utilizado e equivale a dispor no gráfico os quantis de duas variáveis de interesse, de forma que se ambas forem identicamente distribuídas, espera-se verificar um padrão linear dos pontos.

Porém, como este trabalho está tratando de uma distribuição de Poisson, observa-se que a distribuição dos resíduos não é proveniente de uma distribuição normal. Dessa forma, é recomendado o uso de simulação e construção de envelopes simulados (CORDEIRO; DEMÉTRIO, 2008) para verificar a disposição dos pontos dentro deste envelope. Um exemplo de envelope simulado que indica que os resíduos estão dispersos dentro dele, sem apresentar qualquer padrão sistemático, dando indício de que o modelo está bem ajustado se encontra na Figura 3.

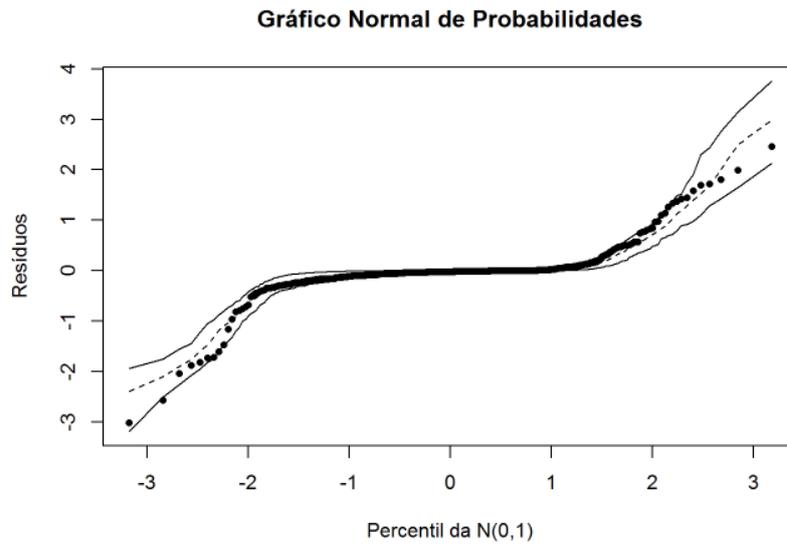


Figura 3: Exemplo do uso de envelope simulado

Sendo assim, se o gráfico se encontrar sem qualquer padrão sistemático, como na Figura 3, tem-se que o pressuposto é aceito.

- Homocedasticidade

Para avaliar a homocedasticidade dos dados neste trabalho, foi verificado o gráfico dos valores ajustados *versus* os resíduos estudentizados. Sendo assim, caso o gráfico possua padrões nos pontos, temos que a hipótese de homocedasticidade não é satisfeita, mas caso os pontos se comportem de forma aleatória, a homocedasticidade é satisfeita. Um exemplo onde o pressuposto de homocedasticidade é aceito, é visto na Figura 4.

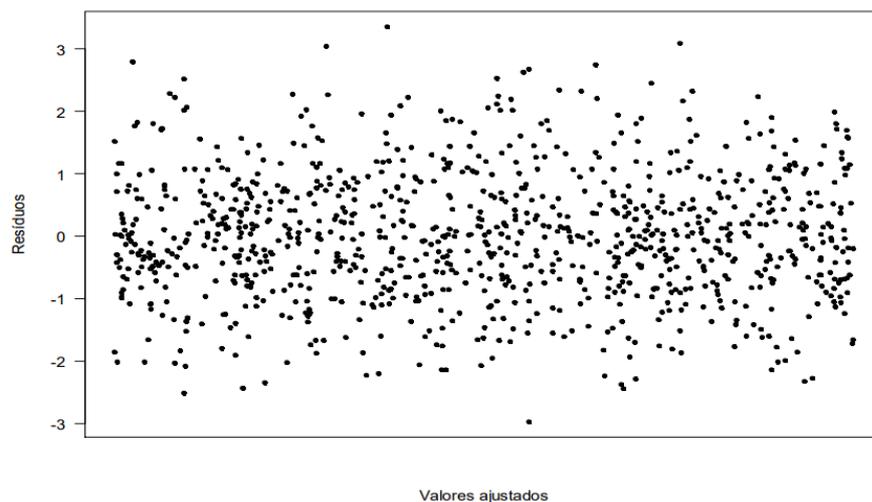


Figura 4: Exemplo de homocedasticidade aceita

- Observações Influentes

Para a análise de observações influentes, foi utilizado a distância de Cook, onde segundo (DOBSON; BARNETT, 2018) é da forma

$$D_i = \frac{1}{p} \left( \frac{h_{ii}}{1 - h_{ii}} \right) r_i^2,$$

em que  $h_{ii}$  é o  $i$ -ésimo elemento da diagonal da matriz de projeção  $H = X(X^T X)^{-1} X^T$  e  $r_i^2$  são os resíduos padronizados. Dessa forma, caso seja observado valores acima de 1, haverá indícios de que estas observações atrapalhem o ajuste do modelo, indicando que as mesmas devem ser retiradas e o modelo deva ser reajustado.

### 3.3.2.4 Critério de Informação de Akaike (AIC)

O critério de Informação de Akaike, segundo (DOBSON; BARNETT, 2018) e utilizado neste trabalho, é uma estatística de qualidade de ajuste baseada na função de verossimilhança, com penalização para o número de parâmetros estimados. Sendo assim, o AIC é dado da forma:

$$AIC = -2l(\hat{\boldsymbol{\pi}}; \mathbf{y}) + 2p,$$

onde  $p$  é o número de parâmetros estimados.

Porém, como observou-se problemas de superdispersão neste trabalho, localizado na subseção 3.3.2.7, este critério teve de ser corrigido segundo (GASPARRINI; ARMSTRONG; KENWARD, 2010), de modo que o AIC foi ajustado por meio de quasi-verossimilhança, dado por

$$QAIC = -2L(\hat{\boldsymbol{\theta}}) + 2\hat{\phi}k,$$

onde  $L$  é o log-verossimilhança do modelo ajustado com os parâmetros  $\hat{\boldsymbol{\theta}}$  e  $\hat{\phi}$  é o parâmetro de sobredispersão estimado, enquanto  $k$  é o número de parâmetros.

Sendo assim, o modelo com menor valor de  $AIC$  ou  $QAIC$  é considerado o que possui melhor ajuste.

### 3.3.2.5 Coeficiente de Determinação

- Coeficiente de Determinação Múltipla

Este coeficiente denotado por  $R^2$  serve para avaliar o quão bem o modelo ajustado utilizado representa os dados observados, ou seja, este coeficiente avalia como a proporção da variação total da variável resposta  $Y_b$  é explicado pelo modelo ajustado.

É dado da forma:

$$R^2 = \frac{SQR_{eg}}{SQT},$$

onde  $SQR_{eg}$  é a soma dos quadrados da regressão e  $SQT$  é a soma dos quadrados totais.

Este coeficiente varia de entre 0 e 1, de forma a ser representado em porcentagem, ou seja,  $0\% \leq R^2 \leq 100\%$ . Então quanto mais próximo de 1 (100%) estiver o  $R^2$ , melhor será o ajuste do modelo.

- Coeficiente de Determinação Múltipla Ajustado

Embora o  $R^2$  aumente ao adicionar uma nova covariável ao modelo, isso não significa que este novo modelo é melhor que o anterior. Sendo assim, para corrigir isto, é definido o **coeficiente de determinação ajustado**, denotado por  $R_a^2$ , dado da forma:

$$R_a^2 = 1 - \left( \frac{n-1}{n-p} \right) (1 - R^2),$$

em que  $R_a^2 < R^2$ , menos quando  $R^2 = 1$ . Além disso, este coeficiente também varia de 0 a 1 e é representado em porcentagem da forma  $0\% \leq R_a^2 \leq 100\%$ .

Sendo assim, em analogia ao  $R^2$ , para o caso de regressão múltipla e utilizado neste estudo, outra estatística de interesse é:

$$pseudo R^2 = \frac{l(\tilde{\boldsymbol{\pi}}; \mathbf{y}) - l(\hat{\boldsymbol{\pi}}; \mathbf{y})}{l(\tilde{\boldsymbol{\pi}}; \mathbf{y})},$$

que representa ganhos proporcionais na função de log-verossimilhança devido aos termos usados do modelo de interesse, quando comparado ao modelo nulo, também variando de 0 a 1 (0% a 100%).

### 3.3.2.6 Multicolinearidade

A multicolinearidade é um problema observado constantemente ao se estimar um modelo de regressão. De acordo com Dobson (2018), é caracterizada por uma quase dependência linear entre as variáveis regressoras, ou seja, trata-se de uma alta correlação linear entre as variáveis explicativas. Sendo assim, a eliminação de algumas variáveis do banco de dados que possuem altas correlações melhora as estimativas dos parâmetros e diminui sua variância, melhorando a capacidade preditiva do modelo.

Para o diagnóstico da multicolinearidade, o Fator da Inflação da Variância (VIF) é uma medida estatística que foi utilizada nos modelos deste trabalho. Essa medida é da forma:

$$VIF_j = \frac{1}{1 - R_j^2}, \quad j = 1, 2, \dots, p - 1, \quad (3.19)$$

onde  $R_j^2$  é o pseudo coeficiente de determinação do modelo de regressão  $j$ , dado que são modelos lineares generalizados, que não possuem uma distribuição normal.

Com isso, quanto mais próximos de 1 os valores de  $VIF_j$  forem, menor a preocupação com a multicolinearidade e seus efeitos. Além disso, valores de  $VIF_j > 5$  são considerados indicadores de sérios problemas de multicolinearidade e dessa forma, as variáveis que apresentarem isto devem ser retiradas do modelo.

### 3.3.2.7 Superdispersão

Diferente de outras distribuições, a distribuição de Poisson possui a propriedade de que  $Y \sim Pois(\lambda) \Rightarrow E(Y) = \lambda = Var(Y)$ . Porém, quando se observa que  $E(Y) < Var(Y)$ , ocorre a superdispersão.

Dessa forma, como se está tratando de forma incorreta a incerteza associada ao modelo, as variâncias serão pobremente estimadas e algumas das consequências relacionadas a esta superdispersão serão avaliações incorretas a respeito dos parâmetros dos modelos, pois algumas de suas causas são variáveis não incluídas no modelo, falta de independência entre as observações entre outros.

Com isso, a solução utilizada neste trabalho para remover este problema de superdispersão foi considerar a utilização de uma forma funcional diferente para o modelo, onde neste caso optou-se pela *Quasi-Poisson* por causa do seu fácil acesso em *softwares*.

De acordo com (HOEF; BOVENG, 2007), o modelo *Quasi-Poisson* pode ser definido da mesma forma para qualquer Modelo Linear Generalizado, de modo que

$$E(Y_{tb}) = \lambda \quad e \quad Var(Y_{tb}) = \phi\lambda,$$

onde  $E(Y_{tb})$  é a esperança de  $Y_{tb}$  e  $Var(Y_{tb})$  é a variância de  $Y_{tb}$ ,  $\lambda$  é o parâmetro da distribuição e  $\phi$  é o parâmetro de dispersão, que para o caso de superdispersão assume valores maiores que 1.

### 3.3.2.8 Natural Cubic Splines

Para suavizar a série ajustando sua tendência e sazonalidade, foi utilizado o conceito de Natural Cubic Splines (PERPEROGLU et al., 2019), de modo que é indicado por uma função polinomial por partes onde neste trabalho foi utilizado a variável tempo para estimar a tendência e sazonalidade. Sendo assim, o domínio dos dados é dividido em intervalos menores, delimitados pelos nós, onde neste trabalho foram utilizados 2 nós por ano, obtendo-se ao todo 20 nós, dado que são 10 anos de estudo. Foi optado por essa escolha de nós visto que obteve-se um melhor ajuste nos modelos ao comparar com outros que possuíam diferentes quantidades de nós.

## 4 Análise dos Resultados

Neste capítulo serão apresentados as análises dos resultados. Estas análises foram realizadas utilizando o *software* estatístico R (R Core Team, 2014) e foi adotado um nível de significância de 5% em toda a análise.

### 4.1 Análise Descritiva

Os dados utilizados neste trabalho foram coletados entre janeiro de 2010 até dezembro de 2020, para as variáveis de internação por asma e focos de queimadas, enquanto as variáveis de temperatura e material particulado fino só possuem dados de janeiro de 2010 até dezembro de 2019, nos locais descritos na seção 3.1. O banco de dados completo de cada um dos biomas estudado neste trabalho (Amazônia e Pantanal) contém respectivamente 120 linhas e 13 colunas, da forma que cada linha representa um ano e mês específico e suas respectivas variáveis de estudo. As variáveis se encontram nas duas bases de dados dos dois biomas em estudo, logo generalizando, a Tabela 2 apresenta o nome destas variáveis.

Tabela 2: Variáveis de estudo

Variáveis	Unidades de Medida	Descrição	Fonte dos dados
Internações por asma	números	Internações por asma em crianças de 0 a 14 anos em cada bioma Amazônia e Pantanal	DATASUS - SIH/SUS
Focos de Queimadas	números	Focos de Queimadas em cada bioma Amazônia e Pantanal	INPE
$PM_{2.5}$ média	$\mu g/m^3$	Média mensal do $PM_{2.5}$ em cada bioma Amazônia e Pantanal	SISAM
$PM_{2.5}$ max	$\mu g/m^3$	Máxima mensal do $PM_{2.5}$ em cada bioma Amazônia e Pantanal	SISAM
$PM_{2.5}$ min	$\mu g/m^3$	Mínima mensal do $PM_{2.5}$ em cada bioma Amazônia e Pantanal	SISAM
$PM_{2.5}$ p90	$\mu g/m^3$	Percentil 90 mensal do $PM_{2.5}$ em cada bioma Amazônia e Pantanal	SISAM
Temperatura média	$^{\circ}C$	Média mensal da Temperatura em cada bioma Amazônia e Pantanal	ECMWF
Temperatura max	$^{\circ}C$	Máxima mensal da Temperatura em cada bioma Amazônia e Pantanal	ECMWF
Temperatura min	$^{\circ}C$	Mínima mensal da Temperatura em cada bioma Amazônia e Pantanal	ECMWF
Temperatura p90	$^{\circ}C$	Percentil 90 mensal da Temperatura em cada bioma Amazônia e Pantanal	ECMWF
Temperatura amplitude	$^{\circ}C$	Amplitude mensal da Temperatura em cada bioma Amazônia e Pantanal	ECMWF

### 4.1.1 Pantanal

Foi analisado primeiramente na Tabela 3, as medidas resumo das variáveis em estudo (Internações por asma, Focos de Queimadas,  $PM_{2.5}$  média e Temperatura média) para o bioma Pantanal.

Tabela 3: Resumo das variáveis de estudo do bioma Pantanal de jan/2010 até dez/2019

Variáveis	Mínimo	Mediana	Média	Máximo	Desvio Padrão
<b>Internações por asma</b>	2 (jan/2016)	10	12,82	67 (mar/2010)	10,02
<b>Focos de Queimadas</b>	2 (mar/2011)	134,50	462,30	4.096 (set/2019)	727,46
<b><math>PM_{2.5}</math> média</b>	4,41 (jul/2013)	9,59	12,84	125,71 (set/2010)	12,97
<b>Temperatura média</b>	21,43 (jun/2016)	26,90	26,40	30,17 (out/2014)	2,02

A Tabela 3 mostra que tanto a variável resposta Internações por Asma quanto a variável explicativa Focos de Queimadas possuem seus valores mínimos de 2, onde para as Internações por Asma, seu mínimo foi obtido em janeiro de 2016 e para Focos de Queimadas foi em março de 2011. Já os valores máximos foram 67, observado em março de 2010, e 4.096, observado em setembro de 2019, respectivamente. O  $PM_{2.5}$  médio e temperatura média tiveram uma amplitude de  $121,30\mu g/m^3$  e  $8,74^{\circ}C$  respectivamente.

Além disso, foi observado que o máximo do  $PM_{2.5}$  médio foi de  $125,71\mu g/m^3$  analisado em setembro de 2010, onde foi 5 vezes maior do que o indicado pelo OMS, que recomenda o limite de  $25\mu g/m^3$  para a média diária (24h) para efeitos na saúde humana.

Para verificar a distribuição das variáveis em estudo, foi realizado um gráfico de histograma com a densidade, onde a mesma é localizada pela linha em vermelho, dado que as variáveis são todas numéricas. Além disso, para a determinação das classes do histograma, foi utilizado a regra de Sturges (LANA et al., 2013), que é dado da forma:  $K = 1 + 3,3 * \log(n)$ , onde  $K$  é o número de classes e  $n$  é a quantidade de indivíduos na base de dados. Após definir  $K = 17$ , foi obtido a Figura 5 e 6.

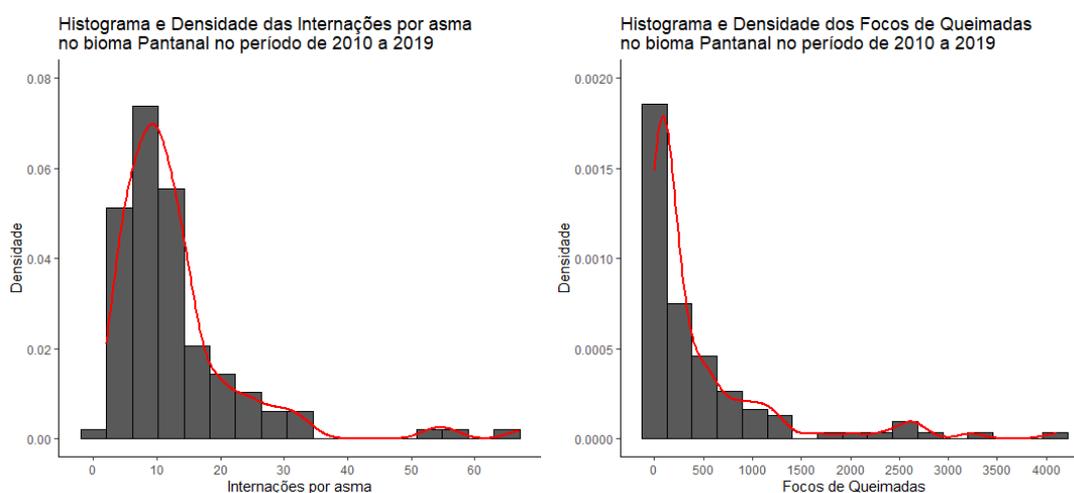


Figura 5: Histograma e Densidade das Internações por Asma e Focos de Queimadas nos municípios que abrangem o bioma do Pantanal de jan/2010 até dez/2019

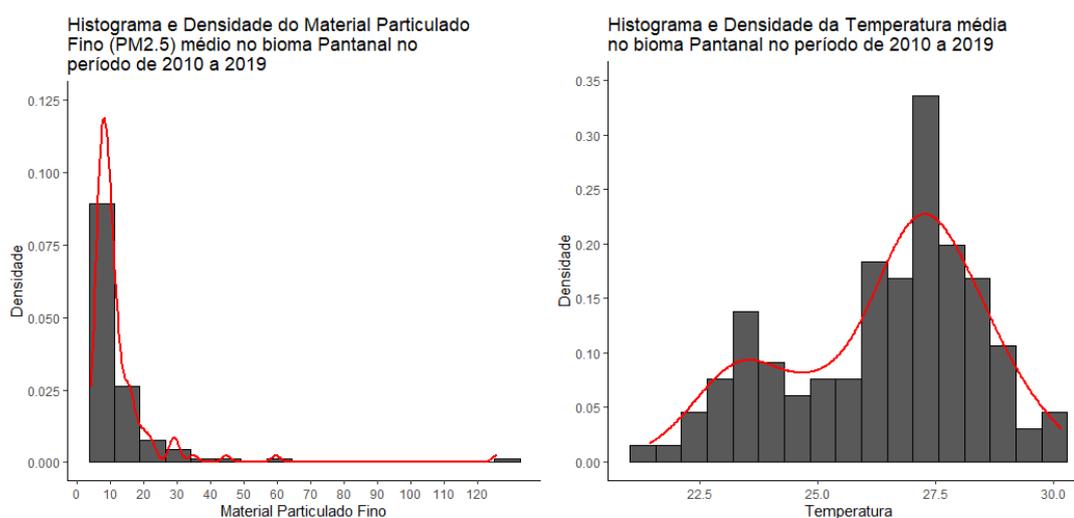


Figura 6: Histograma e Densidade do Material Particulado Fino médio e Temperatura Média nos municípios que abrangem o bioma do Pantanal de jan/2010 até dez/2019

As Figuras 5 e 6 remetem que tanto os Focos de Queimadas quanto o  $PM_{2.5}$  média

possuem uma grande quantidade de medidas com zeros, fazendo com que sua distribuição fique assimétrica a direita.

Além disso, foi observado que a variável resposta Internações por Asma possui uma grande quantidade de dados localizados entre 0 e 15, se assemelhando em partes com as outras distribuições de Focos de Queimadas e  $PM_{2.5}$  médio, dado a assimetria a direita. Já a variável de Temperatura média pode ser identificada possuindo uma distribuição bimodal, dado que possui dois picos de temperatura ao analisar seu histograma e sua densidade.

Para verificar a variabilidade dos meses, foram feitos os boxplots que estão na Figura 7, em que observa-se uma maior quantidade e variabilidade dos Focos de Queimadas nos meses de julho a outubro. Por outro lado, foi analisado que há uma maior quantidade e variabilidade nas Internações por Asma nos meses de março a junho, o que indica que as altas em Internações por Asma ocorrem em momentos diferentes ao se comparar com as altas de Focos de Queimadas. Já para o Material Particulado Fino ( $PM_{2.5}$ ) médio, também foi examinado um aumento entre os meses de julho a outubro, possivelmente causado por conta do aumento das queimadas nesta época. Por fim, ao analisar o boxplot de Temperatura média, há uma queda da mesma durante o inverno e um aumento entre os meses de setembro a janeiro. Os pontos sinalizados em vermelho em cada mês no gráfico representam suas respectivas médias. As Figuras 37 e 38 localizadas no Apêndice 1 indicam os boxplots referentes ao restante das variáveis para o bioma Pantanal.

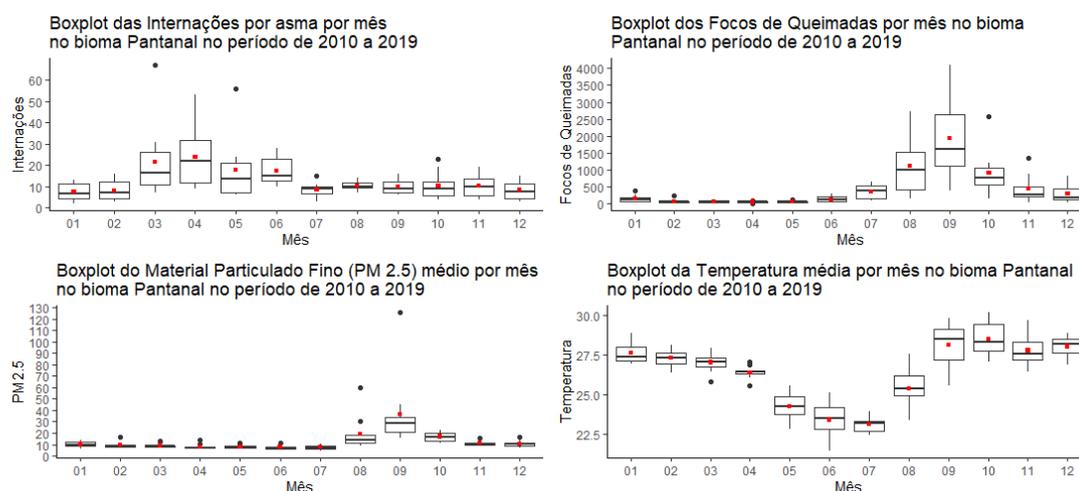


Figura 7: Boxplot das Internações por Asma, Focos de Queimadas,  $PM_{2.5}$  médio e Temperatura média no bioma do Pantanal agrupados por mês de jan/2010 até dez/2019

Na Figura 8 tem-se a série temporal de Internações por Asma e Focos de Queimadas, onde foi observado que os Focos de Queimadas não indicam ter uma forte relação com as

Internações por Asma em indivíduos de 14 anos ou menos nos municípios que abrangem o bioma Pantanal, de modo que os Focos indicam apresentar uma sazonalidade, aumentando todo ano entre julho a outubro, conforme visto na Figura 7. As Internações por Asma, por outro lado, indicam apresentar uma tendência linear decrescente, onde se observa estarem diminuindo a cada novo ano, também observado na Figura 8.

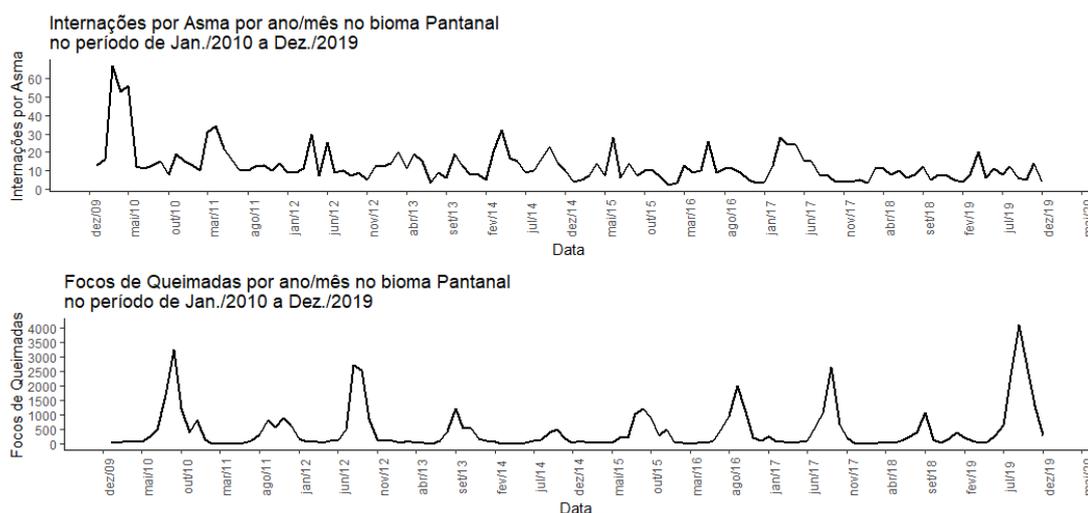


Figura 8: Série Temporal das Internações por Asma e Focos de Queimadas no bioma do Pantanal de jan/2010 até dez/2019

Já na Figura 9 tem-se a série temporal de Material Particulado Fino ( $PM_{2.5}$ ) médio e Temperatura média no bioma Pantanal, onde foi observado que o  $PM_{2.5}$  médio obteve um pico de aproximadamente  $130\mu g/m^3$  no ano de 2010, e após este ano aparenta apresentar certa sazonalidade, aumentando entre agosto a outubro, conforme foi analisado na Figura 7. Para a Temperatura média também foi observada certa sazonalidade porém, por outro lado, indica uma sazonalidade com picos ao fundo, indicando que a Temperatura média diminui nos meses de maio a julho, conforme visto na Figura 7. Além disso, as Figuras 39 e 40 localizadas no Apêndice 1 indicam as séries temporais referentes ao restante das variáveis para o bioma Pantanal.

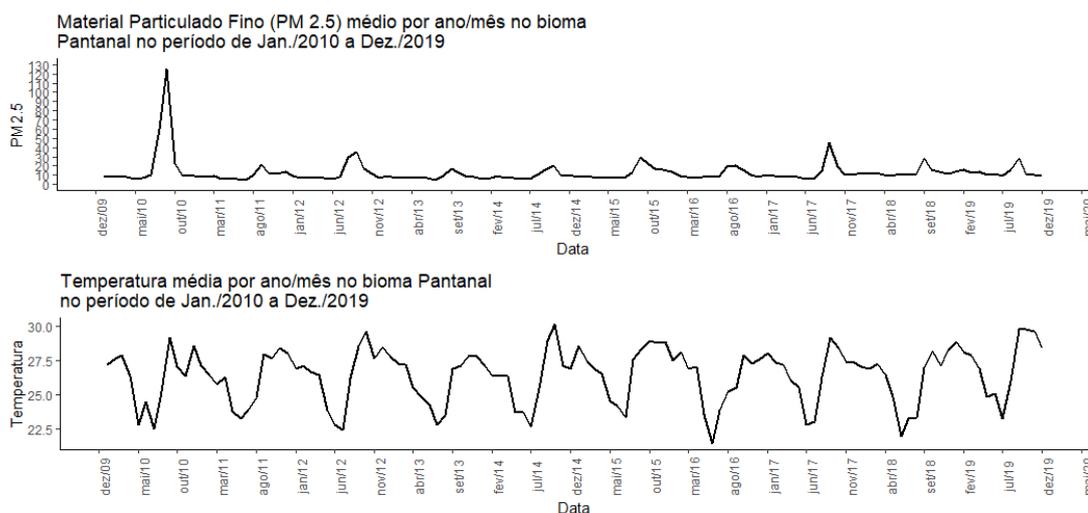


Figura 9: Série Temporal do Material Particulado Fino médio e Temperatura média no bioma Pantanal de jan/2010 até dez/2019

#### 4.1.2 Amazônia

Já para o bioma Amazônia, também foi analisado primeiramente as medidas resumo das variáveis em estudo (Internações por asma, Focos de Queimadas,  $PM_{2.5}$  média e Temperatura média) que são apresentadas na Tabela 4.

Tabela 4: Resumo das variáveis de estudo do bioma Amazônia de jan/2010 até dez/2019

Variáveis	Mínimo	Mediana	Média	Máximo	Desvio Padrão
<b>Internações por asma</b>	608 (jan/2018)	1.422	1.513	3.142 (abr/2010)	586,37
<b>Focos de Calor</b>	432 (fev/2011)	3.490	10.039	55.868 (set/2010)	11.767,44
<b><math>PM_{2.5}</math> média</b>	7,44 (jul/2013)	13,61	15,95	67,04 (set/2010)	8,52
<b>Temperatura média</b>	24,46 (fev/2012)	25,66	25,84	28,45 (set/2015)	0,89

Foi observado nesta Tabela 4 que, diferente do bioma Pantanal, as Internações por Asma e Focos de Queimadas tem seus mínimos sendo 608, observado em janeiro de 2018, e 432, observado em fevereiro de 2011, e máximos sendo 3.142, observado em abril de 2010, e 55.868, analisado em setembro de 2010, respectivamente. Além disso, foi visto que, dado que a diferença entre o mínimo e máximo dos Focos de Queimadas é acima de 50.000, tem-se que sua variância também foi alta (acima de 100.000), dado que seu desvio padrão foi de 727,46, o que também foi visto ao observar o desvio padrão da variável Internações por Asma. Já para o  $PM_{2.5}$  médio e Temperatura média foi observado que os mesmos tiveram uma amplitude de  $59,6\mu\text{g}/\text{m}^3$  e  $3,44^\circ\text{C}$  ao verificar suas mínimas e

máximas respectivamente, o que indica que suas amplitudes são menores ao comparar com o bioma Pantanal.

Além disso, foi observado que o máximo do  $PM_{2.5}$  média foi de  $67,04\mu g/m^3$  analisado em setembro de 2010, assim como para o bioma Pantanal, onde foi 3 vezes maior do que o indicado pelo OMS, que recomenda o limite de  $25\mu g/m^3$  para a média diária (24h) para efeitos na saúde humana.

Também para o bioma Amazônia, para verificar a distribuição das variáveis em estudo, foi necessário realizar um gráfico do histograma com a densidade, onde a mesma é localizada pela linha em vermelho, dado que as variáveis são todas numéricas. Além disso, para a determinação das classes do histograma, também optou-se por utilizar a regra de Sturges (LANA et al., 2013). Após definir  $K = 17$ , foi obtido a Figura 10 e 11.

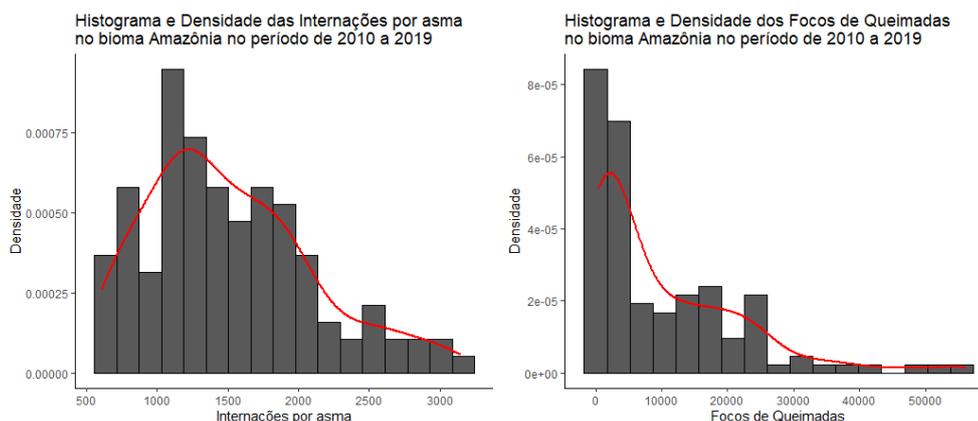


Figura 10: Histograma e Densidade das Interações por Asma e Focos de Queimadas nos municípios que abrangem o bioma da Amazônia de jan/2010 até dez/2019

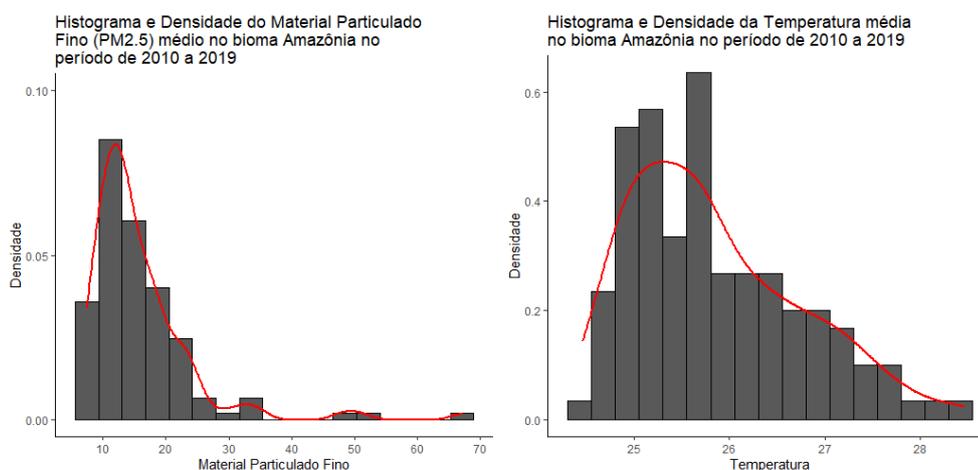


Figura 11: Histograma e Densidade do Material Particulado Fino médio e Temperatura Média nos municípios que abrangem o bioma da Amazônia de jan/2010 até dez/2019

Assim como no bioma Pantanal, as Figuras 10 e 11 remetem que, os Focos de Queimadas possuem uma grande quantidade de medidas com zeros e a variável  $PM_{2.5}$  média também possui uma grande quantidade de valores que estão entre  $0\mu g/m^3$  e  $15\mu g/m^3$ , fazendo com que suas distribuições fiquem assimétricas à direita.

Já a variável resposta Interações por Asma possui uma grande quantidade de dados localizados entre 1000 e 1500 e uma diminuição gradativa ao longo do gráfico. Por fim, mesmo sendo observado que possui uma amplitude menor, a variável de Temperatura média pode se identificar possuindo picos em valores mais baixos e não se assemelhando com uma distribuição bimodal, conforme visto para o bioma Pantanal ao analisar seu histograma e sua densidade.

Além disso, assim como no bioma Pantanal, na Figura 12, foi observado uma maior quantidade e variabilidade dos Focos de Queimadas nos meses de agosto a outubro. Por outro lado, observa-se que há uma maior quantidade e variabilidade de Interações por Asma nos meses de março a junho, ou seja, em meses diferentes do que o analisado para os Focos de Queimadas. Já para o material particulado fino ( $PM_{2.5}$ ) médio, também foi visto que há um aumento entre os meses de agosto a outubro, possivelmente causado por conta do aumento das queimadas nesta época. Por fim, ao analisar o boxplot de Temperatura média, foi observado um aumento nos meses de agosto a outubro, com o período de seca na região. Assim como no gráfico do bioma Pantanal, os pontos em vermelho em cada mês representam suas respectivas médias. As Figuras 41 e 42 localizadas no Apêndice 1 indicam os boxplots referentes ao restante das variáveis para o bioma Amazônia.

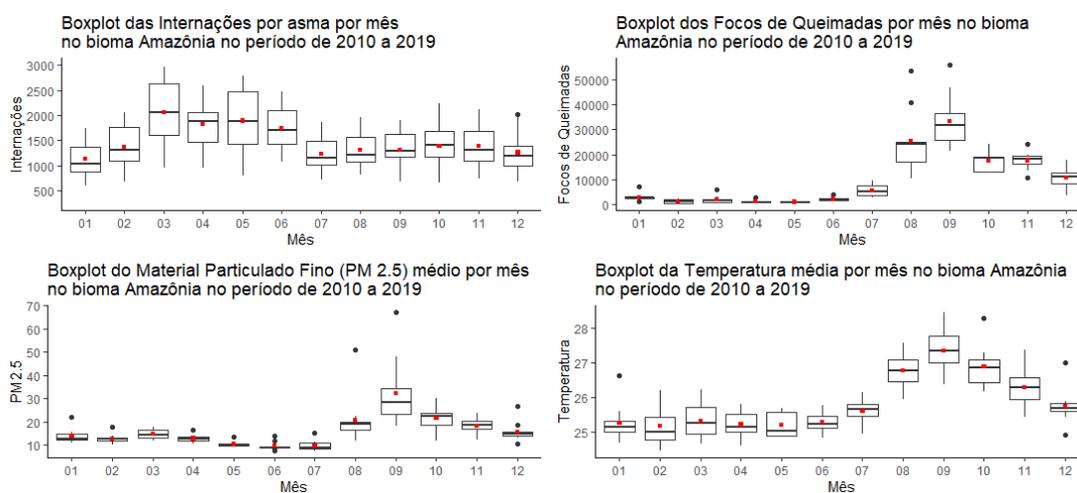


Figura 12: Boxplot das Interações por Asma, Focos de Queimadas,  $PM_{2.5}$  médio e Temperatura média no bioma da Amazônia agrupados por mês de jan/2010 até dez/2019

Por fim, assim como foi realizado para o bioma Pantanal, observa-se na Figura 13 a

série temporal de Internações por Asma e Focos de Queimadas do bioma Amazônia, onde também foi observado que os Focos de Queimadas não indicam ter uma forte relação com as Internações por Asma em indivíduos de 14 anos ou menos nos municípios que abrangem o bioma Amazônia, de modo que os Focos de Queimadas demonstram apresentar uma sazonalidade, aumentando todo ano entre agosto a outubro, conforme visto na Figura 12. As Internações por Asma por outro lado, sinalizam apresentar uma tendência linear decrescente, onde se observa estarem diminuindo a cada novo ano.

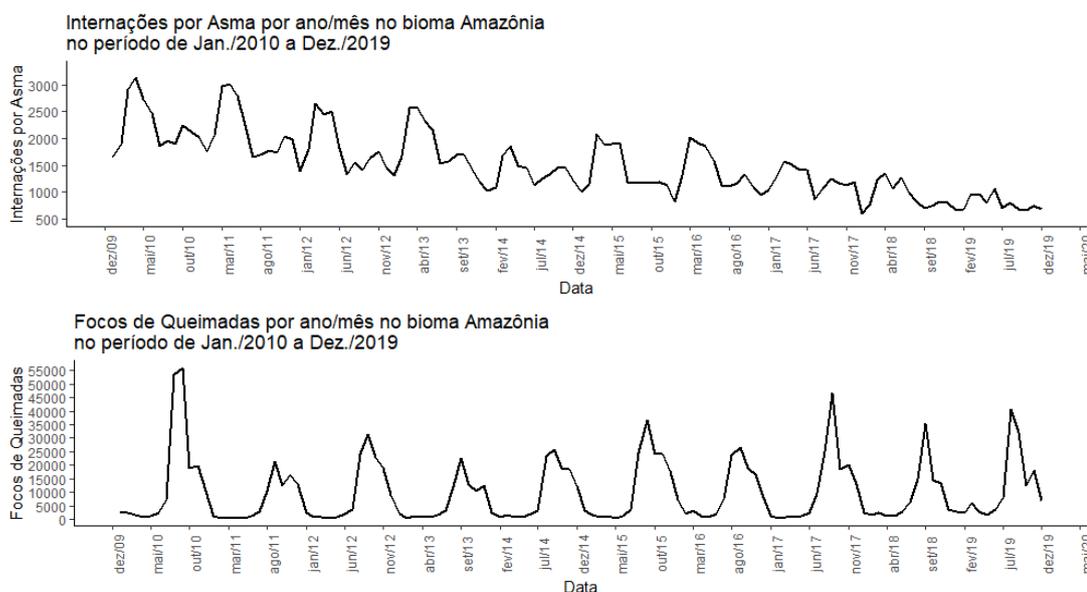


Figura 13: Série Temporal das Internações por Asma e Focos de Queimadas no bioma da Amazônia de jan/2010 até dez/2019

Já na Figura 14 tem-se a série temporal de Material Particulado Fino ( $PM_{2.5}$ ) médio e Temperatura média no bioma Amazônia, onde é observado que  $PM_{2.5}$  médio obteve um pico de aproximadamente  $70\mu g/m^3$  no ano de 2010 e após este ano indica apresentar uma certa sazonalidade, aumentando, assim como no bioma Pantanal, entre agosto a outubro, porém ficando abaixo de  $50\mu g/m^3$ . A Temperatura média também indica uma certa sazonalidade, aumentando durante os meses de agosto a outubro, conforme indicado na Figura 12, que indica que os Focos de Queimadas,  $PM_{2.5}$  e Temperatura média aumentam na mesma época do ano, como pode ser analisado no gráfico. Além disso, as Figuras 43 e 44 localizadas no Apêndice 1 indicam as séries temporais referentes ao restante das variáveis para o bioma Amazônia.

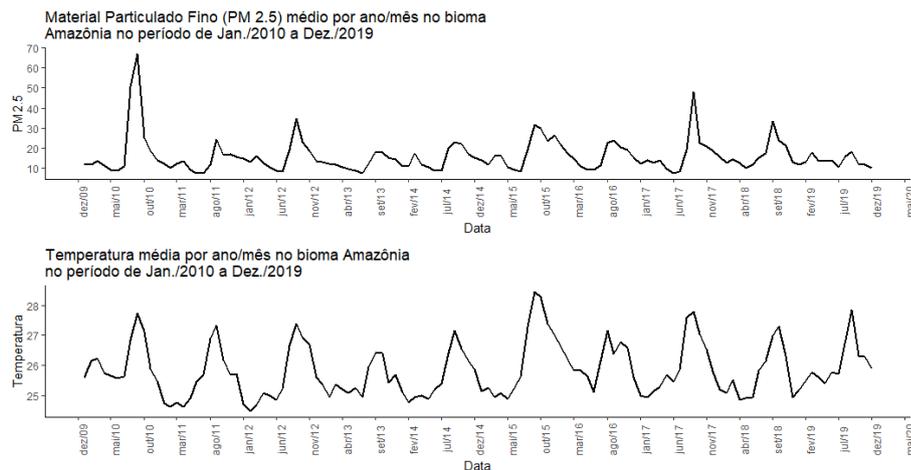


Figura 14: Série Temporal do Material Particulado Fino médio e Temperatura média no bioma da Amazônia de jan/2010 até dez/2019

## 4.2 Séries Temporais

Como dito na Subseção 3.3.1, as duas bases de dados em estudo (bioma da Amazônia e Pantanal) são séries temporais ao longo de um tempo  $t = 1, 2, \dots, n$ , onde  $n$  é o total de meses ao longo dos anos de 2010 à 2019 (120), de modo que as observações  $t$  são dependentes ao longo do tempo, indicando que eventos passados podem influenciar em eventos futuros.

Por conta do ano de 2020 ter demonstrado ser atípico ao comparar com os outros anos devido a pandemia de COVID-19, foi escolhido prever os dados para o ano de 2020 e verificar suas diferenças e qual o resultado previsto pelos modelos para este ano e realizar suas respectivas análises. Dessa forma, foi optado a não realização da previsão para os anos de 2021 em diante.

Sendo assim, o tratamento destas séries temporais desde a retirada de sua tendência e sazonalidade até a previsão para o ano de 2020 e comparação com o mesmo, se encontram nas subseções 4.2.1 e 4.2.2 respectivamente.

### 4.2.1 Pantanal

Inicialmente, foram analisadas as componentes da série temporal da variável resposta Internações por Asma para verificar se há indícios de tendência e sazonalidade ao longo dos anos. Desse modo, a Figura 15 mostra o comportamento desta série ao longo do tempo  $t$ .

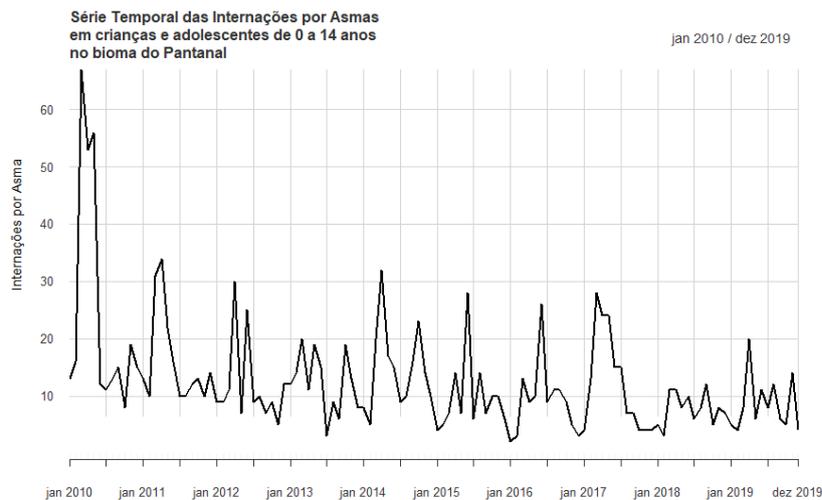


Figura 15: Série Temporal das Internações por Asmas em crianças e adolescentes de 0 a 14 anos no bioma do Pantanal de jan/2010 até dez/2019

Ao analisar a série, foi observado que parece haver uma tendência decrescente e picos sazonais ao longo de cada ano, aumentando e diminuindo em determinadas épocas do ano, conforme analisado na Figura 8, não demonstrando ser estacionária, pois não avança ao longo do tempo aleatoriamente ao redor de uma média constante. Para ratificar isto, a Figura 16 mostra a decomposição aditiva da série temporal das internações por asma, contendo a série completa, seguida de um gráfico só com sua tendência, outro apenas contendo a sazonalidade e o último contendo apenas do erro aleatório. Sendo assim, foi visto que esta decomposição aditiva da série de internações por asma indicou uma tendência decrescente e uma sazonalidade marcante ao longo dos anos. Dessa forma, foi necessário estacionarizar a série, pois observa-se que a mesma possui tendência e sazonalidade.

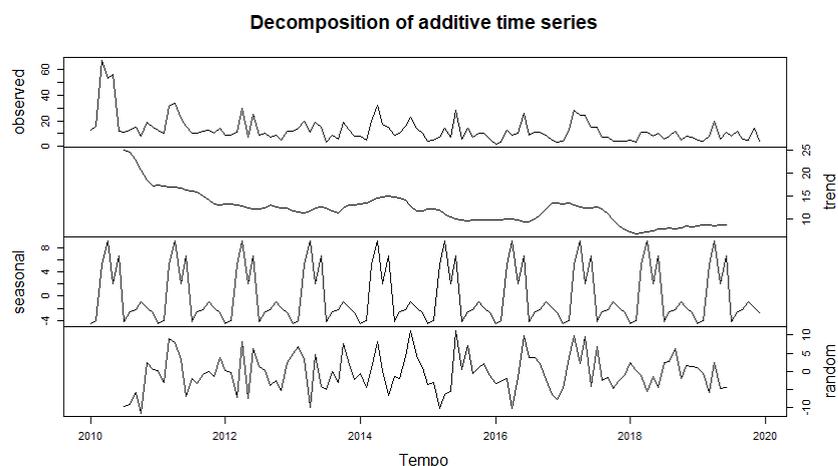


Figura 16: Decomposição aditiva da série temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma do Pantanal de jan/2010 até dez/2019

Esta tendência e sazonalidade da série também foi observada ao verificar os gráficos de autocorrelação e de autocorrelação parcial, conforme a Figura 45 localizada no Apêndice 2. Sendo assim, através do gráfico de autocorrelação, foi visto que a série possui uma tendência decrescente tanto no lag 1 quanto no lag 2, pois seus valores estão acima da linha azul, indicando que a autocorrelação é significativamente diferente de zero. Além disso, o gráfico de Autocorrelação Parcial indica que também há uma certa sazonalidade na série, conforme indicado através da decomposição da mesma, de modo que os valores vão subindo e descendo ao longo dos lags.

Para retirar a tendência da série deve-se diferenciar a mesma quantas vezes forem necessárias. Neste caso, na Figura 45 localizada no Apêndice 2 foi observado que a autocorrelação no lag 1 e 2 é significativamente diferente de zero. Com isso, essa tendência teve de ser retirada diferenciando a série primeiramente para o lag 1 (primeira diferença) e verificando se a série se tornou estacionária através do teste de Dickey-Fuller indicado na subseção 3.3.1.7. O p-valor obtido foi  $< 0,001$ , rejeitando a hipótese nula e indicando que a série sem tendência é estacionária, onde a Figura 46 também localizada no Apêndice 2, mostra a série após realizar a primeira diferença, onde foi observado que a mesma não mais apresenta a tendência decrescente, logo  $d = 1$  e  $D = 0$ .

Para corroborar esta afirmativa, a Figura 47 localizada no Apêndice 2 contém os gráficos de autocorrelação e autocorrelação parcial da série sem tendência. Com isso, ao analisar o gráfico, ACF e PACF, foi visto que foi retirada a tendência da série sinalizando que a mesma possui média e variância constante em torno de 0 ao longo do tempo, característica de uma série estacionária conforme observado na subseção 3.3.1.1. E ao verificar a Figura 8 e 15, foi observado uma periodicidade que se repete a cada ano de estudo, indicando assim  $\alpha = 12$ .

Com isso, ao analisar a Figura 47 localizada no Apêndice 2, foi observado no gráfico PACF que a série apresenta uma autocorrelação de lag 1, 3, 4, 10 e 11, logo  $p = 1, 2, 3, 4$  ou 5. Já ao analisar o gráfico ACF, foi verificado uma autocorrelação de lag  $q = 1$  ou  $q = 2$  e além disso, para o lag 12 até o lag 24, que representa o ano seguinte foi verificado uma autocorrelação de lag  $Q = 1$ . Sendo assim, os modelos contendo todas as combinações de  $p$ ,  $q$  e  $Q$  estão apresentados na Tabela 5, onde o melhor modelo foi decidido através do menor AIC, o  $d$  foi fixado em 1 e  $\alpha = 12$ , conforme analisado anteriormente.

Tabela 5: Resumo da análise dos modelos de Série Temporal para o bioma do Pantanal

Modelo	Parâmetros	AIC
1	<b>SARIMA(1,1,1)(0,0,1)[12]</b>	841,32
2	<b>SARIMA(2,1,1)(0,0,1)[12]</b>	842,59
3	<b>SARIMA(3,1,1)(0,0,1)[12]</b>	836,00
4	<b>SARIMA(4,1,1)(0,0,1)[12]</b>	837,54
5	<b>SARIMA(5,1,1)(0,0,1)[12]</b>	839,54
6	<b>SARIMA(1,1,2)(0,0,1)[12]</b>	842,99
7	<b>SARIMA(2,1,2)(0,0,1)[12]</b>	841,61
8	<b>SARIMA(3,1,2)(0,0,1)[12]</b>	837,56
9	<b>SARIMA(4,1,2)(0,0,1)[12]</b>	839,54
10	<b>SARIMA(5,1,2)(0,0,1)[12]</b>	841,14

Sendo assim, o melhor modelo que contém o menor AIC foi o modelo 3 que é dado da forma SARIMA(3,1,1)(0,0,1)[12]. Com isso, para comparar o modelo obtido através das análises anteriores com o modelo previsto computacionalmente, foi utilizado, neste caso, a função *auto.arima* do pacote *forecast*, onde a mesma retornou o modelo SARIMA(1,1,1)(0,0,1)[12] com seu AIC sendo de 841,32. Sendo assim, foram comparados os resíduos do modelo 3 e do modelo retornado pela função e os mesmos estão na Figura 48 e 49 respectivamente. Visto que não indica haver autocorrelação e autocorrelação parcial nos resíduos dos dois modelos, optou-se por continuar a análise usando o modelo da função *auto.arima*.

Além disso, é válido destacar que a função *auto.arima* retorna o melhor modelo que contenha o melhor AIC, AICc e BIC, e com isso os modelos analisados e os retornados pela função podem vir a ser diferentes. Dessa forma, para as próximas previsões, também foi utilizado o modelo retornado pela função.

Com isso, foi necessário verificar se os resíduos seguem uma distribuição normal através do teste de Shapiro Wilk, descrito na subseção 3.3.1.10, onde o p-valor encontrado foi  $< 0,001$ , rejeitando a hipótese nula e indicando que há evidências de que os resíduos não seguem uma distribuição normal.

Também foi necessário verificar se os resíduos são independentes e identicamente distribuídos através do teste de Ljung e Box, também descrito na subseção 3.3.1.10. Com isso, o teste foi feito para verificar se há correlações diferentes de zero nas defasagens de 1 até 12 e com o pvalor obtido de 0,3707 não rejeita a hipótese nula, havendo evidências de que não existe autocorrelação nos primeiros 12 lags, indicando assim que os resíduos são i.i.d.

Sendo assim, a Figura 17 indica os valores reais da série de internações por asma

*versus* os valores ajustados para o período de 2010 a 2019, onde foi observado que os valores ajustados se aderem bem aos dados, indicando assim que o modelo SARIMA utilizado possui uma boa aderência aos dados, de modo que os valores ajustados seguem a tendência e sazonalidade descrita através dos valores reais. Além disso, a linha em preto indica os valores observados no período de 2010 a 2020 e a linha em vermelho indica os valores ajustados do modelo para os anos de 2010 a 2019.

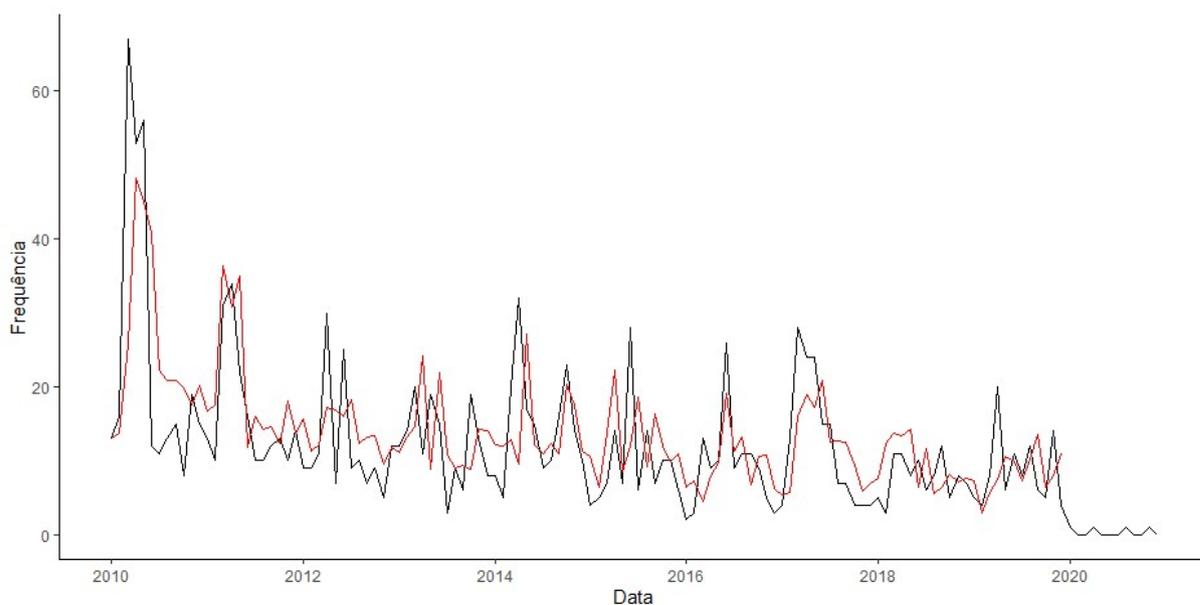


Figura 17: Valores reais *versus* valores ajustados para a série temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma Pantanal para o período de 2010 a 2020

Por fim, a previsão de Internações por Asma em crianças de 0 a 14 anos nos municípios que pertencem ao bioma Pantanal para o ano de 2020 é dada na Figura 18, onde a sombra cinza escuro indica o intervalo de confiança de 95% para os valores previstos, a linha em preto indica os valores reais analisados nos anos de 2010 a 2020 e a linha em azul indica os valores previstos de Internações por Asma para o ano de 2020. Além disso, ao observar o gráfico de previsão e seus valores reais, foi analisado uma enorme diferença entre os mesmos, dado que para os valores reais, foram vistos mínimos em 0 em diversos meses do ano e para os valores previstos foram obtidos mínimos de 6 (jan/2020). Sendo assim, o número observado de Internações por Asma foi muito inferior ao que seria esperado de acordo com o modelo previsto. Este resultado pode sugerir que a pandemia de COVID-19 impactou na ocorrência e/ou registro de internações por outras causas, ou seja, muitos casos sinalizados como Asma podem ter sido computados como COVID para o ano de 2020.

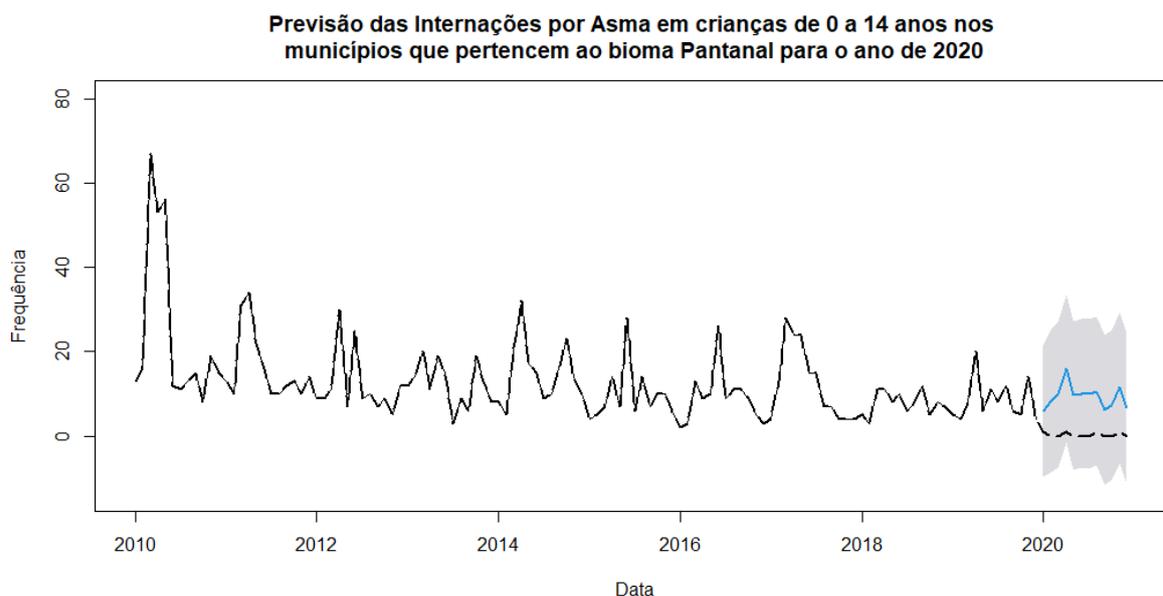


Figura 18: Previsão de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma do Pantanal para o ano de 2020

Além disso, também foi realizada a previsão dos Focos de Queimadas no bioma Pantanal para o ano de 2020 de modo que para verificar qual o melhor modelo para prever estes valores foi utilizado a função *auto.arima* do pacote *forecast*, onde a mesma retornou o modelo SARIMA(1,0,0)(1,1,1)[12]. Após verificar o melhor modelo a ser utilizado, foi necessário analisar a autocorrelação e autocorrelação parcial dos resíduos, onde na Figura 50 localizada no Apêndice 2 indica não haver autocorrelação e autocorrelação parcial nos mesmos.

Todavia, foi necessário verificar se os resíduos seguem uma distribuição normal através do teste de Shapiro Wilk, descrito na subseção 3.3.1.10, onde o p-valor encontrado foi  $< 0,001$ , rejeitando a hipótese nula e indicando que há evidências de que os resíduos não seguem uma distribuição normal.

Também foi necessário verificar se os resíduos são independentes e identicamente distribuídos através do teste de Ljung e Box, também descrito na subseção 3.3.1.10. Com isso, o teste foi feito para verificar se há correlações diferentes de zero nas defasagens de 1 até 12 e com o pvalor obtido de 0,6061 não rejeita a hipótese nula, havendo evidências de que não existe autocorrelação nos primeiros 12 lags, indicando assim que os resíduos são i.i.d.

Desse modo, para analisar a aderência do modelo aos dados, a Figura 19 indica os valores reais da série de Focos de Queimadas *versus* os valores ajustados para o período

de 2010 a 2019, onde foi observado que os valores ajustados se aderem bem aos dados, indicando assim que o modelo SARIMA utilizado possui uma boa aderência aos dados, de modo que os valores ajustados seguem a tendência e sazonalidade descrita através dos valores reais. Além disso, a linha em preto indica os valores observados no período de 2010 a 2020 e a linha em vermelho indica os valores ajustados do modelo para os anos de 2010 a 2019.

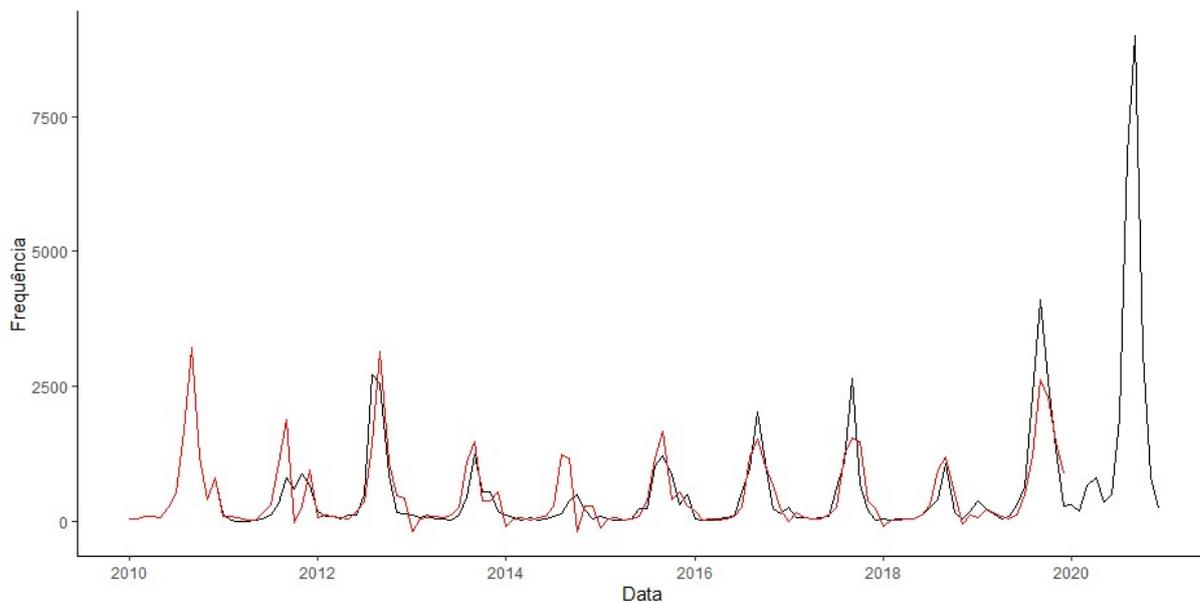


Figura 19: Valores reais *versus* valores ajustados para a série temporal de Focos de Queimadas no bioma Pantanal para o período de 2010 a 2020

Sendo assim, a previsão de Focos de Queimadas no bioma Pantanal para o ano de 2020 é dada na Figura 20, onde a sombra cinza escuro indica o intervalo de confiança de 95% para os valores previstos, a linha em preto indica os valores reais analisados nos anos de 2010 a 2020 e a linha em azul indica os valores previstos de Focos de Queimadas para o ano de 2020. Além disso, ao observar o gráfico de previsão e seus valores reais, foi analisado uma enorme diferença entre os mesmos, dado que para os valores reais, foram vistos máximos em 8990 (set/2020) Focos e para os valores previstos foram obtidos máximos de 1580 (set/2020). Sendo assim, o número observado de Focos de Queimadas foi muito superior ao que seria esperado de acordo com o modelo previsto. Este resultado pode sugerir que a pandemia de COVID-19 impactou nas questões ambientais de forma a haver uma maior liberdade para a prática de queimadas em lugares que antes não poderiam haver queimadas para o ano de 2020.

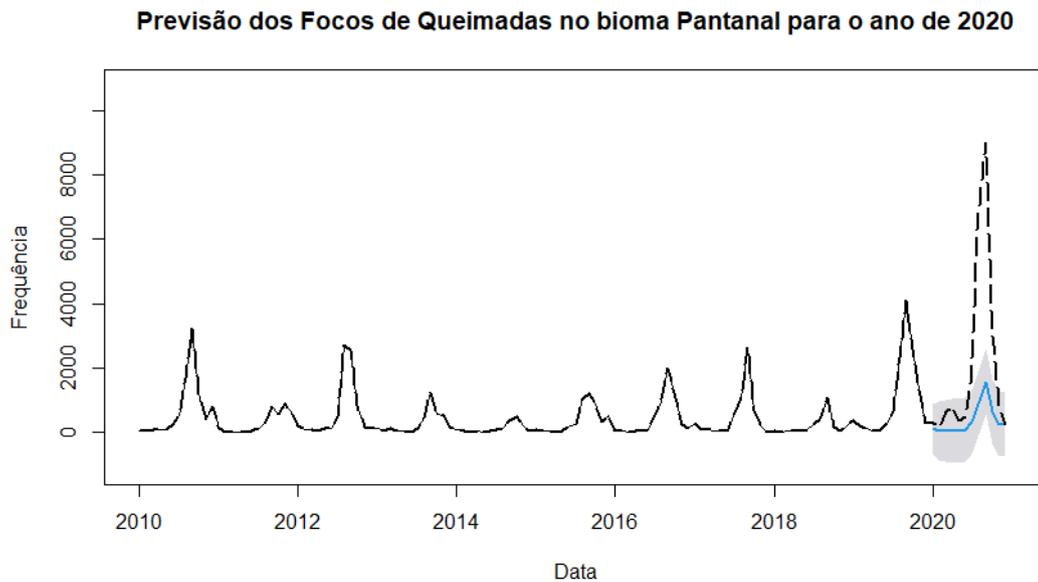


Figura 20: Previsão dos Focos de Queimadas no bioma do Pantanal para o ano de 2020

Por fim, também foi realizada a previsão do Material Particulado Fino ( $PM_{2.5}$ ) médio no bioma Pantanal para o ano de 2020 de modo que para verificar qual o melhor modelo para prever estes valores também foi utilizado a função *auto.arima* do pacote *forecast*, onde a mesma retornou o modelo SARIMA(0,0,1)(2,0,0)[12]. Após verificar o melhor modelo a ser utilizado, foi necessário analisar a autocorrelação e autocorrelação parcial dos resíduos, onde na Figura 51 localizada no Apêndice 2 indica não haver autocorrelação e autocorrelação parcial nos mesmos.

Foi necessário verificar se os resíduos seguem uma distribuição normal através do teste de Shapiro Wilk, descrito na subseção 3.3.1.10, onde o p-valor encontrado foi  $< 0,001$ , rejeitando a hipótese nula e indicando que há evidências de que os resíduos não seguem uma distribuição normal.

Também foi necessário verificar se os resíduos são independentes e identicamente distribuídos através do teste de Ljung e Box, também descrito na subseção 3.3.1.10. Com isso, o teste foi feito para verificar se há correlações diferentes de zero nas defasagens de 1 até 12 e com o pvalor obtido de 0,9979 não rejeita a hipótese nula, havendo evidências de que não existe autocorrelação nos primeiros 12 lags, indicando assim que os resíduos são i.i.d.

Assim como analisado anteriormente, para analisar a aderência do modelo aos dados, a Figura 21 indica os valores reais da série de  $PM_{2.5}$  médio *versus* os valores ajustados para o período de 2010 a 2019, onde foi observado que os valores ajustados se aderem bem

aos dados, de modo que os valores ajustados seguem a tendência e sazonalidade descrita através dos valores reais. Além disso, a linha em preto indica os valores observados no período de 2010 a 2019 e a linha em vermelho indica os valores ajustados do modelo para os anos de 2010 a 2019.

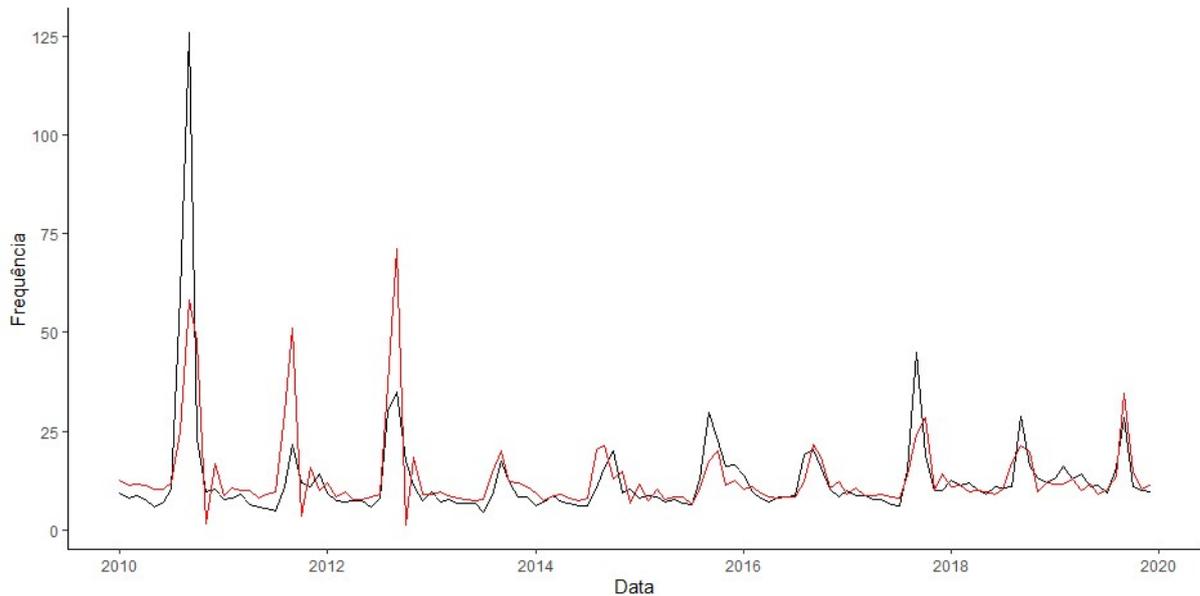


Figura 21: Valores reais *versus* valores ajustados para a série temporal de  $PM_{2.5}$  médio no bioma Pantanal para o período de 2010 a 2019

Sendo assim, a previsão de  $PM_{2.5}$  médio no bioma Pantanal para o ano de 2020 é dada na Figura 22, onde a sombra cinza escuro indica o intervalo de confiança de 95% para os valores previstos, a linha em preto indica os valores reais analisados nos anos de 2010 a 2019 e a linha em azul indica os valores previstos de  $PM_{2.5}$  médio para o ano de 2020. Além disso, ao observar o gráfico de previsão, foi analisado que o mesmo indicou seguir a mesma sazonalidade observada nos anos anteriores, de modo a possuir mínimos de 10,78 (mai/2020) e máximos de 25,11 (set/2020).

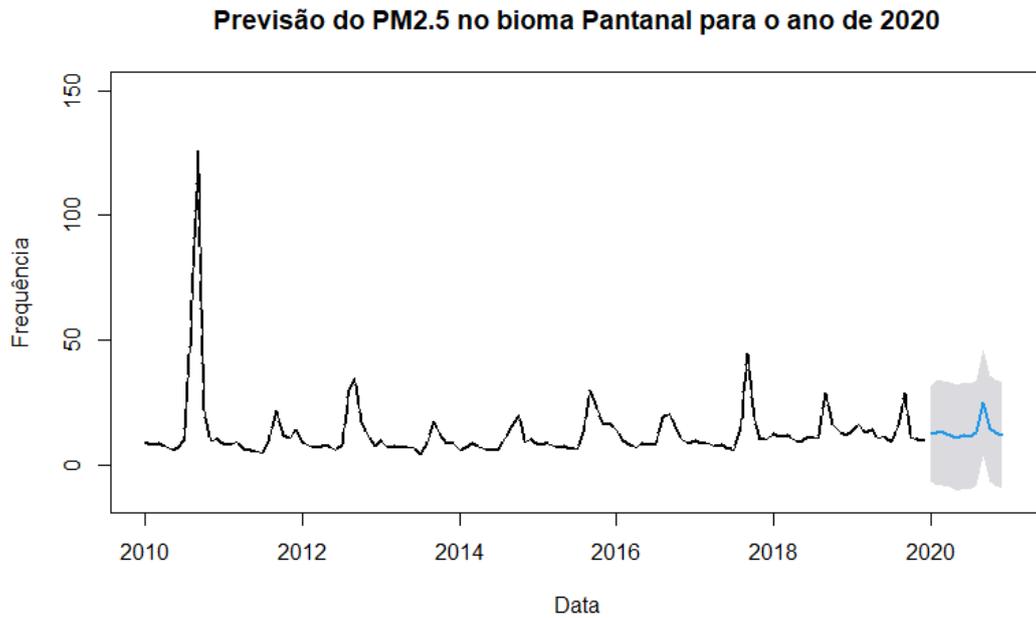


Figura 22: Previsão do Material Particulado Fino médio no bioma do Pantanal para o ano de 2020

### 4.2.2 Amazônia

Inicialmente, assim como foi feito para a série temporal do bioma Pantanal, foi analisado as componentes da série temporal da variável resposta Interações por Asma para verificar se há indícios de tendência e sazonalidade ao longo dos anos. Desse modo, a Figura 23 mostra o comportamento desta série ao longo do tempo  $t$ .

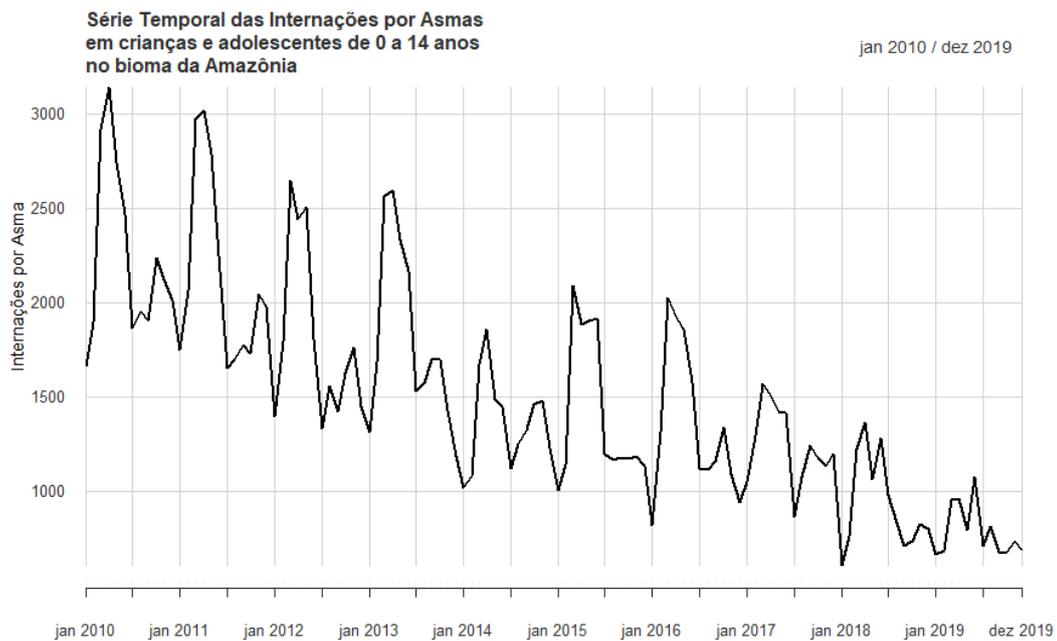


Figura 23: Série Temporal das Internações por Asmas em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia de jan/2010 até dez/2019

Ao analisar a série, também foi observado um indício de tendência decrescente e picos sazonais ao longo de cada ano, aumentando e diminuindo em determinadas épocas do ano, conforme analisado na Figura 13, de modo que a série não demonstra ser estacionária, pois não avança ao longo do tempo aleatoriamente ao redor de uma média constante. Para ratificar isto, a Figura 24 mostra a decomposição aditiva da série temporal das internações por asma. Sendo assim, conforme visto no bioma Pantanal, também observou-se uma tendência decrescente e uma sazonalidade marcante ao longo dos anos, de acordo com esta decomposição aditiva da série. Dessa forma, foi necessário estacionarizar a série, pois observa-se que a mesma possui tendência e sazonalidade.

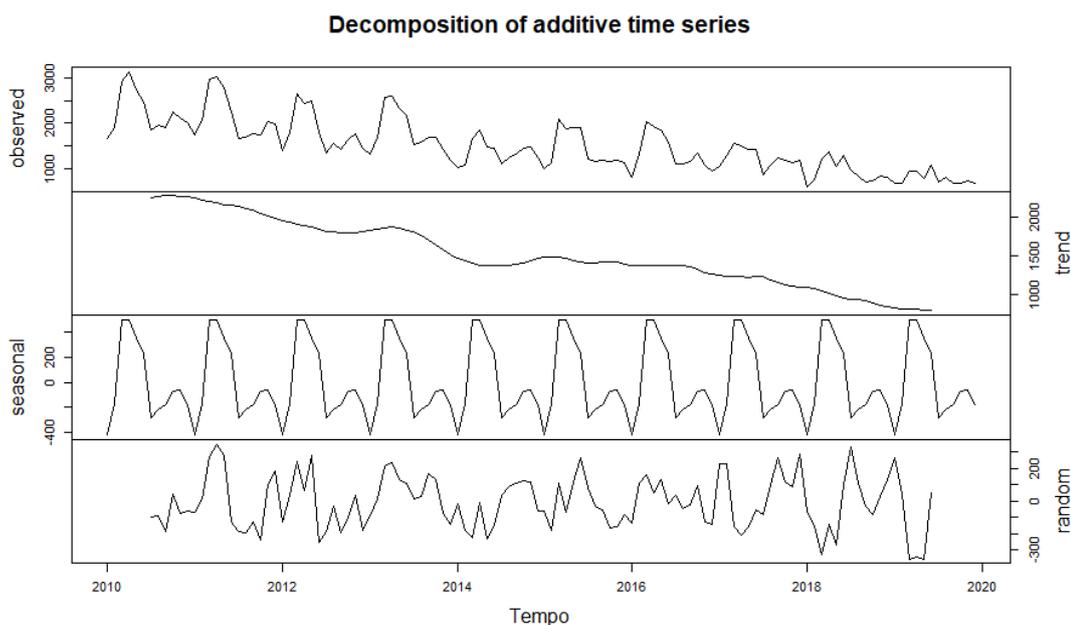


Figura 24: Decomposição aditiva da série temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia de jan/2010 até dez/2019

Esta tendência e sazonalidade da série também foi observada ao verificar os gráficos de autocorrelação e de autocorrelação parcial, conforme a Figura 52 localizada no Apêndice 3. Sendo assim, através do gráfico de autocorrelação, foi visto que, conforme analisado para o bioma Pantanal, a série também possui uma tendência decrescente, de modo que seus lags estão acima da linha tracejada em azul, indicando que a autocorrelação é significativamente diferente de zero. Além disso, foi visto que seus valores apresentam uma diminuição ao longo dos lags. Tanto o gráfico de Autocorrelação quando o gráfico de Autocorrelação Parcial indicam que também há uma certa sazonalidade na série, conforme indicado através da decomposição da mesma, de modo que os valores vão subindo e descendo ao longo dos lags.

Foi primeiro necessário retirar a tendência e analisar seus gráficos de autocorrelação e autocorrelação parcial e depois verificar se ainda há sazonalidade. Para isso, deve-se diferenciar a mesma quantas vezes forem necessárias. Com isso, a tendência teve de ser retirada diferenciando a série primeiramente para o lag 1 (primeira diferença). A Figura 53 localizada no Apêndice 3, mostra a série após realizar a primeira diferença, onde foi observado que a mesma não apresenta mais a tendência decrescente, apresentando apenas uma sazonalidade ao longo dos anos.

Para corroborar esta afirmativa, a Figura 54 localizada no Apêndice 3 contém os gráficos de autocorrelação e autocorrelação parcial da série sem tendência. Com isso, ao

analisar o primeiro gráfico, ACF, foi visto que foi retirada a tendência da série, mas a sua sazonalidade continua visível, dado que os lags estão alternando subindo e descendo ao longo das defasagens, como pode ser analisado também no gráfico de autocorrelação parcial. Sendo assim, também foi necessário retirar a sazonalidade da série, onde ao verificar a Figura 13 e 23, foi observado uma periodicidade que se repete a cada ano de estudo, indicando assim  $\alpha = 12$ .

Para retirar a sazonalidade, a série teve de ser diferenciada no instante  $t - \alpha$ , onde  $t = 120$  visto que são 12 meses e 10 anos, conforme indicado na subseção 3.3.1.9 e  $\alpha = 12$ . Além disso, teve de ser verificado se a mesma se tornou estacionária através do teste de Dickey-Fuller, indicado na subseção 3.3.1.7. O p-valor obtido foi  $< 0,01$ , rejeitando a hipótese nula e indicando que a série sem tendência e sem sazonalidade é estacionária, onde a Figura 55 também localizada no Apêndice 3, mostra a série após retirar a tendência e sazonalidade, onde foi observado que a mesma não apresenta mais estas características e indicando que possui uma média e variância constante em torno do 0 ao longo do tempo, característica de uma série estacionária, conforme descrito na subseção 3.3.1.1 e então  $d = 1$  e  $D = 1$ .

Com isso, ao analisar a Figura 56 localizada no Apêndice 3, foi observado no gráfico PACF que a série apresenta uma autocorrelação de lag 1 e 2, logo  $p = 1$  ou 2 e para o lag 12 até o lag 24, que representa o ano seguinte, foi verificado uma autocorrelação de lag  $P = 1$ . Já ao analisar o gráfico ACF, foi analisado uma autocorrelação de lag  $q = 1, 2$  ou 3 e para o lag 12 até o lag 24 foi verificado uma autocorrelação de lag  $Q = 1$ . Sendo assim, os modelos contendo todas as combinações de  $p, q, P$  e  $Q$  estão apresentados na Tabela 6, onde o melhor modelo foi decidido através do menor AIC, conforme indicado na subseção 3.3.1.10, o  $d$  e  $D$  foram fixados em 1 e  $\alpha = 12$ , conforme analisado anteriormente.

Tabela 6: Resumo da análise dos modelos de Série Temporal para o bioma da Amazônia

Modelo	Parâmetros	AIC
1	<b>SARIMA(1,1,1)(1,1,1)[12]</b>	1423,62
2	<b>SARIMA(1,1,2)(1,1,1)[12]</b>	1425,55
3	<b>SARIMA(1,1,3)(1,1,1)[12]</b>	1427,25
4	<b>SARIMA(2,1,1)(1,1,1)[12]</b>	1433,25
5	<b>SARIMA(2,1,2)(1,1,1)[12]</b>	1426,88
6	<b>SARIMA(2,1,3)(1,1,1)[12]</b>	1427,91

Sendo assim, o melhor modelo dado seu menor AIC foi o modelo 1 que é da forma SARIMA(1,1,1)(1,1,1)[12]. Com isso, para comparar o modelo obtido através das análises anteriores com o modelo previsto computacionalmente, foi utilizado, neste caso, a função

*auto.arima* do pacote *forecast*, onde a mesma retornou o modelo SARIMA(1,0,0)(0,1,1)[12] com seu AIC sendo de 1429,9. Sendo assim, foram comparados os resíduos do modelo 1 e do modelo retornado pela função e os mesmos estão na Figura 57 e 58 respectivamente. Visto que há uma autocorrelação e autocorrelação parcial nos resíduos do modelo da função *auto.arima*, foi optado em continuar as análises usando o modelo 1.

Além disso, é válido destacar que a função *auto.arima* retorna o melhor modelo com base no melhor AIC, AICc e BIC. Sendo assim, para as próximas previsões, optou-se por utilizar esta função.

Além disso, foi necessário verificar se os resíduos seguem uma distribuição normal através do teste de Shapiro Wilk, descrito na subseção 3.3.1.10, onde o p-valor encontrado foi 0,08, não rejeitando a hipótese nula e indicando que há evidências de que os resíduos seguem uma distribuição normal.

Também foi necessário verificar se os resíduos são independentes e identicamente distribuídos através do teste de Ljung e Box, também descrito na subseção 3.3.1.10. Com isso, o teste foi feito para verificar se há correlações diferentes de zero nas defasagens de 1 até 12 e com o pvalor obtido de 0,357 não rejeita a hipótese nula, havendo evidências de que não existe autocorrelação nos primeiros 12 lags, indicando assim que os resíduos são i.i.d.

Sendo assim, conforme visto para o bioma Pantanal, a Figura 25 indica os valores reais da série de internações por asma *versus* os valores ajustados para o período de 2010 a 2019, onde foi observado que os valores ajustados se aderem bem aos dados, indicando assim que o modelo SARIMA utilizado possui uma boa aderência aos dados, de modo que os valores ajustados seguem a tendência e sazonalidade descrita através dos valores reais. Além disso, a linha em preto indica os valores observados no período de 2010 a 2020 e a linha em vermelho indica os valores ajustados do modelo para os anos de 2010 a 2019.

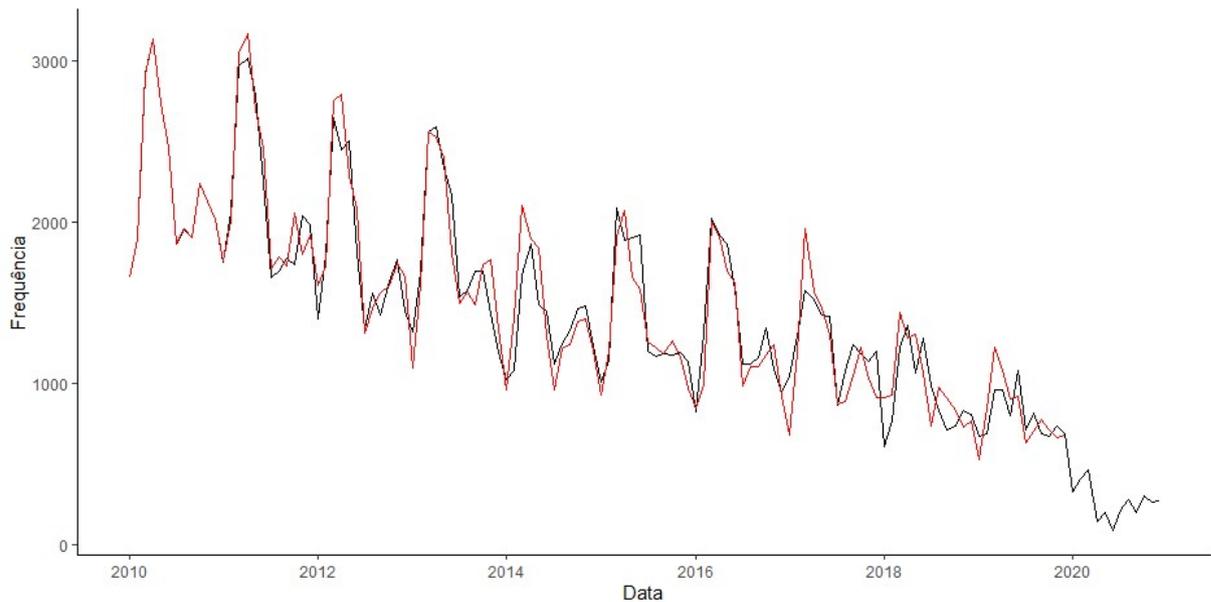


Figura 25: Valores reais *versus* valores ajustados para a série temporal de Interações por Asma em crianças e adolescentes de 0 a 14 anos no bioma Amazônia para o período de 2010 a 2020

Por fim, a previsão de Interações por Asma em crianças de 0 a 14 anos nos municípios que pertencem ao bioma Amazônia para o ano de 2020 é dada na Figura 26, onde a sombra cinza escuro indica o intervalo de confiança de 95% para os valores previstos, a linha em preto indica os valores reais analisados nos anos de 2010 a 2020 e a linha em azul indica os valores previstos de Interações por Asma para o ano de 2020. Além disso, ao observar o gráfico de previsão e seus valores reais, foi analisado uma enorme diferença entre os mesmos, dado que para os valores reais, foram vistos mínimos em 87 em jun/2020 e para os valores previstos foram obtidos mínimos de 417 (jan/2020). Sendo assim, o número observado de Interações por Asma foi muito inferior ao que seria esperado de acordo com o modelo previsto. Este resultado pode sugerir que a pandemia de COVID-19 impactou na ocorrência e/ou registro de interações por outras causas, ou seja, muitos casos sinalizados como Asma podem ter sido computados como COVID para o ano de 2020.

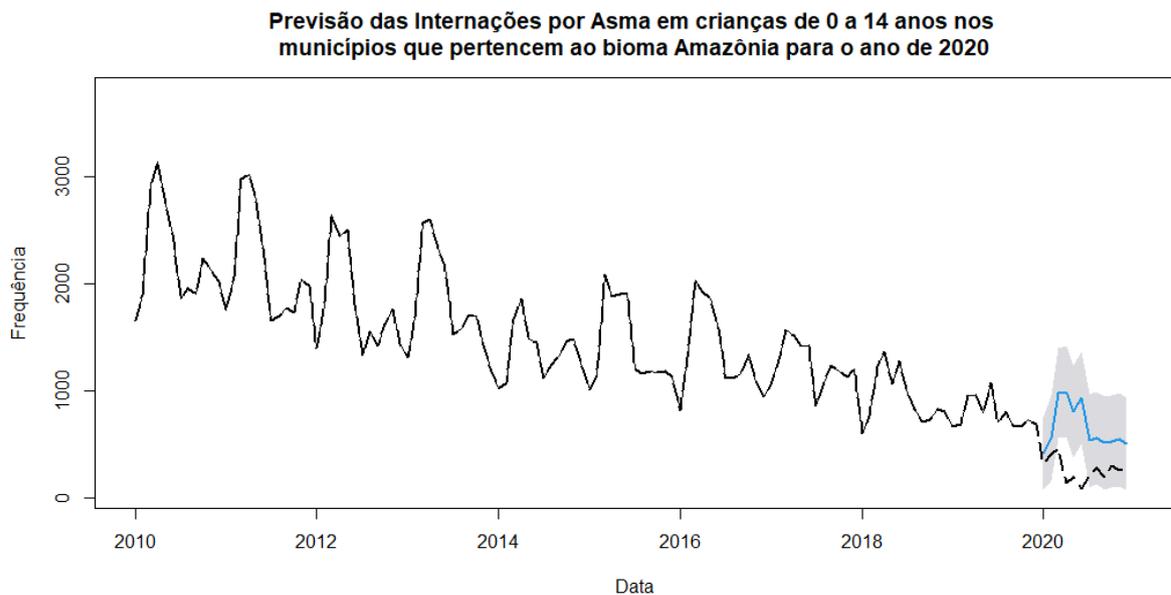


Figura 26: Previsão de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia para o ano de 2020

Além disso, também foi realizada a previsão dos Focos de Queimadas no bioma Amazônia para o ano de 2020 de modo que para verificar qual o melhor modelo para prever estes valores foi utilizado novamente a função *auto.arima* do pacote *forecast*, onde a mesma retornou o modelo SARIMA(1,0,0)(2,1,0)[12]. Após verificar o melhor modelo a ser utilizado, foi necessário analisar a autocorrelação e autocorrelação parcial dos resíduos, onde na Figura 59 localizada no Apêndice 3 indica haver autocorrelação e autocorrelação parcial no lag 12.

Dessa forma, foi necessário verificar se os resíduos são independentes e identicamente distribuídos através do teste de Ljung e Box, descrito na subseção 3.3.1.10. Com isso, o teste foi feito para verificar se há autocorrelações diferentes de zero nas defasagens de 1 até 12 e com o pvalor obtido de 0,6286 não rejeita a hipótese nula, havendo evidências de que não existe autocorrelação nos primeiros 12 lags, indicando assim que os resíduos são i.i.d.

Todavia, foi necessário verificar se os resíduos seguem uma distribuição normal através do teste de Shapiro Wilk, também descrito na subseção 3.3.1.10, onde o p-valor encontrado foi  $< 0,001$ , rejeitando a hipótese nula e indicando que há evidências de que os resíduos não seguem uma distribuição normal.

Para analisar a aderência do modelo aos dados, a Figura 27 indica os valores reais da série de Focos de Queimadas *versus* os valores ajustados para o período de 2010 a 2019,

onde foi observado que os valores ajustados se aderem bem aos dados, de modo que os valores ajustados seguem a tendência e sazonalidade descrita através dos valores reais. Além disso, a linha em preto indica os valores observados no período de 2010 a 2020 e a linha em vermelho indica os valores ajustados do modelo para os anos de 2010 a 2019.

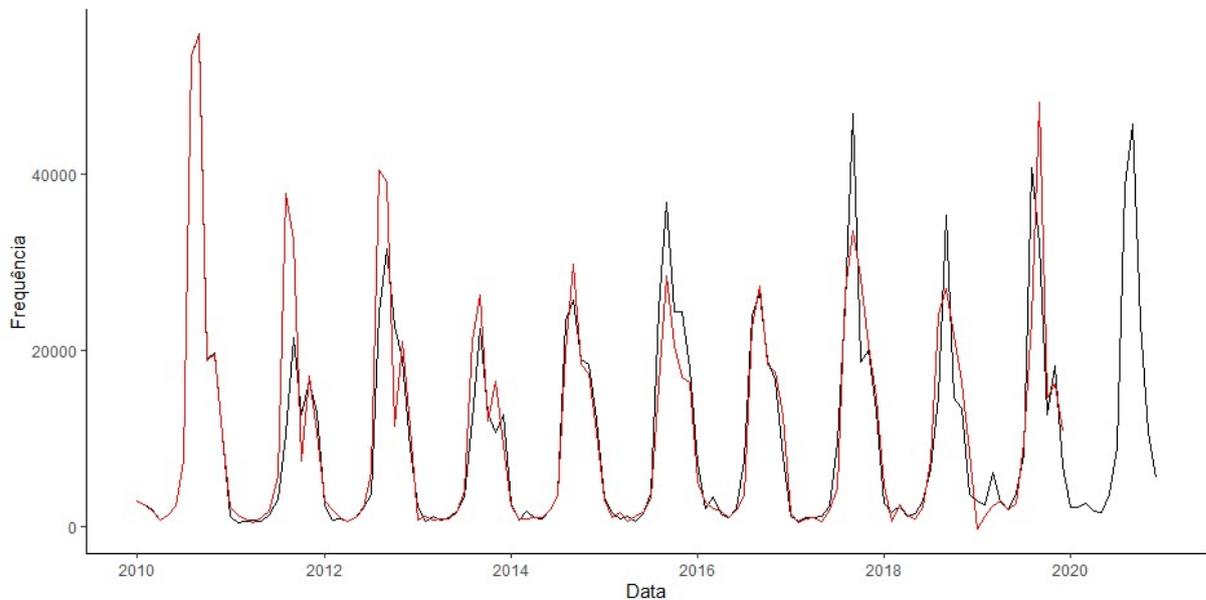


Figura 27: Valores reais *versus* valores ajustados para a série temporal de Focos de Queimadas no bioma Amazônia para o período de 2010 a 2020

Sendo assim, a previsão de Focos de Queimadas no bioma Amazônia para o ano de 2020 é dada na Figura 28, onde a sombra cinza escuro indica o intervalo de confiança de 95% para os valores previstos, a linha em preto indica os valores reais analisados nos anos de 2010 a 2020 e a linha em azul indica os valores previstos de Focos de Queimadas para o ano de 2020. Além disso, ao observar o gráfico de previsão e seus valores reais, foi analisado diferença quase não perceptível, dado que para os valores reais, foram vistos máximos em 45.632 (set/2020) Focos e para os valores previstos foram obtidos máximos de 37.528, também para set/2020. Sendo assim, o número observado de Focos de Queimadas coincidiu ao que seria esperado de acordo com o modelo previsto. Este resultado, diferente do que foi analisado para o bioma Pantanal, pode ser analisado de modo que para o bioma Amazônia, já existia uma liberdade maior para esta prática de queimadas.

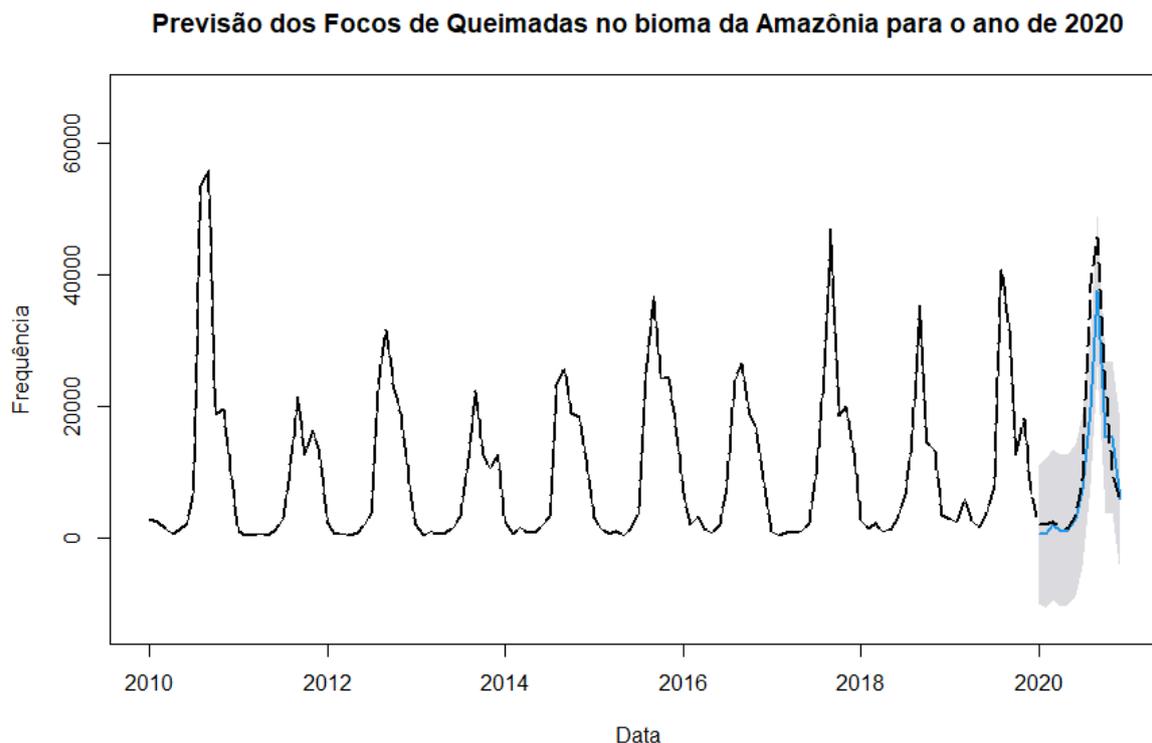


Figura 28: Previsão dos Focos de Queimadas no bioma da Amazônia para o ano de 2020

Por fim, também foi realizada a previsão do Material Particulado Fino ( $PM_{2,5}$ ) médio no bioma Amazônia para o ano de 2020 de modo que para verificar qual o melhor modelo para prever estes valores também foi utilizado a função *auto.arima* do pacote *forecast*, onde a mesma retornou o modelo SARIMA(1,0,0)(1,1,0)[12]. Após verificar o melhor modelo a ser utilizado, foi necessário analisar a autocorrelação e autocorrelação parcial dos resíduos, onde na Figura 60 localizada no Apêndice 2 indica não haver autocorrelação e autocorrelação parcial nos mesmos.

Foi necessário verificar se os resíduos seguem uma distribuição normal através do teste de Shapiro Wilk, descrito na subseção 3.3.1.10, onde o p-valor encontrado foi  $< 0,001$ , rejeitando a hipótese nula e indicando que há evidências de que os resíduos não seguem uma distribuição normal.

Também foi necessário verificar se os resíduos são independentes e identicamente distribuídos através do teste de Ljung e Box, também descrito na subseção 3.3.1.10. Com isso, o teste foi feito para verificar se há autocorrelações diferentes de zero nas defasagens de 1 até 12 e com o pvalor obtido de 0,9095 não rejeita a hipótese nula, havendo evidências de que não existe autocorrelação nos primeiros 12 lags, indicando assim que os resíduos são i.i.d.

Assim como analisado para o bioma Pantanal, para analisar a aderência do modelo aos dados, a Figura 29 indica os valores reais da série de  $PM_{2.5}$  médio *versus* os valores ajustados para o período de 2010 a 2019, onde foi observado que os valores ajustados se aderem bem aos dados, de modo que os valores ajustados seguem a tendência e sazonalidade descrita através dos valores reais. Além disso, a linha em preto indica os valores observados no período de 2010 a 2019 e a linha em vermelho indica os valores ajustados do modelo para os anos de 2010 a 2019.

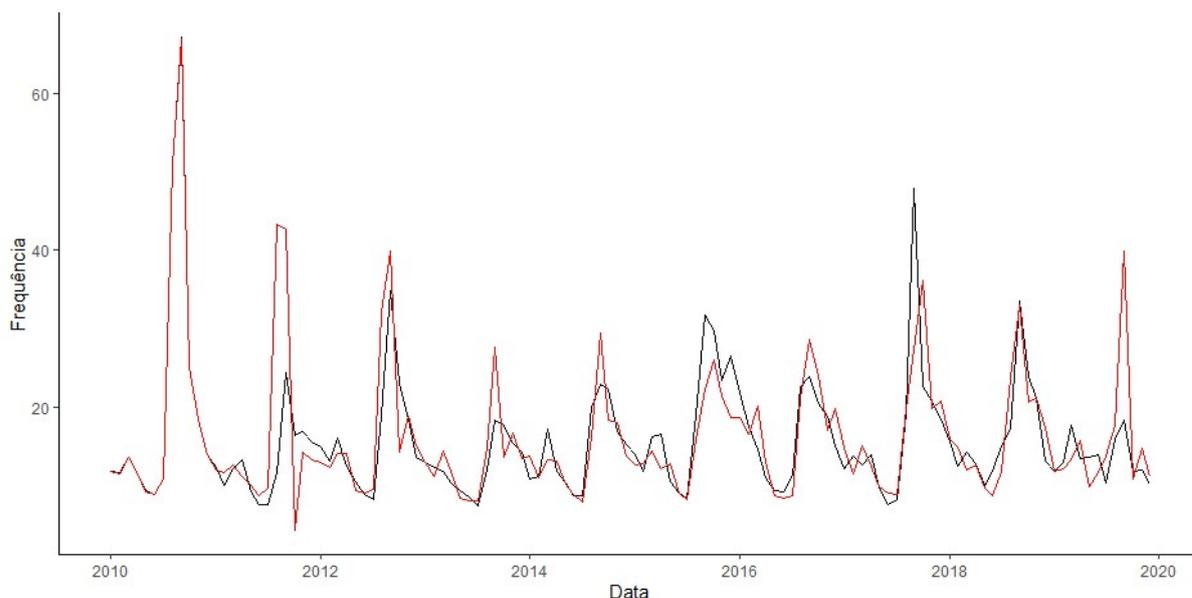


Figura 29: Valores reais *versus* valores ajustados para a série temporal de  $PM_{2.5}$  médio no bioma Amazônia para o período de 2010 a 2019

Sendo assim, a previsão de  $PM_{2.5}$  médio no bioma Amazônia para o ano de 2020 é dada na Figura 30, onde a sombra cinza escuro indica o intervalo de confiança de 95% para os valores previstos, a linha preto indica os valores reais analisados nos anos de 2010 a 2019 e a linha em azul indica os valores previstos de  $PM_{2.5}$  para o ano de 2020. Além disso, ao observar o gráfico de previsão, foi analisado que o mesmo indicou seguir a mesma sazonalidade observada nos anos anteriores, de modo a possuir mínimos de 10,75 (jan/2020) e máximos de 26,43 (set/2020).

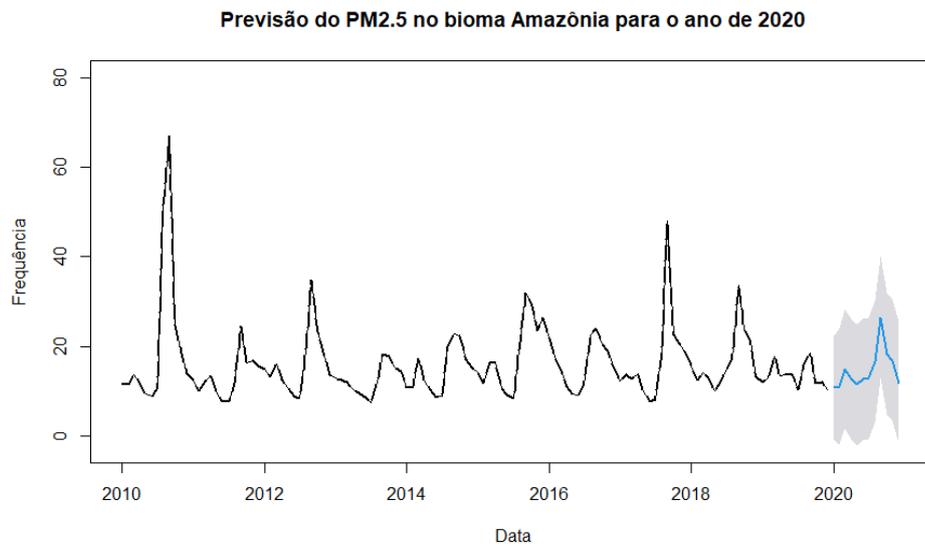


Figura 30: Previsão do Material Particulado Fino médio no bioma da Amazônia para o ano de 2020

### 4.3 Modelos Lineares Generalizados

Como visto no Capítulo 2, dos 7 artigos analisados através da metodologia PRISMA, 1 optou por utilizar os Modelos Aditivos Generalizados a partir da família de distribuição estatística de Poisson para analisar a relação das variáveis explicativas do modelo com a variável resposta em análise, que neste trabalho se trata da variável Interações por Asma.

Dessa forma, inicialmente optou-se por utilizar a família de distribuição de Poisson na estimação dos modelos, mas ao analisar os resultados obtidos na Tabela 3 e 4, foi observado nos dois casos que a média de Interações por Asma se encontrava menor que a Variância da mesma, caracterizando assim a superdispersão, conforme descrito na subseção 3.3.2.7. Sendo assim, visto que esta característica demonstra que se está tratando de forma incorreta a incerteza associada ao modelo, como consequências as estimações incorretas a respeito dos parâmetros dos modelos e falta de independência entre as observações.

Para contornar este problema, foi considerada uma forma funcional diferente para o modelo, para o qual optou-se pela utilização da família *Quasi-Poisson* (HOEF; BOVENG, 2007). Por outro lado, para estimar a tendência e sazonalidade observada na série, foi necessário criar uma variável chamada *tempo* que são valores que vão de 1 até o número total de observações da base de dados, neste caso 120. Após criar esta variável foi utilizado

o conceito de *Natural Cubic Splines* (PERPEROGLOU et al., 2019) de modo a suavizar a série ajustando sua tendência e sazonalidade através da função  $ns()$  com o parâmetro utilizado sendo  $df=2*10$ , onde 10 é o número de anos e 2 são as quantidades de nós por ano.

A utilização da família *Quasi-Poisson* implica que, conforme descrito na subseção 3.3.2.4 e por (GASPARRINI; ARMSTRONG; KENWARD, 2010), a estatística de qualidade de ajuste AIC deve ser ajustada por meio do parâmetro de superdispersão, como proposto pelos autores e definido como QAIC.

Além disso, para tentar melhorar o ajuste do modelo, foi criada uma variável chamada estações, de modo que se o mês fosse 1, 2 ou 3, a estação indicada era "Verão", se o mês fosse 4, 5 ou 6, a estação indicada já passava a ser "Outono" e assim sucessivamente até o mês 12 e a estação "Inverno".

Para verificar qual variável explicativa mais se relaciona com as Internações por Asma individualmente, foram obtidos os resultados que se encontram na Tabela 7, onde cada linha indica uma variável explicativa para os dois biomas em estudo, Amazônia e Pantanal. Nesta Tabela 7 observa-se o resultado do Risco Relativo, Intervalo de Confiança de 95% para o Risco Relativo, Critério de Informação de Akaike Ajustado (QAIC) e o p-valor observado para a variável no modelo em análise. É válido ressaltar que nesta tabela, os modelos foram estimados utilizando a variável *estações* e o ajuste de tendência e sazonalidade pela *spline* com somente uma variável de exposição.

Tabela 7: Resumo da análise do modelo com somente uma variável de exposição - com a variável Estações

Variáveis de Exposição	Amazônia				Pantanal			
	RR	IC(95%)	QAIC	P-valor	RR	IC(95%)	QAIC	P-valor
PM2.5media	1,054	(1,0076;1,1027)	7.062,20	0,024	1,0167	(0,9358;1,1046)	983,45	0,70
PM2.5max	1,0027	(0,9983;1,0071)	7.299,73	0,23	1,00006	(0,9677;1,0335)	984,13	0,99
PM2.5min	1,3965	(1,0402;1,875)	7.085,17	0,029	1,0813	(0,9007;1,2982)	980,75	0,40
PM2.5p90	1,0174	(1,0003;1,0348)	7.138,44	0,049	1,0078	(0,9553;1,0632)	983,75	0,78
tempmedia	1,0387	(0,9743;1,1074)	7.305,93	0,25	1,12	(1,0379;1,2084)	942,85	0,004
tempmax	1,0076	(0,9672;1,0496)	7.384,90	0,72	1,0397	(0,9675;1,1172)	977,97	0,29
tempmin	1,11	(1,0512;1,1721)	6.564,22	< 0,001	1,0552	(1,0101;1,1024)	956,01	0,018
temp90	0,999	(0,9559;1,0461)	7.393,79	0,99	1,051	(0,9771;1,1305)	974,53	0,18
tempamplitude	0,9447	(0,903;0,9882)	7.007,10	0,015	0,9647	(0,9239;1,0073)	970,71	0,11
Focos	1,000004	(0,999999;1,000008)	7.227,50	0,11	1,00004	(0,9998;1,0003)	983,63	0,73

Ao analisar a Tabela 7<sup>1</sup>, nota-se que ao utilizar a variável estações, o QAIC dos modelos para o bioma Amazônia se encontram na faixa de 7.000, enquanto os do bioma Pantanal se encontram por volta de 980. Além disso, para a variável *tempmin* do bioma Amazônia, nota-se que seu risco relativo é 1,11, ou seja, para cada aumento em uma unidade na temperatura mínima, o risco de Internações por Asma aumenta cerca de 11%. Para as variáveis em que os RR estimados se encontram muito próximos de 1 ou que

<sup>1</sup>Os  $\hat{\beta}$ s estimados de  $PM_{2.5}$  foram multiplicados por 10 para realizar o cálculo do RR e IC(RR)

seu Intervalo de Confiança de 95% contenha o valor 1, isso sugere que a variável não é estatisticamente significativa para o modelo ao nível de 5%. Já para as variáveis de poluição, a interpretação é feita com relação a cada aumento de 10 unidades, ou seja, para a variável de poluição *PM2.5media* do bioma Amazônia, observa-se que para um aumento em 10 unidades o risco de Internações por Asma aumenta 5,4%. Assim, as variáveis que demonstraram ser significativas dado seus p-valores menores que 5% foram *PM2.5media*, *PM2.5min*, *PM2.5p90*, *tempmin* e *tempamplitude*.

Entretanto, ao analisar o bioma Pantanal na Tabela 7, nota-se que as variáveis que demonstraram ser significativas dado seus p-valores foram apenas *tempmedia* e *tempmin*, onde seus riscos relativos foram 1,12 e 1,0552 respectivamente. Ou seja, para a variável explicativa *tempmedia*, observa-se que para cada aumento em uma unidade na temperatura média no bioma Pantanal, tem-se que o risco relativo de Internações por Asma aumenta em 12%.

Em contrapartida, para verificar se deve ser utilizada a variável *estações* ou se a utilização da mesma não interfere nas análises de melhores modelos, foram obtidos os resultados que se encontram na Tabela 8, onde foi retirada dos modelos a variável *estações*. Nesta Tabela 8 também observa-se o resultado do Risco Relativo, Intervalo de Confiança de 95% para o Risco Relativo, Critério de Informação de Akaike Ajustado (QAIC) e o p-valor observado para a variável no modelo em análise.

Tabela 8: Resumo da análise do modelo com somente uma variável de exposição - sem a variável *Estações*

Variáveis de Exposição	Amazônia				Pantanal			
	RR	IC(95%)	QAIC	P-valor	RR	IC(95%)	QAIC	P-valor
<b>PM2.5media</b>	0,9535	(0,9071;1,0023)	11.171,33	0,064	0,9311	(0,8513;1,0185)	1.066,37	0,12
<b>PM2.5max</b>	0,9931	(0,9887;0,9976)	10.586,7	0,003	0,9652	(0,9337;0,9978)	1.052,22	0,04
<b>PM2.5min</b>	1,0023	(0,6929;1,45)	11.567,25	0,99	0,9380	(0,7674;1,1466)	1.082,91	0,53
<b>PM2.5p90</b>	0,9802	(0,9622;0,9985)	11.064,37	0,04	0,9514	(0,8986;1,0073)	1.062,87	0,09
<b>tempmedia</b>	0,9076	(0,864;0,9535)	10.100,72	<0,001	1,0525	(0,9714;1,1404)	1.077,05	0,21
<b>tempmax</b>	0,9356	(0,9122;0,9596)	9.254,38	<0,001	0,9387	(0,8883;0,992)	1.052,71	0,027
<b>tempmin</b>	1,1082	(1,0365;1,1849)	10.678,76	0,003	1,0827	(1,0361;1,1314)	1.014,47	<0,001
<b>temp90</b>	0,9241	(0,8984;0,9504)	9.002,09	<0,001	0,9435	(0,8889;1,0014)	1.061,52	0,059
<b>tempamplitude</b>	0,9096	(0,8884;0,9314)	7.465,85	<0,001	0,9344	(0,9051;0,9647)	989,41	<0,001
<b>Focos</b>	0,999993	(0,999990;0,999997)	10.070,10	<0,001	0,9998	(0,9996;0,99995)	1.041,37	0,015

Ao analisar a Tabela 8<sup>2</sup>, nota-se que optando por não utilizar a variável *estações*, o QAIC dos modelos para o bioma Amazônia aumentaram, se encontrando na faixa de 9.000 a 11.000 e os do bioma Pantanal também aumentaram, onde se encontram por volta de 1.000. Além disso, para a variável *tempmin* do bioma Amazônia, nota-se que seu risco relativo é 1,1082, ou seja, para cada aumento em uma unidade na temperatura mínima, o risco de Internações por Asma aumenta cerca de 10,82%.

Já para a variável de poluição *PM2.5max* do bioma Amazônia, observa-se que para

<sup>2</sup>Os  $\hat{\beta}$ 's estimados de  $PM_{2.5}$  foram multiplicados por 10 para realizar o cálculo do RR e IC(RR)

cada aumento em 10 unidades o risco de Internações por Asma diminui cerca de 0,7%, podendo indicar um efeito protetor. As variáveis que demonstraram ser significativas dado seus p-valores menores que 5% foram *PM2.5max*, *PM2.5p90*, *tempmedia*, *tempmax*, *tempmin*, *temp90*, *tempamplitude* e *Focos*. É válido ressaltar também que todas as variáveis explicativas de temperatura deram significativas para o bioma Amazônia.

Além disso, ao analisar o bioma Pantanal na Tabela 8, nota-se que as variáveis que demonstraram ser significativas dado seus p-valores menores que 5% aumentaram consideravelmente ao comparar com a Tabela 7. As variáveis foram *PM2.5max*, *tempmax*, *tempmin*, *tempamplitude* e *Focos*, onde seus riscos relativos foram 0,9652, 0,9387, 1,0827, 0,9344 e 0,9998 respectivamente. Ou seja, para a variável explicativa *Focos*, observa-se que para cada aumento em uma unidade nos Focos de Queimadas no bioma Pantanal, tem-se que o risco de Internações por Asma diminui em aproximadamente 0,02%, indicando uma baixa queda, também podendo indicar um efeito protetor.

Sendo assim, ao comparar a Tabela 8 com a Tabela 7, observou-se que retirar a variável estações aumentou a quantidade de variáveis explicativas que se encontram significativas para o modelo final, mesmo com seu QAIC tendo aumentado, mas em alguns casos os riscos relativos estimados passaram a indicar as variáveis de exposição como fatores de proteção em relação à internação por asma.

Por fim, em uma análise conjunta do QAIC e do p-valor analisado em cada modelo da Tabela 7 e 8, foram realizadas todas as combinações possíveis entre as variáveis, de forma a ter apenas uma variável de poluição, uma de temperatura e Focos, independentemente do seu p-valor e estes melhores modelos foram indicados na Tabela 9 e na Tabela 12, de forma que nestas tabelas, apenas são apresentados os modelos cujo qual pelo menos duas dentre as três variáveis de exposição ambiental indicaram p-valores significativos, seja utilizando a variável estações ou não.

### 4.3.1 Pantanal

Para o modelo completo do bioma Pantanal, utilizando uma variável de poluição, uma de temperatura e a de Focos, foi verificado quais poderiam vir a ser o melhor modelo a ser utilizado analisando o QAIC e seus p-valores, apresentado na Tabela 9, de modo que, em todos os modelos desta Tabela 9 não foi utilizada a variável estações, visto que os mesmos demonstraram ser melhores do que os que continham a variável estações.

Além das variáveis em análise, foram criadas as variáveis indicadoras *pmaltoetempalta*,

que recebe 1 se o valor observado  $PM2.5max$  for maior que o quantil de 75% de sua variável e se a  $tempmax$  observada também for maior que o quantil de 75% de sua variável, e a variável indicadora  $pmbaixoetempbaixa$ , que recebe 1 se o valor observado  $PM2.5min$  for menor que o quantil de 25% de sua variável e se a  $tempmin$  observada também for menor que o quantil de 25% de sua variável.

Na Tabela 9, observa-se o Critério de Informação de Akaike Ajustado (QAIC), os respectivos p-valores na ordem em que as variáveis estão escritas e o pseudo coeficiente de determinação (pseudo  $R^2$ ), conforme descrito na subseção 3.3.2.5.

Tabela 9: Resumo da análise dos melhores modelos completos para o bioma do Pantanal

Modelo	Variáveis de Exposição	Pantanal		
		QAIC	P-valores	Pseudo $R^2$
1	<b>pm2.5media, tempmedia e Focos</b>	980,05	0,86 ; <0,001 ; 0,004	0,5321
2	<b>pm2.5max, tempmedia e Focos</b>	980,38	0,98 ; <0,001 ; 0,001	0,5324
3	<b>pm2.5min, tempmedia e Focos</b>	978,04	0,50 ; <0,001 ; <0,001	0,5373
4	<b>pm2.5min, tempmin e Focos</b>	995,00	0,57 ; 0,001 ; 0,037	0,4997
5	<b>pm2.5p90, tempmedia e Focos</b>	979,96	0,81 ; <0,001 ; 0,002	0,5319
6	<b>pmbaixoetempbaixa e Focos</b>	1003,51	0,004 ; 0,002	0,4568

Foi observado que o modelo que melhor se adequa aos dados contendo todas as variáveis explicativas sendo significativas, foi o modelo 6. Porém seu QAIC foi o maior dentre os demais e seu pseudo  $R^2$  foi de apenas 0,4568, o pior dentre todos os melhores modelos analisados. Sendo assim, escolhendo com base no melhor QAIC e no melhor pseudo  $R^2$ , foi optado por utilizar o modelo 3 como modelo final, que consiste nas variáveis  $pm2.5min$ ,  $tempmedia$  e  $Focos$ , onde conforme informado anteriormente, o  $\hat{\beta}$  estimado de  $PM_{2.5}min$  foi multiplicado por 10 para realizar o cálculo do RR e IC(RR).

Sendo assim, o modelo 3 pode ser representado da seguinte forma, em função dos valores estimados para os parâmetros  $\hat{\beta}$ :

$$\ln(\hat{Y}_i) = P(t_i) - 1,8855 + 0,0848X_{1i} + 0,1831X_{2i} - 0,0005X_{3i}, \quad (4.1)$$

onde

- $P(t_i)$  é a função polinomial em partes da Spline;
- $X_{1i}$  indica o nível de  $PM_{2.5}min$  no tempo  $i$ ;
- $X_{2i}$  indica a temperatura média no tempo  $i$ ;
- $X_{3i}$  indica a quantidade de Focos de Calor no tempo  $i$ .

Além disso, a Tabela 10 indica os valores analisados da multicolinearidade deste modelo 3, como descrito na subseção 3.3.2.6. Nota-se que valores acima de 5 indicam problemas de multicolinearidade no modelo, mas ao analisar a Tabela 10 tem-se que o maior valor observado foi 4,30, indicando não haver multicolinearidade.

Tabela 10: Diagnóstico da Multicolinearidade do modelo 3 por meio do VIF para o bioma do Pantanal

pm2.5min	tempmedia	Focos
2,28	4,30	2,71

Por fim, foi necessário avaliar os resíduos para verificar se as variáveis explicativas estavam especificadas corretamente no modelo. Para isso, conforme descrito na subseção 3.3.2.3, foi necessário montar o gráfico de envelope simulado para verificar se os pontos estão bem dispostos dentro do envelope.

Sendo assim, a Figura 31 indica que os pontos estão bem dispostos dentro deste envelope e permite a continuação da análise.

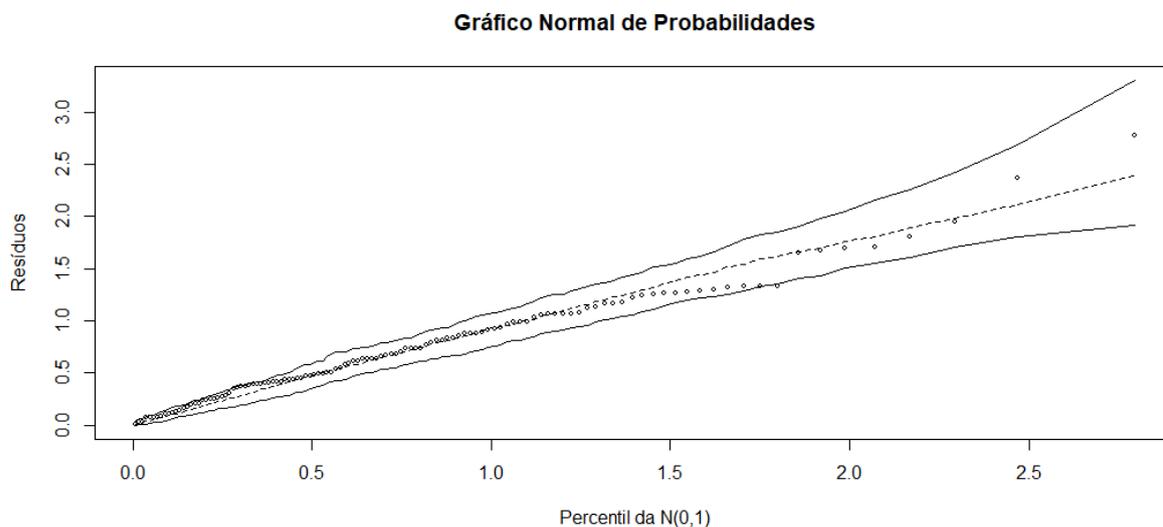


Figura 31: Uso do envelope simulado no modelo 3 para o bioma do Pantanal

Em seguida, verificou-se o pressuposto de homocedasticidade através do gráfico dos valores ajustados *versus* resíduos estudentizados, conforme também foi descrito na subseção 3.3.2.3. A Figura 32 indica que o pressuposto de homocedasticidade é aceito, visto que os pontos não possuem um determinado padrão, se comportando de forma aleatória.

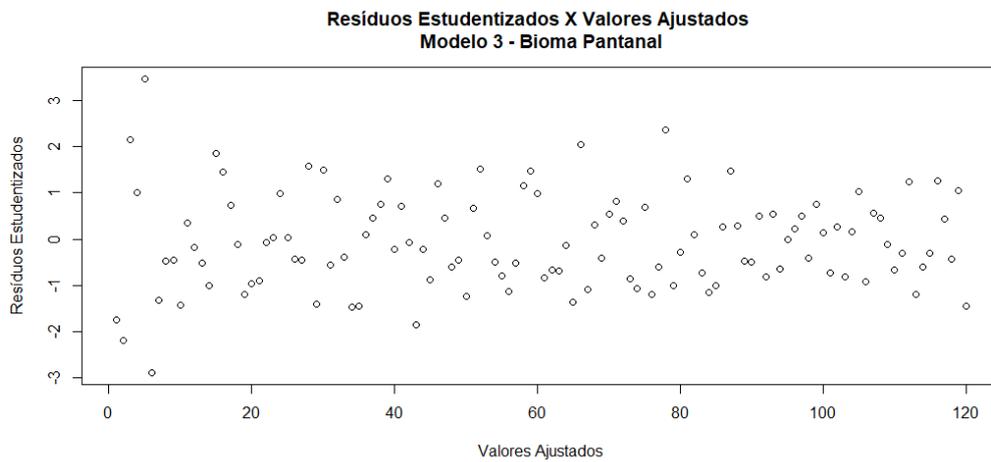


Figura 32: Homocedasticidade dos dados do modelo 3 para o bioma do Pantanal

Para a verificação de possíveis pontos influentes conforme descrito na subseção 3.3.2.3, foi verificado os valores produzidos pela Distância de Cook e analisados na Figura 33. Sendo assim, conforme visto nesta Figura, não há possíveis valores influentes.

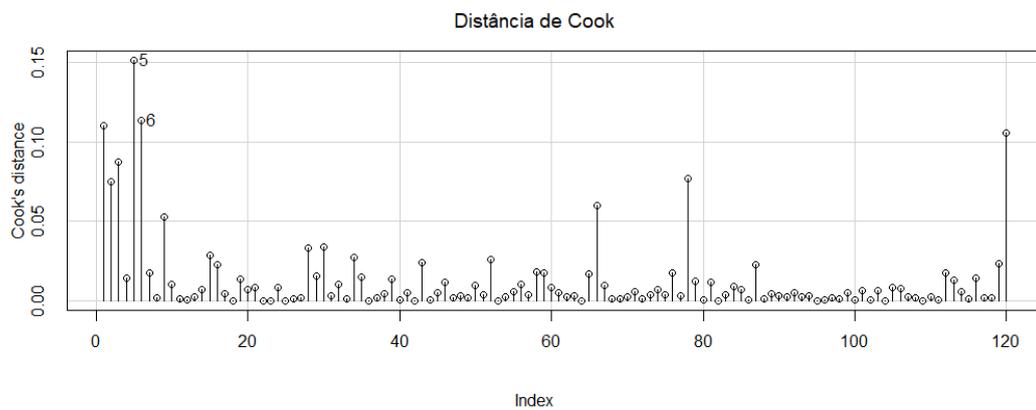


Figura 33: Avaliação de pontos influentes do modelo 3 para o bioma do Pantanal

Por fim, a Tabela 11 indica os riscos relativos e seus respectivos intervalos de confiança de 95% para as variáveis explicativas  $pm_{2.5min}$ ,  $tempmedia$  e  $Focos$  que foram analisadas neste melhor modelo.

Tabela 11: Resumo da análise do melhor modelo para o bioma do Pantanal

Variáveis de Exposição	$\hat{RR}$	IC(95%)
<b>pm2.5min</b>	1.0885	(0,8533;1,3885)
<b>tempmedia</b>	1.2009	(1,092;1,3206)
<b>Focos</b>	0.9995	(0,9992;0,9997)

Ao analisar a Tabela 11, foi observado que para a variável *pm2.5min* seu risco relativo é 1,0885, ou seja, para cada aumento em 10 unidades na concentração de material particulado fino, o risco de Internações por Asma aumenta cerca de 8,85%. Além disso, para a variável *tempmedia*, foi observado que o aumento em uma unidade nesta temperatura média aumenta o risco de Internações por Asma em aproximadamente 20%. Por fim, o aumento em uma unidade nos Focos de Queimadas diminui o risco de Internações por Asma em 0,05%, não demonstrando uma diminuição significativa.

Sendo assim, nota-se que, a temperatura média, quando o modelo é ajustado por *pm2.5min* e *Focos*, é um fator de risco para a Internação por Asma no bioma Pantanal e se mostrou sendo a melhor medida de temperatura nessa associação. De acordo com a fração atribuível calculada a partir do RR estimado, das 1539 internações por asma neste período de 2010 a 2019 no bioma Pantanal, aproximadamente 258 internações podem ser atribuíveis à exposição a aumentos de temperatura.

Já com relação a fração atribuível para o  $PM_{2,5}$ , das 1539 internações por asma neste período de 2010 a 2019 no bioma Pantanal, aproximadamente 125 podem ser atribuíveis à exposição a aumentos de  $PM_{2,5}$ .

Por fim, as Figuras 61 e 62 contém os gráficos de autocorrelação e autocorrelação parcial dos resíduos deste melhor modelo, indicado no Apêndice 4. Sendo assim, observa-se que uma limitação deste melhor modelo foi que, além de ser agrupado pelo bioma e não por seus municípios individualmente e por serem dados agrupados por mês, ele não corrigiu a autocorrelação entre as medidas.

### 4.3.2 Amazônia

Na estimação do modelo completo para o bioma Amazônia, foram utilizados os mesmos procedimentos do bioma Pantanal. Os resultados dos melhores modelos são apresentados na Tabela 12.

Tabela 12: Resumo da análise dos melhores modelos completos para o bioma da Amazônia

Modelo	Variáveis de Exposição	Amazônia		
		QAIC	P-valores	Pseudo $R^2$
1	<b>pm2.5media, tempmedia e Focos</b>	9.407,38	<0,001 ; 0,75 ; 0,002	0,77
2	<b>pm2.5media, tempmax e Focos</b>	8.997,3	0,014 ; 0,024 ; 0,084	0,78
3	<b>pm2.5media, tempmin e Focos</b>	7.765,81	0,026 ; <0,001 ; <0,001	0,82
4	<b>pm2.5media, tempp90 e Focos</b>	8.715,54	0,03 ; 0,0036 ; 0,43	0,79
5	<b>pm2.5max, tempmin e Focos</b>	8.118,71	0,73 ; <0,001 ; <0,001	0,81
6	<b>pm2.5min, tempmedia e Focos</b>	9.516,52	0,003 ; 0,035 ; 0,43	0,76
7	<b>pm2.5min, tempmax e Focos</b>	8.622,00	<0,001 ; <0,001 ; 0,86	0,79
8	<b>pm2.5min, tempmin e Focos</b>	8.120,69	0,78 ; <0,001 ; <0,001	0,81
9	<b>pm2.5min, tempp90 e Focos</b>	8.173,18	<0,001 ; <0,001 ; 0,17	0,80
10	<b>pm2.5min, tempamplitude e Focos</b>	7.190,92	0,03 ; <0,001 ; 0,42	0,83
11	<b>pm2.5p90, tempmedia e Focos</b>	9.755,22	0,02 ; 0,92 ; 0,01	0,76
12	<b>pm2.5p90, tempmin e Focos</b>	7.731,21	0,02 ; <0,001 ; <0,001	0,82
13	<b>pmbaixoetempbaixa e Focos</b>	9.054,11	<0,001 ; <0,001	0,78

Ao analisar a Tabela 12, foi observado que o modelo que melhor se adequa aos dados contendo todas as variáveis explicativas sendo significativas, foi o modelo 12, onde seu QAIC foi o segundo melhor observado, porém seu pseudo  $R^2$  foi de 0,82. Sendo assim, foi optado por utilizar este modelo para o restante da análise do bioma Amazônia, que consiste nas variáveis  $pm2.5p90$ ,  $tempmin$  e  $Focos$ , onde conforme informado anteriormente, o  $\hat{\beta}$  estimado de  $PM_{2.5p90}$  foi multiplicado por 10 para realizar o cálculo do RR e IC(RR).

Sendo assim, o modelo 12 pode ser representado da seguinte forma, em função dos valores estimados para os parâmetros  $\hat{\beta}$ :

$$\ln(\hat{Y}_i) = P(t_i) + 3,682 + 0,0387X_{1i} + 0,1671X_{2i} - 0,00002X_{3i}, \quad (4.2)$$

onde

- $P(t_i)$  é a função polinomial em partes da Spline;
- $X_{1i}$  indica o nível de  $PM_{2.5p90}$  no tempo  $i$ ;
- $X_{2i}$  indica a temperatura mínima no tempo  $i$ ;
- $X_{3i}$  indica a quantidade de Focos de Calor no tempo  $i$ .

Além disso, a Tabela 13 indica os valores analisados da multicolinearidade deste modelo 12, como descrito na subseção 3.3.2.6. Nota-se que valores acima de 5 indicam problemas de multicolinearidade no modelo e ao analisar a Tabela 10 tem-se que o maior valor observado foi 5,87 seguido por 5,56, havendo indício de multicolinearidade no modelo. Para investigar este problema, foi calculado o VIF para o modelo 3, que pela ordem

indicava ser o terceiro melhor modelo dentre os 13 analisados. No entanto, para o modelo 3, também foi identificado valores acima de 5, havendo um 6 para a variável de *pm2.5media*. Sendo assim, visto que o valor da multicolinearidade do modelo 12 não ultrapassa o valor 6, foi optado por continuar as análises com este mesmo modelo 12.

Tabela 13: Diagnóstico da Multicolinearidade do modelo 12 por meio do VIF para o bioma da Amazônia

<b>pm2.5p90</b>	<b>tempmin</b>	<b>Focos</b>
5,87	2,36	5,56

O gráfico de envelope simulado, apresentado na Figura 34 sugere um bom ajuste do modelo.

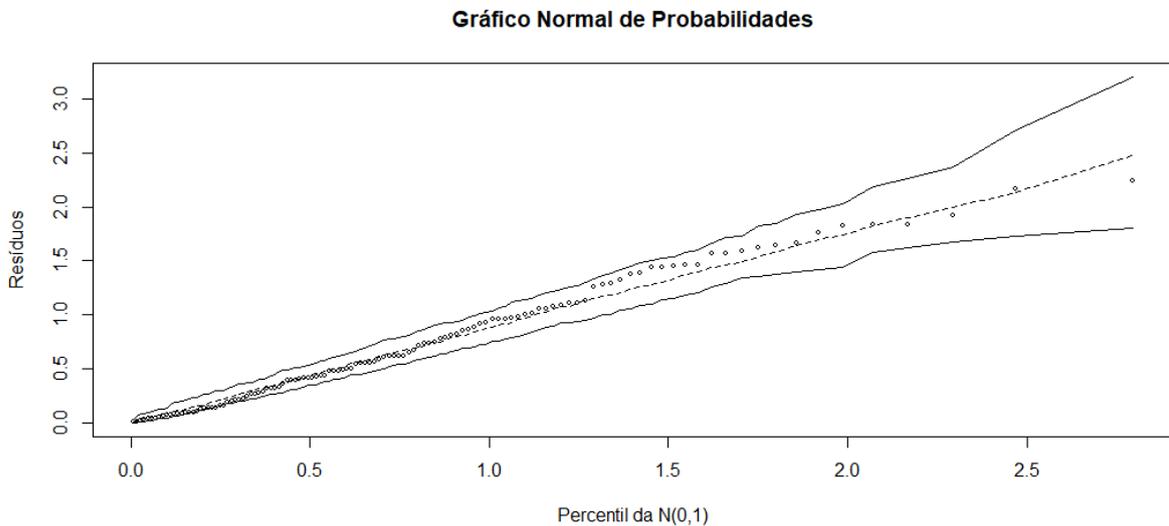


Figura 34: Uso do envelope simulado no modelo 12 para o bioma da Amazônia

Ao dar continuidade na análise, foi necessário verificar o pressuposto de homocedasticidade através do gráfico dos valores ajustados *versus* resíduos estudentizados, conforme também foi descrito na subseção 3.3.2.3. Com isso, a Figura 35 indica que o pressuposto de homocedasticidade é aceito, visto que os pontos não possuem um determinado padrão, se comportando de forma aleatória.

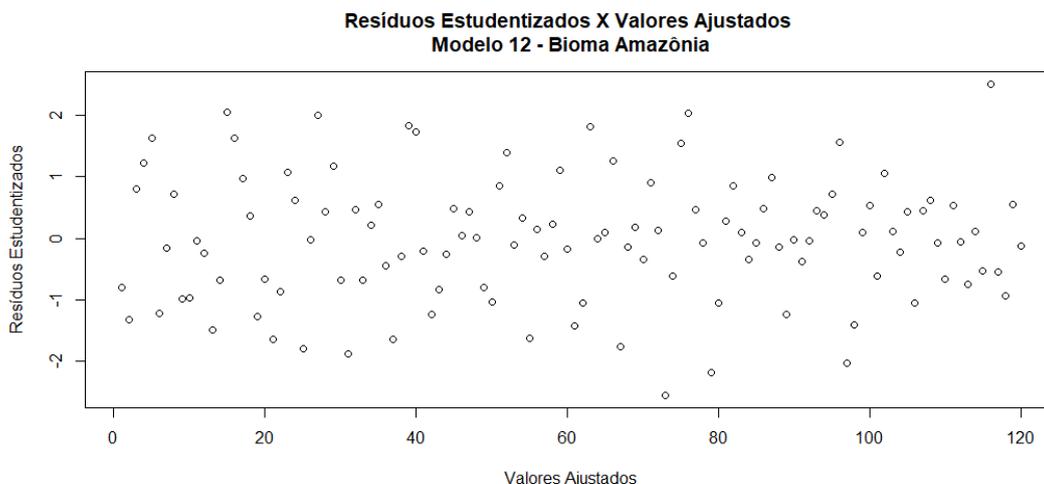


Figura 35: Homocedasticidade dos dados do modelo 12 para o bioma da Amazônia

Para a verificação de possíveis pontos influentes conforme descrito na subseção 3.3.2.3, também foi necessário verificar os valores produzidos pela Distância de Cook e analisados na Figura 36. Sendo assim, conforme visto nesta Figura, não há possíveis valores influentes.

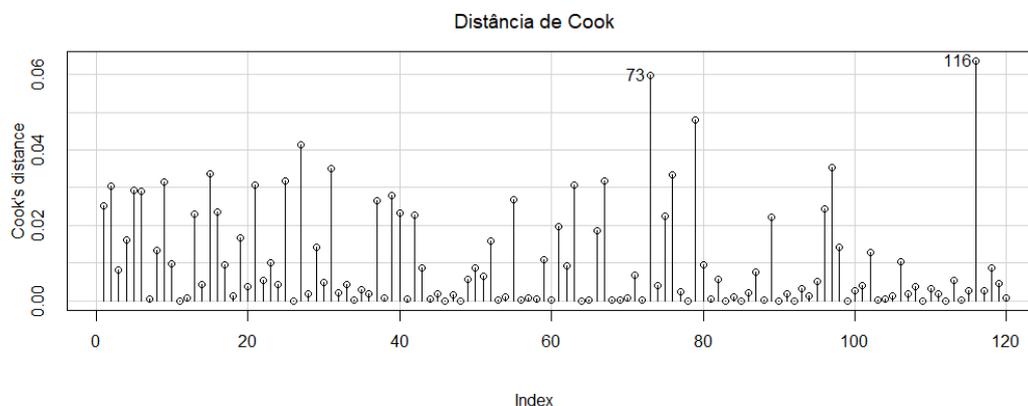


Figura 36: Avaliação de pontos influentes do modelo 12 para o bioma da Amazônia

Por fim, em posse do melhor modelo para o bioma Amazônia, a Tabela 14 indica os riscos relativos e seus respectivos intervalos de confiança de 95% para as variáveis explicativas *pm2.5p90*, *tempmin* e *Focos* que foram analisadas neste melhor modelo.

Tabela 14: Resumo da análise do melhor modelo para o bioma da Amazônia

Variáveis de Exposição	$\hat{RR}$	IC(95%)
<b>pm2.5p90</b>	1,039	(1,0068;1,0732)
<b>tempmin</b>	1,1819	(1,1128;1,2552)
<b>Focos</b>	0,999983	(0,99998;0,99999)

Ao analisar a Tabela 14, foi observado que para a variável *pm2.5p90* seu risco relativo é 1,039, ou seja, para cada aumento em 10 unidades na quantidade do material particulado fino, o risco de Internações por Asma aumenta cerca de 4%. Além disso, para a variável *tempmin*, foi observado que o aumento em uma unidade na temperatura aumenta o risco de Internações por Asma em aproximadamente 18,19%. Por fim, o aumento em uma unidade nos Focos de Queimadas diminui o risco de Internações por Asma em 0,0017%, não demonstrando uma diminuição significativa.

Sendo assim, nota-se que, a temperatura mínima, quando o modelo é ajustado *pm2.5p90* e *Focos*, é um fator de risco para a Internação por Asma no bioma Amazônia e se mostrou sendo a melhor medida de temperatura nessa associação. De acordo com a fração atribuível calculada a partir do RR estimado, das 181.546 internações por asma neste período de 2010 a 2019 no bioma Amazônia, aproximadamente 27.941 internações podem ser atribuíveis à exposição a aumentos da temperatura.

Já com relação a fração atribuível para o  $PM_{2.5}$ , das 181.546 internações por asma neste período de 2010 a 2019 no bioma Amazônia, aproximadamente 6.815 podem ser atribuíveis à exposição a aumentos de  $PM_{2.5}$ .

Por fim, a Figura 63 e 64 refere-se aos gráficos de autocorrelação e autocorrelação parcial dos resíduos deste melhor modelo, indicado no Apêndice 4. Sendo assim, observa-se que uma limitação deste melhor modelo foi que, além de ser agrupado pelo bioma e não por seus municípios individualmente e por serem dados agrupados por mês, ele não corrigiu a autocorrelação entre as medidas.

## 5 Conclusões

Neste trabalho foram comparadas as distribuições temporais das variáveis de exposição ambiental, como Temperatura, Material Particulado Fino e Focos de Calor, quanto à tendência e sazonalidade entre os biomas Amazônia e Pantanal. Como também foram estimados os riscos e impactos dessas exposições sobre as Internações por Asma em crianças e adolescentes com até 14 anos residentes nesses biomas.

No período de estudo, a temperatura média para o bioma Pantanal apresentou uma característica diferente do bioma Amazônia. A temperatura no Pantanal apresentou uma distribuição bimodal, indicando uma frequência maior de meses com temperaturas na faixa de  $23^{\circ}C$  a  $23,5^{\circ}C$  e  $27,5^{\circ}C$ . Por outro lado, para o bioma Amazônia a mesma variável indicou uma distribuição unimodal e com picos entre  $25^{\circ}C$  e  $26^{\circ}C$ . Estes são resultados típicos do clima desses biomas, por exemplo clima quente e úmido da Amazônia e tropical continental, com verão chuvoso e inverno seco, do Pantanal.

Além disso, observou-se que para os dois biomas as três variáveis de exposição ambiental apresentaram dados com valores mais altos, assim como maior variabilidade nos meses de agosto a outubro, coincidindo com o período de seca conforme indicado por (MOUTINHO et al., 2020). Enquanto que as Internações por Asma apresentaram essa maior quantidade e variabilidade entre os meses de março a junho. Ademais, tanto para o bioma do Pantanal quanto para o da Amazônia foi observado nas Figuras 8 e 13 uma tendência decrescente nas Internações por Asma ao longo dos anos.

Com relação aos modelos de séries temporais, foram estimados modelos de previsão de séries temporais para cada uma das variáveis em estudo, excluindo-se a Temperatura, utilizando o modelo SARIMA. Os melhores modelos foram escolhidos dado o menor AIC. Para o bioma Pantanal, a Internação por Asma obteve o melhor modelo sendo SARIMA(3,1,1)(0,0,1)[12], onde a Figura 18 indica a previsão para o ano de 2020, assim como os valores observados para esse ano. Já para os Focos de Calor o modelo obtido foi SARIMA(1,0,0)(1,1,1)[12], onde a Figura 20 indica a previsão para o ano de 2020,

também com os valores observados. Por fim, para o Material Particulado Fino o modelo obtido foi SARIMA(0,0,1)(2,0,0)[12] onde a Figura 22 indica a previsão de  $PM_{2.5}$  também para o ano de 2020.

Já com relação ao bioma Amazônia, o melhor modelo para as Internações por Asma foi SARIMA(1,1,1)(1,1,1)[12], onde a Figura 26 indica a previsão para o ano de 2020 e os valores observados. Para Focos de Calor o modelo obtido foi SARIMA(1,0,0)(2,1,0)[12], onde a Figura 28 indica a previsão para o ano de 2020, também com os valores observados. Por último, para o Material Particulado Fino o modelo obtido foi SARIMA(1,0,0)(1,1,0)[12] onde a Figura 30 indica a previsão de  $PM_{2.5}$  também para o ano de 2020.

Percebe-se destes resultados que os biomas tem as componentes de tendência e sazonalidade das séries temporais estudadas ajustadas por modelos SARIMA com diferentes valores para os parâmetros (p,d,q, P,D,Q). Esta etapa do trabalho permitiu prever os dados de  $PM_{2.5}$ , Focos de Calor e número de internação por Asma para o ano de 2020. Observou-se que o número de focos observados em 2020 foi muito superior ao esperado pelo comportamento da série temporal. Por outro lado, o número observado de internações por asma foi muito inferior ao que seria esperado de acordo com o modelo. Estes resultados podem sugerir que a pandemia de COVID-19 impactou indiretamente na ocorrência e/ou registro de internações por outras causas, assim como nas questões ambientais. Por exemplo, muitos dos casos sinalizados como Asma podem ter sido computados como COVID para este ano.

Quanto a relação entre as variáveis, foram estimados para cada bioma Modelos Lineares Generalizados através da distribuição Quasi-Poisson, tendo como variável resposta o número de internações por asma em crianças e adolescentes com 14 anos ou menos, e como variáveis explicativas as variáveis de exposição ambiental.

O melhor modelo para cada um dos biomas foi selecionado de acordo com: a significância estatística das estimativas dos parâmetros das variáveis explicativas; às medidas do QAIC; e os pseudo  $R^2$ . Na etapa seguinte foi analisada a multicolinearidade, o pressuposto de normalidade por meio do envelope simulado, o pressuposto de homocedasticidade através dos valores ajustados versus resíduos estudentizados e verificado possíveis pontos influentes através da distância de Cook.

Sendo assim, os resultados indicaram que, para o bioma Pantanal, quando a variável de temperatura média mensal aumenta em 1 grau, o risco de internação por asma aumenta em aproximadamente 20%. Ou seja, quando o modelo é ajustado por  $PM_{2.5min}$  e Focos, a temperatura média mensal é um fator de risco para as internações por asma.

Para o bioma Amazônia, quando a variável de temperatura mínima mensal aumenta em 1 grau, o risco de internação por asma aumenta em aproximadamente 18%. Ou seja, quando o modelo é ajustado por  $PM_{2.5p90}$  e Focos, a temperatura mínima mensal é um fator de risco para as internações por asma.

Percebe-se que para os dois biomas a temperatura é um fator de risco para a internação por asma em crianças e adolescentes menores de 14 anos, quando os modelos foram ajustados por  $PM_{2.5}$  e Focos de Calor. Vale destacar que as variáveis explicativas incluídas nos modelos como estimativas da exposição à temperatura e ao  $PM_{2.5}$  foi diferente entre os biomas. Quanto à exposição à temperatura, para o bioma Pantanal a temperatura média foi mais importante no ajuste do modelo, enquanto que para Amazônia foi a temperatura mínima. Da mesma maneira, para as concentrações de  $PM_{2.5}$  no Pantanal foi o valor mínimo do bioma, enquanto que para Amazônia foi o percentil 90. Além das características climáticas distintas entre os biomas, destaca-se também que o número de municípios em cada bioma é diferente. Essas são características que podem explicar os resultados obtidos para as variáveis de exposição.

Quanto as limitações do estudo, é valido destacar que alguns dos Modelos de Séries Temporais estimados neste trabalho não possuíam normalidade nos resíduos e os Modelos Lineares Generalizados estimados neste trabalho não corrigiram a autocorrelação entre as medidas. Além disso, o estudo dos riscos e impactos podem ter limitações uma vez que os dados foram agrupados por biomas e não por seus municípios individualmente, assim como foram agregados por mês.

Sendo assim, recomenda-se um estudo com os dados diários, por município e utilizando modelos espaço-temporais. Espera-se que essas estratégias sejam alternativas melhores para previsões e estimar os riscos e impactos das exposições ambientais sobre as internações por asma em crianças e adolescentes.

## Referências

- ABDON, M. D. M. et al. Desmatamento no bioma pantanal até o ano 2002: relações com a fitofisionomia e limites municipais. *Embrapa Informática Agropecuária-Artigo em periódico indexado (ALICE)*, Revista Brasileira de Cartografia, Rio de Janeiro, v. 59, n. 1, p. 17-24 . . . , 2007.
- ALVES, L. Amazon fires coincide with increased respiratory illnesses in indigenous populations. *The Lancet Respiratory Medicine*, Elsevier, v. 8, n. 11, p. e84, 2020.
- ALVES, R.; DELGADO, C. Processos estocásticos. 1997.
- API, E. *ECMWF API*. 2021. Disponível em: (<https://cds.climate.copernicus.eu/api-how-to>).
- ARAGÃO, L.; JUNIOR, C. S.; ANDERSON, L. O desafio do brasil para conter o desmatamento e as queimadas na amazônia durante a pandemia por covid-19 em 2020: implicações ambientais, sociais e sua governança. *São José dos Campos*, 2020.
- ARBEX, M. A. et al. Air pollution from biomass burning and asthma hospital admissions in a sugar cane plantation area in brazil. *Journal of Epidemiology & Community Health*, BMJ Publishing Group Ltd, v. 61, n. 5, p. 395–400, 2007.
- ARTAXO, P. et al. Química atmosférica na amazônia: a floresta e as emissões de queimadas controlando a composição da atmosfera amazônica. *Acta Amazônica*, SciELO Brasil, 2005.
- BECKER, B. K. Geopolítica da amazônia. *Estudos avançados*, SciELO Brasil, v. 19, n. 53, p. 71–86, 2005.
- BOND, W. J.; WOODWARD, F. I.; MIDGLEY, G. F. The global distribution of ecosystems in a world without fire. *New phytologist*, Wiley Online Library, v. 165, n. 2, p. 525–538, 2005.
- BOX, G. E. et al. *Time series analysis: forecasting and control*. [S.l.]: John Wiley & Sons, 2015.
- CARVALHO, E. R.; ROCHA, H. A. L. Estudos epidemiológicos. *Site do Módulo de Epidemiologia e Bioestatística [Internet]*. Fortaleza: Universidade Federal do Ceará: Faculdade de Medicina, 2008.
- CORDEIRO, G. M.; DEMÉTRIO, C. G. Modelos lineares generalizados e extensões. *Piracicaba: USP*, 2008.
- COUTINHO, A. C.; CORNÉLIO, G. Precisão posicional dos dados de queimada na região do pantanal brasileiro. In: IN: SIMPÓSIO DE GEOTECNOLOGIAS NO PANTANAL, 3., 2010, CÁCERES, MT. ANAIS. *Embrapa Informática Agropecuária-Artigo em anais de congresso (ALICE)*. [S.l.], 2010.

- DATASUS. *Internacao por Asma*. 2021. Disponível em: <http://tabnet.datasus.gov.br/cgi/defctohtm.exe?sih/cnv/nrbr.def>.
- DIAS, C. S. et al. Influência do clima nas hospitalizações por asma em crianças e adolescentes residentes em belo horizonte, minas gerais, brasil. *Ciência & Saúde Coletiva*, SciELO Public Health, v. 25, p. 1979–1990, 2020.
- DICKEY, D. A.; FULLER, W. A. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, Taylor & Francis, v. 74, n. 366a, p. 427–431, 1979.
- DOBSON, A. J.; BARNETT, A. G. *An introduction to generalized linear models*. [S.l.]: CRC press, 2018.
- ECMWF. *ECMWF*. 2021. Disponível em: <https://cds.climate.copernicus.eu/#!/home>.
- ECMWF. *Variaveis Meteorologicas*. 2021. Disponível em: <https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=form>.
- ESPINOSA, M. M.; PRADO, S. M.; GHELLERE, M. Uso do modelo sarima na previsão do número de focos de calor para os meses de junho a outubro no estado de mato grosso. *Ciência e Natura*, v. 32, n. 2, p. 07–21, 2010.
- FERNANDES, V. V. Contribuições sobre o envelope simulado na análise de diagnóstico em modelos de regressão. Universidade Federal de São Carlos, 2019.
- GASPARRINI, A.; ARMSTRONG, B.; KENWARD, M. G. Distributed lag non-linear models. *Statistics in medicine*, Wiley Online Library, v. 29, n. 21, p. 2224–2234, 2010.
- GOUVEIA, N. et al. Hospitalizações por causas respiratórias e cardiovasculares associadas à contaminação atmosférica no município de são paulo, brasil. *Cadernos de Saúde Pública*, SciELO Brasil, v. 22, n. 12, p. 2669–2677, 2006.
- HARRIS, M. B. et al. Estimativas de perda da área natural da bacia do alto paraguai e pantanal brasileiro. *Relatório técnico não publicado. Conservação Internacional, Campo Grande, MS*, 2005.
- HOEF, J. M. V.; BOVENG, P. L. Quasi-poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology*, Wiley Online Library, v. 88, n. 11, p. 2766–2772, 2007.
- HORN, M. G. C. Responsabilidade ambiental nas queimadas urbanas. 2016.
- IBGE. *Municípios Biomas Brasil*. 2019. Disponível em: <https://www.ibge.gov.br/geociencias/cartas-e-mapas/informacoes-ambientais/15842-biomas.html?=&t=downloads>.
- IBGE. *Síntese dos Biomas Amazônia e Pantanal*. 2019. Disponível em: [https://geoftp.ibge.gov.br/informacoes\\_ambientais/estudos\\_ambientais/biomas/documentos/Sintese\\_Descricao\\_Biomas.pdf](https://geoftp.ibge.gov.br/informacoes_ambientais/estudos_ambientais/biomas/documentos/Sintese_Descricao_Biomas.pdf).
- INPE. *Inpe Queimadas*. 2021. Disponível em: <https://queimadas.dgi.inpe.br/queimadas/bdqueimadas#exportar-dados>.

- KAUFMAN, Y. et al. Smoke, clouds, and radiation-brazil (scar-b) experiment. *Journal of Geophysical Research: Atmospheres*, Wiley Online Library, v. 103, n. D24, p. 31783–31808, 1998.
- LANA, M. D. et al. Distribuição diamétrica de *eschweilera ovata* em um fragmento de floresta ombrófila densa-igarassu, pe. *Floresta*, v. 43, n. 1, p. 59–68, 2013.
- LEMES, M. d. C. R.; REBOITA, M. S.; CAPUCIN, B. C. Impactos das queimadas na amazônia no tempo em são paulo na tarde do dia 19 de agosto de 2019. *Revista Brasileira de Geografia Física*, v. 13, n. 3, p. 983–993, 2020.
- LJUNG, G. M.; BOX, G. E. On a measure of lack of fit in time series models. *Biometrika*, Oxford University Press, v. 65, n. 2, p. 297–303, 1978.
- MENDES, M. et al. Impactos das queimadas sobre a saúde da população humana na amazônia maranhense/biomass burning impact on the human population health in the amazon region from maranhão. *Revista de Pesquisa em Saúde*, v. 17, n. 3, 2017.
- MIRANDA, E. E. de; MARTINHO, P. R. R.; CARVALHO, C. A. de. Nota técnica sobre queimadas, desmatamentos e imóveis rurais no bioma amazônia em 2019. *Embrapa Territorial-Nota Técnica/Nota Científica (ALICE)*, Embrapa Territorial: Campinas, 2020., 2020.
- MOHER, D. et al. Preferred reporting items for systematic reviews and meta-analyses: the prisma statement. *Int J Surg*, v. 8, n. 5, p. 336–341, 2010.
- MORAES, E. C. et al. Estudo da dinâmica de queimada no bioma pantanal no período de 2002 a 2015. *SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO*, v. 18, p. 3423–3430, 2017.
- MORETTIN, P. A.; TOLOI, C. Análise de séries temporais. In: *Análise de séries temporais*. [S.l.: s.n.], 2006. p. 538–538.
- MOUTINHO, P. et al. Amazônia em chamas: desmatamento e fogo em tempos de covid-19. *Nota Técnica*, n. 4, 2020.
- NASCIMENTO, L. F. C.; MEDEIROS, A. P. P. d. Internações por pneumonias e queimadas: uma abordagem espacial. *Jornal de Pediatria*, SciELO Brasil, v. 88, n. 2, p. 177–183, 2012.
- NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, Wiley Online Library, v. 135, n. 3, p. 370–384, 1972.
- OLIVEIRA, B. F. A. de et al. Risk assessment of pm 2.5 to child residents in brazilian amazon region with biofuel production. *Environmental health*, BioMed Central, v. 11, n. 1, p. 1–11, 2012.
- OMS. *Material Particulado Fino recomendado*. 2021. Disponível em: <<https://www.who.int/tools/compendium-on-health-and-environment>>.
- PERPEROGLOU, A. et al. A review of spline function procedures in r. *BMC medical research methodology*, BioMed Central, v. 19, n. 1, p. 1–16, 2019.

- PRISMA. *Metodologia PRISMA*. 2020. Disponível em: <http://www.prisma-statement.org/>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2014. Disponível em: <http://www.R-project.org/>.
- REDDINGTON, C. et al. Air quality and human health improvements from reductions in deforestation-related fire in Brazil. *Nature Geoscience*, Nature Publishing Group, v. 8, n. 10, p. 768–771, 2015.
- RIBEIRO, H.; ASSUNÇÃO, J. V. d. Efeitos das queimadas na saúde humana. *Estudos avançados*, SciELO Brasil, v. 16, n. 44, p. 125–148, 2002.
- RODRIGUES, P. C. O. et al. Distribuição espacial das internações por asma em idosos na Amazônia brasileira. *Revista Brasileira de Epidemiologia*, SciELO Public Health, v. 13, p. 523–532, 2010.
- ROSA, A. M. et al. Prevalence of asthma in children and adolescents in a city in the Brazilian Amazon region. *J Bras Pneumol*, v. 35, n. 1, p. 7–13, 2009.
- SALDANHA, C.; BOTELHO, C. Queimadas e suas influências em crianças asmáticas menores de cinco anos atendidas em um hospital público. *Rev. bras. alerg. imunopatol.*, v. 31, n. 3, p. 108–112, 2008.
- SALDANHA, C. T.; SILVA, A. M. C. d.; BOTELHO, C. Variações climáticas e uso de serviços de saúde em crianças asmáticas menores de cinco anos de idade: um estudo ecológico. *Jornal Brasileiro de Pneumologia*, SciELO Brasil, v. 31, n. 6, p. 492–498, 2005.
- SANTOS, A. C. A. Absorção da radiação solar por partículas de aerossóis no pantanal mato-grossense. Universidade Federal de Mato Grosso, 2018.
- SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). *Biometrika*, JSTOR, v. 52, n. 3/4, p. 591–611, 1965.
- SILVA, A. M. C. d. et al. Material particulado originário de queimadas e doenças respiratórias. *Revista de Saúde Pública*, SciELO Public Health, v. 47, p. 345–352, 2013.
- SILVA, P. R. et al. Hospitalização de crianças por asma na Amazônia brasileira: tendência e distribuição espacial. *Jornal de Pediatria*, SciELO Brasil, v. 85, n. 6, p. 541–546, 2009.
- SILVA, P. R. d. S. et al. High risk of respiratory diseases in children in the fire period in western Amazon. *Revista de saúde pública*, SciELO Public Health, v. 50, p. 29, 2016.
- SISAM. *Sisam Poluição*. 2021. Disponível em: <https://queimadas.dgi.inpe.br/queimadas/sisam/v2/dados/download/>.
- XAVIER, J. M. N. *Análise e previsão de séries temporais com modelos ARIMA e análise espectral singular*. Tese (Doutorado), 2016.

# APÊNDICE 1 – Gráficos Análises Descritivas

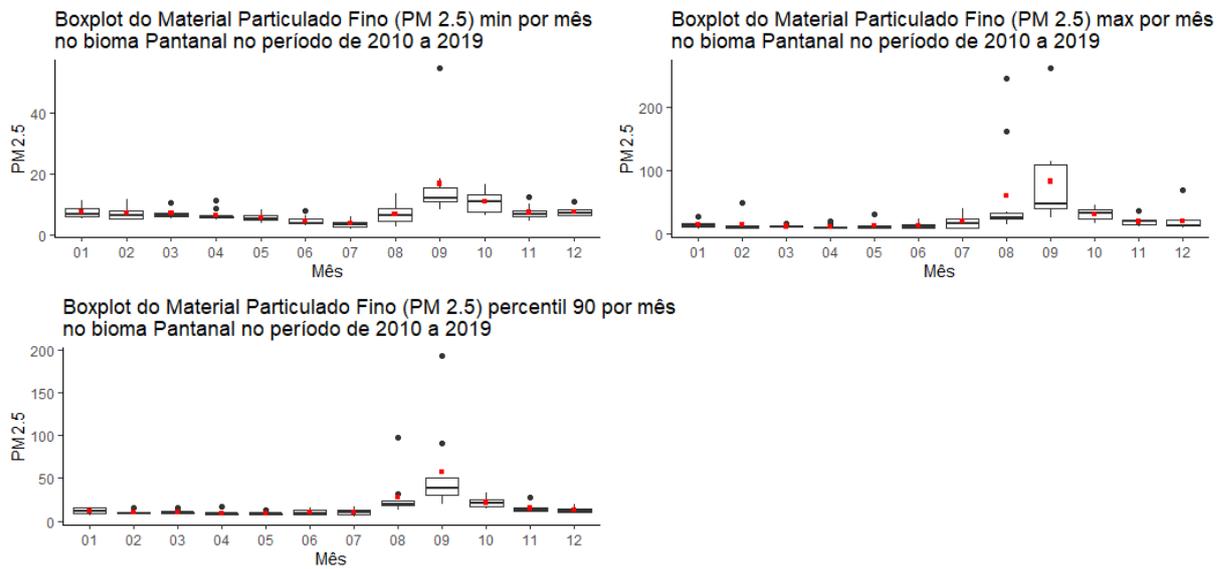


Figura 37: Boxplot das variáveis de  $pm2.5min$ ,  $pm2.5max$  e  $pm2.5p90$  no bioma do Pantanal de jan/2010 até dez/2019

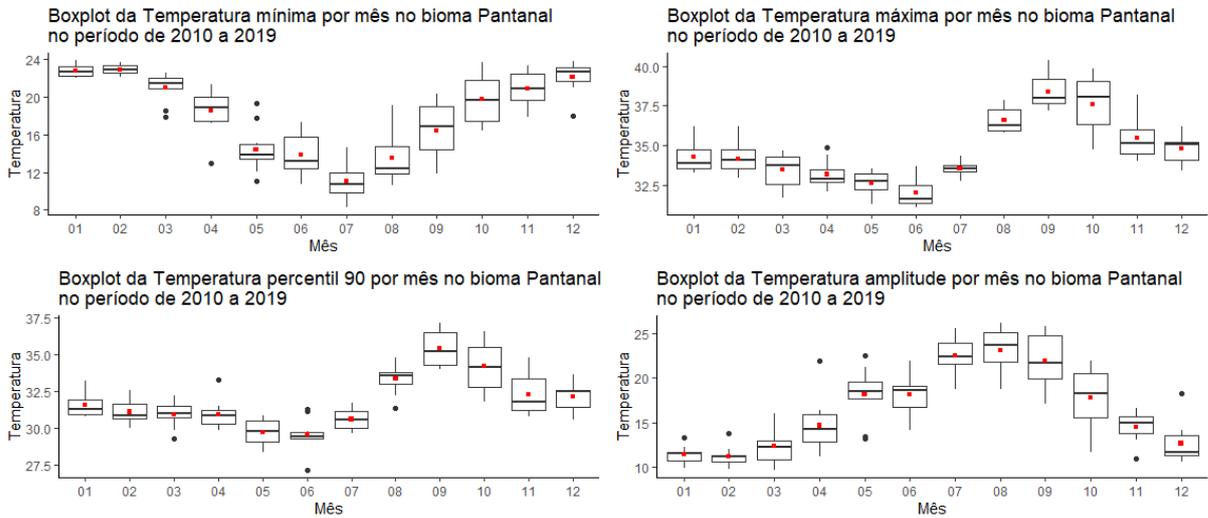


Figura 38: Boxplot das variáveis de *tempmin*, *tempmax*, *temp90* e *tempamplitude* no bioma do Pantanal de jan/2010 até dez/2019

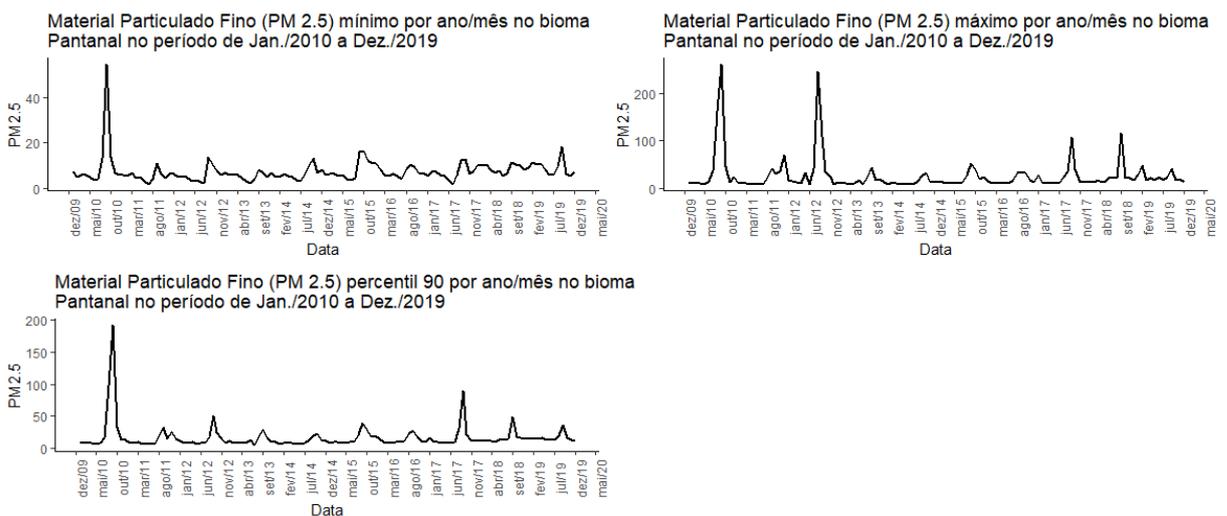


Figura 39: Série Temporal das variáveis de *pm2.5min*, *pm2.5max* e *pm2.5p90* no bioma do Pantanal de jan/2010 até dez/2019

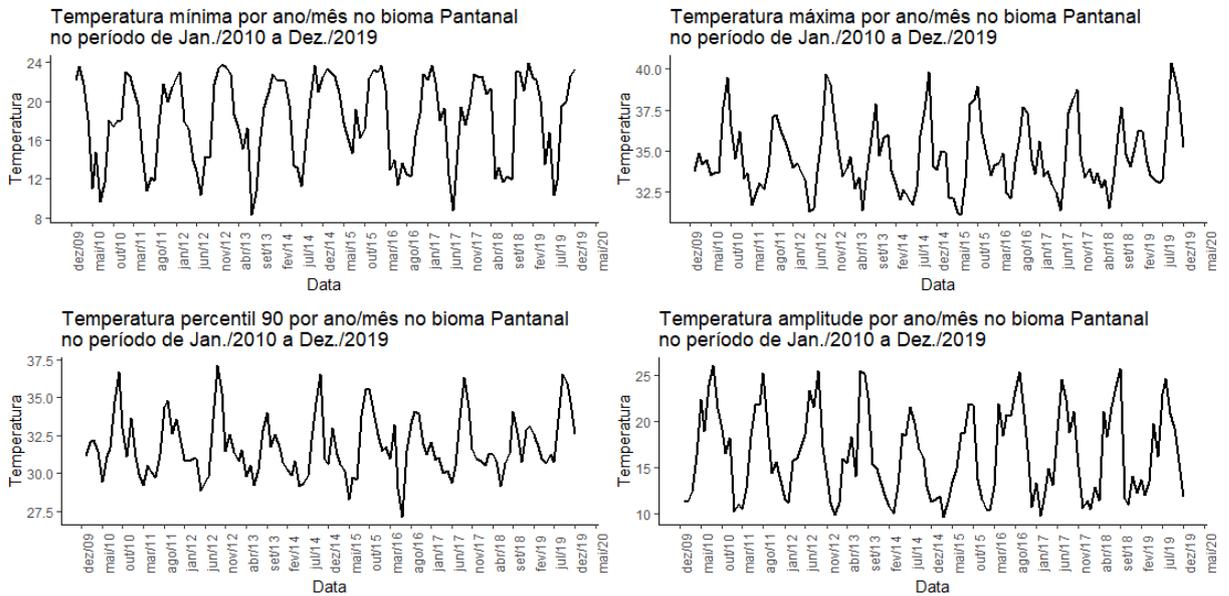


Figura 40: Série Temporal das variáveis de *tempmin*, *tempmax*, *temp90* e *tempamplitude* no bioma do Pantanal de jan/2010 até dez/2019

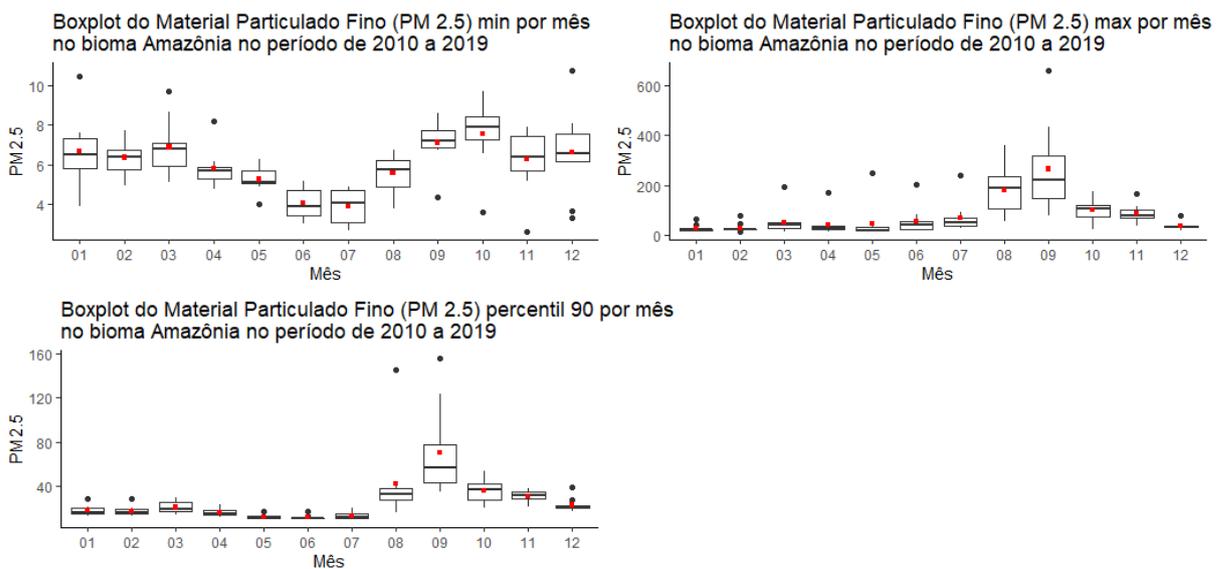


Figura 41: Boxplot das variáveis de *pm2.5min*, *pm2.5max* e *pm2.5p90* no bioma da Amazônia de jan/2010 até dez/2019

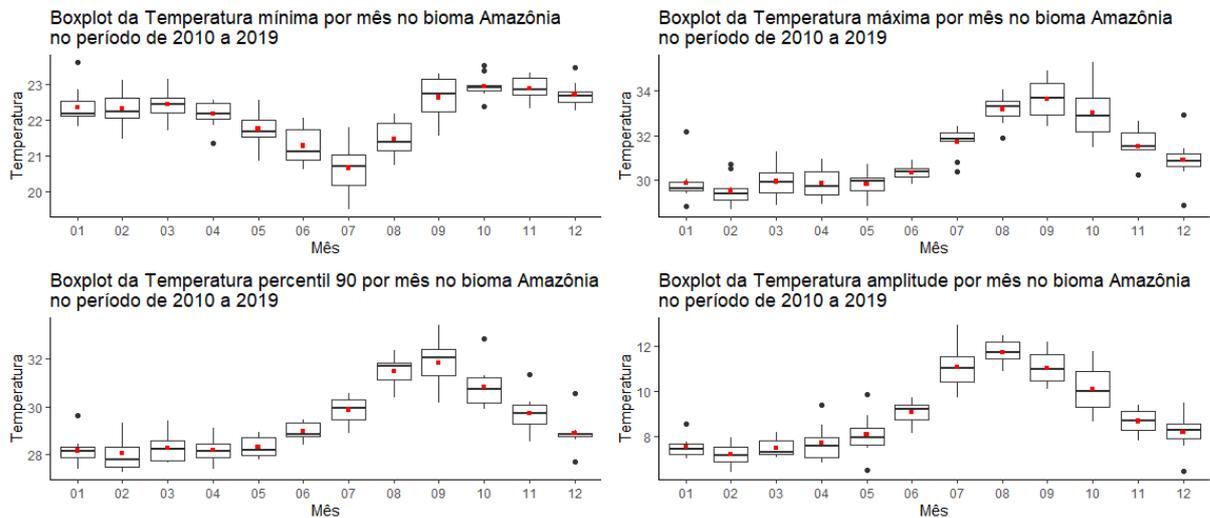


Figura 42: Boxplot das variáveis de *tempmin*, *tempmax*, *temp90* e *tempamplitude* no bioma da Amazônia de jan/2010 até dez/2019

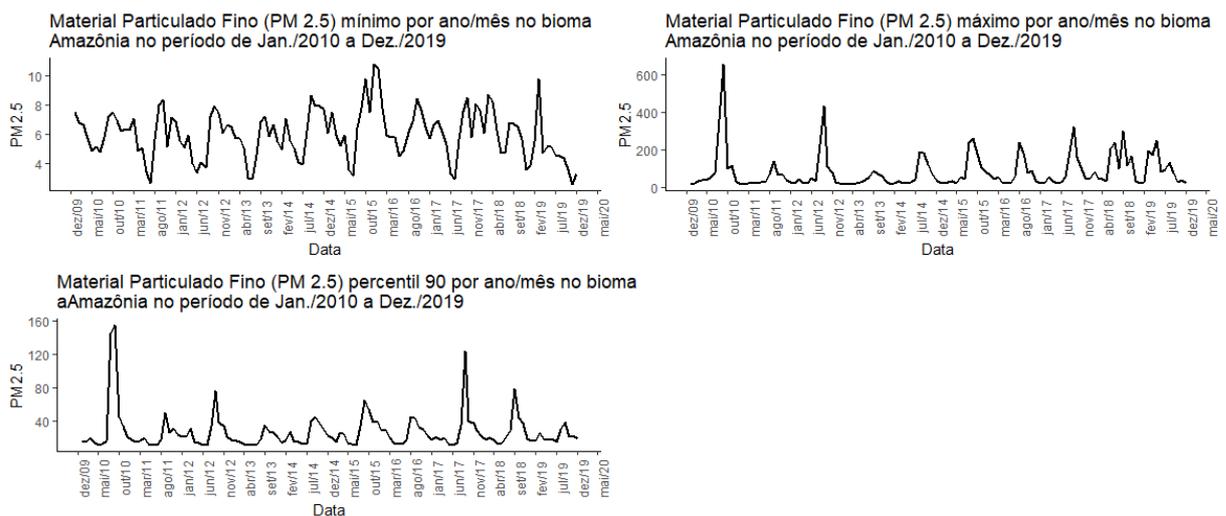


Figura 43: Série Temporal das variáveis de *pm2.5min*, *pm2.5max* e *pm2.5p90* no bioma da Amazônia de jan/2010 até dez/2019

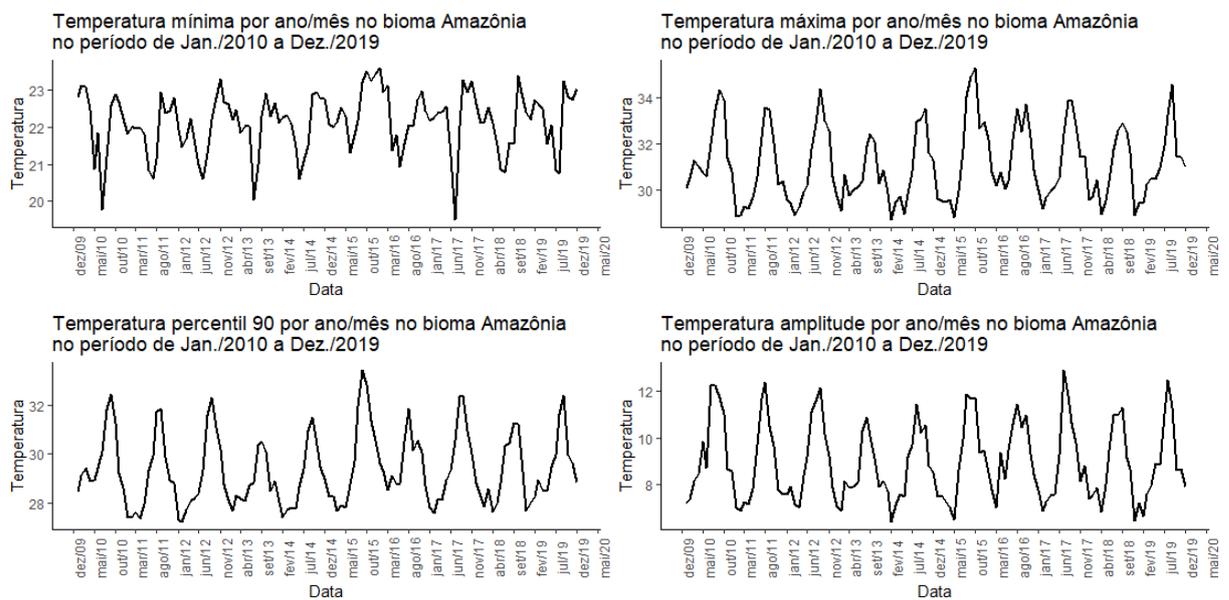


Figura 44: Série Temporal das variáveis de *tempmin*, *tempmax*, *tempp90* e *tempamplitude* no bioma da Amazônia de jan/2010 até dez/2019

## APÊNDICE 2 – Gráficos Séries Temporais Pantanal

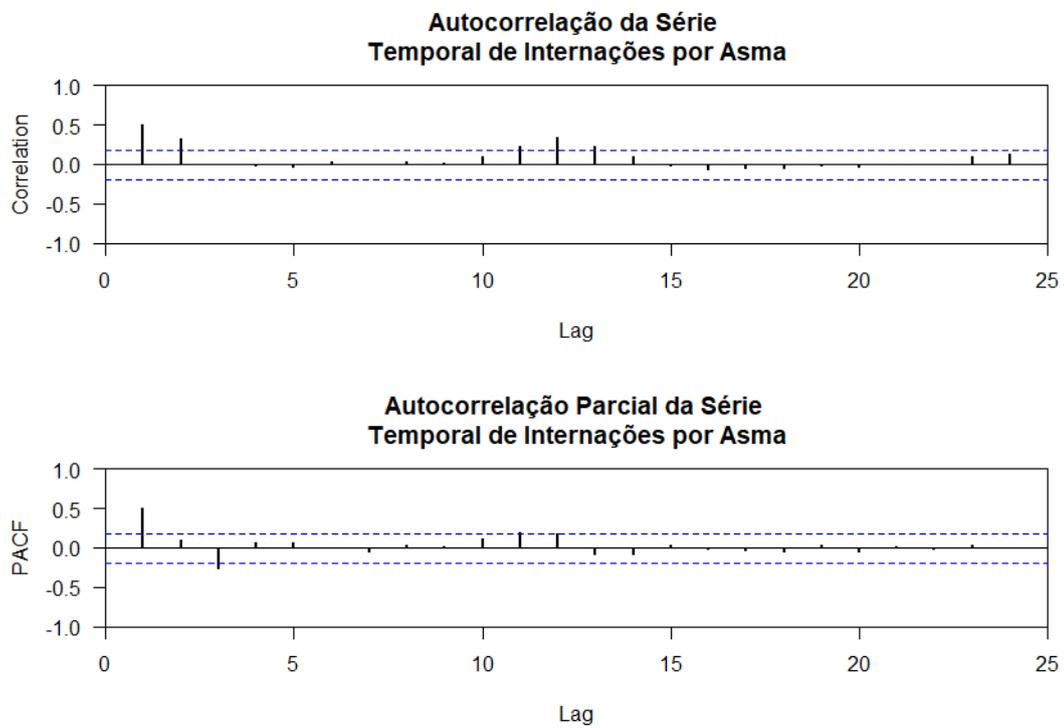


Figura 45: Autocorrelação e Autocorrelação Parcial da série temporal de Interações por Asma em crianças e adolescentes de 0 a 14 anos no bioma do Pantanal de jan/2010 até dez/2019

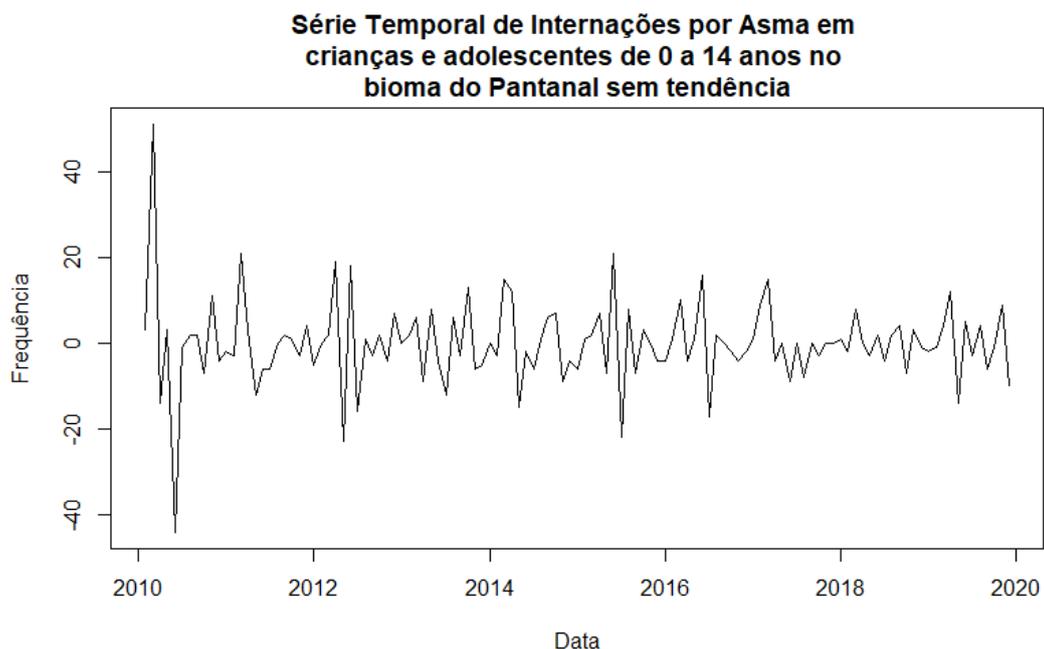


Figura 46: Série Temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma do Pantanal sem tendência

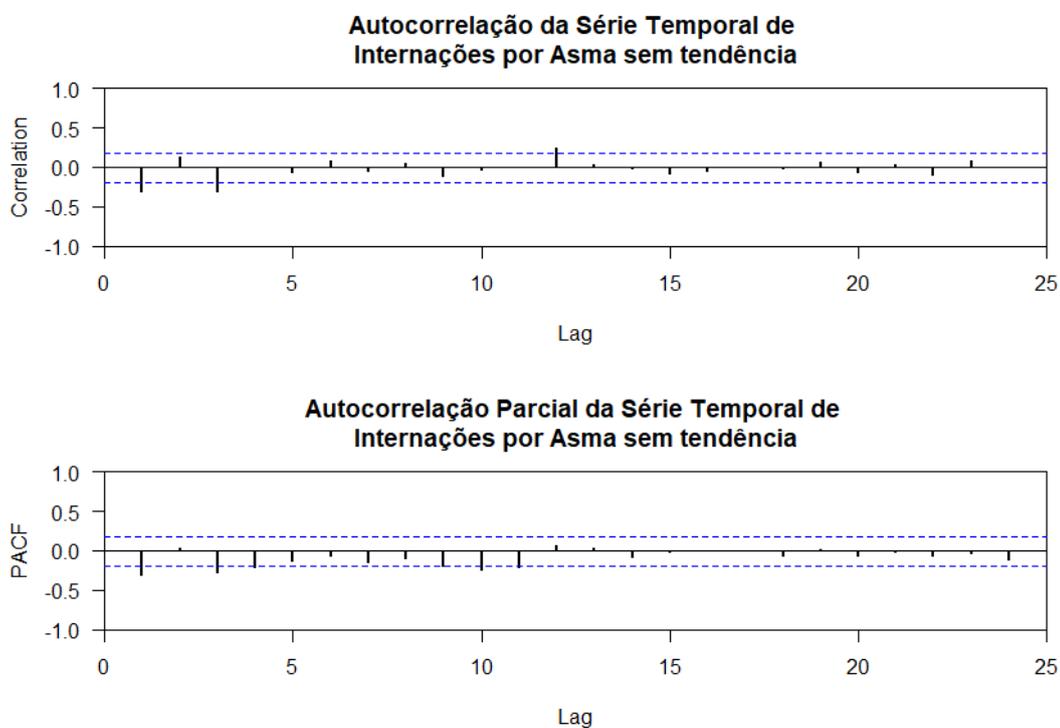


Figura 47: Autocorrelação e Autocorrelação Parcial da série temporal de Internações por Asma no bioma do Pantanal sem tendência

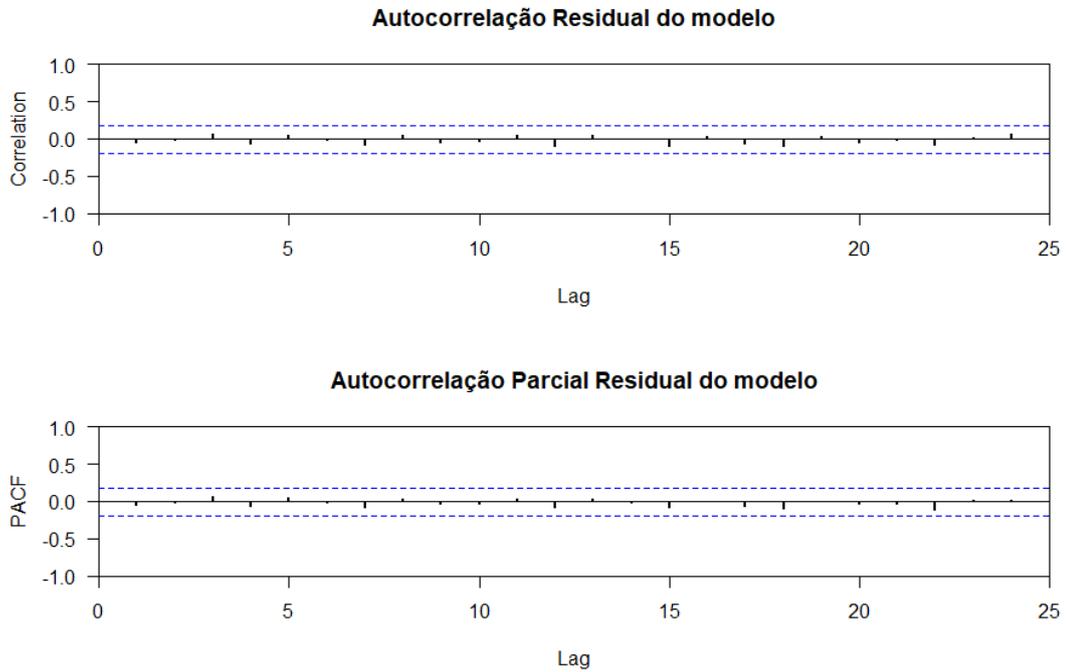


Figura 48: Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Interações por Asma SARIMA(3,1,1)(0,0,1)[12] para o bioma Pantanal

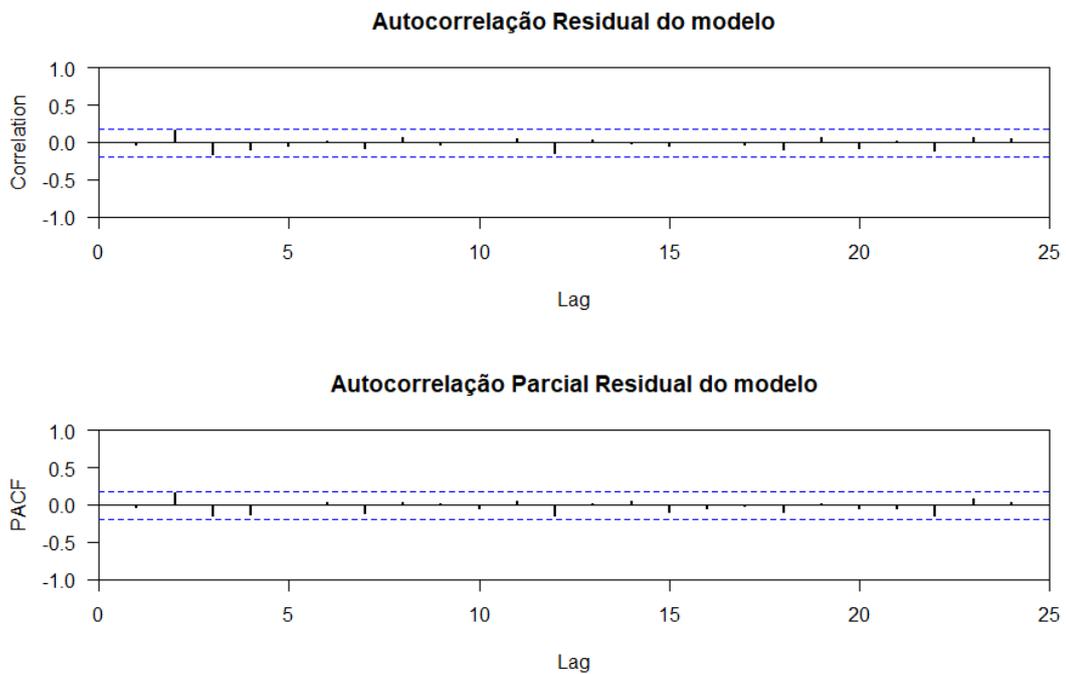


Figura 49: Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Interações por Asma SARIMA(1,1,1)(0,0,1)[12] para o bioma Pantanal

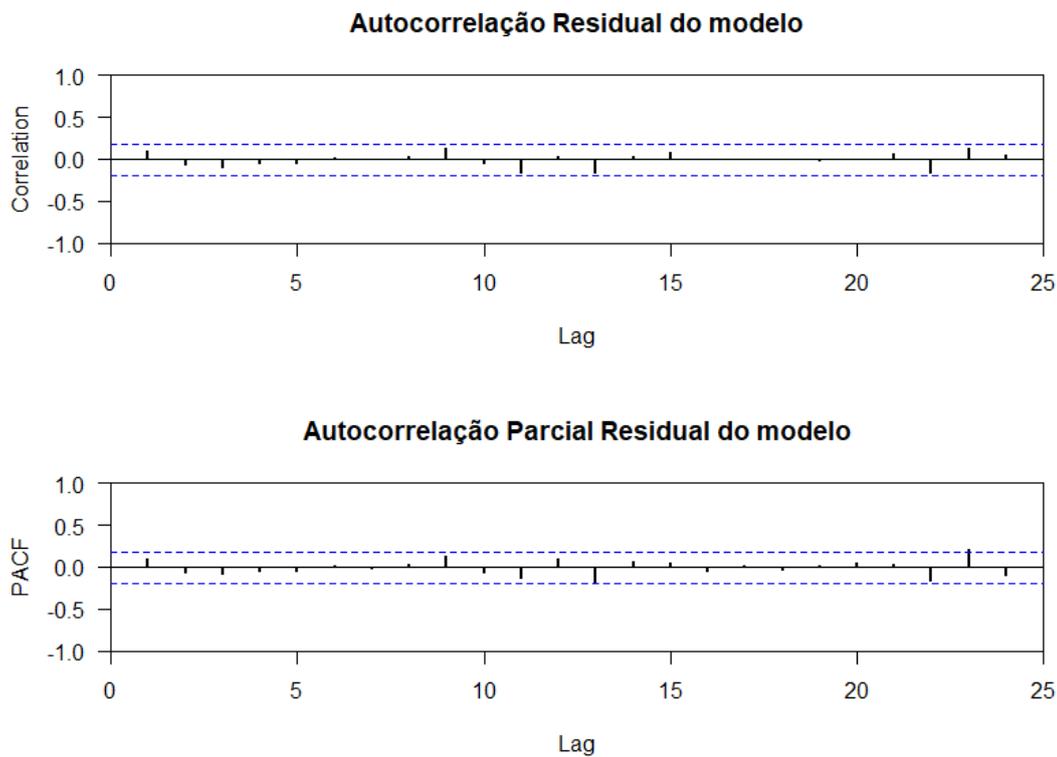


Figura 50: Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Focos de Queimadas SARIMA(1,0,0)(1,1,1)[12] para o bioma Pantanal

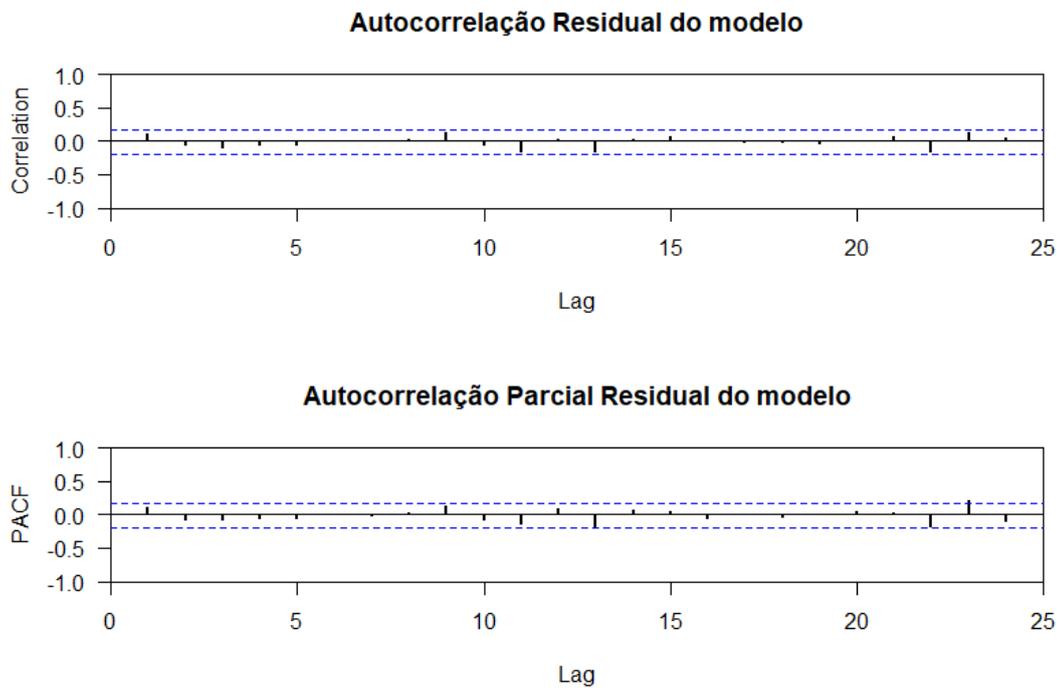


Figura 51: Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Material Particulado Fino médio SARIMA(0,0,1)(2,0,0)[12] para o bioma Pantanal

## APÊNDICE 3 – Gráficos Séries Temporais Amazônia

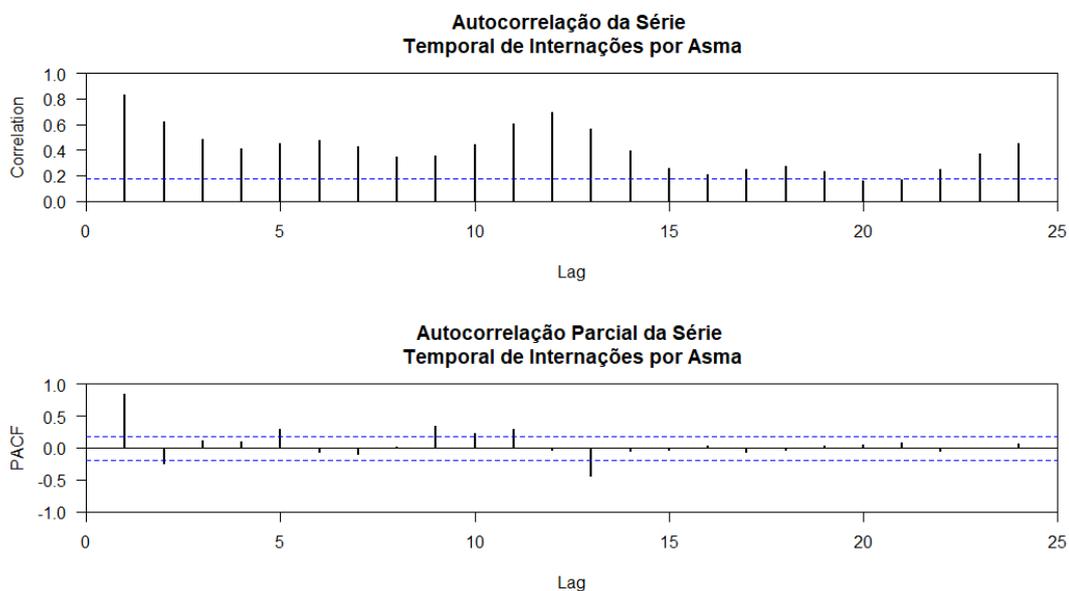


Figura 52: Autocorrelação e Autocorrelação Parcial da série temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia de jan/2010 até dez/2019

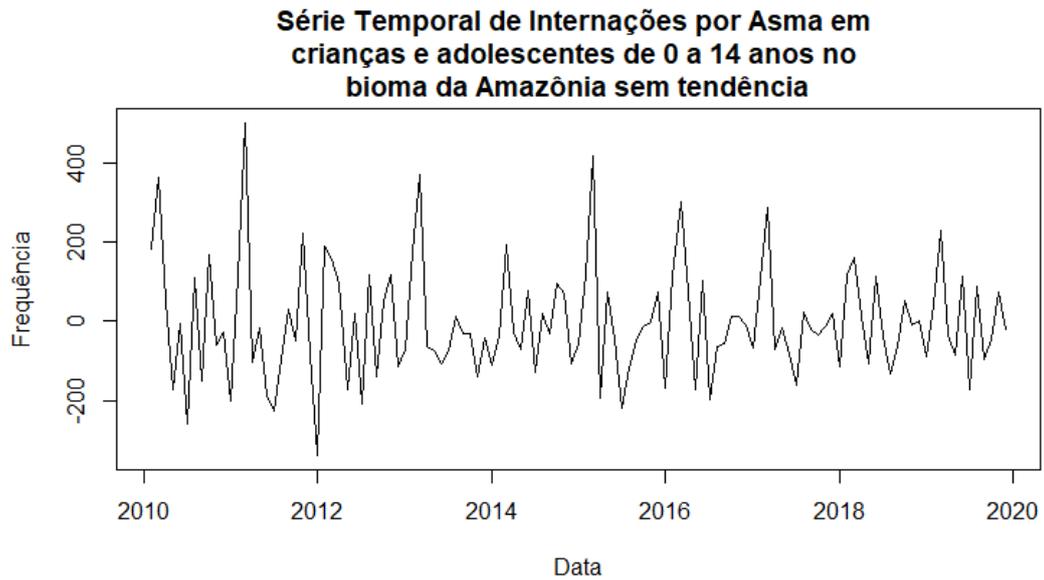


Figura 53: Série Temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia sem tendência

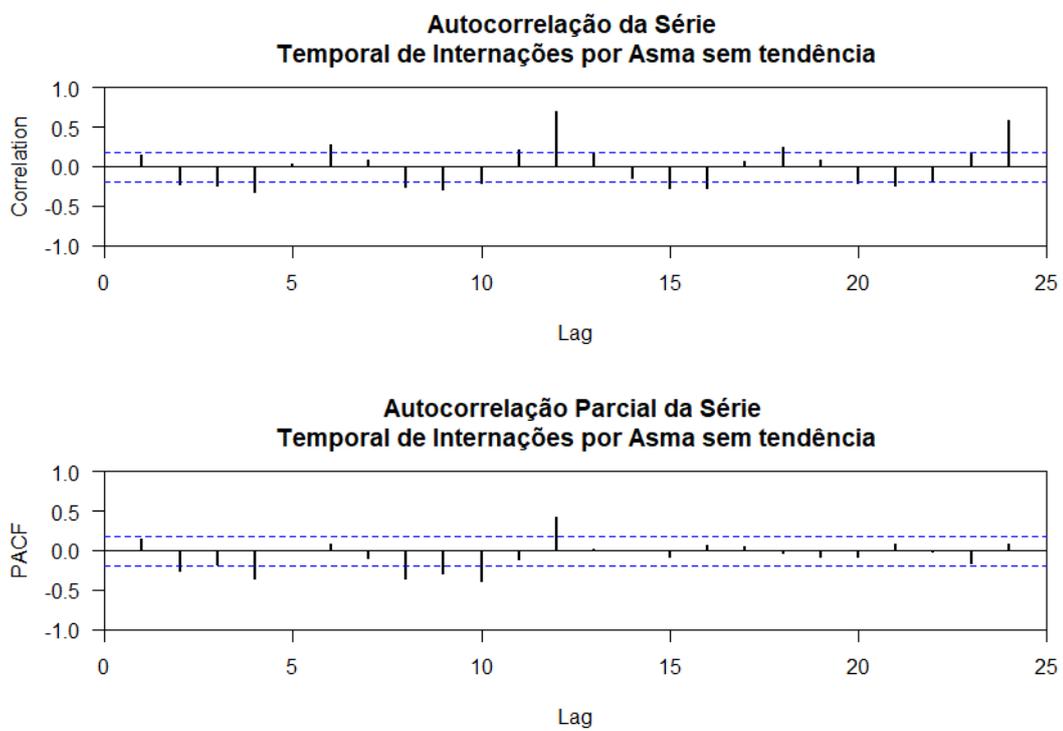


Figura 54: Autocorrelação e Autocorrelação Parcial da série temporal de Internações por Asma no bioma da Amazônia sem tendência

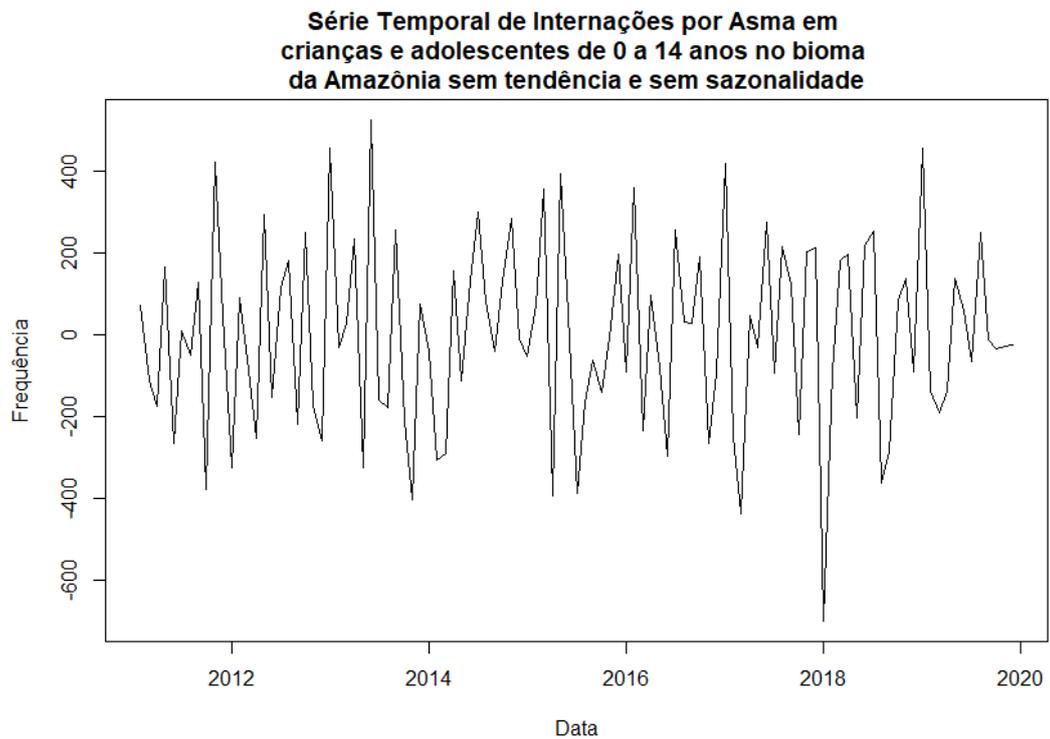


Figura 55: Série Temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia sem tendência e sazonalidade

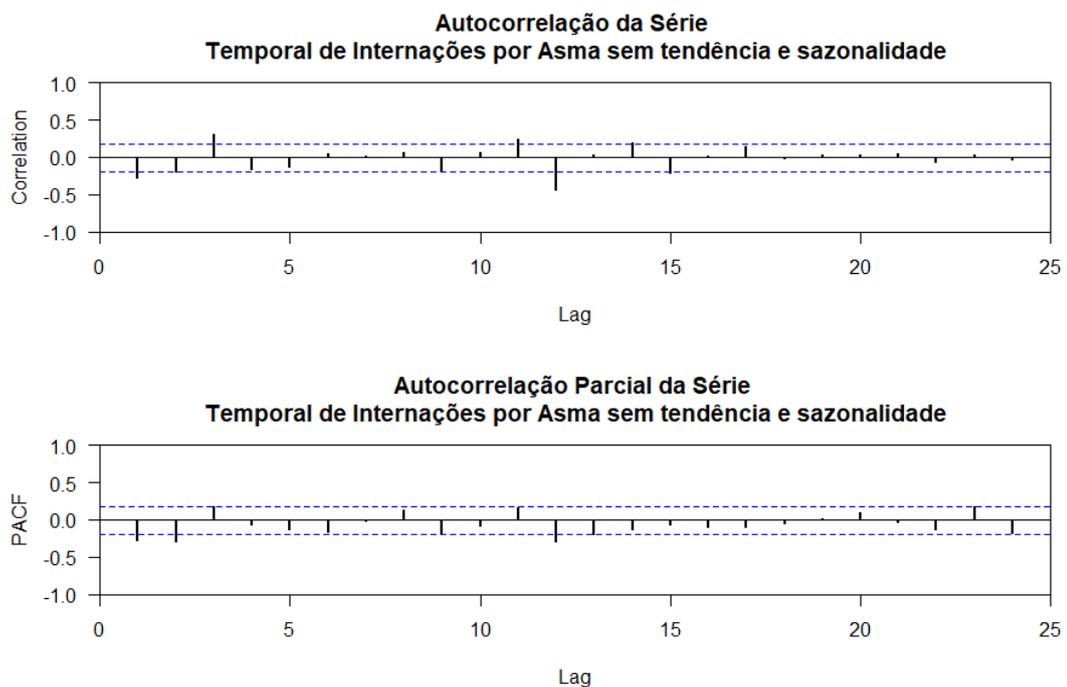


Figura 56: Autocorrelação da Série Temporal de Internações por Asma em crianças e adolescentes de 0 a 14 anos no bioma da Amazônia sem tendência e sazonalidade

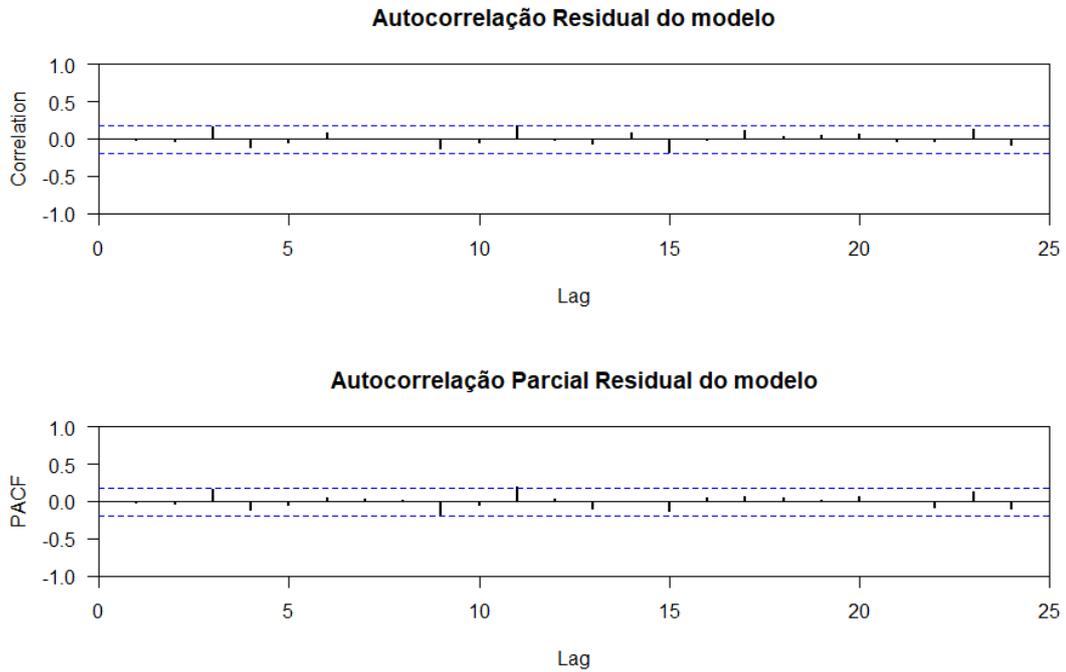


Figura 57: Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Interações por Asma SARIMA(1,1,1)(1,1,1)[12] para o bioma da Amazônia

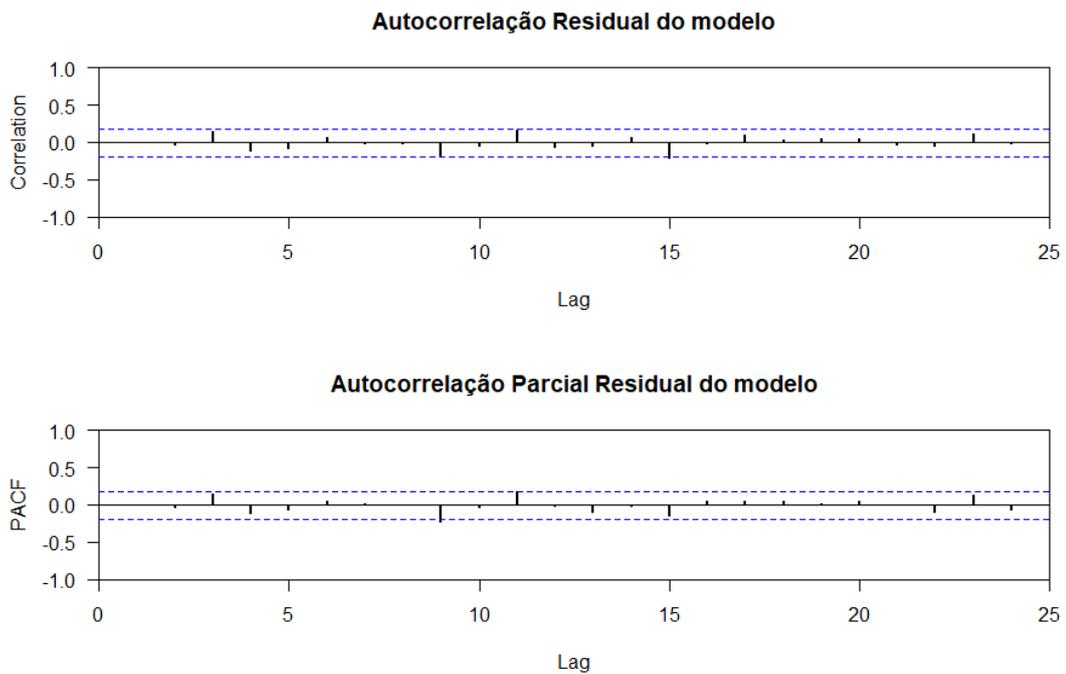


Figura 58: Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Interações por Asma SARIMA(1,0,0)(0,1,1)[12] para o bioma da Amazônia

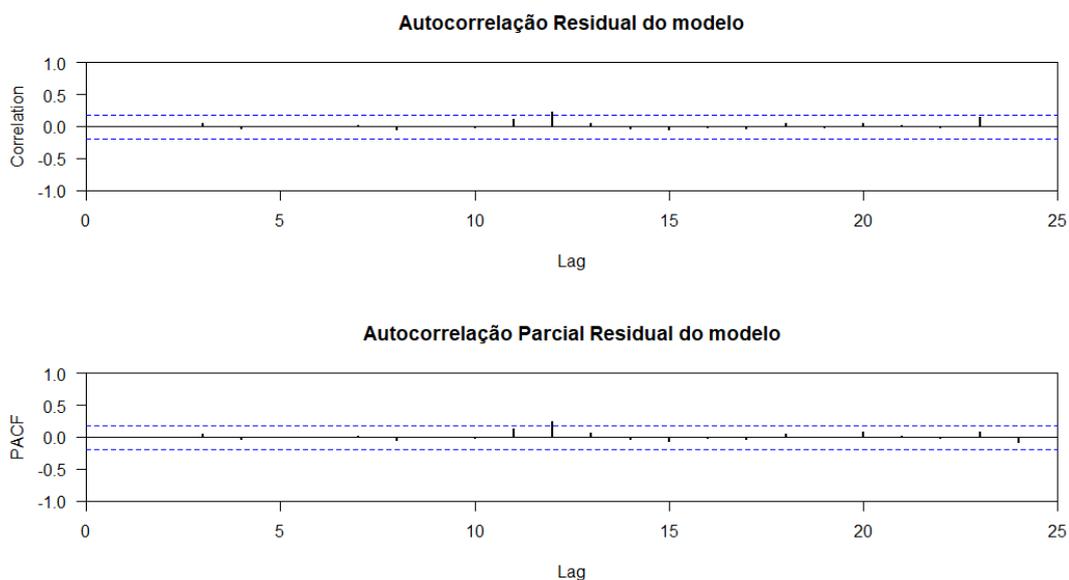


Figura 59: Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Focos de Queimadas SARIMA(1,0,0)(2,1,0)[12] para o bioma da Amazônia

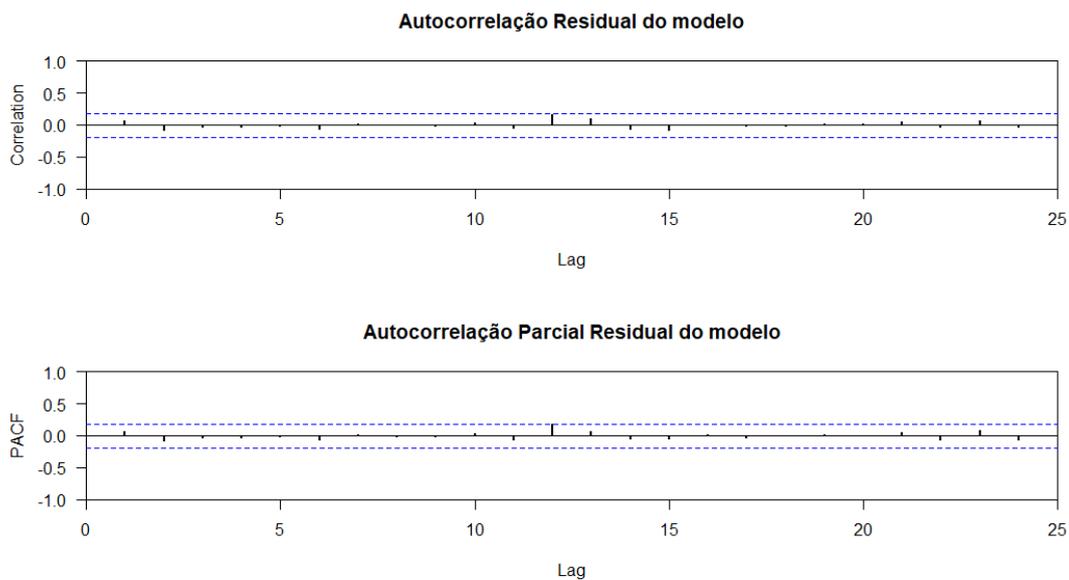


Figura 60: Autocorrelação e Autocorrelação Parcial dos resíduos do modelo de Material Particulado Fino médio SARIMA(1,0,0)(1,1,0)[12] para o bioma da Amazônia

## APÊNDICE 4 – Gráficos Modelagem

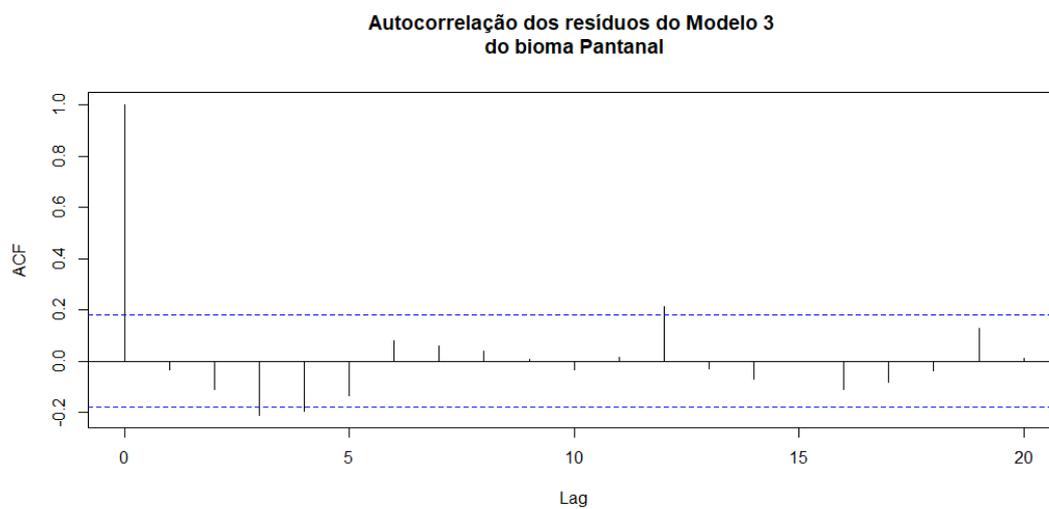


Figura 61: Autocorrelação dos resíduos do modelo 3 do bioma Pantanal

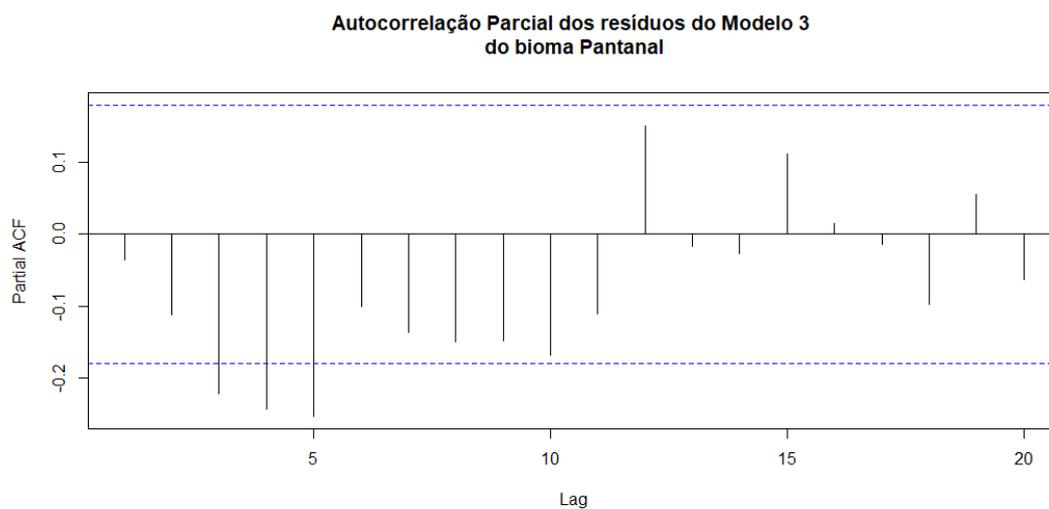


Figura 62: Autocorrelação Parcial dos resíduos do modelo 3 do bioma Pantanal

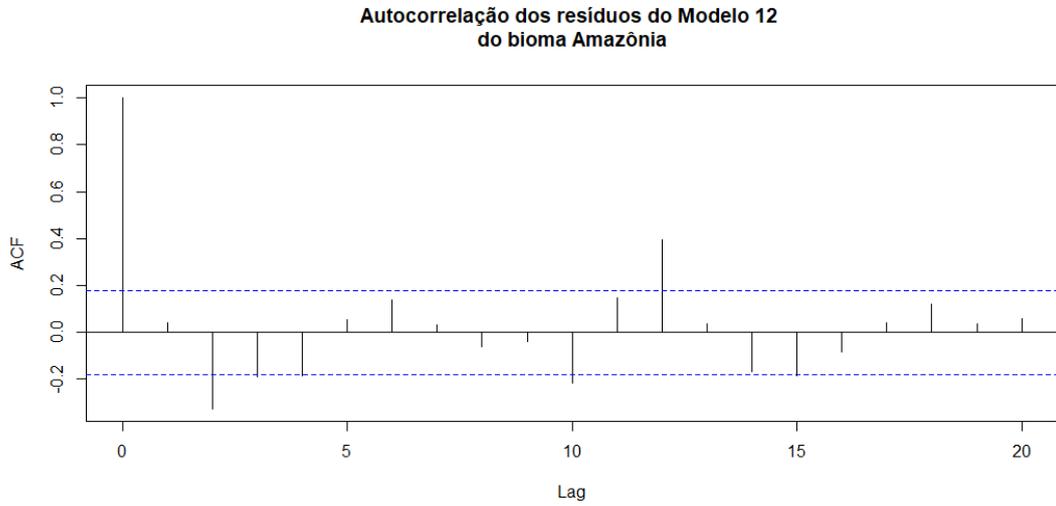


Figura 63: Autocorrelação dos resíduos do modelo 12 do bioma Amazônia

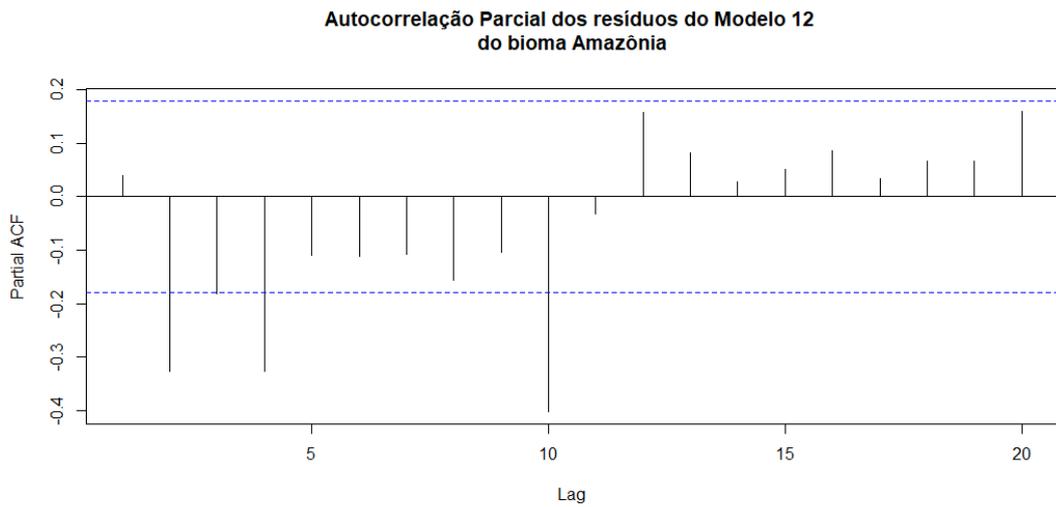


Figura 64: Autocorrelação Parcial dos resíduos do modelo 12 do bioma Amazônia

## APÊNDICE 5 – Variável de Temperatura

1. Primeiramente, foi necessário entrar no site do European Centre for Medium-Range Weather Forecasts (ECMWF) e criar um login no site e depois buscar pelo nome ”*ERA5-Land hourly data from 1981 to present*”.
2. Após realizar a busca, clicou-se no primeiro resultado, que coincide com a busca feita anteriormente. A nova página que é carregada está sendo mostrada na Figura 65. Na parte de *Overview*, há toda a explicação do ERA5 Land e uma explicação detalhada de cada variável de possível escolha na parte de *Download data*.

The screenshot shows the website interface for ERA5-Land data. At the top, there are logos for Copernicus, ECMWF, and Climate Change Service. A navigation menu includes 'Home', 'Search', 'Datasets', 'Applications', 'Toolbox', 'Support', and 'Live'. The main heading is 'ERA5-Land hourly data from 1981 to present'. Below this, there are tabs for 'Overview', 'Download data', 'Quality assessment', and 'Documentation'. The 'Overview' tab is active, displaying a text block about the dataset and a world map titled 'Mean 2m temperature for February 2019 from ERA5-Land'. A sidebar on the right contains 'Contact', 'Licence', 'Publication date', and 'Resource updated'.

Figura 65: Overview e Download data ERA5 Land site ECMWF

Ao mudar de guia e selecionar *Download data*, aparecem as variáveis disponíveis para escolha. No caso deste trabalho, foram selecionadas as seguintes variáveis:

- *2m temperature*, do campo *Temperature*.
- Os anos de 2010 a 2019, do campo *Year*, de forma que pela alto uso computacional, selecionou-se um ano por requisição.
- Dois meses por vez, da forma: *January, February, ... , November, December*”, do campo *Month*.

- Todos os dias disponíveis no campo *Day*.
  - Todas as horas disponíveis no campo *Time*.
  - Para o campo de *Geographical area*, foi selecionado o item *Sub-region extraction*, de modo que foi colocado o valor 10 em *North*, -80 em *West*, -30 em *East* e -35 em *South*.
  - No campo *Format* foi selecionado o *NetCDF (experimental)*.
3. Após fazer a seleção conforme indicado e ter feito o *login* no site, pode-se clicar em *Submit Form* para baixar os dados no próprio navegador ou pode-se clicar em *Show API request*, de modo que devem ser feitas algumas modificações para utilizar esta API no *software R*, visto que a requisição é dada através da linguagem *Python*. Uma documentação relacionada a esta API se encontra em API. É importante destacar que caso seja a primeira vez em que o *download* seja realizado, deve-se aceitar os termos de uso que aparecem no campo *Terms of use*.
  4. Um exemplo de uma requisição feita selecionando as variáveis indicadas no passo 2, porém apenas sendo selecionado o ano de 2010 e os meses de Janeiro e Fevereiro, se encontra na Figura 66.

```
import cdsapi

c = cdsapi.Client()

c.retrieve(
    'reanalysis-era5-land',
    {
        'variable': '2m_temperature',
        'year': '2010',
        'month': [
            '01', '02',
        ],
        'day': [
            '01', '02', '03',
            '04', '05', '06',
            '07', '08', '09',
            '10', '11', '12',
            '13', '14', '15',
            '16', '17', '18',
            '19', '20', '21',
            '22', '23', '24',
            '25', '26', '27',
            '28', '29', '30',
            '31',
        ],
        'time': [
            '00:00', '01:00', '02:00',
            '03:00', '04:00', '05:00',
            '06:00', '07:00', '08:00',
            '09:00', '10:00', '11:00',
            '12:00', '13:00', '14:00',
            '15:00', '16:00', '17:00',
            '18:00', '19:00', '20:00',
            '21:00', '22:00', '23:00',
        ],
        'area': [
            10, -80, -35,
            -30,
        ],
        'format': 'netcdf',
    },
    'download.nc')
```

Figura 66: Exemplo de uma requisição realizada pela API do site ECMWF