



MARCELO SÁVIO RAMOS

**MODELAGEM DE VALORES EXTREMOS NA AVALIAÇÃO
DA QUALIDADE DO AR PÓS-ROMPIMENTO DA
BARRAGEM EM BRUMADINHO, MG**

**LAVRAS – MG
2022**

MARCELO SÁVIO RAMOS

**MODELAGEM DE VALORES EXTREMOS NA AVALIAÇÃO DA QUALIDADE DO
AR PÓS-ROMPIMENTO DA BARRAGEM EM BRUMADINHO, MG**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

Prof^ª. Dr^ª. Thelma Sáfadi
Orientadora

**LAVRAS – MG
2022**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Ramos, Marcelo Sávio

Modelagem de valores extremos na avaliação da qualidade
do ar pós-rompimento da barragem em Brumadinho, MG /
Marcelo Sávio Ramos. – Lavras : UFLA, 2022.

104 p. : il.

Orientadora: Prof^ª. Dr^ª. Thelma Sáfyadi.

Dissertação (mestrado acadêmico)- Universidade Federal
de Lavras, 2022.

Bibliografia.

1. Teoria de valores extremos. 2. Estacionariedade. 3.
Valor em risco. I. Sáfyadi, Thelma. II. Título.

MARCELO SÁVIO RAMOS

**MODELAGEM DE VALORES EXTREMOS NA AVALIAÇÃO DA QUALIDADE DO
AR PÓS-ROMPIMENTO DA BARRAGEM EM BRUMADINHO, MG**

**MODELING EXTREME VALUES IN THE EVALUATION OF AIR QUALITY AFTER
THE DAM FAILURE IN BRUMADINHO, MG**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Mestre.

APROVADA em 31 de Maio de 2022.

Prof. Dr. Luiz Alberto Beijo UNIFAL
Prof. Dr. Renato Ribeiro de Lima UFLA

Prof^ª. Dr^ª. Thelma Sáfadi
Orientadora

**LAVRAS – MG
2022**

AGRADECIMENTOS

Agradeço primeiramente a Deus, por estar presente todos os dias na minha vida, pelas oportunidades e a capacidade de concluir esse trabalho.

À minha orientadora, Prof^a. Thelma, sempre atenciosa e paciente na orientação e por todo apoio para a realização desta pesquisa.

À Universidade Federal de Lavras, por me proporcionar a oportunidade de fazer pós-graduação e por oferecer ótimas instalações e ensino gratuito de qualidade.

Aos professores e colegas do Departamento de Estatística pelos ensinamentos e toda ajuda nesta minha trajetória.

Aos membros da banca, Prof. Luiz e Prof. Renato pela disposição em ler este trabalho e sugestões para o enriquecimento deste trabalho.

Aos amigos de Joinville, especialmente a Tatiana, pelo incentivo, carinho e companheirismo em todos os momentos.

Às demais pessoas que conheci nesta bela e calorosa cidade que me acolheu, os meus sinceros agradecimentos.

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

*"As nuvens mudam sempre de posição,
mas são sempre nuvens no céu.
Assim devemos ser todo dia, mutantes,
porém leais com o que pensamos e sonhamos;
Lembre-se, tudo se desmancha no ar,
menos os pensamentos."
(Paulo Beleki)*

RESUMO

Atualmente existem estudos que indicam um aumento significativo da poluição do ar em todo o mundo, o que pode causar sérios danos ambientais e à saúde humana. Dentre os inúmeros fatores que contribuem para esse aumento, a atividade de mineração possui uma participação relevante ao longo do processo de beneficiamento do minério. Em casos de acidentes, como o rompimento de uma barragem de rejeitos, esses problemas podem ser acentuados. Nesse sentido, este trabalho teve como objetivo avaliar a qualidade do ar antes e após o rompimento da Barragem I da Mina Córrego do Feijão, no município de Brumadinho (MG). Para isso, foi utilizada a metodologia de valores extremos aplicada às séries de máximos de partículas totais em suspensão (PTS) e material particulado (MP₁₀), ambas em $\mu\text{g}/\text{m}^3$. Para o ajuste de um modelo, foram utilizadas duas abordagens de valores extremos: Distribuição Generalizada de Valores Extremos (GEV) e Distribuição Generalizada de Pareto (GPD). Além disso, os parâmetros estimados para cada modelo foram obtidos pelo método da máxima verossimilhança. Uma vez obtidas as estimativas dos parâmetros, foram calculadas as probabilidades de excederem o limite legal, além do Valor em Risco (VaR), para verificar a qualidade do ar junto aos níveis padrões já estabelecidos. Os resultados obtidos mostraram altas probabilidades de exposição aos poluentes, considerando os padrões de qualidade do ar, em ambos os períodos. No entanto, a probabilidade de maiores concentrações de MP₁₀ diminuiu pós-rompimento da barragem. Para as concentrações de PTS, houve aumento na probabilidade de ocorrência de extremos.

Palavras-chave: Poluentes atmosféricos. Danos ambientais. Rejeitos de mineração. Valor em Risco. Distribuição Generalizada de Valores Extremos. Distribuição Generalizada de Pareto.

ABSTRACT

Currently there are studies that indicate a significant increase in air pollution around the world, which can cause serious environmental damage and human health. Among the numerous factors that contribute to this increase, the mining activity has a relevant partition throughout the ore beneficiation process. In cases of accidents, such as the rupture of a tailings dam, these problems can be accentuated. In this sense, this work aimed to evaluate the air quality before and after the rupture of Dam I of the Córrego do Feijão Mine, in the municipality of Brumadinho (MG). For this, the extreme values methodology was applied to the series of maximums of total suspended particles (TSP) and particulate matter (PM₁₀) is used, both in $\mu\text{g}/\text{m}^3$. For the adjustment of a model, two approaches of extreme values were used: Generalized Extreme Value Distribution (GEV) and Generalized Pareto Distribution (GPD). Furthermore, the parameters estimated for each model were obtained by the maximum likelihood method. Once the parameter estimates were obtained, the probabilities of exceeding the legal limit were calculated, in addition to the Value at Risk (VaR), to verify the air quality at the established standard levels. The results obtained showed high probabilities of exposure to pollutants, considering air quality standards, in both periods. However, the probability of higher PM₁₀ concentrations decreased after the dam failure. For TSP concentrations, there was an increase in the probability of extreme occurrences.

Keywords: Atmospheric pollutants. Environmental damages. Mining tailings. Value at Risk. Generalized Extreme Value Distribution. Generalized Pareto Distribution.

LISTA DE FIGURAS

Figura 2.1 – Número de minas/unidades produtoras em operação no Brasil.	16
Figura 2.2 – Método de alteamento a montante.	17
Figura 2.3 – Os efeitos dos poluentes atmosféricos na saúde humana.	19
Figura 2.4 – Amostrador de grande volume.	21
Figura 2.5 – Materiais particulados comparados com um fio de cabelo ($\sim 70 \mu\text{m}$) e um grão de areia ($\sim 90 \mu\text{m}$).	24
Figura 3.1 – Gráfico amplitude <i>versus</i> média.	35
Figura 3.2 – Função densidade de probabilidade da distribuição GEV para $\xi = -0,40$ (Weibull), $\xi = 0$ (Gumbel) e $\xi = 0,40$ (Fréchet), com $\mu = 2$ e $\sigma = 20$	42
Figura 3.3 – Função densidade de probabilidade da distribuição GP para $\xi = -0,40$ (Pareto comum ou Beta), $\xi \rightarrow 0$ (Exponencial) e $\xi = 0,40$ (Pareto tipo II), com $\sigma_u = 2$	48
Figura 3.4 – Gráfico de vida média residual.	50
Figura 4.1 – Localização e mapa do município de Brumadinho, MG.	62
Figura 5.1 – Gráficos de concentração de PTS, em $\mu\text{g}/\text{m}^3$, do município de Brumadinho antes do rompimento (a) e após o rompimento (b) da barragem.	66
Figura 5.2 – Gráficos de concentração diária de PTS, em $\mu\text{g}/\text{m}^3$, do município de Brumadinho antes (a) e após (b) o rompimento da barragem.	67
Figura 5.3 – Gráficos das séries imputadas de concentrações de PTS, em $\mu\text{g}/\text{m}^3$, antes (a) e após (b) o rompimento da barragem.	68
Figura 5.4 – Série completa (observações horárias) da concentração de PTS, em $\mu\text{g}/\text{m}^3$ (a); Função de autocorrelação para a v.a. concentração de PTS (b); Série de 175 máximos de PTS (blocos de 100 observações) (c); Função de autocorrelação para os máximos de PTS (d).	69
Figura 5.5 – Série de 93 máximos de PTS, em $\mu\text{g}/\text{m}^3$, antes do rompimento (blocos de 100 observações) (a); Função de autocorrelação para a v.a. 93 máximos de PTS (b); Série de 82 máximos de PTS, em $\mu\text{g}/\text{m}^3$, após o rompimento (blocos de 100 observações) (c); Função de autocorrelação para a v.a. 82 máximos de PTS (d).	70

Figura 5.6 – Probabilidades ((a) e (c)) e quantis ((b) e (d)) para a distribuição GEV ajustada as séries de PTS, em $\mu\text{g}/\text{m}^3$, antes e após o rompimento da barragem, respectivamente.	73
Figura 5.7 – Gráfico de vida média residual para os dados de PTS, em $\mu\text{g}/\text{m}^3$	74
Figura 5.8 – Estimação dos parâmetros escala (a) e forma (b) para 240 limiares diferentes para os dados horários de PTS.	75
Figura 5.9 – Série de 397 máximos de PTS, em $\mu\text{g}/\text{m}^3$, com limiar $u = 210$ (a); Função de autocorrelação para a v.a. 397 máximos de PTS (b).	75
Figura 5.10 – Série de 108 máximos de PTS, em $\mu\text{g}/\text{m}^3$, antes do rompimento (com limiar $u = 210$) (a); Função de autocorrelação para a v.a. 108 máximos de PTS (b); Série de 289 máximos de PTS, em $\mu\text{g}/\text{m}^3$, após o rompimento (com limiar $u = 210$) (c); Função de autocorrelação para a v.a. 289 máximos de PTS (d).	76
Figura 5.11 – Probabilidades ((a) e (c)) e quantis ((b) e (d)) para a distribuição GP ajustada as séries de PTS, em $\mu\text{g}/\text{m}^3$, antes e após o rompimento da barragem, respectivamente.	78
Figura 5.12 – Gráficos de concentração de MP_{10} , em $\mu\text{g}/\text{m}^3$, do município de Brumadinho antes do rompimento (a) e após o rompimento (b) da barragem.	80
Figura 5.13 – Gráficos de concentração diária de MP_{10} , em $\mu\text{g}/\text{m}^3$, do município de Brumadinho antes (a) e após (b) o rompimento da barragem.	81
Figura 5.14 – Gráficos das séries imputadas de concentrações de MP_{10} , em $\mu\text{g}/\text{m}^3$, antes (a) e após (b) a ruptura da barragem.	82
Figura 5.15 – Série completa (observações horárias) da concentração de MP_{10} , em $\mu\text{g}/\text{m}^3$ (a); Função de autocorrelação para a v.a. concentração de MP_{10} (b); Série de 208 máximos de MP_{10} (blocos de 84 observações) (c); Função de autocorrelação para os máximos de MP_{10} (d).	83
Figura 5.16 – Série de 111 máximos de MP_{10} , em $\mu\text{g}/\text{m}^3$, antes do rompimento (blocos de 84 observações) (a); Função de autocorrelação para a v.a. 111 máximos de MP_{10} (b); Série de 97 máximos de MP_{10} , em $\mu\text{g}/\text{m}^3$, após o rompimento (blocos de 84 observações) (c); Função de autocorrelação para a v.a. 97 máximos de MP_{10} (d).	84

Figura 5.17 – Probabilidades ((a) e (c)) e quantis ((b) e (d)) para a distribuição GEV ajustada as séries de MP_{10} , em $\mu g/m^3$, antes e após o rompimento da barragem, respectivamente.	87
Figura 5.18 – Gráfico de vida média residual para os dados de MP_{10} , em $\mu g/m^3$	88
Figura 5.19 – Estimação dos parâmetros escala (a) e forma (b) para 150 limiares diferentes para os dados horários de MP_{10}	89
Figura 5.20 – Série de 531 máximos de MP_{10} , em $\mu g/m^3$, com limiar $u = 90$ (a); Função de autocorrelação para a v.a. 531 máximos de MP_{10} (b).	89
Figura 5.21 – Série de 301 máximos de MP_{10} , em $\mu g/m^3$, antes do rompimento (com limiar $u = 80$) (a); Função de autocorrelação para a v.a. 301 máximos de MP_{10} (b); Série de 230 máximos de MP_{10} , em $\mu g/m^3$, após o rompimento (com limiar $u = 80$) (c); Função de autocorrelação para a v.a. 230 máximos de MP_{10} (d).	90
Figura 5.22 – Probabilidades ((a) e (c)) e quantis ((b) e (d)) para a distribuição GP ajustada as séries de MP_{10} , em $\mu g/m^3$, antes e após o rompimento da barragem, respectivamente.	92
Figura 1 – Padrões nacionais de qualidade do ar e diretrizes da Organização Mundial da Saúde (OMS).	104

LISTA DE TABELAS

Tabela 2.1 – Índices de qualidade do ar.	20
Tabela 5.1 – Estatísticas descritivas para a concentração de PTS ($\mu\text{g}/\text{m}^3$).	66
Tabela 5.2 – Quantidade acumulada em horas que a concentração de PTS ultrapassou o limite de $240 \mu\text{g}/\text{m}^3$, suas respectivas proporções e teste Z.	67
Tabela 5.3 – Resultados do teste de Cox-Stuart e teste de Fisher.	71
Tabela 5.4 – Estimativas dos parâmetros, seus respectivos erros padrões e $\text{IC}_{95\%}$ para a distribuição GEV e Gumbel ajustada a série de máximos de PTS, antes e após o rompimento da barragem.	71
Tabela 5.5 – Resultados do TRV.	72
Tabela 5.6 – Resultados do teste de Cox-Stuart e teste de Fisher.	76
Tabela 5.7 – Estimativas dos parâmetros, seus respectivos erros padrões e $\text{IC}_{95\%}$ para a distribuição GP e Exponencial ajustada a série de máximos de PTS, antes e após o rompimento da barragem.	77
Tabela 5.8 – Resultados do TRV.	77
Tabela 5.9 – Níveis de risco e valores ao risco para v.a. máximos de PTS, em $\mu\text{g}/\text{m}^3$	79
Tabela 5.10 – Estatísticas descritivas para a concentração de MP_{10} ($\mu\text{g}/\text{m}^3$).	80
Tabela 5.11 – Quantidade acumulada em horas que a concentração de MP_{10} ultrapassou o limite de $50 \mu\text{g}/\text{m}^3$, suas respectivas proporções e teste Z.	81
Tabela 5.12 – Resultados do teste de Cox-Stuart e teste de Fisher.	85
Tabela 5.13 – Estimativas dos parâmetros, seus respectivos erros padrões e $\text{IC}_{95\%}$ para a distribuição GEV e Gumbel ajustada a série de máximos de MP_{10} , antes e após o rompimento da barragem.	85
Tabela 5.14 – Resultados do TRV.	86
Tabela 5.15 – Resultados do teste de Cox-Stuart e teste de Fisher.	90
Tabela 5.16 – Estimativas dos parâmetros, seus respectivos erros padrões e $\text{IC}_{95\%}$ para a distribuição GP e Exponencial ajustada a série de máximos de MP_{10} , antes e após o rompimento da barragem.	91
Tabela 5.17 – Resultados do TRV.	91
Tabela 5.18 – Níveis de risco e valores ao risco para v.a. máximos de MP_{10} , em $\mu\text{g}/\text{m}^3$	93

SUMÁRIO

1	INTRODUÇÃO	14
2	REVISÃO BIBLIOGRÁFICA	16
2.1	A mineração no Brasil e seus efeitos	16
2.2	Poluentes e qualidade do ar	18
2.2.1	Partículas totais em suspensão (PTS)	21
2.2.2	Material particulado (MP₁₀)	22
2.2.3	Material particulado (MP_{2,5})	23
2.2.4	Ozônio (O₃)	25
2.2.5	Monóxido de carbono (CO)	26
2.2.6	Dióxido de nitrogênio (NO₂)	27
2.2.7	Dióxido de enxofre (SO₂)	27
2.3	Métodos estatísticos aplicados nos estudos de poluição do ar	28
3	REFERENCIAL TEÓRICO	31
3.1	Imputação de dados	31
3.2	Séries temporais	34
3.2.1	Estacionariedade	36
3.2.2	Teste de tendência	37
3.2.3	Teste de sazonalidade	38
3.2.4	Função de autocorrelação	38
3.3	Teoria de valores extremos (TVE)	39
3.3.1	Valores extremos pelo método de blocos	40
3.3.1.1	Distribuição generalizada de valores extremos (GEV)	41
3.3.1.2	Estimação dos parâmetros da distribuição GEV	43
3.3.2	Valores extremos pelo método <i>Peaks Over Threshold</i> (POT)	46
3.3.2.1	Distribuição generalizada de Pareto (GPD)	47
3.3.2.2	Escolha do limiar u	48
3.3.2.3	Estimação dos parâmetros da GPD	50
3.3.3	Qualidade do ajuste	52
3.3.4	Critérios de seleção	54
3.3.5	Probabilidade de ocorrência de valores extremos	55
3.3.6	Período de retorno	56

3.4	Valor em risco (VaR)	58
4	MATERIAL E MÉTODOS	61
4.1	Dados	61
4.2	Métodos	63
5	RESULTADOS	66
5.1	Série de partículas totais em suspensão (PTS)	66
5.1.1	Método de blocos	69
5.1.2	Método dos excessos acima de um limiar u	74
5.1.3	Cálculo do VaR	79
5.2	Série de material particulado (MP ₁₀)	80
5.2.1	Método de blocos	83
5.2.2	Método dos excessos acima de um limiar u	88
5.2.3	Cálculo do VaR	93
6	DISCUSSÃO	95
7	CONCLUSÕES	97
	REFERÊNCIAS	98
	APENDICE A – Padrões de qualidade do ar	104

1 INTRODUÇÃO

Nos últimos anos foram divulgadas várias ocorrências de eventos que acabaram resultando em um aumento expressivo da poluição, seja ela em uma determinada região ou cidade, ocasionando assim inúmeros danos ambientais, dentre eles a redução da qualidade do ar. Segundo Amorim, Conceição e Ferreira (2016), a extração de minérios é uma delas, que é agravada em casos de rompimento de barragens de rejeitos, provocando inúmeros danos ambientais e sociais.

Em Minas Gerais, de acordo com Lacaz, Porto e Pinheiro (2017), a lista de casos mais recentes teve início em 1986, quando a barragem do grupo Itaminas se rompeu, em Itabirito, matando sete pessoas. Em 2001, a barragem de Rio Verde, em Nova Lima, matou cinco trabalhadores. Em 2003, a barragem da Indústria Cataguases de Papel se rompeu e atingiu os rios Pomba e Paraíba do Sul, deixando 600 mil pessoas sem água. Em 2007, a barragem da mineradora Rio Pomba Cataguases se rompeu em Miraf, espalhando cerca de 2 milhões de m^3 de rejeitos, desalojando mais de 4 mil moradores e afetando quatro municípios. Já em 2014, em Itabirito, os rejeitos de minério de ferro atingiram diversos cursos d'água da região.

A penúltima dessa lista, foi a barragem do Fundão no município de Mariana, em 2015, que se rompeu e a lama destruiu vilarejos, percorreu 663 km ao longo dos rios Gualaxo do Norte, Carmo e Doce, deixando cerca de 1,2 milhões de pessoas sem água, além de afetar a vida de 35 municípios em Minas Gerais e quatro no Espírito Santo (ES). Por fim, em 2019 a barragem do Córrego do Feijão, em Brumadinho, onde cerca de 11,7 milhões de m^3 de rejeito percorreram 8,5 km até atingir o rio Paraopeba, estendendo-se por mais de 300 km ao longo desse, em direção ao rio São Francisco.

Além disso, também podem ser citados os danos à qualidade do ar, pois se sabe que são causados principalmente por dois fatores. O primeiro deles, ocorre devido a processos naturais, e que geralmente nada pode ser feito, como por exemplo, a emissão de gases por meio da erupção de um vulcão. Já o outro, é provocado por ação humana, e que em muitos casos pode ser evitado ou, pelo menos, minimizado. Alguns exemplos que podem ser citados incluem o aumento do uso de veículos automotores, incêndios causados em áreas florestais e os processos de eliminação de resíduos por fábricas, usinas e mineradoras.

O aumento de poluentes no ar afeta principalmente a qualidade de vida de todo e qualquer ser vivo, provocando vários problemas de saúde, à sociedade e ao meio ambiente. De acordo com Soares *et al.* (2018), os principais poluentes causadores de problemas de saúde são

aqueles denominados “material particulado” (MP_{10} e $MP_{2,5}$), o monóxido de carbono e o ozônio. Especialmente, quando se refere à liberação de particulados, estes materiais podem causar graves doenças respiratórias e cardiovasculares, pois conforme os autores anteriormente citados, as partículas inaláveis podem atingir com mais facilidade as vias respiratórias inferiores (traqueia, brônquios e alvéolos), dificultando trocas gasosas.

Dessa forma, este estudo será de grande relevância, pois a partir das informações obtidas dos níveis máximos de poluentes liberados antes e após o rompimento da barragem do Córrego do Feijão, será possível fazer um levantamento da qualidade do ar na região que foi afetada, realizando assim algumas comparações. Além disso, verificar se esses valores máximos de emissão de poluentes ultrapassaram os níveis padrões de qualidade do ar, que são regulamentados por órgãos governamentais especializados, como o Conselho Nacional do Meio Ambiente (CONAMA) e a Organização Mundial da Saúde (OMS).

Objetivo Geral

A pesquisa teve como objetivo geral, avaliar a qualidade do ar anteriormente e posteriormente ao rompimento da Barragem I da Mina Córrego do Feijão, na região de Brumadinho, MG. Para isso, foram utilizados os valores máximos de emissão de partículas totais em suspensão (PTS) e de material particulado (MP_{10}).

Objetivos específicos

Os objetivos específicos deste estudo foram:

- a) analisar a aplicabilidade da Teoria de Valores Extremos (TVE) em índices de poluição no município de Brumadinho;
- b) identificar qual a distribuição de eventos extremos tem melhor desempenho para descrever o comportamento de cada poluente;
- c) comparar as concentrações de PTS e MP_{10} , em $\mu\text{g}/\text{m}^3$, que foram liberadas no ar durante os períodos: antes e pós rompimento da barragem;
- d) calcular o *Value-at-Risk*, ou seja, Valor ao Risco (VaR) para cada um dos poluentes, após o ajuste de uma distribuição.

2 REVISÃO BIBLIOGRÁFICA

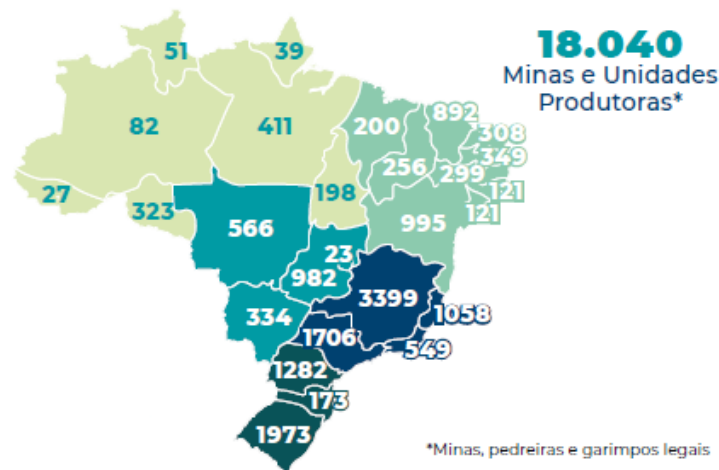
Nesta seção serão apresentadas algumas informações sobre a extração de minérios e seu impacto na sociedade. Além disso, quais são os principais poluentes liberados no ar devido a essa extração, suas possíveis consequências e as razões para o aumento da poluição atmosférica e, por fim, algumas pesquisas na área.

2.1 A mineração no Brasil e seus efeitos

A atividade socioeconômica de extração de minério no Brasil iniciou-se no século XVII, e a principal forma de extração ocorre por meio das minas a céu aberto, ocasionando assim impactos ao meio ambiente e à sociedade. Segundo Thomé e Passini (2018), a mineração encontra-se fortemente atrelada à história brasileira contemporânea, pois foi a responsável pela ocupação deste território e tem sido essencial para o crescimento econômico do país, seja no período colonial, através da produção do ouro e diamantes, ou pela extração e produção do ferro, no Império (de 1822 a 1889) e no início da República (de 1889 a 1930).

De acordo com dados do Instituto Brasileiro de Mineração – IBRAM (2016), o Estado de Minas Gerais, com 3399 minas em operação, é considerado o Estado com a maior quantidade de unidades em funcionamento e o maior produtor de minério de ferro, extraíndo mais de 180 milhões de toneladas/ano, sendo o responsável por mais da metade da produção nacional de minerais metálicos. A Figura 2.1 apresenta a quantidade de minas em funcionamento no Brasil, sendo em primeiro lugar o Estado de Minas Gerais e o Rio Grande do Sul o segundo maior, com 1973 unidades.

Figura 2.1 – Número de minas/unidades produtoras em operação no Brasil.



Fonte: Adaptado de IBRAM, 2020.

O processo de extração de minérios acaba gerando resíduos e com isso é preciso fazer o descarte em lugar apropriado. No Brasil, especialmente em Minas Gerais, grande parte desses rejeitos é armazenada em barragens. Quanto à quantidade delas, conforme a listagem do ano de 2020 da Fundação Estadual do Meio Ambiente – FEAM (2020) de Minas Gerais, o Estado possui 404 barragens classificadas como pertencentes à mineração, sendo aproximadamente 50 delas construídas pelo método de alteamento a montante (Figura 2.2), utilizado nas barragens de Mariana e Brumadinho.

Figura 2.2 – Método de alteamento a montante.



Fonte: Adaptado de IBRAM, 2016.

O método de alteamento à montante possui algumas vantagens, principalmente econômicas, como a redução do tempo e o custo empregado na construção, pois se utiliza de solos, estéreis ou mesmo o próprio rejeito. A esse respeito, Araujo (2006, p. 14) informa que:

Embora seja o mais utilizado pela maioria das mineradoras, o método de montante apresenta um baixo controle construtivo, tornando-se crítico principalmente em relação à segurança. O agravante neste caso está ligado ao fato de os alteamentos serem realizados sobre materiais previamente depositados e não consolidados. Assim, sob condição saturada e estado de compactação fofo, estes rejeitos (granulares) tendem a apresentar baixa resistência ao cisalhamento e susceptibilidade à liquefação por carregamentos dinâmicos e estáticos. Neste método construtivo ainda existe uma dificuldade na implantação de um sistema interno de drenagem eficiente para controlar o nível d'água dentro da barragem, constituindo um problema adicional, com reflexos na estabilidade da estrutura.

Assim, conforme Amorim, Conceição e Ferreira (2016), a atividade mineradora é um dos setores básicos da economia e de enorme contribuição para o desenvolvimento social. Porém, é uma atividade potencialmente degradadora do meio ambiente, por mais que se desenvolva dentro dos melhores padrões de controle ambiental, pois sempre haverá um impacto residual. Em situações, como no vazamento ou rompimento da barragem, os impactos são enormes e prejudicam vários setores da economia, pois diferente de um vazamento de água, os rejeitos acabam contaminando os solos e rios, dificultando a produção de culturas, pesca, transporte, entre outros.

Sendo assim, entre os mais variados problemas socioambientais provocados pela extração de minério, por parte das mineradoras, estão a poluição do ar e dos recursos hídricos, transformação das paisagens, destruição da flora e da fauna, ocupação desordenada do solo, consumo excessivo de energia elétrica, esgotamento e abandono a céu aberto de minas exploradas (AMORIM; CONCEIÇÃO; FERREIRA, 2016).

Quanto à poluição do ar, alguns estudos citam vários tipos de poluentes que são liberados na atmosfera devido à mineração, dentre os quais destacam-se os óxidos de carbono (CO e CO₂), os óxidos de nitrogênio (NO_x), os óxidos de enxofre (SO_x) e os particulados. Também é mencionado na literatura que o material particulado se destaca por estar relacionado a quase todas as atividades e etapas da mineração e, portanto, apresenta grande potencial poluidor.

Como o estudo tratará da poluição do ar, na seção seguinte serão apresentados com maiores detalhes quais poluentes são liberados no ar devido a exploração de minérios, quais as consequências na natureza e para o ser humano, além de informações a respeito da qualidade do ar.

2.2 Poluentes e qualidade do ar

De acordo com a Fundação Estadual do Meio Ambiente – FEAM (2016), a qualidade do ar sofre interferência direta da distribuição e intensidade das emissões de poluentes atmosféricos provenientes das mais variadas fontes. Os efeitos do aumento descontrolado de poluentes do ar podem se manifestar na saúde da população, na vegetação e na fauna, sobre os materiais que possuem metais em sua composição, sobre as propriedades da atmosfera passando pela redução da visibilidade e alteração da acidez das águas da chuva.

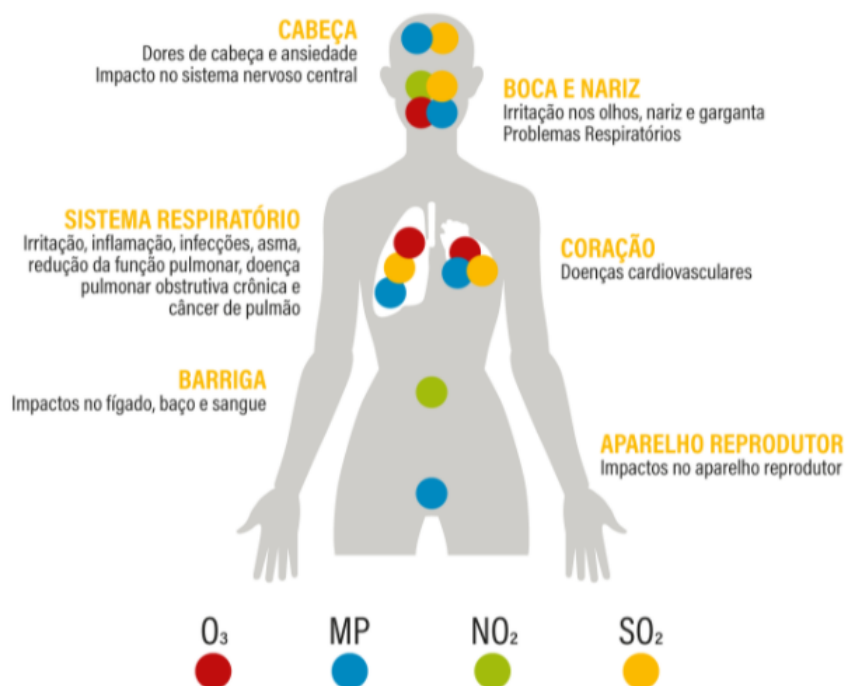
A Resolução CONAMA nº 491 de 19/11/2018 (BRASIL, 2018), define poluente atmosférico como "qualquer forma de matéria em quantidade, concentração, tempo ou outras características, que tornem ou possam tornar o ar impróprio ou nocivo à saúde, inconveniente ao bem-estar público, danoso aos materiais (monumentos, estruturas metálicas e condutores elétricos), à fauna e flora ou prejudicial à segurança, ao uso e gozo da propriedade ou às atividades normais da comunidade".

Segundo Silva *et al.* (2020), a qualidade do ar é dividida em dois tipos: os padrões primários e os secundários. Os considerados padrões primários de qualidade do ar, são as concentrações de poluentes que, ultrapassadas, poderão afetar a saúde da população. Já os padrões secundários são aqueles em que as concentrações de poluentes abaixo das quais se

prevê o mínimo efeito adverso sobre o bem-estar da população, assim como o mínimo dano à fauna, à flora, aos materiais e ao meio ambiente em geral.

Dessa forma, a exposição prolongada do indivíduo a esses poluentes afeta a saúde, principalmente aquelas ligadas a doenças respiratórias, decorrentes dos materiais particulados que são liberados através da mineração. Isso ocorre, pois segundo Amorim, Conceição e Ferreira (2016), o sistema de defesa natural do organismo humano, como o muco, pêlos nasais e macrófagos não conseguem reter partículas de tamanho médio em suspensão no ar, de forma a atingir os alvéolos. Uma vez nos pulmões, essas partículas se depositam ali, causando doenças comuns como por exemplo a tosse, dificuldade em respirar e dor no peito, ou doenças mais graves, como silicose, bronquite, asbestose, entre outras. Como pode ser visto na Figura 2.3, os efeitos desses materiais particulados vão além das doenças respiratórias ¹.

Figura 2.3 – Os efeitos dos poluentes atmosféricos na saúde humana.



Fonte: Adaptado de WRI BRASIL, 2018.

Neste sentido, é necessário que se tenha uma medida base que indique se certa concentração de poluente ultrapassou os padrões, podendo assim afetar a população e o meio ambiente. Para isso, foi criado o Índice de Qualidade do Ar (IQA_r), que tem como objetivo principal permitir que as pessoas entendam a qualidade do ar local. É uma ferramenta matemática usada para converter a concentração medida de diferentes poluentes em um único valor adimensional para

¹ Disponível em <https://wribrasil.org.br/pt/blog/2018/07/qual-o-impacto-da-poluicao-do-ar-na-saude>.

comparação com a concentração legal. Ainda segundo FEAM (2016), pode ser obtido através da equação 2.1:

$$IQAr = I_{inicial} + \left(\frac{I_{final} - I_{inicial}}{C_{final} - C_{inicial}} \right) \times (C - C_{inicial}) \quad (2.1)$$

em que,

$I_{inicial}$ é o valor do índice correspondente a concentração inicial;

I_{final} é o valor do índice correspondente a concentração final;

$C_{inicial}$ é concentração inicial da faixa onde encontra-se a concentração medida;

C_{final} é concentração final da faixa onde encontra-se a concentração medida;

C é a concentração medida.

Na Tabela 2.1 são apresentadas as quantidades dos principais poluentes, legislados pela Resolução CONAMA nº 491/2018, que podem prejudicar a saúde da população e o meio ambiente, e seus respectivos limites legais, obtidos através da equação (2.1), indicando assim a qualidade do ar desejada.

Tabela 2.1 – Índices de qualidade do ar.

Qualidade do ar	Índice	Níveis de episódios críticos de poluição do ar	PTS	MP ₁₀	MP _{2,5}	O ₃	CO	NO ₂	SO ₂
			($\mu\text{g}/\text{m}^3$) 24h	($\mu\text{g}/\text{m}^3$) 24h	($\mu\text{g}/\text{m}^3$) 24h	($\mu\text{g}/\text{m}^3$) 8h	(ppm) 8h	($\mu\text{g}/\text{m}^3$) 1h	($\mu\text{g}/\text{m}^3$) 24h
			$C_{inicial} - C_{final}$						
Boa	0 - 40		0 - 240	0 - 50	0 - 25	0 - 100	0 - 9	0 - 200	0 - 20
Moderada	41 - 80		>240 - 285	>50 - 100	>25 - 50	>100 - 130	>9 - 11	>200 - 240	>20 - 40
Ruim	81 - 120	Atenção	>285 - 330	>100 - 150	>50 - 75	>130 - 160	>11 - 13	>240 - 320	>40 - 365
Muito ruim	121 - 200	Alerta	>330 - 375	>150 - 250	>75 - 125	>160 - 200	>13 - 15	>320 - 1130	>365 - 800
Péssima	201 - 400	Emergência	>375 - 900	>250 - 600	>125 - 300	>200 - 800	>15 - 50	>1130 - 3750	>800 - 2620

Fonte: Adaptado de CETESB, 2019.

Alves e Freitas (2021) reiteram que embora a concentração de poluentes seja inferior aos padrões de qualidade do ar, existe uma estreita relação entre a poluição atmosférica e a mortalidade por doenças respiratórias, sendo as pessoas mais vulneráveis as que já possuem uma determinada doença do sistema respiratório, bem como crianças e idosos.

No entanto, vários estudos como os de Oliveira e Kummrow (2014) e Klumpp *et al.* (2001), relatam que os possíveis efeitos não dependem somente da concentração e da duração da exposição, mas também de outros fatores como o clima, o estado nutricional, a idade e a exposição à mistura complexa de poluentes.

Sendo assim, são apresentados nos tópicos seguintes os principais poluentes que são liberados através da extração mineral e que estão presentes na Resolução CONAMA nº 491/2018. Além disso, algumas informações sobre cada um deles, como por exemplo, outras principais

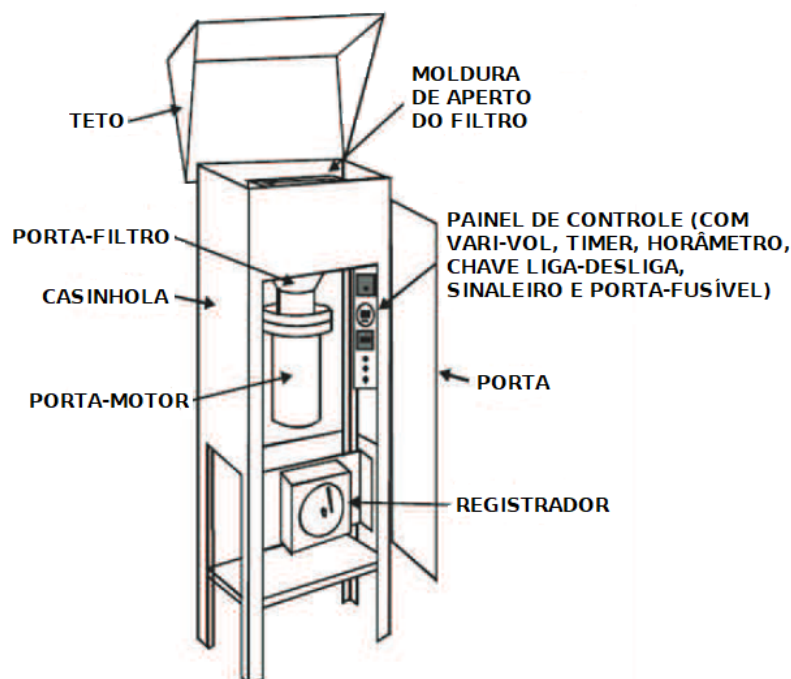
fontes de emissão, quais os principais efeitos à saúde e ao meio ambiente, entre outras informações que sejam relevantes.

2.2.1 Partículas totais em suspensão (PTS)

As partículas totais em suspensão são aquelas produzidas principalmente pela mistura de material sólido e líquido, que são liberados por processos industriais, veículos motorizados, poeira de rua ressuspensa e pela queima de biomassa. Essas partículas possuem tamanho menor que $50 \mu\text{m}$ (micrômetros) e ficam suspensas no ar na forma de poeira, neblina, aerossol, fumaça, fuligem, entre outras.

O equipamento utilizado para fazer a amostragem de PTS liberado na atmosfera é o Amostrador de Grande Volume (AGV), também denominado de AGV-PTS (Figura 2.4). O amostrador não tem controlador. Portanto, a taxa de fluxo muda durante o processo de amostragem e é registrada pelo registrador contínuo (SOARES *et al.*, 2015).

Figura 2.4 – Amostrador de grande volume.



Fonte: Adaptado de Soares *et al.*, 2015.

O processo de obtenção do número de partículas retidas no amostrador é o seguinte: um filtro limpo, corretamente identificado e pesado, é colocado no equipamento, juntamente com a carta gráfica (função de apenas registrar eventuais anormalidades). Em seguida, anota-se a leitura de tempo inicial do dispositivo. Após 24 horas, é removido cuidadosamente o filtro

sujo e a carta gráfica, além de anotar a leitura de tempo final. De posse da massa de material particulado retido no filtro e do volume total de ar, calcula-se a concentração de PTS a partir da expressão (2.2):

$$C = \frac{M_f - M_i}{V} \times 10^6, \quad (2.2)$$

em que:

C é a concentração de PTS (em $\mu\text{g}/\text{m}^3$);

M_f é a massa final do filtro (g);

M_i é a massa inicial do filtro (g);

V é o volume total de ar (m^3);

10^6 é o fator de conversão de g para μg .

A concentração de PTS, em $\mu\text{g}/\text{m}^3$, obtida através da equação (2.2) deve ser comparada com os padrões de qualidade do ar. Sendo que parte destas partículas, conforme Soares *et al.* (2015), são inaláveis e podem causar problemas à saúde. Outra parte pode afetar negativamente a qualidade de vida da população, ficando presas no sistema respiratório superior (nariz, faringe, laringe e traqueia), prejudicando as atividades normais da comunidade.

Já quando inaladas, especialmente em grandes concentrações, essas partículas causam efeitos significativos em pessoas com doenças pulmonares, como asma e bronquite. Além disso, no meio ambiente podem ocasionar danos à vegetação, redução da visibilidade, contaminação do solo, além de causar danos estéticos, como manchas e danificações de rochas e outros materiais (OLIVEIRA; KUMMROW, 2014).

Os particulados são classificados conforme o seu tamanho e geralmente representam diferentes fontes e características. Como visto, o PTS é um dos presentes nesta classificação, além dos materiais particulados MP_{10} e $\text{MP}_{2,5}$, que são apresentados nos tópicos seguintes.

2.2.2 Material particulado (MP_{10})

O MP_{10} é também chamado de partículas inaláveis grossas, devido ao seu diâmetro estar entre 2,5 e 10 μm . Consiste de uma mistura de partículas sólidas e líquidas suspensas no ar, na forma de poeira, neblina, aerossol, fumaça, fuligem, etc. O método utilizado para medir a concentração de material particulado na atmosfera é a radiação beta. O carbono 14 (C_{14}) é utilizado como fonte de radiação de baixa energia por meio de um monitor e a concentração é medida a cada hora (FEAM, 2016).

De acordo com Alves e Freitas (2021), este é um dos principais poluentes emitidos a partir das atividades de mineração, devido a desagregação mecânica do solo pelo uso de explosivos e pela escavação, além da movimentação de máquinas e equipamentos, o tráfego de veículos pesados na frente de lavras, pela ação do vento nas pilhas de estéril e nos taludes descobertos das frentes de lavra, nas plantas de beneficiamento de minério, entre outros.

Entre as formas de emissão do poluente, estão os processos de combustão das demais indústrias e veículos automotores, na construção e demolição, pulverização de praguicidas, bioaerossóis (como o pólen) e cinzas vulcânicas (OLIVEIRA; KUMMROW, 2014).

Uma das principais características do material particulado é a acidez, causada pela presença de sulfatos e nitratos. Conforme Alves e Freitas (2021), essa característica é extremamente prejudicial à saúde humana, pois assim que esse material particulado altamente ácido atingir o trato respiratório, as trocas gasosas entre o sangue e o ar absorvido na respiração serão prejudicadas.

Logo, as partículas inaláveis (MP_{10}) estão associadas ao aumento de atendimentos hospitalares e mortes prematuras. Além disso, contribuem para o aumento da incidência de doenças respiratórias, como a bronquite e a asma, principalmente em crianças e idosos.

De acordo com o estudo de Gouveia *et al.* (2006), no município de São Paulo, o acréscimo de $10 \mu\text{g}/\text{m}^3$ na concentração de material particulado inalável está associado, em média, a um aumento de 4,6% no número de crianças hospitalizadas com asma, 4,3% para doença pulmonar obstrutiva crônica em idosos, bem como para doenças do sistema circulatório e doenças isquêmicas do coração que, respectivamente, são de 1,0% e 1,5%, também em idosos. Já quanto ao meio ambiente, estão os danos à vegetação, à redução da visibilidade e à contaminação do solo.

Mesquita (2005) reitera que dentre os danos ambientais provocados pelos materiais particulados, os danos ao clima são os mais preocupantes, pois afetam os animais, os rios, as florestas, o mecanismo do ciclo d'água, o aumento do efeito estufa e do aquecimento global, além de alterar drasticamente todo o meio ambiente por meio de fortes chuvas e enchentes.

2.2.3 Material particulado ($MP_{2,5}$)

O $MP_{2,5}$ é também denominado de partículas inaláveis finas, devido ao seu diâmetro aerodinâmico ser inferior ou igual a $2,5 \mu\text{m}$. A sua composição é a mesma que apresentada anteriormente, para o caso do MP_{10} , porém são menores e podem permanecer no ar e percor-

rer longas distâncias. A Figura 2.5 demonstra essa diferença de tamanho entre os materiais particulados, comparando com um fio de cabelo e um grão de areia².

Figura 2.5 – Materiais particulados comparados com um fio de cabelo ($\sim 70 \mu\text{m}$) e um grão de areia ($\sim 90 \mu\text{m}$).



Fonte: Adaptado de EPA, 2021.

Estudos indicam que quanto menores são essas partículas, maiores os prejuízos à saúde. Oliveira e Kummrow (2014), declaram que as partículas menores que $10 \mu\text{m}$ ficam retidas nas vias aéreas superiores e podem ser depositadas na árvore traqueobrônquica, que após a exposição são eliminadas entre 1 e 20 dias. Enquanto as menores que $2,5 \mu\text{m}$ se depositam na porção terminal dos brônquios e nos alvéolos, sendo eliminadas muito mais lentamente.

Segundo Costa (2012), praticamente qualquer metal pode ser encontrado no material particulado e muitos deles possuem grande potencial tóxico ou pró-oxidante. No caso do MP_{2,5} são metais derivados de fontes de combustão como o petróleo e carvão que em contato com o meio ambiente causa danos à vegetação, deterioração da visibilidade e contaminação do solo e da água.

Porém, é possível reduzir a emissão de determinado material particulado e, mesmo que muito lentamente, trará enorme benefício à saúde humana e ao meio ambiente a longo prazo. Alguns exemplos, como o uso de combustíveis mais limpos como o álcool, melhores técnicas de combustão, ou ainda, através do uso de separadores mecânicos, por gravidade ou centrifugação (câmaras de poeira), precipitadores eletrostáticos, lavadores na saída das chaminês e filtros de tela poderão minimizar os efeitos causados por estas partículas (OLIVEIRA; KUMMROW, 2014).

² Disponível em <https://www.epa.gov/pm-pollution>.

2.2.4 Ozônio (O₃)

O ozônio é um gás incolor, inodoro nas concentrações ambientais e o principal componente da névoa fotoquímica mais conhecido como *smog*. De acordo com Oliveira e Kummrow (2014), o termo *smog* surgiu da associação das palavras inglesas *smoke* (fumaça) + *fog* (néblina), sendo que existem dois tipos, o redutor e o oxidante. O *smog* redutor é rico em óxidos de enxofre (SO_x) e fuligem (MP) derivados principalmente da queima de carvão. Já o *smog* oxidante é rico em óxidos de nitrogênio (NO_x), aldeídos, ozônio e nitratos de peroxiacila (PAN), sendo mais propenso à formação em cidades com tráfego pesado, clima seco e ensolarado.

O O₃ é considerado um composto muito ativo quimicamente, pois é formado a partir de reações químicas na atmosfera, originadas pela presença de dióxido de nitrogênio (NO₂) e radiação proveniente do sol. Além do dióxido de nitrogênio, os hidrocarbonetos (HC), também conhecidos como compostos orgânicos voláteis (COV), são importantes precursores do ozônio (MONTE; ALBUQUERQUE; REISEN, 2017).

Um dos principais métodos empregados para medir a concentração de O₃ no ar atmosférico é o fotométrico, por meio do analisador modelo O₃41M. O equipamento utiliza a tecnologia de fotometria de radiação ultravioleta (UV) e mede a quantidade de ozônio no ar a partir de emissões de 0,5 ppb até 1000 ppb. O detector consiste em um fotômetro de célula dupla e o tempo de resposta do dispositivo é de 20 segundos (SILVA *et al.*, 2014).

Segundo Oliveira e Kummrow (2014), os principais sintomas que o O₃ apresenta no ser humano são a irritação nos olhos e vias respiratórias, diminuição da capacidade pulmonar e alergias. Em casos de exposição a altas concentrações, pode resultar em sensações de aperto no peito, tosse e chiado na respiração. Além de estar associado ao expressivo aumento de admissões hospitalares.

De acordo com a Companhia Ambiental do Estado de São Paulo – CETESB (2019), em relação aos impactos ambientais, o ozônio causa danos à vegetação natural, as plantações agrícolas e as plantas ornamentais, pois age como inibidor da fotossíntese, além de produzir lesões nas folhas das plantas. Também pode danificar materiais com presença de metais, como o ferro, chumbo e arsênico, devido ao seu alto poder oxidante.

Liu *et al.* (2013) consideram que os dois poluentes do ar que mais preocupam em relação à saúde humana são o ozônio, pois os males causados por ele ocorrem na faixa de ar perto da superfície terrestre, onde o gás é muito tóxico, e os materiais particulados, que causam diversas enfermidades.

Nas próximas seções são apresentados os demais poluentes, como o monóxido de carbono, o dióxido de nitrogênio e o dióxido de enxofre, que além de serem prejudiciais a saúde humana e ao meio ambiente, contribuem para o aumento e aparição dos materiais particulados.

2.2.5 Monóxido de carbono (CO)

O monóxido de carbono é um gás incolor, inodoro, insípido e extremamente tóxico. Ele é muito utilizado como agente redutor no processamento de minérios e na obtenção do hidrogênio a partir da água, ou ainda como matéria-prima para a produção de ácido acético, plásticos, metanol e formiatos (CETESB, 2020b).

A principal fonte de emissão do monóxido de carbono é devido à combustão incompleta de combustíveis fósseis, como o carvão mineral, o petróleo e pela decomposição dos resíduos sólidos urbanos (alimentos, vidro, metal, plásticos, embalagens, podas de árvores e rejeitos em geral). Além desses, há outras fontes como em veículos automotores, sistemas de aquecimento, usinas termelétricas a carvão, queima de biomassa e tabaco. Por outro lado, Martins Júnior (2010) declara que metade das emissões de CO é liberado através do desflorestamento e, portanto, a extração mineral contribui em parte, pois para a exploração de minérios é necessário a derrubada de uma determinada área florestal.

O método de infravermelho não dispersivo (NDIR) é utilizado para medir a concentração de CO liberado no ar. Ele é caracterizado pela alta precisão, tempo de resposta relativamente curto, taxa de consumo de amostra moderada, baixa sensibilidade cruzada e fácil manuseio. O analisador é usado como um monitor contínuo para detectar a absorção de CO na faixa de luz infravermelha (FEAM, 2016).

Uma vez inalado, de moderada a alta quantidade, o gás será rapidamente absorvido pelos pulmões e se combinará de forma estável com a hemoglobina durante sua circulação, impedindo o transporte de oxigênio e causando hipóxia tecidual. Por isso, entre os efeitos estão a perda de consciência e visão, prejuízos no aprendizado e na capacidade de trabalho e aumento da mortalidade por infarto cardíaco agudo entre idosos (OLIVEIRA; KUMMROW, 2014). Se inalado em quantidades menores, este poluente pode provocar dores de cabeça e tonturas. No entanto, altas concentrações desse gás podem ser fatais.

2.2.6 Dióxido de nitrogênio (NO₂)

O dióxido de nitrogênio é um gás marrom avermelhado, com odor forte e muito irritante. Pode levar à formação de ácido nítrico (HNO₃), que é um dos componentes da chuva ácida, responsável por danos à vegetação e às colheitas. Além disso, a produção de nitratos, contribui para o aumento das partículas inaláveis na atmosfera (OLIVEIRA; KUMMROW, 2014).

Esse gás é liberado através de processos de combustão envolvendo veículos automotores, indústrias, usinas termelétricas (óleo, gás, carvão) e incineração. Um dos métodos de medição do NO₂ é por quimiluminescência, ou seja, através da emissão de luz originada pela oxidação de NO em presença de ozônio. Assim, a concentração é obtida calculando a diferença entre as concentrações de NO_x e NO (FEAM, 2016).

Os principais efeitos tóxicos do NO₂, conforme Oliveira e Kummrow (2014), estão relacionados ao seu poder oxidante e são: irritação das vias aéreas, enfisema pulmonar a longo prazo, aumento da sensibilidade à asma e bronquite e edema pulmonar em intoxicações agudas. Além disso, os efeitos são mais acentuados quando expostos às crianças, prejudicando a função pulmonar e, em casos de gestantes, complicações na gravidez, ou até mesmo provocando aborto.

2.2.7 Dióxido de enxofre (SO₂)

O dióxido de enxofre (SO₂) é um gás incolor, com forte odor e altamente solúvel. Quando em contato com superfícies úmidas, como por exemplo o vapor d'água, ele pode se transformar em trióxido de enxofre (SO₃) e passando rapidamente a ácido sulfúrico (H₂SO₄), sendo um dos principais constituintes da chuva ácida (CETESB, 2020a). Portanto, o contato dela em certos materiais, como monumentos, estruturas metálicas e condutores elétricos, causa a corrosão e também danos à vegetação e agricultura.

Além disso, é liberado na atmosfera a partir de atividades industriais que processam materiais contendo enxofre, como termelétricas, fabricação de fertilizantes, fundição de alumínio e aço, produção de ácido sulfúrico e papel. O poluente está, sempre também, presente na combustão de combustíveis fósseis (carvão), queima de óleo combustível, refinaria de petróleo, veículos a diesel e gases vulcânicos (OLIVEIRA; KUMMROW, 2014).

Para medir a concentração de dióxido de enxofre liberado no ar atmosférico, é utilizado o método da fluorescência por radiação ultravioleta (UV). O método baseia-se na excitação da molécula de SO₂ por UV, onde o analisador funciona em regime contínuo, medindo assim a concentração de forma praticamente instantânea (FEAM, 2016).

Esse poluente é um importante precursor dos sulfatos, sendo um dos principais componentes das partículas inaláveis (MP₁₀). Assim, uma vez inalado, em especial no caso de grandes concentrações, provoca desconforto na respiração, alteração na defesa dos pulmões, agravamento de doenças respiratórias e cardiovasculares, e também irritação no nariz, garganta e pulmões causando tosse, falta de ar, chiado no peito, catarro e crises de asma (CETESB, 2020a).

2.3 Métodos estatísticos aplicados nos estudos de poluição do ar

Cada um dos poluentes que foram aqui apresentados tem sua parcela de contribuição, quanto a qualidade de vida humana e ambiental. Sendo que quanto maior a presença deles no ar atmosférico, maiores serão os danos. Além disso, muito desses poluidores possuem grande relação entre si para que se desenvolvam ou até mesmo se intensifiquem no ar.

No caso da mineração, com o rompimento de uma barragem de rejeitos pode haver efeitos a curto e longo prazos, se estendendo por centenas de quilômetros do local de origem. Na pesquisa de Vergilio *et al.* (2021), eles demonstraram o impacto do rompimento da barragem de rejeitos do Fundão no período de 15 dias e seis meses após a ruptura, onde o material particulado em suspensão, composto principalmente por metais, é transportado pelo vento até a foz do rio Doce e o oceano. Após seis meses, com a deposição, a concentração de metal na água diminuiu e os níveis no sedimento aumentaram.

Além do trabalho citado anteriormente, existem muitos outros relacionados à poluição do ar, com objetivos diversos. Dentre eles, o estudo de Almeida *et al.* (2019) cujo objetivo era comparar a qualidade do ar de duas localidades da Região Metropolitana do Rio de Janeiro, sendo uma impactada pelas emissões veiculares (Taquara) e a outra pelas emissões industriais (Jardim Primavera). Para isso, foram utilizados dados de óxidos de nitrogênio (NO_x, NO₂ e NO), MP₁₀, ozônio, monóxido de carbono e dióxido de enxofre. Ao final, concluíram que as estações de monitoramento alocadas em ambientes urbanos e industriais apresentaram resultados distintos. Com destaque no aumento da concentração de NO₂ em Taquara. Já em Jardim Primavera houve a diminuição gradativa das concentrações de ozônio.

O trabalho de Liska e Beijo (2012) teve como objetivo analisar os níveis máximos de MP₁₀ e O₃ nas cidades de Cubatão e Paulínia (SP), bem como calcular as probabilidades deles excederem os limites legais. Para tanto, foi feito o ajuste das distribuições Generalizada de Valores Extremos (GVE) e Gumbel e utilizado o teste Kolmogorov-Smirnov, que indicaram

uma boa aderência das distribuições às séries de dados. Os resultados obtidos, mostraram que as maiores probabilidades dos poluentes MP_{10} e O_3 ultrapassarem os níveis críticos ocorrem no inverno e verão, respectivamente.

Em contrapartida, Martins Júnior (2010) analisou índices elevados de CO e MP_{10} na cidade de São Paulo. Dessa maneira, foi utilizada a metodologia de valores extremos (Método dos Blocos) aplicada a séries temporais estacionárias e independentes. Após o ajuste, foi calculado o Valor ao Risco (VaR), para cada poluente. Como resultado, o autor obteve que o risco de uma pessoa ficar exposta ao monóxido de carbono é de aproximadamente 64% e ao MP_{10} é de 58%, considerando-se os limites legais.

No trabalho de Soares *et al.* (2018), o objetivo foi avaliar o impacto na qualidade do ar após implantação do *Bus Rapid Service* (BRS) em Copacabana, RJ. Para tal, foram utilizadas observações de monitoramento do MP_{10} , SO_2 e CO de 2010 a 2016. A metodologia utilizada para as comparações foi o teste *t-Student* com nível de confiança de 95%, além de algumas técnicas de análise multivariada aplicadas às variáveis meteorológicas, para verificar se houve influência de alguma delas. De maneira geral, as emissões dos três poluentes foram reduzidas, indicando que esse novo sistema de transporte trouxe benefícios para a qualidade do ar na região.

Além disso, vários métodos de séries temporais são utilizados, seja para descrever comportamentos, efetuar previsões ou comparações. O trabalho de Pinto, Reisen e Monte (2018) objetivou modelar e prever a concentração média diária de material particulado inalável, na Região da Grande Vitória (ES), aplicando modelos da classe ARIMA. Para isso, utilizaram dados de MP_{10} e de variáveis meteorológicas (velocidade do vento, umidade relativa, precipitação pluvial e temperatura). Em comparação com os modelos ajustados, o modelo SARIMAX com as variáveis velocidade do vento e precipitação pluvial teve melhor desempenho quanto a predição.

Por outro lado, Costa e Sáfadi (2010) analisaram se as séries de máximos de partículas inaláveis grossas observadas nas estações Centro e Santo Amaro, em São Paulo, foram geradas pelo mesmo processo aleatório. Para a análise, utilizaram-se do teste das somas acumuladas, do teste de igualdade das funções de autocorrelação e do procedimento de diferença de séries. As autoras concluíram por meio dos três métodos, que as séries não foram geradas pelo mesmo processo, e portanto, uma série não pode ser usada para explicar o comportamento da outra.

Ao se utilizar dados de poluição, geralmente existem valores ausentes, que ocorrem devido a falhas no equipamento. Para isso, Razak, Zubairi e Yunus (2014) utilizaram três métodos de imputação, em que o algoritmo *expectation-maximization* (EM) foi considerado o melhor, pois apresentou menor valor da raiz do erro quadrático médio (RMSE). Em seguida, o método EM foi aplicado ao conjunto de dados de concentração de MP_{10} para as monções do sudoeste e nordeste em Petaling Jaya e Seberang Perai, Malásia. As distribuições utilizadas no ajuste foram a Weibull, Lognormal, Gama e Gumbel para descrever o comportamento das séries. A distribuição Weibull forneceu o melhor ajuste para os dados de monções do sudoeste de Petaling Jaya e a Lognormal para Seberang Perai. Quanto às monções do nordeste de ambas as áreas, a distribuição Gama é a que melhor descreve a série de dados.

Já a proposta de Hadeed *et al.* (2020) era apenas determinar quais entre os métodos de imputação existentes teve o melhor desempenho no monitoramento de curto prazo. Para este fim, foi utilizado medidas de $MP_{2,5}$ em intervalos de 1 minuto ao longo de 24 horas de 20 domicílios de uma comunidade rural do norte do Arizona. Em resumo, concluíram que os métodos Markov, aleatório e média foram os mais indicados para concentrações médias de 1 dia e, no caso de substituição minuto a minuto, o método Markov foi a melhor abordagem.

3 REFERENCIAL TEÓRICO

Nos tópicos a seguir, serão apresentados alguns métodos de imputação para o preenchimento de valores ausentes, caso que ocorre constantemente em série de dados de poluentes. Após, são apresentados os principais conceitos da teoria de séries temporais e alguns testes, com o objetivo de verificar algumas pressuposições. Por fim, a teoria de valores extremos e suas principais distribuições, além de algumas medidas da qualidade do ajuste, critérios para selecionar o melhor modelo e como obter o Valor em Risco (VaR).

3.1 Imputação de dados

Em diversas áreas do conhecimento o pesquisador se depara com valores ausentes em um banco de dados, podendo ser encontrados em uma ou mais variáveis de interesse. Essa ausência de registro pode ser vista em dados meteorológicos, epidemiológicos, sociais, de poluentes, entre outros. São vários os fatores que podem influenciar na ausência ou no erro de dados. Dentre eles podem ser citados: a ausência do observador, as falhas nos sensores, a perda das anotações ou das transcrições dos registros pelos operadores, o indivíduo ou unidade (domicílio, empresa, escola, etc.) que não responde algum item da pesquisa, além de falhas nos equipamentos ou nas transmissões dos dados (TEODORO, 2019).

A substituição ou o preenchimento de valores ausentes é conhecida como imputação de dados, que consiste em estimar tais valores a partir de informações disponíveis no conjunto de dados. Para isso, existem inúmeros métodos de imputação e determinar a abordagem mais adequada para um conjunto de dados com observações incompletas é uma tarefa delicada, pois o uso inadequado pode levar a conclusões erradas.

Segundo Harrell Jr. (2015), a escolha entre os métodos de imputação deve ser feita de acordo com a proporção de dados faltantes presente na variável de interesse, obtida através da divisão entre o número de ausentes e o tamanho da amostra, como segue:

- a) se proporção $\leq 0,05$, então pode ser usada a imputação única ou fazer a análise somente com os dados completos;
- b) se $0,05 < \text{proporção} < 0,15$, a imputação única pode ser utilizada sem problemas, entretanto o uso da imputação múltipla é recomendado;
- c) se proporção $\geq 0,15$, então a imputação múltipla é indicado na maior parte dos modelos.

Conforme Rubin (1987), na imputação única o processo é feito somente uma vez, como por exemplo, onde se obtém a média ou a mediana dos dados presentes na série e, em seguida aplica-se a medida em cada uma das lacunas. Já a imputação múltipla é uma técnica que visa estimar dados faltantes por meio de processos iterativos, onde é gerado m bancos de dados distintos e completos.

Além disso, a escolha do método a ser utilizado também deve obedecer o mecanismo de geração desses dados faltantes. Rubin (1987) define três mecanismos teóricos gerais utilizados na literatura. São eles:

- a) perdas completamente ao acaso (*Missing Completely at Random* - MCAR);
- b) perdas ao acaso (*Missing at Random* - MAR);
- c) perdas não-aleatórias (*Not Missing at Random* - NMAR).

Para o caso MCAR, os dados faltantes não possuem relação com qualquer outra variável medida, nem com a própria variável em falta, ou seja, aconteceu apenas devido a fatores inesperados. Por exemplo, medições laboratoriais perdidas devido à queda de um tubo de ensaio, ou em uma pesquisa onde o entrevistado omitiu sua resposta a uma pergunta por motivos não relacionados à resposta que eventualmente teria dado ou a qualquer outra de suas características (RUBIN, 1987; HARRELL JR., 2015).

Na MAR os dados faltantes dependem apenas das informações observadas, disponíveis para análise e correlacionadas com a variável que possui valores ausentes. Como exemplo, considera-se uma pesquisa na qual as mulheres têm menos probabilidade de fornecer o seu peso, mas a probabilidade de responder é independente de seu peso atual, mas considerando o sexo (masculino ou feminino) (RUBIN, 1987; HARRELL JR., 2015).

Por último, em que os dados são NMAR, a probabilidade de dados faltantes depende também das informações não registradas, ou seja, a probabilidade de se ter dados faltantes varia por razões desconhecidas. Por exemplo, suponha que em uma pesquisa, algumas pessoas com sobrepeso não relataram o seu peso. Portanto, essa falta de resposta ao peso depende de seu próprio peso (RUBIN, 1987; HARRELL JR., 2015).

Conforme Carvalho *et al.* (2017), o método de imputação múltipla, em áreas como a climatologia e a ambiental, tem levado a melhores resultados comparado as técnicas de krigagem e cokrigagem ordinária, apresentando um erro quadrático médio menor e estimativas melhores, pelo fato de estimar os dados faltantes por meio de processos iterativos. Sendo assim, Rubin

(1987) classifica o processo da imputação múltipla em três etapas principais: a imputação, a análise, e o agrupamento, descritas a seguir:

- a) A análise inicia-se com dados observados e dados incompletos. A partir disso, são obtidos m (em geral, m entre 3 e 10) bancos de dados completos por meio de técnicas adequadas de imputação, substituindo os valores ausentes por valores plausíveis;
- b) No segundo passo, o objetivo é analisar separadamente os m bancos através de um método estatístico tradicional, ou seja, estimar os parâmetros de interesse, como: média, correlação, variância, entre outros, de cada banco de dados que foi gerado na etapa anterior;
- c) No último passo os m resultados obtidos nas etapas anteriores são agrupados em uma estimativa pontual final acrescida do desvio padrão, por meio de regras de agrupamento simples, conhecida simplesmente como “Regras de Rubin”.

As Regras de Rubin podem ser usadas independente da escolha do método da imputação múltipla e a idéia descrita por Rubin (1987) é apresentada a seguir.

A estimativa combinada, ou seja, a média geral do parâmetro de interesse (\bar{Q}) será dada pela média das estimativas individuais, expressa por:

$$\bar{Q} = \frac{1}{m} \sum_{i=1}^m \hat{Q}_i, \quad (3.1)$$

em que \hat{Q}_i é a estimativa do i -ésimo parâmetro considerado. Para o erro padrão são duas as fontes de variação: a variação dentro das imputações e a variação entre imputações.

A primeira é obtida a partir da média das m variâncias do conjunto de dados imputados por meio da expressão (3.2):

$$\bar{U} = \frac{1}{m} \sum_{i=1}^m \hat{U}_i, \quad (3.2)$$

sendo \hat{U}_i a variância do m -ésimo conjunto de dados imputados.

A outra variação é calculada através da equação (3.3), dada por:

$$B = \frac{1}{m-1} \sum_{i=1}^m (\hat{Q}_i - \bar{Q})^2. \quad (3.3)$$

Logo, a variância combinada total é representada da seguinte forma:

$$T = \bar{U} + \left(1 + \frac{1}{m}\right) B, \quad (3.4)$$

onde $(1 + \frac{1}{m})$ é a correção referente ao número de imputações.

Posteriormente, pode-se realizar testes de hipóteses e construir intervalos de confiança para a média (\bar{Q}) por meio de uma distribuição *t-Student*, ou seja,

$$\frac{(\bar{Q} - Q)}{SE} \sim t_{v_m}, \quad (3.5)$$

com $v_m = (m - 1) \left[\frac{1 + \bar{U}}{(1 + m)^{-1} B} \right]^2$ graus de liberdade, sendo SE o valor do desvio padrão combinado e Q o valor da média real da variável em estudo (RUBIN, 1987).

Dentre os métodos de imputação múltipla que podem ser encontrados na literatura (RUIZ-CÁRDENAS; KRAINSKI, 2011); (BUUREN, 2018); (MORRIS; WHITE; ROYSTON, 2014), os mais utilizados são: o método de médias preditivas (*Predictive Mean Matching - PMM*), o método da regressão linear Bayesiana, a regressão logística (dados binários ou qualitativos), as florestas aleatórias (Random Forests), os modelos Markovianos, entre outros.

Quanto às métricas utilizadas para avaliar a eficiência dos métodos, estão: a raiz do erro quadrático médio (REQM), o coeficiente de correlação de Pearson e o índice de concordância ou exatidão de Willmott. Para mais informações sobre as mesmas, podem ser vistas nos trabalhos de Willmott *et al.* (1985) e em Dancey e Reidy (2017).

3.2 Séries temporais

Uma série temporal, também considerada uma série histórica ou então cronológica, segundo Morettin e Toloi (2006), é definida como qualquer conjunto de observações coletadas sequencialmente ao longo do tempo. Além disso, a sua principal característica é que as observações vizinhas são dependentes.

Quanto ao tipo de série, podem ser discretas ou contínuas. No caso discreto, as observações são feitas em tempos específicos, ou seja, igualmente espaçados. Um exemplo que pode ser expresso é o número de exportações de café mensalmente. A série é contínua quando as observações são obtidas continuamente em um determinado intervalo de tempo, por exemplo, o registro das marés durante um dia na praia de Copacabana, se a unidade de tempo for a hora, considera-se $T = [0, 24]$.

A aplicação de séries temporais ocorre nas mais diferentes áreas, como na economia (preços diários de ações, taxa de emprego e desemprego), na medicina (níveis de eletrocardiograma), na epidemiologia (casos mensais de sarampo, casos semanais de febre amarela), na meteorologia (temperatura e precipitação diária), entre outras.

Um modelo clássico de decomposição de uma série temporal Z_t , com $t = 1, 2, \dots, N$, pode ser expressa como a soma, ou produto de três componentes, conforme equações (3.6) e (3.7), respectivamente:

$$Z_t = T_t + S_t + a_t \quad (3.6)$$

ou

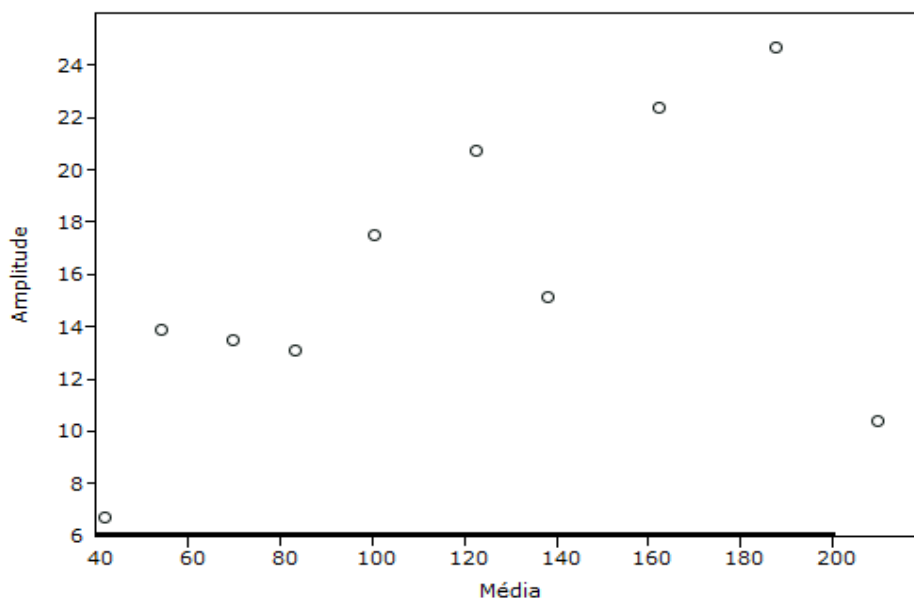
$$Z_t = T_t \cdot S_t \cdot a_t, \quad (3.7)$$

em que, T_t é a componente tendência; S_t é a componente sazonal; a_t é a componente aleatória com média zero e variância constante σ_a^2 .

A equação (3.6) é chamada de modelo aditivo e sua utilização é pertinente quando a componente sazonal (S_t) varia independentemente das demais componentes. Entretanto, se tal componente for dependente, o mais indicado é o modelo multiplicativo (equação (3.7)).

Em alguns casos é necessário realizar a transformação de um modelo multiplicativo em aditivo como, por exemplo, quando há uma grande variação nos dados, e para isso aplica-se o logaritmo na expressão (3.7) com o objetivo de reduzir esse efeito da variância. Uma maneira simples de saber se a série necessita de alguma transformação é utilizar o gráfico amplitude *versus* média (Figura 3.1). Ao se obter uma reta paralela ao eixo das abcissas e pontos espalhados em torno dela, indica que a variância é homogênea (MORETTIN; TOLOI, 2006).

Figura 3.1 – Gráfico amplitude *versus* média.



Fonte: Autoria própria, 2022.

3.2.1 Estacionariedade

Antes de iniciar-se a definição de estacionariedade de uma série temporal é preciso ter a definição clara do que seria um processo estocástico. Assim, é apresentada a definição de um processo estocástico conforme Morettin e Toloi (2006).

Definição: Um processo estocástico é uma família $Z = \{Z(t), t \in T\}$, tal que, para cada $t \in T$, $Z(t)$ é uma variável aleatória (v.a.), sendo T um conjunto arbitrário. Em outras palavras, são processos controlados por leis probabilísticas.

Agora, quanto a estacionariedade de uma série, pode ser basicamente interpretada como aquela que não apresenta tendência, nem sazonalidade e com variância finita. Portanto, é uma série que se desenvolve no tempo, de forma aleatória, ao redor de uma média constante. Ainda, são duas as formas de estacionariedade: fraca (de segunda ordem, ou ampla) e estrita (ou forte).

Primeiramente, consideram-se t_1, t_2, \dots, t_n elementos quaisquer de um conjunto arbitrário T , então as distribuições finito-dimensionais para todo $n \geq 1$ são dadas por:

$$F(z_1, \dots, z_n; t_1, \dots, t_n) = P\{Z(t_1) \leq z_1, \dots, Z(t_n) \leq z_n\}. \quad (3.8)$$

Sendo assim, Morettin e Toloi (2006) definem um processo estocástico $Z = \{Z(t), t \in T\}$ como estritamente estacionário se todas as distribuições finito-dimensionais continuam as mesmas sob translações no tempo, ou seja,

$$F(z_1, \dots, z_n; t_1 + \tau, \dots, t_n + \tau) = F(z_1, \dots, z_n; t_1, \dots, t_n), \quad (3.9)$$

para quaisquer $t_1, t_2, \dots, t_n, \tau$ de T .

Por outro lado, um processo estocástico $Z = \{Z(t), t \in T\}$ é dito fracamente estacionário ou estacionário de segunda ordem se satisfazer as três condições,

- i) $E\{Z(t)\} = \mu(t) = \mu$, constante, para todo $t \in T$;
- ii) $E\{Z^2(t)\} < \infty$, para todo $t \in T$;
- iii) $\gamma(t_1, t_2) = \text{Cov}\{Z(t_1), Z(t_2)\}$ é uma função de $|t_1 - t_2|$.

Portanto, a estacionariedade é uma das pressuposições que devem ser satisfeitas, quando se pretende aplicar alguns modelos como, por exemplo, o modelo auto-regressivo integrado de

médias móveis (ARIMA) ou modelos de valores extremos, que será visto na seção 3.3. Entretanto, existem métodos para séries não estacionárias, que não serão vistos nesta dissertação, mas caso seja de interesse podem ser encontrados em Coles (2001). Sendo assim, nos próximos tópicos serão apresentados alguns testes para verificar se uma série é estacionária.

3.2.2 Teste de tendência

Conforme visto anteriormente, a série temporal dada pela equação (3.6) é escrita como a soma de três componentes, sendo uma delas a tendência, representada por T_t . Na literatura, podem ser encontrados vários testes com o objetivo de verificar a existência dessa componente. Dentre eles, têm-se: o teste das sequências, o teste de Mann-Kendall, o teste de run, o teste de raiz unitária, o teste de Cox-Stuart, entre outros. Para maiores detalhes a respeito de alguns testes aqui mencionados, pode ser visto em Paiva e Sáfiadi (2021).

O teste de Cox-Stuart, também chamado de teste do sinal, dentre os citados é o mais utilizado, além de ser de fácil manuseio. Devido a esses motivos, são descritos a seguir os passos para a sua aplicação.

Considerando uma série com N observações, primeiramente é feita a divisão da série em duas partes, obtendo assim duas subséries. Em seguida, formam-se pares de tal forma que, o primeiro valor de cada uma das subséries representa o primeiro par, depois é feito com o segundo valor de cada uma delas, formando o segundo par, e assim por diante até o último valor, ou seja, $(Z_1, Z_{c+1}), \dots, (Z_{N-c}, Z_N)$. Caso o número de observações seja par $c = \frac{N}{2}$, mas se o número de observações for ímpar $c = \frac{N+1}{2}$.

Após formado os pares, é feita a comparação entre as observações de cada par (Z_i, Z_{i+c}) , atribuindo o sinal positivo (+) caso $Z_i < Z_{i+c}$ e o sinal negativo (-) caso contrário, ou seja, $Z_i > Z_{i+c}$, eliminando-se os empates.

Por fim, adota-se n ao número de pares em que $Z_i \neq Z_{i+c}$. Logo após, compara-se estatisticamente a quantidade de sinais positivos e negativos. As hipóteses a serem testadas são:

$$\begin{cases} H_0 : P(Z_i < Z_{i+c}) = P(Z_i > Z_{i+c}), \forall i : \text{ não existe tendência;} \\ H_1 : P(Z_i < Z_{i+c}) \neq P(Z_i > Z_{i+c}), \forall i : \text{ existe tendência.} \end{cases}$$

Logo, para um valor de $n \leq 20$ a regra de decisão baseia-se na distribuição binomial, com parâmetros n e $p = \frac{1}{2}$. Se $n > 20$, utiliza-se a aproximação da normal, com $\mu = np$ e $\sigma^2 = npq$, com $p = q = \frac{1}{2}$. Então, seja T_2 o número de pares com o sinal (+), se $T_2 \geq n - t$,

sendo t um valor tabelado, rejeita-se H_0 ao nível de significância α , ou seja, a série apresenta tendência (MORETTIN; TOLOI, 2006). Ou ainda, a hipótese nula é rejeitada quando o valor- p for inferior ao nível de significância α .

3.2.3 Teste de sazonalidade

Retornando à equação (3.6), além da componente tendência, pode haver a presença de sazonalidade, sendo necessária a verificação. Para tanto, existem testes que podem ser utilizados como alternativa para avaliar a presença dessa componente, como o teste de Kruskal-Wallis, o teste de Friedman e o teste de Fisher.

Os primeiros dois testes mencionados podem ser vistos com mais detalhes em Morettin e Tolo (2006). Já o teste de Fisher, segundo Paiva e Sáfadi (2021), além de ser utilizado para verificar a presença da componente sazonal, foi proposto inicialmente para testar o maior período, ou seja, detectar grandes periodicidades. Como hipóteses para serem testadas, têm-se:

$$\begin{cases} H_0 : \text{não existe sazonalidade;} \\ H_1 : \text{existe sazonalidade.} \end{cases}$$

A estatística do teste é dada por:

$$g = \frac{\max[I_p]}{\sum_{p=1}^{N/2} I_p}, \quad (3.10)$$

em que I_p é o valor do periodograma no período p e N é o número de observações da série. Enquanto que o valor crítico z_α é calculado por:

$$z_\alpha = 1 - \left(\frac{\alpha}{n}\right)^{\frac{1}{n-1}}, \quad (3.11)$$

em que α é um dado nível de significância e $n = \frac{N}{2}$.

Caso $g > z_\alpha$, então se rejeita H_0 , ou seja, a série apresenta sazonalidade correspondente ao período p .

3.2.4 Função de autocorrelação

Uma outra característica ao se utilizar a teoria clássica de valores extremos, objeto deste estudo, é de que a série de dados seja independente. No entanto, existem técnicas que são

utilizadas caso os valores sejam dependentes, porém não será foco desta pesquisa. Caso seja de interesse do leitor, pode ser visto com mais detalhes em Coles (2001).

A função de autocorrelação (FAC), representada graficamente, é usada para verificar a correlação entre as observações. Então, caso a correlação for nula indicará que a série é independente. Logo, a função de autocorrelação de um processo estacionário é definida como:

$$\rho_k = \frac{\text{Cov}(Z_t, Z_{t+k})}{\sqrt{\text{Var}(Z_t)\text{Var}(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0}, \quad k \in \mathbb{Z} \quad (3.12)$$

onde,

γ_k é a função de autocovariância de lag k ;

γ_0 é a variância da série temporal Z_t , com $t = 1, 2, \dots, N$.

3.3 Teoria de valores extremos (TVE)

A teoria de valores extremos (TVE) é considerada um ramo da estatística que permite a análise de eventos que apresentam baixa frequência relativa e alto impacto, particularmente chamados de eventos extremos. Sua aplicação é encontrada nas mais diferentes áreas como, por exemplo: na análise de eventos hidrológicos, no cálculo de seguros, para controle de tráfego, na engenharia para predição de resistência de materiais, e atualmente está sendo utilizado em dados de poluição do ar.

Os valores extremos (máximos ou mínimos), segundo Silva (2008), são eventos raros e que ocorrem nas caudas das distribuições, situados longe do centro da distribuição dos dados. Estes valores, em muitos casos são descartados da análise, devido ao número pequeno deles presente em uma série. De acordo com Lopo *et al.* (2015), a maioria das teorias estatísticas concentra-se na análise do centro da distribuição, prestando pouca atenção à cauda. Em muitos casos, o comportamento da cauda é diferente do comportamento da maioria das caudas nas distribuições tradicionais, como a Normal e Gama.

Ao se fazer uso de tal metodologia, conforme Coles (2001), existem duas abordagens principais para tratar variáveis aleatórias com valores extremos: o método de blocos e o método *Peaks Over Threshold* (POT). Pelo método de blocos, os valores de máximo (ou mínimo) são extraídos a cada determinado período (semana, mês, ano) e o teorema de Fisher-Tippet a ser visto na seção 3.3.1, garantem que esses valores convergem para uma das três distribuições de valores extremos, sendo representada unicamente pela Distribuição Generalizada de Valores

Extremos (GEV). Enquanto que a abordagem POT, em geral, se faz uso da Distribuição Generalizada de Pareto (GPD) aos valores que excedem um limiar suficientemente alto (para máximo) ou baixo (para mínimo). Nos tópicos a seguir, cada método será descrito em detalhes.

3.3.1 Valores extremos pelo método de blocos

Seja X_1, X_2, \dots, X_n uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d) e com função de distribuição F_X . Ao dispor a sequência em ordem crescente, tem-se $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ como sendo as estatísticas de ordem. Sendo assim, $X_{(n)} = M_n = \max(X_1, X_2, \dots, X_n)$ é o máximo das n primeiras variáveis aleatórias. Logo, a função de distribuição de M_n é dada por:

$$\begin{aligned}
 F_{M_n}(x) &= P(M_n \leq x) \\
 &= P[\max(X_1, X_2, \dots, X_n) \leq x] \\
 &= P(X_1 \leq x, X_2 \leq x, \dots, X_n \leq x) \\
 &= P(X_1 \leq x) \times P(X_2 \leq x) \times \dots \times P(X_n \leq x), \quad (\text{por independência}) \\
 &= \prod_{i=1}^n P(X_i \leq x) \\
 &= \prod_{i=1}^n [F_X(x_i)], \quad (\text{por definição}) \\
 F_{M_n}(x) &= [F_X(x)]^n. \tag{3.13}
 \end{aligned}$$

Como a função de distribuição F_X é desconhecida, ao fazer a aproximação de F_X^n , por distribuições limites com $n \rightarrow \infty$, tem-se que a distribuição assintótica de M_n é degenerada. Então, para contornar isso, Fisher e Tippett (1928) assumiram as constantes de normalização $a_n > 0$ e $b_n \in \mathbb{R}$, de tal forma que:

$$\lim_{n \rightarrow \infty} P\left(M_n^* = \frac{M_n - b_n}{a_n} \leq x\right) = F_X^n(a_n x + b_n) \rightarrow G(x), \tag{3.14}$$

em que, $G(x)$ é uma função de distribuição não-degenerada.

Teorema de Fisher-Tippett (FISHER; TIPPETT, 1928; GNEDENKO, 1943): seja (X_n) uma sequência de variáveis aleatórias i.i.d. Se existem constantes padronizadoras $a_n > 0$ e $b_n \in \mathbb{R}$ e alguma função distribuição não degenerada G tal que $M_n^* = \frac{M_n - b_n}{a_n}$ converge em distribuição para G , então a função $G(x)$ pertence a alguma das três famílias de valores extremos, dadas por:

i) Distribuição Gumbel (Tipo I)

$$G(x) = \exp \left\{ -\exp \left[-\left(\frac{x-b}{a} \right) \right] \right\}, \quad -\infty < x < \infty; \quad (3.15)$$

ii) **Distribuição Fréchet (Tipo II)**

$$G(x) = \begin{cases} 0, & x \leq b, \\ \exp \left\{ -\left(\frac{x-b}{a} \right)^{-\alpha} \right\}, & x > b \text{ e } \alpha > 0; \end{cases} \quad (3.16)$$

ii) **Distribuição Weibull (Tipo III)**

$$G(x) = \begin{cases} \exp \left\{ -\left[-\left(\frac{x-b}{a} \right)^\alpha \right] \right\}, & x < b \text{ e } \alpha > 0, \\ 1, & x \geq b, \end{cases} \quad (3.17)$$

para os parâmetros $a > 0$ e $b \in \mathbb{R}$. Por outro lado, Coles (2001) declara que uma das principais dificuldades é saber qual das três distribuições seria a mais adequada para uma determinada série de dados. Com isso, em 1955, Jenkinson propôs uma forma generalizada dessas distribuições, ou seja, a unificação das três famílias de valores extremos em uma única distribuição, chamada de Distribuição Generalizada de Valores Extremos (GEV), que será vista a seguir.

3.3.1.1 Distribuição generalizada de valores extremos (GEV)

Com a unificação das três distribuições, definidas pelo Teorema de Fisher-Tippet, obteve-se a distribuição GEV, proposta por Jenkinson (1955), de um único parâmetro $\xi = \frac{1}{\alpha}$, além dos parâmetros μ e σ . A função de distribuição de probabilidade acumulada é dada por:

$$F_{GEV}(x) = \exp \left\{ -\left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}, \quad (3.18)$$

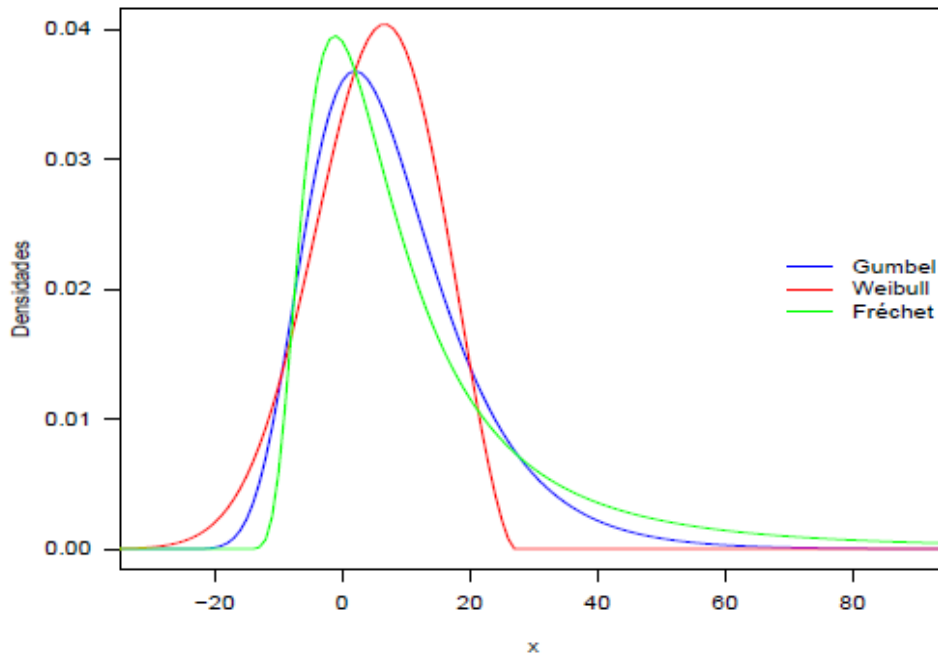
definida em $-\infty < x < \mu - \frac{\sigma}{\xi}$ para $\xi < 0$, $-\infty < x < \infty$ para $\xi \rightarrow 0$ e $\mu - \frac{\sigma}{\xi} < x < +\infty$ para $\xi > 0$, em que μ , σ e ξ são os parâmetros localização, escala e forma, respectivamente, com $\mu, \xi \in \mathbb{R}$ e $\sigma > 0$.

De acordo com Coles (2001), as distribuições de valores extremos de Gumbel, Fréchet e Weibull correspondem aos casos particulares da expressão (3.18) em que $\xi = 0$, $\xi > 0$ e $\xi < 0$, respectivamente.

As diferenças entre as três distribuições podem ser vistas graficamente, em que cada uma exibe um comportamento característico em suas caudas extremas (inferior e superior). Para tanto, foram construídos os gráficos (Figura 3.2) da função densidade de probabilidade da

GEV para $\xi = -0,40$ (Weibull), $\xi = 0$ (Gumbel) e $\xi = 0,40$ (Fréchet), com $\mu = 2$ e $\sigma = 20$. Pode-se observar que o parâmetro ξ determina a natureza das caudas da distribuição, ou seja, a velocidade com que as caudas se aproximam de zero.

Figura 3.2 – Função densidade de probabilidade da distribuição GEV para $\xi = -0,40$ (Weibull), $\xi = 0$ (Gumbel) e $\xi = 0,40$ (Fréchet), com $\mu = 2$ e $\sigma = 20$.



Fonte: Autoria própria, 2022.

Para o caso particular, em que aplicando-se o limite de $F_{GEV}(x)$, equação (3.18), com $\xi \rightarrow 0$ se tem a função de distribuição acumulada de Gumbel, que é expressa por:

$$F_{GU}(x) = \exp \left[-\exp \left(-\frac{x-\mu}{\sigma} \right) \right], \quad (3.19)$$

definida em $-\infty < x < \infty$, em que μ e σ são os parâmetros de localização e escala, respectivamente, com $\mu \in \mathbb{R}$ e $\sigma > 0$.

Ao derivar a expressão (3.18) em relação a x , obtém-se a função densidade de probabilidade da distribuição GEV, dada por:

$$f_{GEV}(x; \mu, \sigma, \xi) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \exp \left\{ - \left[1 + \xi \left(\frac{x-\mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}, \quad (3.20)$$

definida em $-\infty < x < \mu - \frac{\sigma}{\xi}$ para $\xi < 0$ e $\mu - \frac{\sigma}{\xi} < x < +\infty$ para $\xi > 0$.

Agora para o caso particular, ou seja quando aplicado o limite de f_{GEV} com $\xi \rightarrow 0$, derivando-se a expressão (3.19) em relação a x , obtém-se a função densidade de probabilidade da distribuição Gumbel, dada por:

$$f_{GU}(x; \mu, \sigma) = \frac{1}{\sigma} \exp \left\{ - \left(\frac{x - \mu}{\sigma} \right) - \exp \left[- \left(\frac{x - \mu}{\sigma} \right) \right] \right\}, \quad (3.21)$$

definida em $-\infty < x < \infty$.

3.3.1.2 Estimação dos parâmetros da distribuição GEV

Coles (2001) afirma que diversos métodos foram propostos para realizar inferências sobre os parâmetros μ , σ e ξ da distribuição generalizada de valores extremos (GEV). Dentre eles têm-se: técnicas gráficas baseadas em gráficos de probabilidade, estimadores baseados no método dos momentos, mínimos quadrados ordinários, métodos Bayesianos, mínimos quadrados generalizados, método dos momentos L e da máxima verossimilhança. Este último, consiste em estimar os valores de diferentes parâmetros do modelo estatístico de modo a maximizar a probabilidade de ocorrência da amostra.

Entre as técnicas aqui citadas, segundo Smith (1985), os métodos baseados em verossimilhança são preferidos devido à teoria dos estimadores de máxima verossimilhança ser bem compreendida e as inferências serem facilmente modificadas ao incorporar modelos com estruturas mais complexas. Ainda de acordo com o autor, após realizados inúmeros estudos quanto ao comportamento assintótico dos estimadores de máxima verossimilhança para a distribuição GEV, verificou-se que as condições de regularidade não são necessariamente satisfeitas e obteve os seguintes resultados:

- a) quando $\xi > -0,5$, os estimadores de máxima verossimilhança são completamente regulares;
- b) quando $-1 \leq \xi \leq -0,5$, os estimadores de máxima verossimilhança existem, mas não são regulares;
- c) quando $\xi < -1$, os estimadores de máxima verossimilhança não existem.

Smith (1985) faz uma breve observação quanto a situação em que $\xi < -0,5$, afirmando que tal caso é extremamente raro para dados ambientais, pois correspondem a distribuições com cauda superior muito curta e finita.

Casella e Berger (2001) mencionam que os estimadores de máxima verossimilhança apresentam boas propriedades assintóticas, tais como: invariância, eficiência, consistência e suficiência. Sendo assim, a função de verossimilhança é definida a seguir.

Definição: Seja uma amostra X_1, X_2, \dots, X_n de n variáveis aleatórias independentes e identicamente distribuídas (i.i.d), com função densidade de probabilidade $f(x_i, \theta)$ e $\mathbf{x} = (x_1, x_2, \dots, x_n)$ uma série de observações. Então, a função de verossimilhança é dada por:

$$L(\theta | \mathbf{x}) = \prod_{i=1}^n f(x_i | \theta), \quad (3.22)$$

onde θ é o vetor de parâmetros e \mathbf{x} o vetor de observações.

Desta maneira, satisfeita a pressuposição de independência, ao aplicar a função densidade de probabilidade da GEV (expressão (3.20)) na função de verossimilhança (expressão (3.22)) com $\theta = (\mu, \sigma, \xi)'$, tem-se a seguinte expressão:

$$L(\mu, \sigma, \xi | \mathbf{x}) = \frac{1}{\sigma^n} \prod_{i=1}^n \left\{ \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \right\} \exp \left\{ \sum_{i=1}^n \left\{ - \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \right\}, \quad (3.23)$$

para $\xi < 0$ e $x_i < \mu - \frac{\sigma}{\xi}$, ou seja, $\mu - \frac{\sigma}{\xi} > x_{(n)}$ ou para $\xi > 0$ e $x_i > \mu - \frac{\sigma}{\xi}$, ou seja, $\mu - \frac{\sigma}{\xi} < x_{(1)}$.

Para facilitar a derivação da função de verossimilhança, primeiramente é aplicado o logaritmo na expressão (3.23), obtendo-se a função Log-verossimilhança dada por:

$$\begin{aligned} l(\mu, \sigma, \xi; \mathbf{x}) &= \ln[L(\mu, \sigma, \xi | \mathbf{x})] \\ &= \ln \left[\frac{1}{\sigma^n} \prod_{i=1}^n \left\{ \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\left(\frac{1+\xi}{\xi}\right)} \right\} \exp \left\{ \sum_{i=1}^n \left\{ - \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\} \right\} \right] \\ &= -n \ln(\sigma) - \left(\frac{1+\xi}{\xi} \right) \sum_{i=1}^n \ln \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^n \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \\ &= \sum_{i=1}^n \left\{ -\ln(\sigma) - \left(\frac{1+\xi}{\xi} \right) \ln \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right] - \left[1 + \xi \left(\frac{x_i - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right\}, \quad (3.24) \end{aligned}$$

Os estimadores de máxima verossimilhança de μ , σ e ξ são obtidos maximizando-se a expressão (3.24), ou seja, efetuando a derivada de primeira ordem em relação aos três parâmetros e igualando a zero, e em seguida resolvendo o seguinte sistema de equações não-lineares,

$$\begin{cases} \frac{\partial l}{\partial \mu}(\mu, \sigma, \xi) = \frac{1}{\hat{\sigma}} \sum_{i=1}^n \left\{ \frac{1 + \hat{\xi} - w_i^{-\frac{1}{\hat{\sigma}}}}{w_i} \right\} = 0; \\ \frac{\partial l}{\partial \sigma}(\mu, \sigma, \xi) = -\frac{n}{\hat{\sigma}} + \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n \left\{ \frac{(x_i - \hat{\mu}) \left[1 + \hat{\xi} - w_i^{-\frac{1}{\hat{\sigma}}} \right]}{w_i} \right\} = 0; \\ \frac{\partial l}{\partial \xi}(\mu, \sigma, \xi) = \sum_{i=1}^n \left\{ \left(1 - w_i^{-\frac{1}{\hat{\sigma}}} \right) \left[\frac{\ln(w_i)}{\hat{\xi}^2} - \left(\frac{x_i - \hat{\mu}}{\hat{\xi} \hat{\sigma} w_i} \right) \right] - \frac{(x_i - \hat{\mu})}{\hat{\sigma} w_i} \right\} = 0, \end{cases} \quad (3.25)$$

em que $w_i = 1 + \hat{\xi} \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right)$.

No entanto, o sistema de equações não-lineares (3.25), em geral, não possui solução analítica, sendo necessário o uso de algum método iterativo (ou computacional). Sendo assim, na literatura podem ser encontrados diferentes métodos numéricos, como o algoritmo de Nelder-Mead (NM), Newton-Raphson, Gradiente Conjugado (CG), *Simulating Annealing* (SANN), entre outros, que podem ser vistos na função *optim* do pacote *stats* do R (R Core Team, 2021).

Dentre eles, o método de Nelder-Mead é, provavelmente, o mais utilizado método de busca direta, sendo as principais vantagens de seu uso, a não necessidade de calcular as derivadas da função para fazer a otimização, pode ser usado para funções descontínuas, além de ser utilizado para encontrar o ponto ótimo de uma vasta gama de funções (DINIZ-EHRHARDT; LOPES; PEDROSO, 2010).

Para o caso particular da distribuição GEV quando $\xi = 0$, ou seja, a distribuição Gumbel representada pela expressão (3.21), ao aplicá-la na expressão (3.22) com $\theta = (\mu, \sigma)'$, obtém-se a função de verossimilhança da distribuição Gumbel, dada por:

$$\begin{aligned} L(\mu, \sigma | \mathbf{x}) &= \prod_{i=1}^n f(x_i | \mu, \sigma) = \prod_{i=1}^n \left\{ \frac{1}{\sigma} \exp \left\{ - \left(\frac{x_i - \mu}{\sigma} \right) - \exp \left[- \left(\frac{x_i - \mu}{\sigma} \right) \right] \right\} \right\} \\ &= \frac{1}{\sigma^n} \exp \left\{ \sum_{i=1}^n \left[- \left(\frac{x_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^n \exp \left[- \left(\frac{x_i - \mu}{\sigma} \right) \right] \right\}. \end{aligned} \quad (3.26)$$

Após, aplica-se o logaritmo na equação (3.26), obtendo a função Log-verossimilhança da distribuição Gumbel,

$$\begin{aligned}
l(\mu, \sigma; \mathbf{x}) &= -n \ln(\sigma) + \sum_{i=1}^n \left[-\left(\frac{x_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^n \exp \left[-\left(\frac{x_i - \mu}{\sigma} \right) \right] \\
&= \sum_{i=1}^n \left\{ -\ln(\sigma) - \left(\frac{x_i - \mu}{\sigma} \right) - \exp \left[-\left(\frac{x_i - \mu}{\sigma} \right) \right] \right\}. \tag{3.27}
\end{aligned}$$

Por fim, maximizando-se a expressão (3.27), os estimadores de máxima verossimilhança de μ e σ são obtidos através do seguinte sistema de equações não-lineares:

$$\begin{cases} \frac{\partial l}{\partial \mu}(\mu, \sigma) = -\frac{1}{\hat{\sigma}} \left\{ \left[\sum_{i=1}^n \exp \left(-\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \right] - n \right\} = 0; \\ \frac{\partial l}{\partial \sigma}(\mu, \sigma) = -\frac{n}{\hat{\sigma}} + \sum_{i=1}^n \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}^2} \right) \left[1 - \exp \left(-\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \right] = 0. \end{cases} \tag{3.28}$$

Como no caso da distribuição GEV, o sistema de equações (3.28) não possui solução analítica, então utiliza-se o algoritmo de Nelder-Mead para obtenção de uma solução numérica. Uma observação que precisa ser exposta sobre o método, é a não convergência para funções estritamente convexas em \mathbb{R}^2 que pode ocorrer, pois o limite da sequência dos vértices pode ser um ponto não estacionário (DINIZ-EHRHARDT; LOPES; PEDROSO, 2010).

3.3.2 Valores extremos pelo método *Peaks Over Threshold* (POT)

O método *Peaks Over Threshold* (POT), de acordo com Martins Júnior, Sáfadi e Carvalho (2013), também chamado de método dos excessos (pontos excedentes) além do limiar u , é utilizado com o propósito de reduzir a perda de informação, seja pelo motivo de haver mais de um valor extremo no mesmo bloco, ou seja pela escassez de observações.

Além disso, segundo Alves (2007), a metodologia POT é considerada bastante útil em aplicações práticas devido ao seu uso mais eficiente das observações mais elevadas. Como é o caso de dados de poluição, em que para cada um existe um limite estabelecido por lei.

A abordagem POT consiste em obter, a partir de um conjunto de amostras x_1, x_2, \dots, x_n , um novo conjunto de observações excedentes. Os eventos extremos são identificados após definido um limiar ótimo chamado de u , para o qual tem-se um conjunto de excedências ($x_i : x_i > u$). Em seguida, denota-se estes excedentes por $x_{(1)}, x_{(2)}, \dots, x_{(k)}$, e os excessos acima deste limiar são representados por:

$$y_j = x_{(j)} - u, \quad j = 1, 2, \dots, k. \tag{3.29}$$

No t3pico 3.3.2.2, ser3o fornecidas informa33es mais detalhadas sobre como escolher o limiar u , que 3 obtido graficamente com o aux3lio da fun33o da m3dia amostral dos excessos.

Pickands (1975), afirma que se y_j , com $j = 1, 2, \dots, k$, sendo os valores acima de um limiar suficiente grande, ent3o esse pode ser ajustado em uma Distribu33o Generalizada de Pareto, que ser3 vista em seguida. Al3m disso, se o limiar 3timo u , pode ser determinado, ent3o o conjunto de dados pertence ao dom3nio de atra33o das distribu33es de valores extremos, ou seja, a distribu33o de Pareto comum, ou distribu33o exponencial, ou ainda a distribu33o Pareto tipo II, que tamb3m ser3o apresentadas na pr3xima se33o.

3.3.2.1 Distribu33o generalizada de Pareto (GPD)

Seja uma amostra X_1, X_2, \dots, X_n de vari3veis aleat3rias i.i.d com fun33o de distribu33o $F(x)$, a diferen3a entre os valores X_i e um dado limiar u , representado por $Y = \{X_i - u \mid X_i > u\}$ 3 denominado de excesso sobre um limiar, na se33o 3.3.2, e segue uma distribu33o generalizada de Pareto, dada por:

$$G(y) = \begin{cases} 1 - \left[1 + \xi \left(\frac{y}{\sigma_u} \right) \right]^{-\frac{1}{\xi}}, & \text{se } \xi \neq 0; \\ 1 - \exp\left(-\frac{y}{\sigma_u}\right), & \text{se } \xi = 0, \end{cases} \quad (3.30)$$

onde u , $\sigma_u > 0$ e ξ s3o os par3metros de posi33o fixado, escala e forma, respectivamente. A equa33o (3.30) 3 v3lida para $0 \leq y < \infty$ se $\xi > 0$, $0 \leq y < -\frac{\sigma_u}{\xi}$ se $\xi < 0$ e $y \in \mathbb{R}$ se $\xi = 0$.

Coles (2001) afirma que a fam3lia de distribu33es definida acima 3 chamada fam3lia generalizada de Pareto. Assim como as distribu33es GEV s3o as distribu33es limite para o m3ximo, as do tipo GPD s3o as formas param3tricas para distribu33es limite de excessos. As distribu33es generalizadas de Pareto s3o da forma Exponencial ($\xi = 0$), Pareto tipo II ($\xi > 0$) e Pareto comum ou Beta ($\xi < 0$). Uma vez que a distribu33o exponencial 3 um caso especial da distribu33o GP, ao aplicar a restri33o que o par3metro de forma tende a zero.

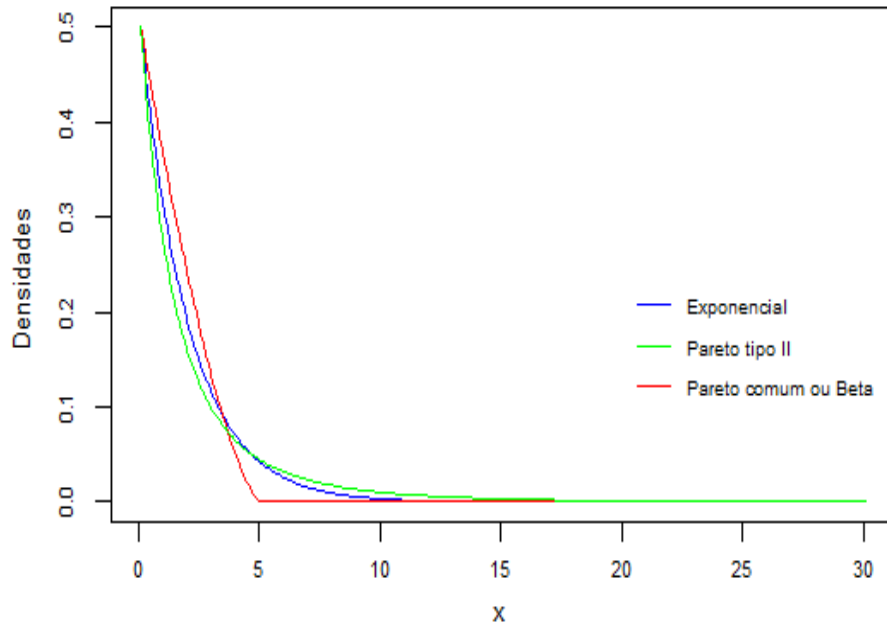
Para obter a fun33o densidade de probabilidade da distribu33o GP e para o caso particular, em que se aplica o limite de $G(y)$ com $\xi \rightarrow 0$, deriva-se a fam3lia generalizada de Pareto, dada pela equa33o (3.30), em rela33o a vari3vel y , adquirindo a seguinte express3o:

$$g(y) = \begin{cases} \frac{1}{\sigma_u} \left[1 + \xi \left(\frac{y}{\sigma_u} \right) \right]^{-\frac{1+\xi}{\xi}}, & \text{se } \xi \neq 0; \\ \frac{1}{\sigma_u} \exp\left(-\frac{y}{\sigma_u}\right), & \text{se } \xi = 0, \end{cases} \quad (3.31)$$

definida em $0 \leq y < \infty$ se $\xi > 0$, $0 \leq y < -\frac{\sigma_u}{\xi}$ se $\xi < 0$ e $y \in \mathbb{R}$ se $\xi = 0$.

A Figura 3.3, apresenta os gráficos da função de distribuição GP para $\xi = -0,40$ (Pareto comum ou Beta), ξ tendendo a zero (exponencial) e $\xi = 0,40$ (Pareto tipo II), todas com $\sigma_u = 2$. Observa-se que assim como na GEV o parâmetro ξ é quem determina as caudas da distribuição.

Figura 3.3 – Função densidade de probabilidade da distribuição GP para $\xi = -0,40$ (Pareto comum ou Beta), $\xi \rightarrow 0$ (Exponencial) e $\xi = 0,40$ (Pareto tipo II), com $\sigma_u = 2$.



Fonte: Autoria própria, 2022.

Além disso, existe também uma relação analítica entre a função de distribuição GP (expressão (3.30)) e a função de distribuição F da GEV correspondente a expressão (3.18), dada por:

$$G_{\xi}(x) = 1 + \ln(F_{\xi}(x)), \quad \ln(F_{\xi}(x)) > -1, \quad (3.32)$$

Tal relação explica o fato das densidades da GPD possuírem cauda extrema assintoticamente equivalente às de uma GEV (MORITA; BUENO; PIRES, 2008).

3.3.2.2 Escolha do limiar u

De acordo com Coles (2001), como na escolha do tamanho do bloco, a escolha do limiar u é uma etapa fundamental na modelagem de excessos por uma distribuição generalizada de Pareto, pois conforme a escolha do valor de u implicará em uma amostra com maior viés ou com alta variância. Neste caso, se o limiar é muito baixo pode não garantir a convergência

dos excessos, levando a um viés. Por outro lado, se o limiar for muito alto produzirá poucos excessos com os quais o modelo pode ser estimado, resultando no aumento da variância.

Para isso, estão disponíveis na literatura (BEHRENS; LOPES; GAMERMAN, 2004) alguns métodos que auxiliam na escolha desse valor. Aqui serão apresentados dois deles: o gráfico de vida média residual (ou média dos excessos) e o de avaliação da estabilidade das estimativas dos parâmetros.

Segundo Coles (2001), o primeiro método é baseado na média da distribuição generalizada de Pareto. Então, se Y tem uma distribuição Generalizada de Pareto (GP) com parâmetros σ_u e ξ , a média é representada por:

$$E[Y] = \frac{\sigma_u}{1 - \xi}, \quad (3.33)$$

sendo $\xi < 1$, e caso contrário, ou seja $\xi \geq 1$, a média é infinita. Agora, supondo que a distribuição GP é válida como modelo para os excessos de um limiar u_0 gerado por uma série X_1, X_2, \dots, X_n , da qual um termo arbitrário é denotado por X , então:

$$E[X - u_0 \mid X > u_0] = \frac{\sigma_{u_0}}{1 - \xi}, \quad (3.34)$$

onde $\xi < 1$, e σ_{u_0} como sendo o parâmetro de escala para os excessos do limiar u_0 . Mas se a distribuição GP é válida para os excessos do limiar u_0 , deve igualmente ser válida para todos os limiares $u > u_0$. Portanto, considerando $u > u_0$ tem-se:

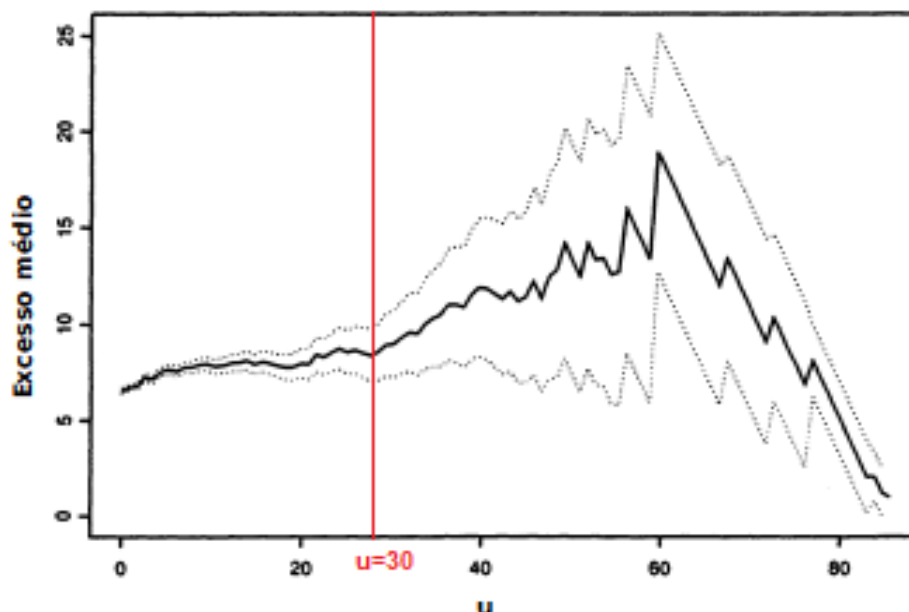
$$E[X - u \mid X > u] = \frac{\sigma_u}{1 - \xi} = \frac{\sigma_{u_0} + \xi u}{1 - \xi}. \quad (3.35)$$

Então, $E[X - u \mid X > u]$ é uma função linear de u , além de ser a média dos excessos do limiar u , para o qual a média amostral dos excessos fornece uma estimativa empírica. De acordo com a expressão (3.35), espera-se que essas estimativas mudem linearmente com u , em níveis de u para os quais a distribuição GP é apropriada. Logo, o gráfico de vida residual média é obtido por:

$$\left\{ \left(u, \frac{1}{n_u} \sum_{i=1}^{n_u} (x_{(i)} - u) \right) : u < x_{max} \right\}, \quad (3.36)$$

em que $x_{(1)}, x_{(2)}, \dots, x_{(n_u)}$ são as observações que excedem u e x_{max} é o maior entre os valores observados. Para mais detalhes de como obter esse resultado, pode ser visto em Silva (2008). Na Figura 3.4 é apresentado um exemplo de gráfico de vida média residual.

Figura 3.4 – Gráfico de vida média residual.



Fonte: Adaptado de COLES, 2001.

Após obtido o gráfico da média de excessos, Coles (2001) declara que o processo de interpretação não é simples, mas que o limiar é definido sobre o valor de u que diferencia a curva linear e não linear (ou aproximadamente linear). Complementando, Silva (2008) sugere que o valor do limiar u seja o menor valor de limiar a partir do qual a função das médias dos excessos amostral segue aproximadamente uma função afim. Como exemplo, utilizando a Figura 3.4 tem-se que a partir de $u_0 \cong 30$ a função das médias dos excessos possui comportamento próximo de uma função de 1º grau até $u \cong 60$ e depois decai.

O segundo procedimento, para seleção de limiares, em muitos casos, é utilizado como uma técnica complementar, devido a análise gráfica da média dos excessos ser subjetiva. O método consiste em ajustar a distribuição GP para uma gama de limiares acima de um nível u_0 , e em seguida identificar uma estabilidade nas estimativas dos parâmetros. Para o caso das estimativas do parâmetro forma (ξ), essas devem ser constantes, enquanto que as estimativas de σ_u devem apresentar comportamento linear para valores de $u > u_0$ (LEMOS *et al.*, 2020).

3.3.2.3 Estimação dos parâmetros da GPD

Dentre as várias técnicas de estimação dos parâmetros, para o caso da distribuição GP, o método da máxima verossimilhança é o mais aconselhável. Sendo assim, satisfeita a pressuposição de que os excessos acima de um limiar u são i.i.d, substitui-se a função densidade de

probabilidade da GP na função de verossimilhança (expressão (3.22)), em que $\theta = (\sigma_u, \xi)'$, obtendo a seguinte função de verossimilhança da distribuição generalizada de Pareto como sendo:

$$\begin{aligned} L(\sigma_u, \xi | \mathbf{y}) &= \prod_{i=1}^n \frac{1}{\sigma_u} \left(1 + \xi \frac{y_i}{\sigma_u}\right)^{-\left(\frac{1+\xi}{\xi}\right)} \\ &= \frac{1}{\sigma_u^n} \sum_{i=1}^n \left(1 + \xi \frac{y_i}{\sigma_u}\right)^{-\left(\frac{1+\xi}{\xi}\right)}, \end{aligned} \quad (3.37)$$

em que \mathbf{y} é o vetor de observações acima do limiar u . Em seguida, aplica-se o logaritmo natural na expressão (3.37), obtendo a função Log-verossimilhança da distribuição GP,

$$\begin{aligned} l(\sigma_u, \xi; \mathbf{y}) &= \ln[L(\sigma_u, \xi | \mathbf{y})] \\ &= -n \ln(\sigma_u) - \left(\frac{1+\xi}{\xi}\right) \sum_{i=1}^n \ln \left(1 + \xi \frac{y_i}{\sigma_u}\right). \end{aligned} \quad (3.38)$$

Derivando-se a expressão (3.38) em relação aos parâmetros σ_u e ξ e igualando as suas derivadas a zero obtém-se o seguinte sistema de equações:

$$\begin{cases} \frac{\partial l}{\partial \sigma_u}(\sigma_u, \xi) = -\frac{n}{\hat{\sigma}_u} + \left(\frac{1+\hat{\xi}}{\hat{\xi}}\right) \frac{\hat{\xi}}{\hat{\sigma}_u^2} \sum_{i=1}^n \frac{y_i}{\left(1 + \hat{\xi} \frac{y_i}{\hat{\sigma}_u}\right)} = 0; \\ \frac{\partial l}{\partial \xi}(\sigma_u, \xi) = \frac{1}{\hat{\xi}^2} \sum_{i=1}^n \ln \left(1 + \hat{\xi} \frac{y_i}{\hat{\sigma}_u}\right) - \left(\frac{1+\hat{\xi}}{\hat{\xi}}\right) \sum_{i=1}^n \frac{y_i}{\hat{\sigma}_u \left(1 + \hat{\xi} \frac{y_i}{\hat{\sigma}_u}\right)} = 0. \end{cases} \quad (3.39)$$

O sistema de equações (3.39) não possui solução analítica e, portanto, utiliza-se o algoritmo de Nelder-Mead implementado no *software* R (R Core Team, 2021).

Para o caso particular da distribuição GP com $\xi = 0$, ou seja, para a distribuição exponencial, aplica-se a sua função densidade de probabilidade na expressão (3.22), em que $\theta = \sigma_u$, obtendo-se a função de verossimilhança da distribuição exponencial dada por:

$$\begin{aligned} L(\sigma_u | \mathbf{y}) &= \prod_{i=1}^n \frac{1}{\sigma_u} \exp\left(-\frac{y_i}{\sigma_u}\right) \\ &= \frac{1}{\sigma_u^n} \sum_{i=1}^n \exp\left(-\frac{y_i}{\sigma_u}\right), \end{aligned} \quad (3.40)$$

e cujo logaritmo é,

$$l(\sigma_u; \mathbf{y}) = -n \ln(\sigma_u) - \frac{1}{\sigma_u} \sum_{i=1}^n y_i. \quad (3.41)$$

Portanto, o estimador de máxima verossimilhança para σ_u é obtido após efetuar a derivada primeira da expressão (3.41) em relação a σ_u e igualando a zero, ou seja,

$$\frac{dl(\sigma_u)}{d\sigma_u} = -\frac{n}{\hat{\sigma}_u} + \frac{\sum_{i=1}^n y_i}{\hat{\sigma}_u^2} = 0. \quad (3.42)$$

Resolvendo a equação acima (expressão (3.42)) obtém-se a seguinte solução:

$$-\frac{n}{\hat{\sigma}_u} + \frac{\sum_{i=1}^n y_i}{\hat{\sigma}_u^2} = 0 \Rightarrow \frac{n}{\hat{\sigma}_u} = \frac{\sum_{i=1}^n y_i}{\hat{\sigma}_u^2}$$

$$\hat{\sigma}_u = \frac{\sum_{i=1}^n y_i}{n}. \quad (3.43)$$

Assim sendo, a estimativa para o parâmetro escala (σ_u) será a média dos máximos, com n observações.

3.3.3 Qualidade do ajuste

Ao realizar o ajuste de uma distribuição de probabilidade a um conjunto de dados, ou seja, obter as estimativas para os parâmetros de um modelo, pretende-se que essa distribuição represente adequadamente aquele conjunto de observações. Para isso, na literatura existem alguns testes de aderência como o Qui-Quadrado, Filliben, Kolmogorov-Smirnov e o Anderson-Darling, que caso seja de interesse do leitor, podem ser vistos em Naghettini e Pinto (2007). Na parte de análise deste estudo, será empregado o último teste aqui mencionado.

De acordo com Naghettini e Pinto (2007), o teste de Anderson-Darling (AD) apresenta-se como uma boa alternativa, pois tenta atribuir maior significado aos extremos das distribuições, em que observações maiores (ou menores) podem alterar a qualidade do ajuste. Esse teste, tal como o de Kolmogorov-Smirnov (KS), baseia-se na diferença entre as funções de probabilidades acumuladas teórica e empírica, mas com uma função de peso que enfatiza as discrepâncias em ambas as caudas.

Para a realização do teste, primeiramente considera-se $X_{(1)}, X_{(2)}, \dots, X_{(n)}$ uma série de dados observados e ordenados de forma crescente. A função de distribuição teórica é definida por $F(x_{(i)})$ e a função de distribuição acumulada empírica de X é dada por:

$$\hat{F}(x_{(i)}) = \frac{i}{n+1}, \quad i = 1, 2, \dots, n. \quad (3.44)$$

Em seguida, define-se um nível α de significância, além das seguintes hipóteses:

$$\begin{cases} H_0 : F(x) = \hat{F}(x); \\ H_1 : F(x) \neq \hat{F}(x). \end{cases}$$

Calcula-se então a sua estatística de teste, representada por A^2 , que é descrita conforme equação (3.45):

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) \{ \ln [F(x_{(i)})] + \ln [1 - F(x_{(n-i+1)})] \}, \quad (3.45)$$

onde $\hat{F}(x)$ e $F(x)$ são as distribuições empírica e teórica, respectivamente. Portanto, se a estatística de teste $A^2 < A^2_{crit, \alpha}$ (tabelado), então não existe indícios para rejeitar H_0 , ou seja, a série de dados segue a distribuição teórica. A hipótese nula também não será rejeitada quando o valor- p for superior ao nível de significância α .

Além dos testes estatísticos, uma ferramenta utilizada para avaliar o ajuste da distribuição é por meio dos gráficos *pp-plot* (gráfico probabilidade-probabilidade) e *qq-plot* (gráfico quantil-quantil). No entanto, em alguns casos, a aplicação dos métodos gráficos isoladamente pode levar a conclusões errôneas. Assim sendo, após a avaliação gráfica, recomenda-se realizar alguns dos testes aqui citados para confirmar o que se vê nos gráficos.

O gráfico *pp-plot* é formado pelos pares,

$$\left[\frac{i}{n+1}, F(x_{(i)}) \right], \quad i = 1, 2, \dots, n,$$

onde $F(x_{(i)})$ é o valor obtido após aplicação do valor observado na função teórica.

O gráfico *qq-plot* é formado pelos pontos de coordenadas,

$$\left[F^{-1} \left(\frac{i}{n+1} \right), x_{(i)} \right], \quad i = 1, 2, \dots, n,$$

em que $F^{-1}(\cdot)$ é a função inversa da distribuição teórica e $x_{(i)}$ os valores observados. Em ambos os casos, os pontos devem estar próximos a uma reta com inclinação 1, ou seja, quanto mais distantes da reta, menos adequada é a distribuição proposta.

3.3.4 Critérios de seleção

Após verificar quais modelos se ajustaram a série de dados, é feita a escolha de qual modelo melhor descreve os valores observados. Para isso, são utilizados alguns testes para comparar esses modelos ajustados.

O teste da razão de verossimilhança é utilizado para testar dois modelos, desde que um deles seja um caso especial do outro, ou seja, para o método dos blocos, tem-se que a distribuição Gumbel é obtida através da aplicação do limite com $\xi \rightarrow 0$ na distribuição GEV, ou para o método POT, a distribuição exponencial é obtida a partir da aplicação do limite com $\xi \rightarrow 0$ na distribuição GP.

Com isso, verifica-se se a presença do parâmetro ξ acrescentam-se informações significativas a respeito do comportamento de tal variável aleatória. Sendo assim, as hipóteses a serem testadas são:

$$\begin{cases} H_0 : \xi = 0; \\ H_1 : \xi \neq 0. \end{cases}$$

De acordo com Hosking (1984), seja (x_1, x_2, \dots, x_n) uma série de n observações, $l(\hat{\theta}_{\text{GEV}})$ e $l(\hat{\theta}_{\text{GU}})$ (ou, $l(\hat{\theta}_{\text{GP}})$ e $l(\hat{\theta}_{\text{Exp}})$) os máximos do logaritmo da função de verossimilhança das distribuições GEV (expressão 3.24) e Gumbel (expressão 3.27) (ou, GP (expressão 3.38) e Exponencial (expressão 3.41)), em que $\hat{\theta}_{\text{GEV}} = (\hat{\mu}, \hat{\sigma}, \hat{\xi})'$ e $\hat{\theta}_{\text{GU}} = (\hat{\mu}, \hat{\sigma})'$ (ou, $\hat{\theta}_{\text{GP}} = (\hat{\sigma}_u, \hat{\xi})'$ e $\hat{\theta}_{\text{Exp}} = \hat{\sigma}_u$) são os vetores de estimativas de máxima verossimilhança. Para as distribuições pelo método dos blocos, a estatística do teste é dada por:

$$T_{RL} = -2 [l(\hat{\theta}_{\text{GU}}) - l(\hat{\theta}_{\text{GEV}})], \quad (3.46)$$

e para as distribuições pelo método POT, tem-se:

$$T_{RL} = -2 [l(\hat{\theta}_{\text{Exp}}) - l(\hat{\theta}_{\text{GP}})], \quad (3.47)$$

ambas tem distribuição assintótica χ^2 com 1 grau de liberdade.

De modo a obter uma estimativa mais aproximada de uma distribuição assintótica, Hosking (1984) sugeriu uma modificação no teste, obtendo a seguinte reformulação:

$$T_{RL}^* = \left(1 - \frac{2,8}{n}\right) T_{RL}. \quad (3.48)$$

Por fim, para se testar a hipótese H_0 contra H_1 , o valor da estatística T_{RL}^* deve ser comparado com o valor tabelado da distribuição qui-quadrado com 1 grau de liberdade (diferença de 1 parâmetro entre os modelos). Se $T_{RL}^* \geq \chi_{\alpha,1}^2$, então rejeita-se H_0 , ou seja, os modelos com parâmetro ξ acrescentam informações, e portanto não podem ser desprezados.

Após obter o modelo que melhor descreve a série de máximos, pode-se então calcular a probabilidade de ocorrência de um valor extremo máximo de interesse, além do período de retorno com o seu respectivo nível de retorno, que são apresentados nas próximas seções.

3.3.5 Probabilidade de ocorrência de valores extremos

Para obter as probabilidades de ocorrência de valores extremos máximos do modelo, com base no método de blocos, a seguinte expressão é usada:

$$P(X > x) = 1 - F_{\text{GEV}}(x) |_{\theta=\hat{\theta}} = 1 - \exp \left\{ - \left[1 + \hat{\xi} \left(\frac{x - \hat{\mu}}{\hat{\sigma}} \right) \right]^{-\frac{1}{\hat{\xi}}} \right\}. \quad (3.49)$$

Para o caso em que $\xi = 0$, a probabilidade é dada por:

$$P(X > x) = 1 - F_{\text{GU}}(x) |_{\theta=\hat{\theta}} = 1 - \exp \left\{ - \exp \left(- \frac{x - \hat{\mu}}{\hat{\sigma}} \right) \right\}, \quad (3.50)$$

em que x representa o limite acima do qual se pretende calcular a probabilidade de ocorrência de um evento.

Para se obter a probabilidade de ocorrência de um valor maior ou igual a x , dado que esse valor seja maior que um limiar u , utiliza-se a expressão (3.51),

$$\begin{aligned} P(X - u > x | X > u) &= 1 - G_{\text{GP}}(x) |_{\theta=\hat{\theta}} = 1 - \left\{ 1 - \left[1 + \hat{\xi} \left(\frac{x - u}{\hat{\sigma}_u} \right) \right]^{-\frac{1}{\hat{\xi}}} \right\} \\ &= \left[1 + \hat{\xi} \left(\frac{x - u}{\hat{\sigma}_u} \right) \right]^{-\frac{1}{\hat{\xi}}}. \end{aligned} \quad (3.51)$$

Quando $\xi = 0$, a probabilidade é dada através da cumulativa da exponencial:

$$\begin{aligned}
P(X - u > x \mid X > u) &= 1 - G_{\text{Exp}}(x) \mid_{\theta=\hat{\theta}} = 1 - \left[1 - \exp\left(-\frac{x-u}{\hat{\sigma}_u}\right) \right] \\
&= \exp\left(-\frac{x-u}{\hat{\sigma}_u}\right).
\end{aligned} \tag{3.52}$$

3.3.6 Período de retorno

Em muitos casos, além do interesse na probabilidade de um determinado evento máximo acontecer, também se tem o interesse de saber quanto tempo levará para que o evento ocorra novamente. Esse é denominado período de retorno, conhecido também como tempo de recorrência, denotado por τ e definido como o inverso da probabilidade de um evento ser igualado ou superado, ou seja,

$$\tau = \frac{1}{P(X \geq x)} = \frac{1}{1 - F(x)}. \tag{3.53}$$

O nível de retorno (x_p) para o caso da GEV, associado ao período de retorno τ , é obtido através da solução de,

$$\int_{-\infty}^{x_p} f(\hat{\theta}) dx = 1 - p, \tag{3.54}$$

em que $p = \frac{1}{\tau}$, chegando assim na seguinte expressão:

$$F(x_p) = 1 - p. \tag{3.55}$$

Aplicando a função inversa de F em ambos os lados da equação (3.55), obtém-se:

$$\hat{x}_p = F^{-1}(1 - p) = \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left\{ 1 - [-\ln(1 - p)]^{-\hat{\xi}} \right\}, \tag{3.56}$$

para $\hat{\xi} \neq 0$. Considerando o limite da equação (3.56) com $\hat{\xi} \rightarrow 0$, ou seja, para o caso da distribuição Gumbel, tem-se:

$$\hat{x}_p = F^{-1}(1 - p) = \hat{\mu} - \hat{\sigma} \ln[-\ln(1 - p)]. \tag{3.57}$$

Supondo que uma distribuição GP com parâmetros σ e ξ é um modelo apropriado para os excessos de um limiar u para uma variável X , então para $x > u$, tem-se:

$$P(X > x \mid X > u) = \left[1 + \xi \left(\frac{x-u}{\sigma} \right) \right]^{-\frac{1}{\xi}}. \tag{3.58}$$

Após manipular a expressão, obtém-se:

$$P(X > x) = \zeta_u \left[1 + \xi \left(\frac{x-u}{\sigma} \right) \right]^{-\frac{1}{\xi}}, \quad (3.59)$$

onde $\zeta_u = P(X > u)$. Assim, o nível x_m que é excedido, em média, uma vez a cada m observações, é dada pela seguinte resolução:

$$\begin{aligned} \zeta_u \left[1 + \xi \left(\frac{x_m - u}{\sigma} \right) \right]^{-\frac{1}{\xi}} &= \frac{1}{m} \\ \left[1 + \xi \left(\frac{x_m - u}{\sigma} \right) \right]^{-\frac{1}{\xi}} &= \frac{1}{m\zeta_u} \\ 1 + \xi \left(\frac{x_m - u}{\sigma} \right) &= \left(\frac{1}{m\zeta_u} \right)^{-\xi} \\ \xi \left(\frac{x_m - u}{\sigma} \right) &= (m\zeta_u)^\xi - 1 \\ x_m - u &= \frac{\sigma}{\xi} \left[(m\zeta_u)^\xi - 1 \right]. \end{aligned}$$

Logo,

$$x_m = u + \frac{\sigma}{\xi} \left[(m\zeta_u)^\xi - 1 \right], \quad (3.60)$$

dado que m seja suficientemente grande para garantir que $x_m > u$, e com $\xi \neq 0$. Se $\xi = 0$, ou seja, a distribuição exponencial, tem-se:

$$\zeta_u \left[\exp \left(-\frac{x_m - u}{\hat{\sigma}_u} \right) \right] = \frac{1}{m}. \quad (3.61)$$

Reorganizando fica,

$$x_m = u + \sigma \ln(m\zeta_u), \quad (3.62)$$

desde que m seja suficientemente grande. Para se obter a estimativa \hat{x}_m requer a substituição das estimativas de máxima verossimilhança de σ e ξ na equação (3.60) ou a substituição da estimativa de máxima verossimilhança de σ na equação (3.62). A estimativa de ζ_u , ou seja, a probabilidade de uma observação exceder o limiar u , terá o seguinte estimador natural,

$$\hat{\zeta}_u = \frac{k}{n}, \quad (3.63)$$

a proporção da amostra de pontos que excede u , sendo k o número de observações acima do limiar e n o número total de observações, tal que o número de excedências seguem uma distribuição Binomial(n, ζ_u) (COLES, 2001).

3.4 Valor em risco (VaR)

O valor em risco, ou do inglês *Value at Risk* (VaR), é uma medida amplamente utilizada nas instituições financeiras como um indicador de gestão de risco para auxiliar na tomada de decisão, com o objetivo de sempre minimizar as perdas. Segundo Marimoutou, Raggad e Trabelsi (2009), o VaR é definido como a perda máxima que ocorrerá em uma carteira de investimentos (ações, seguros, etc.) com certo grau de confiança em um determinado período de tempo e envolve a estimativa do quantil da distribuição dos retornos.

Martins Júnior (2010) faz uma observação quanto ao uso dessa medida para a análise de dados de poluição, em que relata a necessidade de fazer algumas alterações. Uma delas é o fato de que para o estudo de poluentes o mais importante é o valor absoluto e não o valor das variações em certo período. O VaR para dados de poluição representa basicamente a probabilidade (ou, o risco) de se ultrapassar um determinado valor legalmente arbitrado. Assim, conforme Tsay (2010), o procedimento para se obter o VaR utilizando a Teoria de Valores Extremos, através do método de blocos é descrito a seguir.

Seja p^* uma probabilidade, ou nível de risco que se admita ocorrer e x_n^* o quantil do máximo subperíodo em questão, então:

$$p^* = \begin{cases} \exp \left[- \left(1 + \frac{\xi(x_n^* - \mu)}{\sigma} \right)^{-\frac{1}{\xi}} \right], & \text{se } \xi \neq 0; \\ \exp \left[- \exp \left(- \left(\frac{x_n^* - \mu}{\sigma} \right) \right) \right], & \text{se } \xi = 0, \end{cases} \quad (3.64)$$

em que, $1 + \frac{\xi(x_n^* - \mu)}{\sigma} > 0$ se $\xi \neq 0$. Aplicando o logaritmo natural em ambos os lados de cada equação, obtém-se:

$$\ln(p^*) = \begin{cases} - \left[1 + \frac{\xi(x_n^* - \mu)}{\sigma} \right]^{-\frac{1}{\xi}}, & \text{se } \xi \neq 0; \\ - \exp \left[- \left(\frac{x_n^* - \mu}{\sigma} \right) \right], & \text{se } \xi = 0. \end{cases} \quad (3.65)$$

Assim, os quantis são dados por:

$$x_n^* = \begin{cases} \mu - \frac{\sigma}{\xi} \left\{ 1 - [-\ln(1 - p^*)]^{-\xi} \right\}, & \text{se } \xi \neq 0; \\ \mu - \sigma \ln[-\ln(1 - p^*)], & \text{se } \xi = 0. \end{cases} \quad (3.66)$$

Dada uma probabilidade p^* , o quantil x_n^* representado pela expressão (3.66) é o VaR baseado na teoria de valores extremos para os máximos do subperíodo.

A relação entre o máximo de um bloco e os valores observados de uma série original pode ser vista a seguir:

$$p^* = P(x_{n,i} > x_n^*) = 1 - P(x_{n,i} \leq x_n^*) = 1 - [P(x_t \leq x_n^*)]^n,$$

segue que

$$1 - p^* = [1 - P(x_t > x_n^*)]^n. \quad (3.67)$$

Com $p = P(x_t > x_n^*)$, tem-se a seguinte relação,

$$1 - p^* = (1 - p)^n. \quad (3.68)$$

A partir dessa relação entre probabilidades, quando substituída na expressão (3.66), é possível obter a estimativa do VaR, representada por:

$$\widehat{\text{VaR}}_p = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}} \left\{ 1 - [-n \ln(1 - p)]^{-\hat{\xi}} \right\}, & \text{se } \xi \neq 0; \\ \hat{\mu} - \hat{\sigma} \ln[-n \ln(1 - p)], & \text{se } \xi = 0, \end{cases} \quad (3.69)$$

em que n é o tamanho do subperíodo.

O procedimento para encontrar o VaR utilizando a abordagem POT, conforme Tsay (2010) e Marimoutou, Raggad e Trabelsi (2009), envolve a apresentação da distribuição generalizada de Pareto, e em seguida a distribuição exponencial.

Para o caso da distribuição GP, sabe-se que a função de distribuição de excesso condicional $F_u(y)$ é definida como,

$$F_u(y) = P(X - u \leq y \mid X > u), \quad 0 \leq y \leq x_F - u, \quad (3.70)$$

em que X é a variável aleatória, u é um determinado limite, $y = X - u$ são os excessos e $x_F \leq \infty$ é o ponto final de F . Ao reescrever a expressão (3.70), tem-se:

$$\begin{aligned}
F_u(y) &= \frac{F(u+y) - F(u)}{1 - F(u)} \\
&= \frac{F(x) - F(u)}{1 - F(u)}.
\end{aligned} \tag{3.71}$$

Isolando $F(x)$ da equação (3.71), obtém-se:

$$F(x) = [1 - F(u)]F_u(y) + F(u). \tag{3.72}$$

Substituindo F_u pela distribuição GP e $F(u)$ pela estimativa $\frac{n-N_u}{n}$, onde n é o número total de observações e N_u o número de observações acima do limiar u , logo a expressão a ser resolvida é:

$$\begin{aligned}
\hat{F}(x) &= \left(1 - \frac{n-N_u}{n}\right) \left\{1 - \left[1 + \hat{\xi} \left(\frac{x-u}{\hat{\sigma}}\right)\right]^{-\frac{1}{\hat{\xi}}}\right\} + \left(\frac{n-N_u}{n}\right) \\
&= \frac{N_u}{n} \left[1 - \left(1 + \hat{\xi} \left(\frac{x-u}{\hat{\sigma}}\right)\right)^{-\frac{1}{\hat{\xi}}}\right] + \left(1 - \frac{N_u}{n}\right) \\
&= \frac{N_u}{n} - \frac{N_u}{n} \left(1 + \hat{\xi} \left(\frac{x-u}{\hat{\sigma}}\right)\right)^{-\frac{1}{\hat{\xi}}} + 1 - \frac{N_u}{n},
\end{aligned}$$

que ao simplificar chega no seguinte resultado:

$$\hat{F}(x) = 1 - \frac{N_u}{n} \left(1 + \hat{\xi} \left(\frac{x-u}{\hat{\sigma}}\right)\right)^{-\frac{1}{\hat{\xi}}}. \tag{3.73}$$

Por fim, ao inverter a equação (3.73) para uma dada probabilidade p , a estimativa do VaR de uma distribuição GP pode ser obtida por,

$$\widehat{\text{VaR}}_p = u + \frac{\hat{\sigma}}{\hat{\xi}} \left\{ \left[\frac{n}{N_u} (1-p) \right]^{-\hat{\xi}} - 1 \right\}. \tag{3.74}$$

No caso de $\hat{\xi} \rightarrow 0$, ou seja, a distribuição exponencial, a estimador do VaR é o seguinte:

$$\widehat{\text{VaR}}_p = u + \hat{\sigma} \ln \left(\frac{n}{N_u} (1-p) \right). \tag{3.75}$$

4 MATERIAL E MÉTODOS

4.1 Dados

Para este estudo foram utilizados dados horários de poluição atmosférica proveniente do município de Brumadinho, MG. O município (latitude $20^{\circ}07'06''S$, longitude $44^{\circ}12'04''W$ e altitude de 881 m) está localizado na região Metropolitana de Belo Horizonte a 49 km de distância da capital mineira, com uma área de 639,434 km² e população total de 41.208 habitantes (IBGE, 2021). O clima predominante é o subtropical úmido (*Cwa*), conforme a classificação de Köppen, com temperatura média anual de 21 °C e precipitação média anual de 1267 mm, sendo considerado dezembro o mês mais chuvoso e o inverno correspondente ao período mais seco.

A principal atividade produtiva de Brumadinho é a mineração. Como principais áreas de extração de minerais têm-se a Mina do Tejuco (mineração Tejucana), a Mina Bocaina (empresa Mineral do Brasil), a mina da MIB (Mineração Ibrité), a Mina Serra Azul (Mineração Ipê), a mina da Ferrous, a Mina de Jangada e a Mina Córrego do Feijão. Além disso, possui várias áreas de extração de areia, localizadas principalmente ao longo do rio Paraopeba, atendendo ao mercado metropolitano (AMPLO ENGENHARIA E GESTÃO DE PROJETOS, 2019).

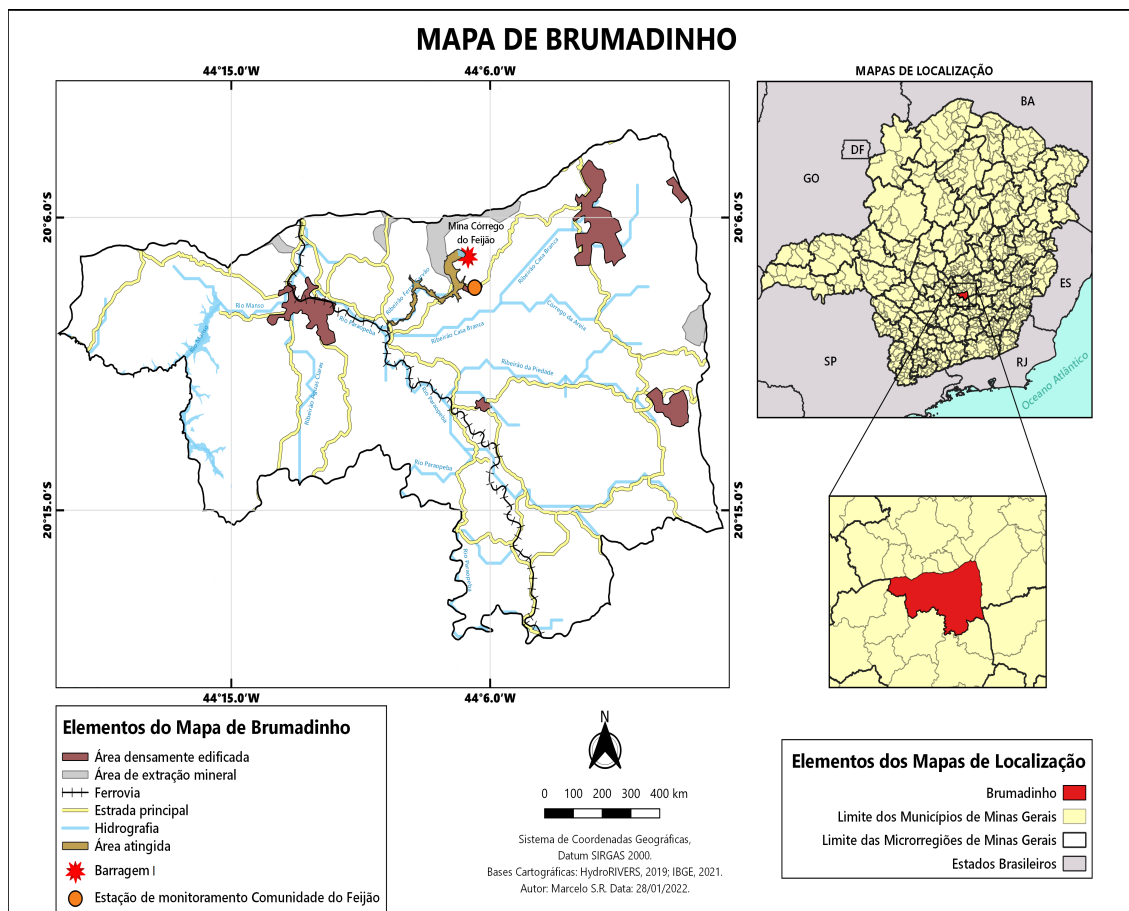
Na produção agrícola o município destaca-se no cultivo de cana-de-açúcar, tomate, mandioca e milho, laranja, limão e banana. De acordo com Santos e Farias (2014), pela sua proximidade com a capital, faz com que seja um importante produtor e fornecedor de produtos hortifrutigranjeiros para atender ao mercado da Região Metropolitana de Belo Horizonte (RMBH).

A região possui três estações de monitoramento de qualidade do ar, sendo que a escolhida foi a Estação Comunidade do Feijão, situada no povoado de Córrego do Feijão, nas seguintes coordenadas geográficas: latitude $20^{\circ}08'9.13''S$, longitude $44^{\circ}06'32.19''W$ e altitude de 854 m. Todos os dados foram obtidos no *site* da Fundação Estadual do Meio Ambiente (FEAM) de Minas Gerais. A FEAM é responsável por desenvolver e implementar políticas públicas relacionadas às mudanças climáticas, energias renováveis, qualidade do ar e solo, além da gestão de efluentes líquidos e sólidos, visando proteger e melhorar a qualidade ambiental de Minas Gerais (FEAM, 2021).

A escolha da Estação Comunidade do Feijão se deu pelo fato de estar localizada próxima ao local do rompimento da barragem e das atividades de mineração, além de apresentar maior número de observações registradas. A Figura 4.1 apresenta a localização do município, da

estação de monitoramento e da Barragem I no Complexo da Mina Córrego do Feijão, que está situada no Quadrilátero Ferrífero, na bacia hidrográfica do Ribeirão Ferro-Carvão.

Figura 4.1 – Localização e mapa do município de Brumadinho, MG.



Fonte: Elaborado pelo autor, 2022.

Na Figura 4.1, observa-se, também, que o município de Brumadinho possui vários rios e ribeirões, dentre eles o principal é o rio Paraopeba, que atravessa a região no sentido sul - norte, e que recebeu parte dos rejeitos de minério devido ao rompimento da Barragem I. Conforme Carvalho *et al.* (2021), em 25 de janeiro de 2019, foram liberados para jusante cerca de 12 milhões de metros cúbicos de rejeitos, sedimentos e detritos, que assoreou o vale do ribeirão Ferro-Carvão e alcançou a foz no rio Paraopeba (Figura 4.1), onde formou um depósito.

Dentre os poluentes monitorados, foram escolhidos as partículas totais em suspensão (PTS) e o material particulado (MP_{10}), ambos em $\mu g/m^3$, pois são as que estavam disponíveis no *site* da FEAM. As séries foram divididas e identificadas em antes e depois do rompimento da barragem. A primeira foi denominada série antes do rompimento, que é referente ao período de 01 de janeiro de 2018 a 24 de janeiro de 2019, constituindo 9336 observações horárias (considerando os valores ausentes). A segunda foi a série depois do rompimento, que corresponde

ao período de 25 de janeiro de 2019 a 31 de dezembro de 2019, totalizando 8184 observações horárias (incluso as ausentes).

4.2 Métodos

Primeiramente, foi feita uma análise exploratória dos dados originais de PTS e MP₁₀, ou seja, da série com falhas (valores faltantes). Foram obtidas algumas medidas descritivas para cada poluente, além dos gráficos. Para auxiliar na análise optou-se por gráficos com a mediana e alguns percentis, para cada dia da semana, com intuito de analisar o comportamento de cada poluente, antes e após o rompimento da barragem.

Após, para efetuar as comparações das médias entre os períodos (antes e após o rompimento da barragem), para cada poluente, utilizou-se o teste *t-Student*, com nível de significância de 95%, em que a hipótese nula considera que em média as concentrações do poluente em questão são iguais. Além disso, pelo teste *Z* foram feitas as comparações das proporções de observações que ultrapassaram o limite legal, sendo que a hipótese nula considera que a proporção de poluente entre os períodos é igual. Para uma melhor compreensão dos testes citados, recomenda-se ver o livro de Morettin e Bussab (2017).

Em seguida, se fez uso de métodos de imputação de dados às séries incompletas, de forma a completar cada uma com valores plausíveis, obtendo séries completas, a fim de darem mais consistência ao estudo. Para isso, inicialmente foi aplicada a metodologia de imputação múltipla, descrita na seção 3.1, por meio do método *predictive mean matching* (PMM) e que está implementado no pacote *mice* (BUUREN; GROOTHUIS-OUDSHOORN, 2011). A escolha desse método se deu, pois de acordo com os estudos de Oleques *et al.* (2020) e Pinto, Reisen e Lima (2016), a PMM é uma boa alternativa para se realizar o preenchimento de valores faltantes em séries de dados ambientais e/ou atmosféricos.

Em primeiro momento, após serem realizadas as imputações, foi feita uma análise preliminar com as duas abordagens de valores extremos, tanto com as séries originais como com as séries imputadas. Sendo assim, pelo método de blocos, as observações máximas foram obtidas através da série completa, ou seja, a série com dados imputados. Já pelo método POT, os máximos foram retirados da série original (ou, com falhas). No entanto, com o objetivo de se obter valores máximos de mesmo tamanho de bloco e limiar u , de ambos os períodos, foi proposto a união das séries de dados antes e após o rompimento.

Em seguida, foram obtidos os valores máximos de cada poluente, aplicando a abordagem de blocos à série completa, ou seja, a série com dados originais e imputados. Para isso, foi dividida a série em intervalos contendo um determinado número de observações, formando vários blocos e então retirou-se o maior valor em cada um desses blocos, formando assim uma série de máximos.

Para o caso em que se utilizou a abordagem POT, esta foi aplicada aos dados originais, ou seja, a série com valores faltantes. Desta forma, os valores máximos obtidos foram aqueles que excedem um determinado limiar u , sendo considerado como limite, em que somente os valores acima dele é que irão compor a série de máximos. Para a escolha do melhor valor de limiar utilizou-se o gráfico de vida média residual (seção 3.3.2.2), selecionando três possíveis valores de u que satisfizessem a estabilidade das estimativas dos parâmetros σ_u e ξ .

De posse das concentrações máximas de cada poluente, foi então feita a separação das observações conforme cada período, formando duas séries de máximos (antes e depois do rompimento da barragem) para cada poluente. Ao utilizar a teoria de valores extremos, uma das pressuposições que deve ser satisfeita é a da independência dos dados. Então, para verificar se as observações máximas de cada poluente não são correlacionadas, utilizou-se o gráfico da função de autocorrelação, descrita no tópico 3.2.4. Em geral, séries longitudinais possuem correlação e é pela busca por séries independentes que indicará a escolha do tamanho dos blocos ou do limiar u .

Portanto, para cada série de máximos foi definido um período inicial e a cada análise foi se aumentando o tamanho dos blocos, com o objetivo de eliminar a correlação entre os máximos e assim obter séries independentes. No caso da abordagem POT, dentre os três possíveis valores para o limiar u , escolheu-se como limiar ótimo o menor valor tal que a função das médias dos excessos apresente comportamento linear, além disso os máximos obtidos a partir desse limiar devem ser independentes.

Após analisada a independência de cada série de poluentes, foi verificada a existência ou não de tendência por meio do teste de Cox-Stuart (teste do sinal), de acordo com o procedimento descrito na seção 3.2.2. Para verificar a presença de sazonalidade nas séries, utilizou-se o teste de Fisher conforme apresentado na seção 3.2.3.

Visto que os valores máximos de cada poluente satisfazem os pressupostos de independência e estacionariedade, foi então realizado o ajuste da distribuição generalizada de valores extremos (GEV) e da distribuição generalizada de Pareto (GPD), apresentadas nas seções 3.3.1

e 3.3.2, às séries de máximos. A estimação dos parâmetros para os dois tipos de abordagem se fez através do método da máxima verossimilhança, por meio do algoritmo de Nelder-Mead, descritos nos tópicos 3.3.1.2 e 3.3.2.3.

Em seguida, com posse dos valores estimados de cada parâmetro, foi verificada a qualidade dos modelos ajustados por meio do teste de Anderson-Darling, exibido na seção 3.3.3, além dos gráficos *pp-plot* e *qq-plot*. Por fim, descartados os modelos que porventura não obtiverem um bom ajuste, é então selecionado o modelo que melhor descreve cada série de extremos. Para tanto, foi aplicado o teste da razão de verossimilhança (para modelos aninhados), descrito na seção 3.3.4.

Desta forma, obtidos os parâmetros μ , σ e ξ caso a distribuição seja pelo método de blocos, ou σ_u e ξ caso a distribuição for pelo método POT, foi possível obter os quantis substituindo-se as estimativas em sua respectiva função de distribuição. Assim, pôde-se calcular o VaR conforme o procedimento descrito na seção 3.4, dado o nível de significância e diferentes períodos. Além disso, foi calculada a probabilidade de exceder o limite legal de cada poluente, de acordo com a seção 3.3.5. Continuando, com base nesses valores foi possível avaliar se com o rompimento da barragem em Brumadinho houve o agravamento na qualidade do ar.

Todas as análises, ajustes e gráficos foram feitos utilizando-se o *software R* (R Core Team, 2021), por meio do pacote básico *stats* e de outros que são descritos a seguir: *openair* (CARSLAW; ROPKINS, 2012) para a construção dos gráficos da concentração de poluente em cada dia da semana; *mice* (BUUREN; GROOTHUIS-OUDSHOORN, 2011) para a realização das imputações; *randtests* (CAEIRO; MATEUS, 2014) para o teste de tendência; *evmix* (HU; SCARROTT, 2018) para a escolha do limiar u ; *evd* (STEPHENSON, 2002), *extRemes* (GILLELAND; KATZ, 2016) e *POT* (RIBATET; DUTANG, 2022) para o ajuste dos modelos; *gofstest* (FARAWAY *et al.*, 2021) para o teste de aderência; *VaRES* (CHAN; NADARAJAH; AFUECHETA, 2016) para auxiliar no cálculo do VaR.

5 RESULTADOS

5.1 Série de partículas totais em suspensão (PTS)

Para uma melhor compreensão da concentração de PTS, em $\mu\text{g}/\text{m}^3$, liberada na atmosfera do município de Brumadinho antes (de 01/01/2018 a 24/01/2019) e após (de 25/01/2019 a 31/12/2019) o rompimento da barragem de rejeitos, apresentam-se na Tabela 5.1 algumas medidas descritivas utilizando-se das séries de dados incompletas.

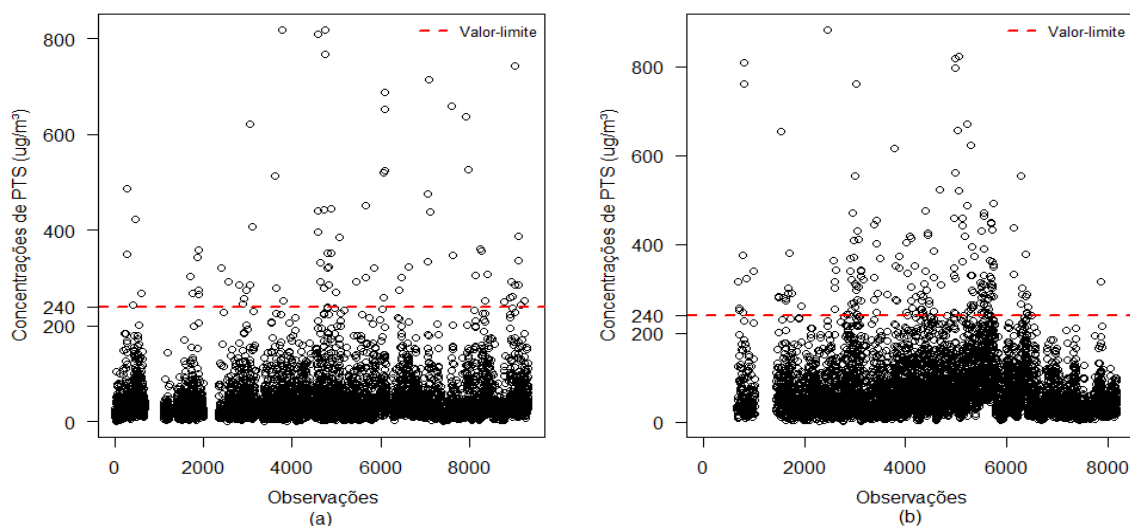
Tabela 5.1 – Estatísticas descritivas para a concentração de PTS ($\mu\text{g}/\text{m}^3$).

Localidade	Nº de observações faltantes	Proporção de valores faltantes (%)	Média	Mediana	Desvio padrão	Máximo
Brumadinho (antes do rompimento)	2377	25,46	44,29	29,00	53,97	818,00
Brumadinho (após o rompimento)	1855	22,67	73,33	51,00	72,47	883,00

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

A partir das medidas apresentadas na Tabela 5.1, percebe-se que a média da concentração de PTS foi estatisticamente superior em Brumadinho após o rompimento da barragem, pois o teste *t-Student* ($t = -25,986$ e valor- $p = 2,2 \times 10^{-16}$), com nível de confiança de 95%, indicou que as médias das concentrações são diferentes. Além disso, o valor máximo após o rompimento foi levemente superior ao valor máximo do período anterior à ruptura da barragem. Uma outra forma de se analisar descritivamente as séries é por meio da análise gráfica, representada na Figura 5.1.

Figura 5.1 – Gráficos de concentração de PTS, em $\mu\text{g}/\text{m}^3$, do município de Brumadinho antes do rompimento (a) e após o rompimento (b) da barragem.



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Pelos gráficos apresentados na Figura 5.1, observa-se que as concentrações máximas de PTS estão em torno de 700 a 800 $\mu\text{g}/\text{m}^3$. Quanto à quantidade acumulada em horas que a concentração de PTS ultrapassou o limite de 240 $\mu\text{g}/\text{m}^3$ e suas respectivas proporções, antes e após o rompimento, podem ser vistas na Tabela 5.2.

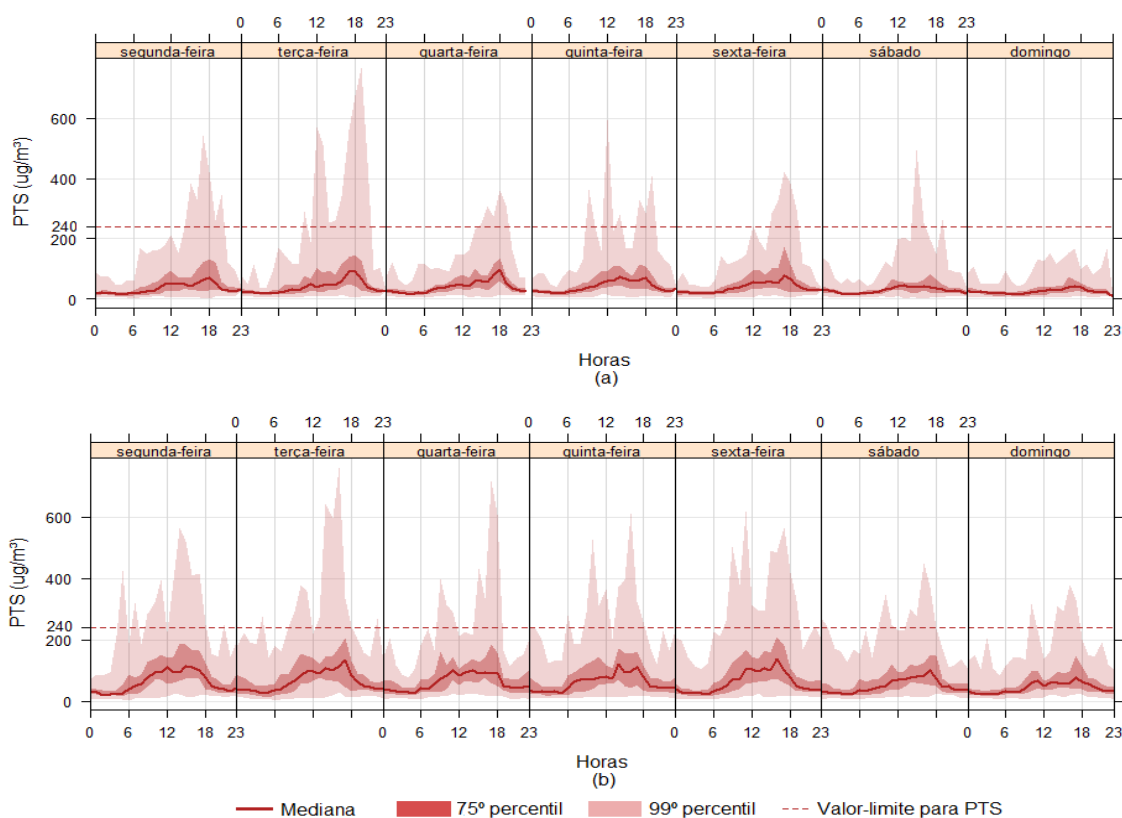
Tabela 5.2 – Quantidade acumulada em horas que a concentração de PTS ultrapassou o limite de 240 $\mu\text{g}/\text{m}^3$, suas respectivas proporções e teste Z.

Localidade	Nº de observações acima do limite (240 $\mu\text{g}/\text{m}^3$)	Nº total de observações	Proporção de observações acima do limite legal (%)	Teste Z	valor-p
Brumadinho (antes do rompimento)	77	6959	1,106%	-7,6759	$8,20 \times 10^{-15}$
Brumadinho (após o rompimento)	188	6329	2,97%		

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Ao observar a Tabela 5.2, percebe-se que a proporção de observações que ultrapassaram o limite legal após o rompimento foi superior, pois pelo teste Z o valor-p foi menor que o nível de significância $\alpha = 0,05$, rejeitando a hipótese nula de que as proporções são iguais entre si. Enquanto isso, na Figura 5.2 são apresentados os gráficos de concentração diária de PTS liberada a cada hora (mediana e percentis), antes e depois do rompimento da barragem.

Figura 5.2 – Gráficos de concentração diária de PTS, em $\mu\text{g}/\text{m}^3$, do município de Brumadinho antes (a) e após (b) o rompimento da barragem.

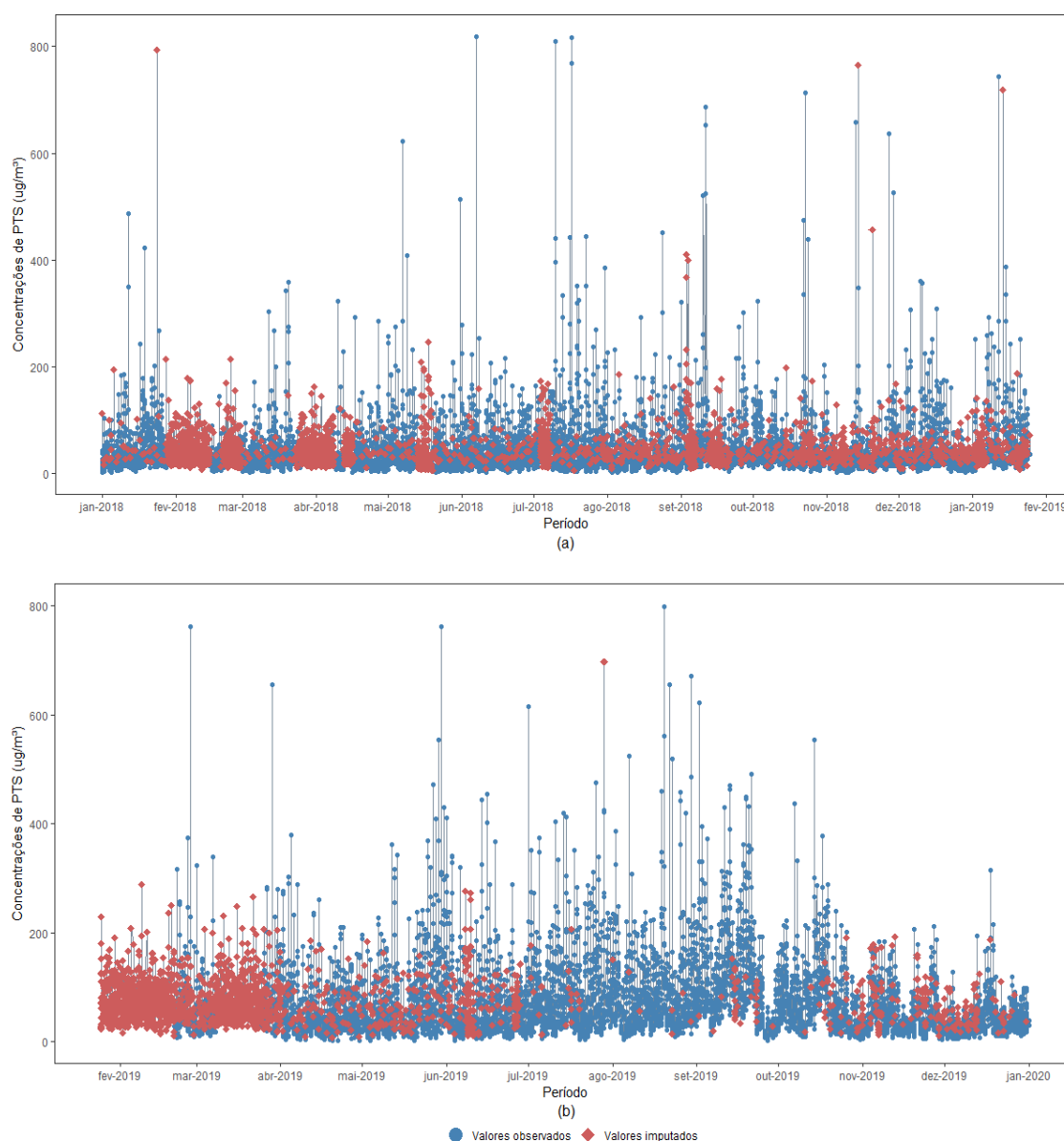


Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Na Figura 5.2 (a) observa-se que as maiores concentrações de PTS aconteceram entre as 12 e 18 horas, com excessão do domingo que em nenhum momento houve registro superior aos $240 \mu\text{g}/\text{m}^3$, determinado como limite legal. Já na Figura 5.2 (b), percebe-se que em vários horários a concentração de PTS ultrapassou o limite, além de ter ocorrido em todos os dias da semana, o que não havia acontecido antes do rompimento da barragem.

Em seguida, para que fosse possível obter os valores máximos das séries de concentração de PTS, em $\mu\text{g}/\text{m}^3$, foram realizadas imputações utilizando o método *predictive mean matching* (PMM). As Figuras 5.3 (a) e (b) representam graficamente as séries imputadas para as concentrações de PTS antes e após o rompimento da barragem, respectivamente.

Figura 5.3 – Gráficos das séries imputadas de concentrações de PTS, em $\mu\text{g}/\text{m}^3$, antes (a) e após (b) o rompimento da barragem.



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

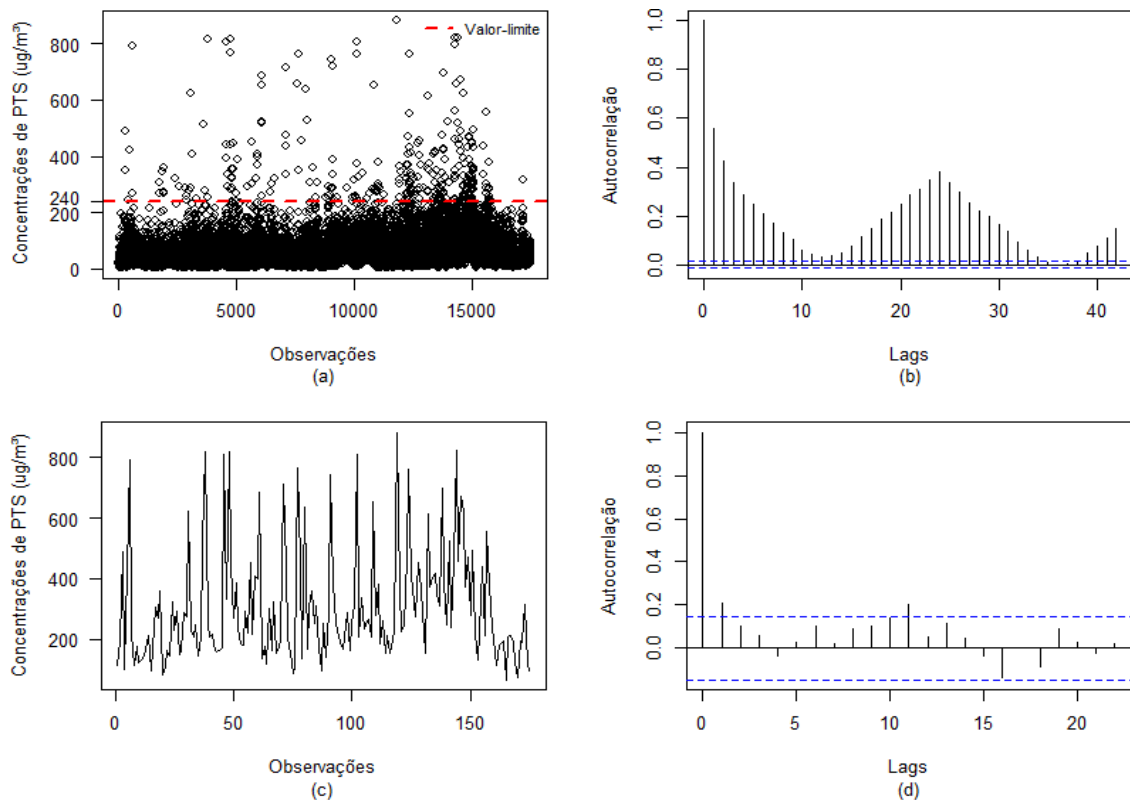
5.1.1 Método de blocos

Após feitas as imputações, foram obtidos os máximos das séries de PTS pelo método dos blocos. No entanto, para que ambas as séries, antes e após o rompimento, tivessem o mesmo tamanho de bloco foi necessário realizar a união delas (Figura 5.4 (a)).

Uma das principais exigências é de que a série de máximos seja independente. Sendo assim, observa-se na Figura 5.4 (b) que existe uma forte correlação, uma vez que existe correlação estatisticamente diferente de zero em vários *lags*. Inicialmente, dividiu-se a série em blocos menores (24 a 84 observações), mas os valores máximos obtidos apresentaram correlação.

Portanto, com o objetivo de eliminar a correlação, a partir da série completa (Figura 5.4 (a)) formaram-se blocos de 100 observações, que corresponde aproximadamente 4 dias. Em seguida, retirou-se a observação máxima em cada bloco, compondo assim a série de máximos de PTS, que pode ser vista na Figura 5.4 (c). Analisando a função de autocorrelação representada na Figura 5.4 (d), apenas em 2 *lags* existem correlações significativas, podendo ser considerados os dados não correlacionados, já que não excederam o limite de 10% do número total de *lags*.

Figura 5.4 – Série completa (observações horárias) da concentração de PTS, em $\mu\text{g}/\text{m}^3$ (a); Função de autocorrelação para a v.a. concentração de PTS (b); Série de 175 máximos de PTS (blocos de 100 observações) (c); Função de autocorrelação para os máximos de PTS (d).

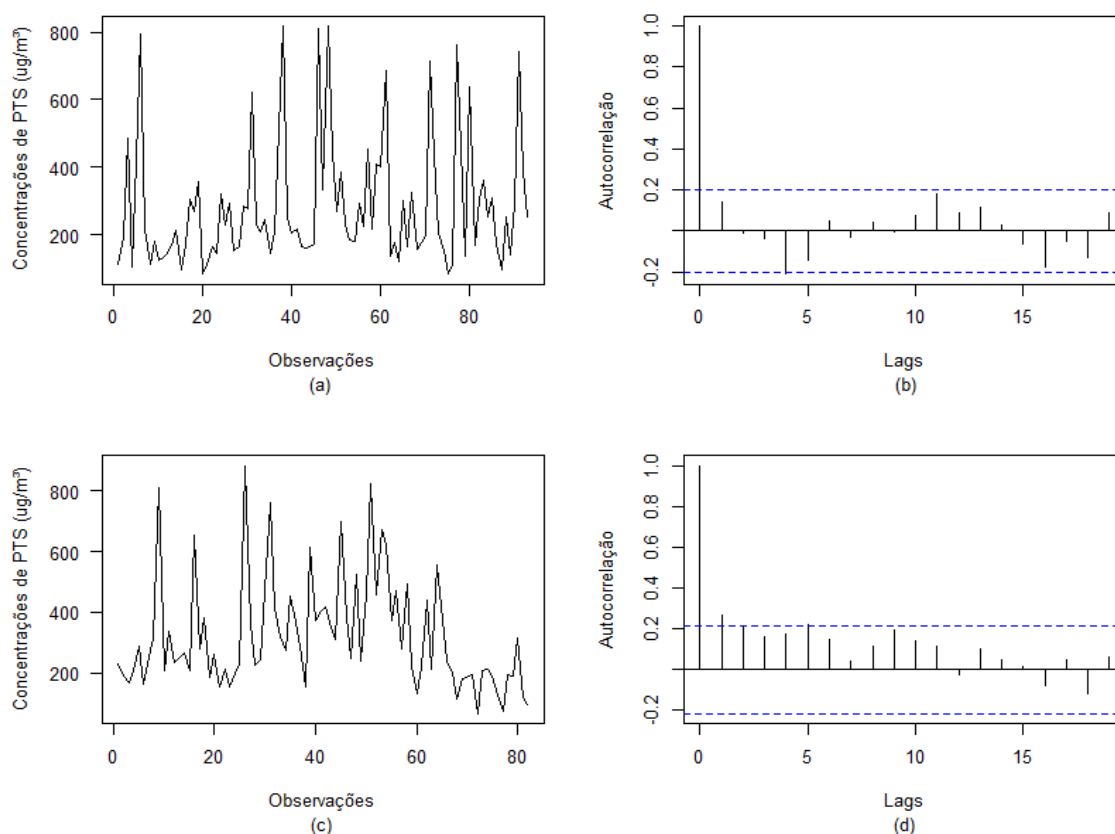


Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Uma vez obtidas as concentrações máximas de PTS, em $\mu\text{g}/\text{m}^3$, em seguida foi feita a separação dessas observações formando duas séries de máximos. Dessa forma, a primeira delas como sendo a série antes do rompimento (Figura 5.5 (a)), com um total de 93 observações e a outra sendo a série após o rompimento da barragem (Figura 5.5 (c)), possuindo 82 observações.

Após, foi verificado a partir das funções de autocorrelação, apresentadas nas Figuras 5.5 (b) e (d), que a maioria dos valores das funções de autocorrelação se encontra dentro do intervalo (linha tracejada em azul), indicando independência. Além disso, não apresentaram nenhum comportamento característico, tendência e/ou sazonalidade, indicando que as séries de máximos de PTS, em $\mu\text{g}/\text{m}^3$, antes e após o rompimento sejam estacionárias.

Figura 5.5 – Série de 93 máximos de PTS, em $\mu\text{g}/\text{m}^3$, antes do rompimento (blocos de 100 observações) (a); Função de autocorrelação para a v.a. 93 máximos de PTS (b); Série de 82 máximos de PTS, em $\mu\text{g}/\text{m}^3$, após o rompimento (blocos de 100 observações) (c); Função de autocorrelação para a v.a. 82 máximos de PTS (d).



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Para confirmar este fato, foram realizados o teste de Cox-Stuart e o teste de Fisher, com nível de confiança de 95%. Pelo teste de Cox-Stuart, verificou-se a presença ou não da componente tendência e quanto ao teste de Fisher, este foi utilizado para analisar a presença de sazonalidade. Na Tabela 5.3 são apresentados os resultados dos testes para cada período.

Tabela 5.3 – Resultados do teste de Cox-Stuart e teste de Fisher.

Série	Teste de Cox-Stuart		Teste de Fisher	
	Estatística	valor- <i>p</i>	Estatística	valor- <i>p</i>
Antes do rompimento da barragem	0,5388	0,5900	0,1210	0,1387
Depois do rompimento da barragem	1,6577	0,0974	0,1546	0,0596

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Com base nos resultados vistos na Tabela 5.3, pode-se concluir que as séries de concentrações máximas de PTS não apresentaram as componentes tendência e sazonal, pois os valores-*p* dos testes foram maiores que o nível de significância de 5%. Sendo assim, tanto a série de máximos de PTS com 93 observações (antes do rompimento da barragem), como a com 82 observações (depois do rompimento da barragem) satisfazem aos pressupostos de independência e estacionaridade.

Para a estimação dos parâmetros referente aos dados representados nas Figuras 5.5 (a) e (c) foi utilizado o método da máxima verossimilhança, por meio do procedimento iterativo de Nelder-Mead. As estimativas e seus respectivos erros padrões, além do intervalo de confiança (IC) de 95% são apresentados na Tabela 5.4.

Tabela 5.4 – Estimativas dos parâmetros, seus respectivos erros padrões e IC_{95%} para a distribuição GEV e Gumbel ajustada a série de máximos de PTS, antes e após o rompimento da barragem.

Série	Modelo	Parâmetro	Estimativa	Erro padrão	IC _{95%}
Antes do rompimento da barragem	GEV	μ	190,4302	11,4737	[167,9421; 212,9182]
		σ	94,3136	10,3073	[74,1117; 114,5155]
		ξ	0,3828	0,1076	[0,1719; 0,5936]
	Gumbel	μ	208,9811	12,6480	[184,1915; 233,7707]
		σ	117,2075	10,4251	[96,7747; 137,6403]
Depois do rompimento da barragem	GEV	μ	232,3643	14,5420	[203,8621; 260,8666]
		σ	115,9116	11,4040	[93,5605; 138,2627]
		ξ	0,1692	0,0920	[-0,0111; 0,3495]
	Gumbel	μ	242,8742	14,4772	[214,4993; 271,2490]
		σ	125,2551	11,3824	[102,9460; 147,5641]

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Uma vez obtidas as estimativas dos parâmetros, foi então realizado o teste da razão de verossimilhança (TRV) para a escolha dos melhores modelos, ou seja, avaliar se o parâmetro

ξ era igual ou diferente de zero. Os resultados do TRV, para cada período (antes e após o rompimento), são apresentados na Tabela 5.5.

No entanto, ao observar o intervalo de confiança de 95% para o parâmetro ξ (Tabela 5.4), percebe-se que o zero não pertence ao intervalo para a série de máximos de PTS antes do rompimento e, portanto, $\xi \neq 0$. Já a série após o rompimento, é possível observar que o zero pertence ao intervalo, indicando um possível melhor ajuste da distribuição Gumbel aos dados.

Tabela 5.5 – Resultados do TRV.

Série	Teste da razão de verossimilhança	
	Estatística	valor- <i>p</i>
Antes do rompimento da barragem	18,1564	$1,51 \times 10^{-5}$
Depois do rompimento da barragem	3,7748	0,0481

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

De acordo com a Tabela 5.5, observou-se para ambas as séries que o valor-*p* $< 0,05$, então se rejeitou a hipótese H_0 , ou seja, o parâmetro ξ é diferente de zero. Portanto, o modelo indicado para as séries de concentrações máximas de PTS foi a distribuição GEV.

Desta forma, o modelo para a série de máximos de PTS, em $\mu\text{g}/\text{m}^3$, antes do rompimento pode ser escrito como:

$$F_{GEV}(x) = \exp \left\{ - \left[1 + 0,3828 \left(\frac{x - 190,4302}{94,3136} \right) \right]^{-\frac{1}{0,3828}} \right\}, \quad (5.1)$$

com $x \in [0; +\infty)$. Para a série de máximos de PTS, em $\mu\text{g}/\text{m}^3$, depois do rompimento, o modelo pode ser expresso pela seguinte equação:

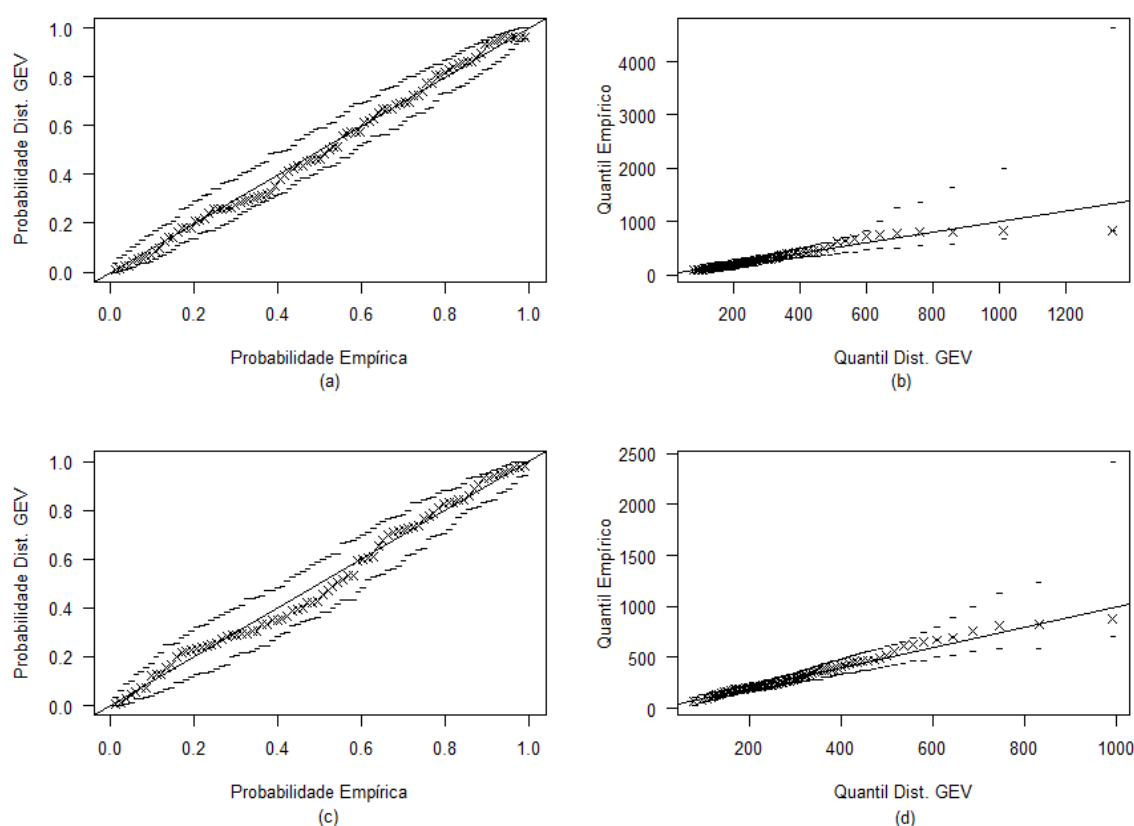
$$F_{GEV}(x) = \exp \left\{ - \left[1 + 0,1692 \left(\frac{x - 232,3643}{115,9116} \right) \right]^{-\frac{1}{0,1692}} \right\}, \quad (5.2)$$

com $x \in [0; +\infty)$.

Após estimados os parâmetros, foi então analisado o desempenho de cada modelo por meio dos gráficos (Figura 5.6) de probabilidade-probabilidade (*pp-plot*) e quantil-quantil (*qq-plot*).

Pelas Figuras 5.6 (a) e (b), referente ao ajuste da distribuição GEV às observações máximas de PTS antes do rompimento da barragem, como para os gráficos representados nas Figuras 5.6 (c) e (d), que são relacionados ao ajuste da GEV as observações máximas de PTS após o rompimento, se tem indícios de um bom ajuste dos modelos aos dados observados. Em ambos os casos, os pontos estão próximos da reta de referência, além de manterem-se dentro do envelope de confiança.

Figura 5.6 – Probabilidades ((a) e (c)) e quantis ((b) e (d)) para a distribuição GEV ajustada as séries de PTS, em $\mu\text{g}/\text{m}^3$, antes e após o rompimento da barragem, respectivamente.



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Porém, as conclusões a respeito da qualidade do ajuste não podem ser baseadas somente em uma análise visual. Para isso, foi utilizado o teste de aderência de Anderson-Darling, no qual a estatística do teste é obtida por meio das distribuições teórica e empírica.

A estatística do teste de Anderson Darling para a distribuição GEV ajustada a série de PTS antes do rompimento da barragem foi: $A = 1,8051$ e valor- $p = 0,7184$, então não se rejeita a hipótese H_0 , ou seja, $F(x) = \hat{F}(x)$. Já para a série de PTS após o rompimento da barragem, a estatística do teste obtida foi: $A = 1,3584$ e valor- $p = 0,8854$, logo também a série de máximos segue a distribuição teórica.

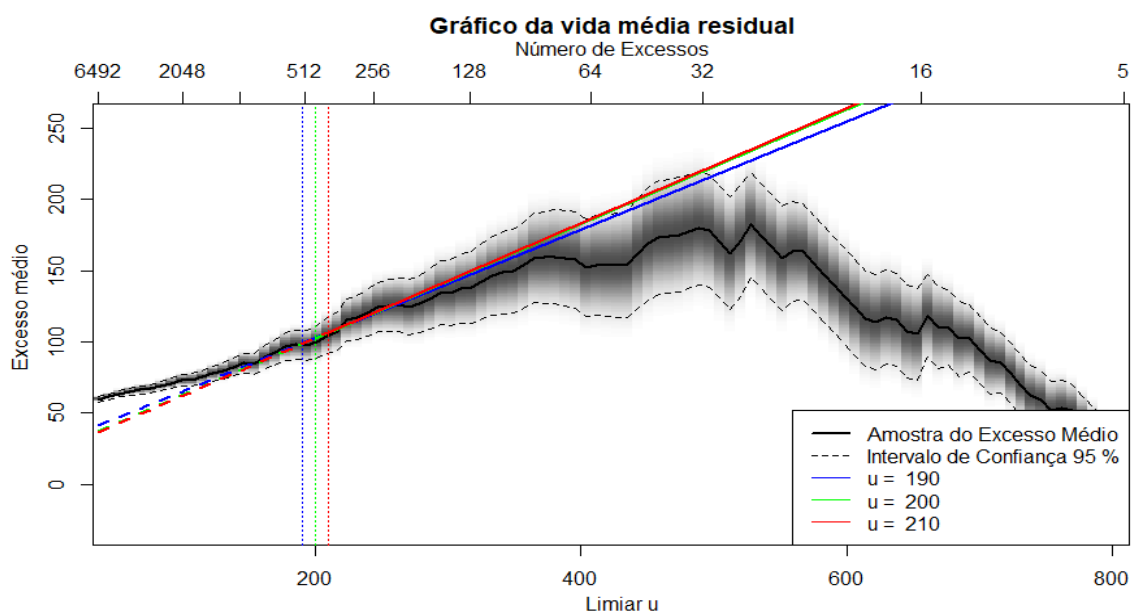
Uma vez que foram ajustados alguns modelos e analisado a sua adequabilidade aos dados de PTS, obteve-se a probabilidade de ter concentrações acima do limite de $240 \mu\text{g}/\text{m}^3$. Ao aplicar as estimativas dos parâmetros da distribuição GEV (Tabela 5.4) na expressão 3.49, as probabilidades encontradas antes e após o rompimento da barragem, foram de: 46,18% e 60,80%, respectivamente. Sendo assim, o poluente PTS apresentou maior probabilidade de superar o nível estabelecido pelo CONAMA após o rompimento da barragem.

5.1.2 Método dos excessos acima de um limiar u

Ao se fazer uso do método dos excessos acima de um limiar u , optou-se em utilizar os dados originais, ou seja, sem a necessidade de ser feito o preenchimento dos valores ausentes por meio de uma técnica de imputação. Essa escolha se deu pelo fato que a série de particulado não precisa estar completa para que sejam obtidos os valores máximos. Além disso, a partir de um estudo preliminar, constatou-se que a série original obteve melhores resultados.

Desta maneira, a primeira etapa é a escolha do limiar u , sendo considerada em alguns casos uma tarefa difícil. Para isso, com o objetivo de facilitar essa escolha foi utilizado o gráfico de vida média residual (ou, gráfico da média de excessos). Na Figura 5.7 está representado o mesmo referente as concentrações de PTS, em $\mu\text{g}/\text{m}^3$, para antes do rompimento da barragem.

Figura 5.7 – Gráfico de vida média residual para os dados de PTS, em $\mu\text{g}/\text{m}^3$.

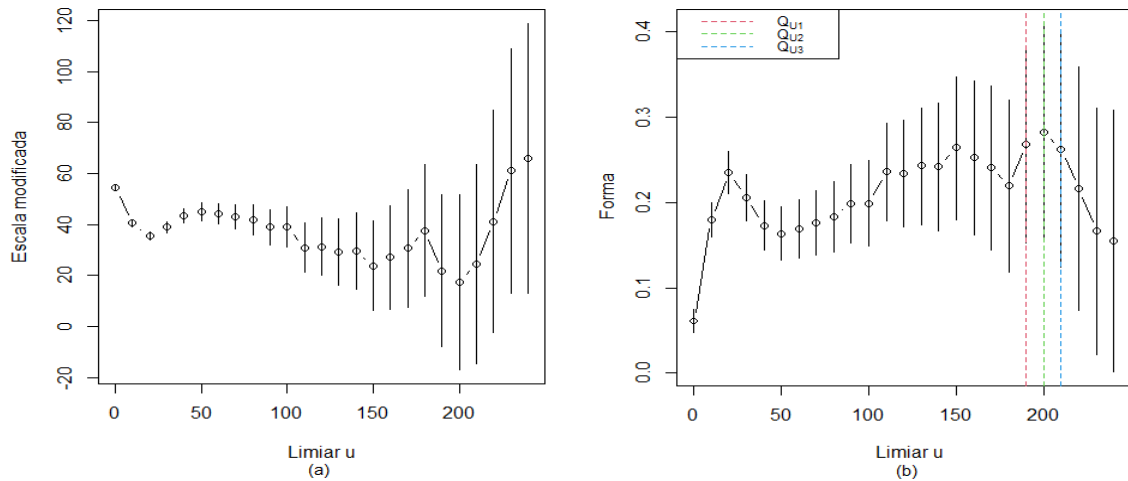


Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Para tornar essa escolha mais simples, também foi utilizada uma técnica complementar, em que se ajusta a distribuição GP para uma gama de limiares e procura a estabilidade das

estimativas dos parâmetros. Na Figura 5.8 são apresentados os gráficos para os parâmetros escala ($\hat{\sigma}_u$) e forma ($\hat{\xi}$) em função de 240 diferentes limiares.

Figura 5.8 – Estimação dos parâmetros escala (a) e forma (b) para 240 limiares diferentes para os dados horários de PTS.

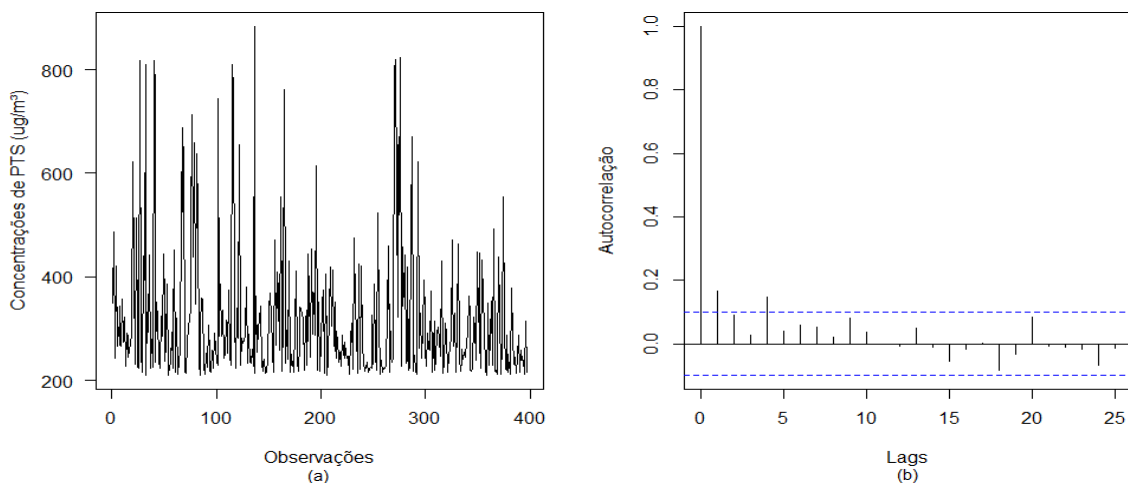


Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Pela Figura 5.8 (a) observa-se que a partir de $u = 190$, o parâmetro σ_u possui comportamento linear e o parâmetro ξ , Figura 5.8 (b), é aproximadamente constante. Sendo assim, optou-se em avaliar qual dentre esses três limiares ($u_1 = 190$, $u_2 = 200$ e $u_3 = 210$) deve ser utilizado para obter as concentrações máximas de PTS.

No entanto, se alcançou a independência dos dados somente com $u = 210$, conforme pode se observar pelo gráfico da função de autocorrelação (Figura 5.9 (b)). A Figura 5.9 (a) representa graficamente a série de máximos de PTS obtida a partir do limiar.

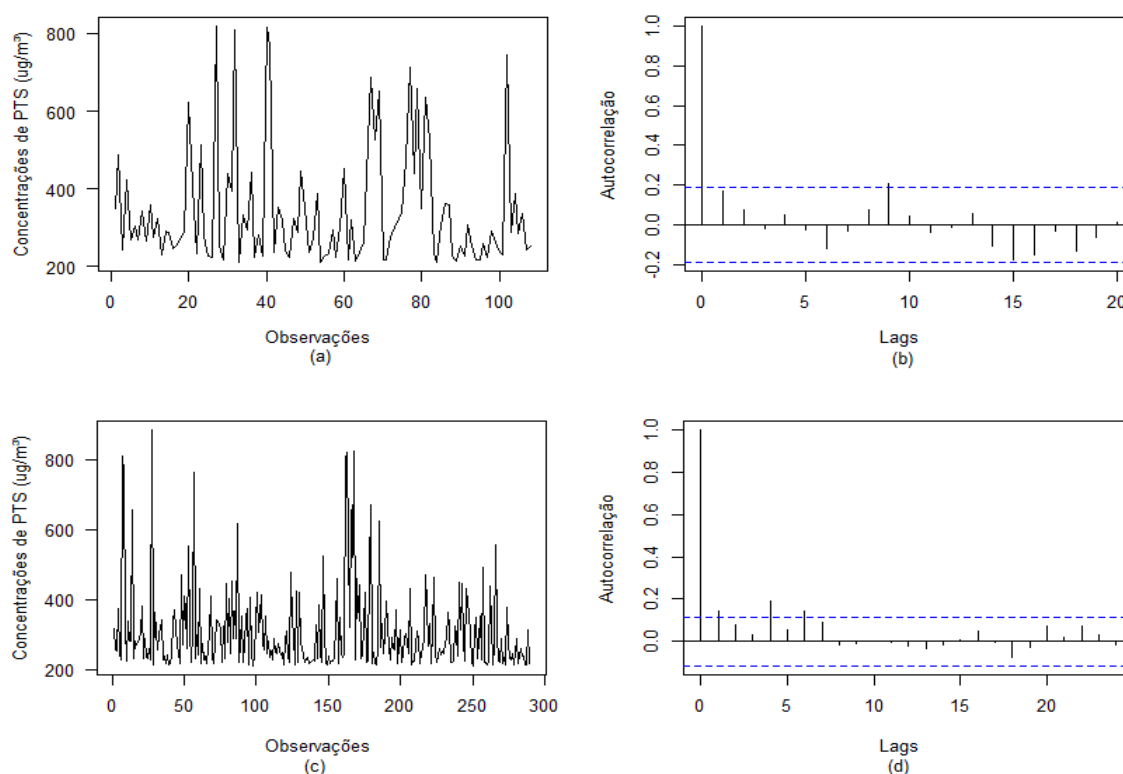
Figura 5.9 – Série de 397 máximos de PTS, em $\mu\text{g}/\text{m}^3$, com limiar $u = 210$ (a); Função de autocorrelação para a v.a. 397 máximos de PTS (b).



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Obtidas as concentrações máximas de PTS, em seguida foram separadas de acordo com cada período. Logo, a série antes do rompimento apresentou 108 máximos (Figura 5.10 (a)) e a série após o rompimento com um total de 289 máximos (Figura 5.10 (c)). Os gráficos das funções de autocorrelação (Figuras 5.10 (b) e (d)), indicaram que os dados de ambas as séries são não correlacionados.

Figura 5.10 – Série de 108 máximos de PTS, em $\mu\text{g}/\text{m}^3$, antes do rompimento (com limiar $u = 210$) (a); Função de autocorrelação para a v.a. 108 máximos de PTS (b); Série de 289 máximos de PTS, em $\mu\text{g}/\text{m}^3$, após o rompimento (com limiar $u = 210$) (c); Função de autocorrelação para a v.a. 289 máximos de PTS (d).



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Após verificado o pressuposto de independência, na Tabela 5.6 encontram-se os resultados dos testes de Cox-Stuart e de Fisher, para analisar a estacionariedade.

Tabela 5.6 – Resultados do teste de Cox-Stuart e teste de Fisher.

Série	Teste de Cox-Stuart		Teste de Fisher	
	Estatística	valor- p	Estatística	valor- p
Antes do rompimento da barragem	0,6667	0,5050	0,0823	0,5682
Depois do rompimento da barragem	0,9848	0,3247	0,0358	0,7852

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

A partir dos resultados da Tabela 5.6, tem-se que as séries de concentrações máximas de PTS não apresentaram as componentes tendência e sazonal, com o nível de confiabilidade de 95%. Portanto, foram satisfeitas as pressuposições de independência e estacionaridade para ambas as séries de máximos do poluente PTS.

Sendo assim, obteve-se as estimativas dos parâmetros referente as observações vistas nas Figuras 5.10 (a) e (c) utilizando o método da Máxima Verossimilhança. Na Tabela 5.7 encontram-se as estimativas e seus respectivos erros padrões e o intervalo de confiança de 95%.

Tabela 5.7 – Estimativas dos parâmetros, seus respectivos erros padrões e IC_{95%} para a distribuição GP e Exponencial ajustada a série de máximos de PTS, antes e após o rompimento da barragem.

Série	Modelo	Parâmetro	Estimativa	Erro padrão	IC _{95%}
Antes do rompimento da barragem	GPD	σ_u	102,3238	15,4703	[72,0025; 132,6450]
		ξ	0,2344	0,1188	[0,0016; 0,4672]
	Exponencial	σ_u	131,7222	12,6749	[106,8799; 156,5646]
Depois do rompimento da barragem	GPD	σ_u	73,4787	6,8223	[60,1072; 86,8501]
		ξ	0,2457	0,0733	[0,1021; 0,3893]
	Exponencial	σ_u	96,3218	5,6659	[85,2168; 107,4268]

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Após obter as estimativas dos parâmetros para cada modelo, foi verificado qual dentre eles melhor descreve os valores máximos de PTS. Esta escolha foi baseada no teste da razão de verossimilhança e os resultados são apresentados na Tabela 5.8, para ambas as séries.

Tabela 5.8 – Resultados do TRV.

Série	Teste da razão de verossimilhança	
	Estatística	valor- <i>p</i>
Antes do rompimento da barragem	3,9841	0,0459
Depois do rompimento da barragem	14,2264	0,0002

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Na Tabela 5.8, analisando os valores-*p*, observou-se que são menores que $\alpha = 0,05$, indicando que $\xi \neq 0$, e portanto, a distribuição GP foi a escolhida. Enfim, o modelo para a série de máximos de PTS, em $\mu\text{g}/\text{m}^3$, antes do rompimento pode ser expressa pela equação:

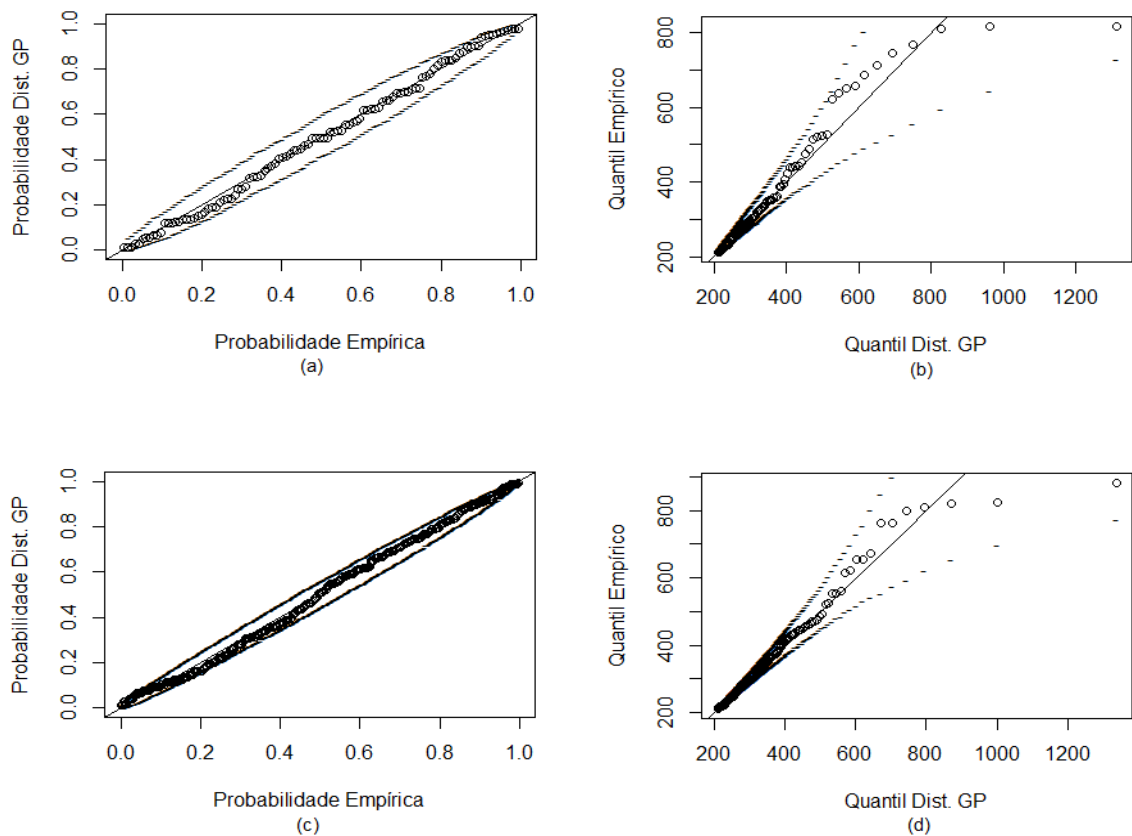
$$G_{GP}(y) = 1 - \left[1 + 0,2344 \left(\frac{y}{102,3238} \right) \right]^{-\frac{1}{0,2344}}, \quad (5.3)$$

com $y \in [0; +\infty)$ e $u = 210$. Para a série de máximos de PTS, em $\mu\text{g}/\text{m}^3$, depois do rompimento, o modelo é dado a seguir:

$$G_{GP}(y) = 1 - \left[1 + 0,2457 \left(\frac{y}{73,4787} \right) \right]^{-\frac{1}{0,2457}}, \quad (5.4)$$

com $y \in [0; +\infty)$ e $u = 210$. Em posse das estimativas dos parâmetros, a partir da Figura 5.11 foi analisado o desempenho de cada modelo utilizando os gráficos *pp-plot* e *qq-plot*.

Figura 5.11 – Probabilidades ((a) e (c)) e quantis ((b) e (d)) para a distribuição GP ajustada as séries de PTS, em $\mu\text{g}/\text{m}^3$, antes e após o rompimento da barragem, respectivamente.



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Por meio dos gráficos *pp-plot* (Figuras 5.11 (a) e (c)) observa-se que os pontos estão próximos da reta. Para os gráficos *qq-plot* (Figuras 5.11 (b) e (d)) alguns pontos apresentam-se um pouco distantes da reta, mas ainda se encontram dentro do intervalo de confiança. No entanto, mesmo havendo indícios que os modelos ajustados sejam adequados aos dados, é necessário a realização do teste de aderência de Anderson Darling, para a validação dos modelos.

A estatística do teste de Anderson Darling para a distribuição GP ajustada a série de PTS antes do rompimento da barragem foi: $A = 1,4914$ e $\text{valor-}p = 0,8606$, então não se rejeita a

hipótese H_0 , ou seja, $F(x) = \hat{F}(x)$. Para a série de PTS após o rompimento da barragem, a estatística do teste obtida foi: $A = 2,4258$ e valor- $p = 0,6173$, logo também a série de máximos segue a distribuição teórica.

Por fim, aplicando-se as estimativas do parâmetros (Tabela 5.7) da distribuição GP na equação 3.51, tem-se que a probabilidade de ocorrer uma concentração máxima superior ao limite de $240 \mu\text{g}/\text{m}^3$ antes e depois do rompimento foi de $67,77\%$ e $75,31\%$, respectivamente.

5.1.3 Cálculo do VaR

O valor em risco (VaR) para concentração de PTS, por meio das duas abordagens da teoria de valores extremos, foi obtido a partir das equações 3.69 e 3.74, apresentando os seguintes resultados na Tabela 5.9.

Tabela 5.9 – Níveis de risco e valores ao risco para v.a. máximos de PTS, em $\mu\text{g}/\text{m}^3$.

P	Antes do rompimento ($\mu\text{g}/\text{m}^3$)		Depois do rompimento ($\mu\text{g}/\text{m}^3$)	
	VaR pela GEV	VaR pela GPD	VaR pela GEV	VaR pela GPD
0,2	6,7592	15,8450	7,2447	58,9244
0,3	36,3277	20,5090	47,5922	63,8604
0,4	62,2564	39,1925	78,4919	69,7637
0,5	89,5948	51,4122	100,5531	77,0406
0,6	100,7356	60,3392	118,2729	86,4020
0,7	123,0953	67,2944	133,6941	99,2536
0,8	135,0737	72,9474	148,0926	118,9808
0,9	152,3981	77,6820	162,6864	157,6080
0,999	538,7619	603,5951	558,6457	675,6578

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

A partir da Tabela 5.9, observa-se que ao aumentar o nível de risco (P), considerando as duas abordagens de valores extremos, têm-se maiores valores em risco (VaR) para o período pós-rompimento. Para o poluente PTS antes do rompimento, ao utilizar a distribuição GEV, a probabilidade de se atingir o nível crítico de $240 \mu\text{g}/\text{m}^3$ foi de $99,38\%$. Já pela distribuição GP, a probabilidade foi de exatamente $98,83\%$ que se alcance o limite legal.

Para a série referente ao período pós-rompimento, por meio da GEV ao supor um risco de $99,9\%$ tem-se que, ultrapassado, o limite legal. Sendo mais preciso, a uma probabilidade de $99,07\%$ é o risco de se atingir o limite legal para este poluente. Ao utilizar a GPD, novamente o valor crítico está entre $0,9$ e $0,999$, ou seja, a uma probabilidade de $96,90\%$, atingindo, assim, o limite legal de concentrações de partículas totais em suspensão.

5.2 Série de material particulado (MP₁₀)

Para a concentração de MP₁₀, em $\mu\text{g}/\text{m}^3$, tem-se na Tabela 5.10 algumas medidas descritivas a partir da série incompleta em cada situação, ou seja, antes (período de 01/01/2018 a 24/01/2019) e após (período de 25/01/2019 a 31/12/2019) o rompimento da barragem.

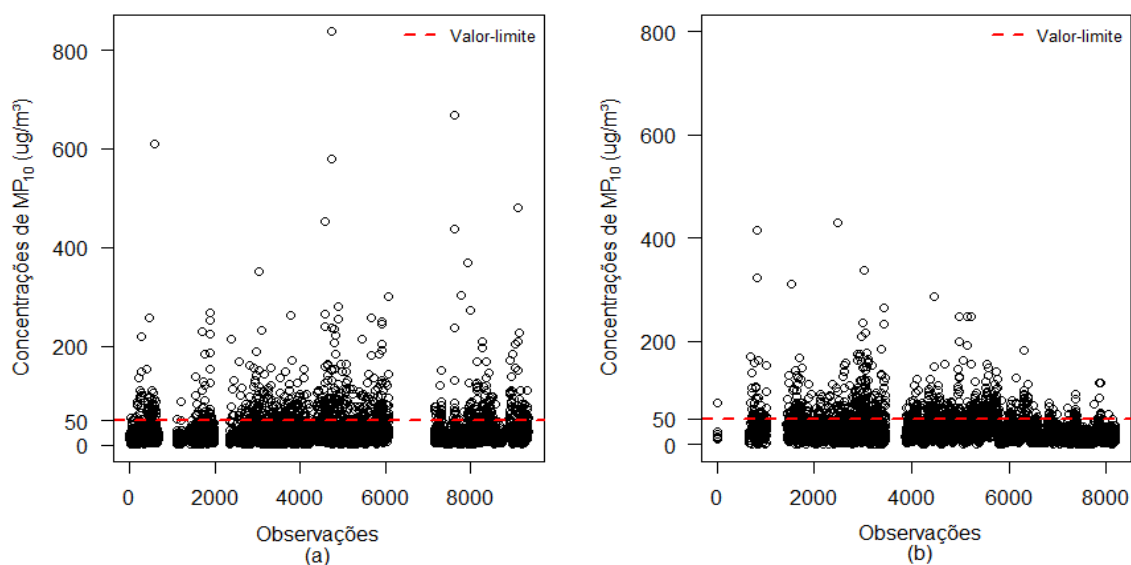
Tabela 5.10 – Estatísticas descritivas para a concentração de MP₁₀ ($\mu\text{g}/\text{m}^3$).

Localidade	Nº de observações faltantes	Proporção de valores faltantes (%)	Média	Mediana	Desvio padrão	Máximo
Brumadinho (antes do rompimento)	3170	33,95	27,84	19,00	34,97	837,00
Brumadinho (após o rompimento)	2180	26,64	28,15	20,84	26,89	431,00

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Diferente do poluente PTS, a partir da Tabela 5.10, observa-se que a média da concentração de MP₁₀ não foi estatisticamente superior nem inferior após o rompimento da barragem em Brumadinho, pois pelo teste *t-Student* ($t = -0,5365$ e $\text{valor-}p = 0,5916$), com nível de confiança de 95%, têm-se que as médias das concentrações são iguais. No entanto, a maior concentração registrada deste poluente foi anteriormente ao rompimento. Graficamente, a concentração de MP₁₀ para antes e após o rompimento da barragem é representada na Figura 5.12.

Figura 5.12 – Gráficos de concentração de MP₁₀, em $\mu\text{g}/\text{m}^3$, do município de Brumadinho antes do rompimento (a) e após o rompimento (b) da barragem.



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Pela Figura 5.12 (a), ou seja, antes do rompimento da barragem percebe-se que em alguns horários a concentração de MP₁₀ liberada no ar foi superior a $400 \mu\text{g}/\text{m}^3$, chegando a

ultrapassar os $800 \mu\text{g}/\text{m}^3$ uma única vez. A respeito da quantidade acumulada por hora que o poluente ultrapassou o limite legal de $50 \mu\text{g}/\text{m}^3$ e as respectivas proporções, considerando os valores disponíveis de cada série (antes e após o rompimento), são apresentadas na Tabela 5.11.

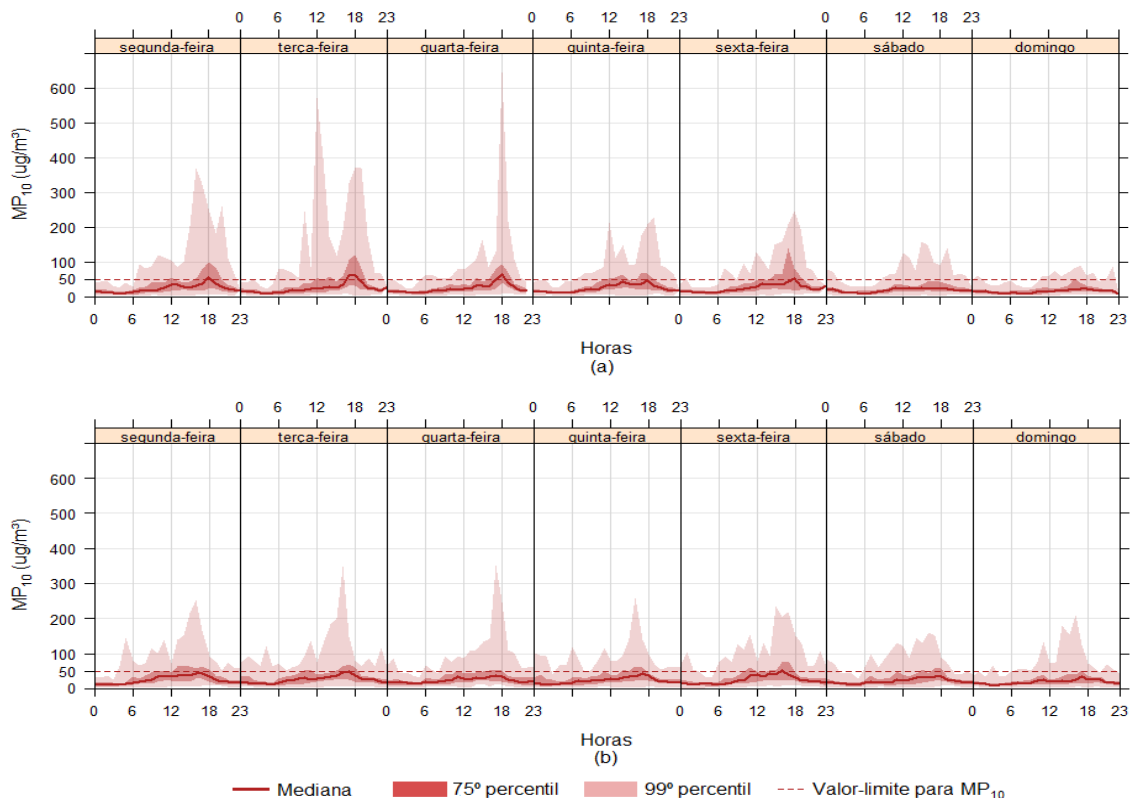
Tabela 5.11 – Quantidade acumulada em horas que a concentração de MP_{10} ultrapassou o limite de $50 \mu\text{g}/\text{m}^3$, suas respectivas proporções e teste Z.

Localidade	Nº de observações acima do limite ($240 \mu\text{g}/\text{m}^3$)	Nº total de observações	Proporção de observações acima do limite legal (%)	Teste Z	valor-p
Brumadinho (antes do rompimento)	743	6166	12,05%	-1,8688	0,0308
Brumadinho (após o rompimento)	791	6004	13,17%		

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Pelo resultado do teste Z, visto na Tabela 5.11, verifica-se que a proporção de observações acima do limite legal após o rompimento da barragem foi considerada superior ao período antes do rompimento, pois o valor-p foi menor que o nível de significância $\alpha = 0,05$, rejeitando-se a hipótese H_0 , indicando que a proporção posterior é maior que a anterior. Em seguida, foi analisado o comportamento do poluente para cada dia da semana e em cada uma das situações, representado na Figura 5.13.

Figura 5.13 – Gráficos de concentração diária de MP_{10} , em $\mu\text{g}/\text{m}^3$, do município de Brumadinho antes (a) e após (b) o rompimento da barragem.

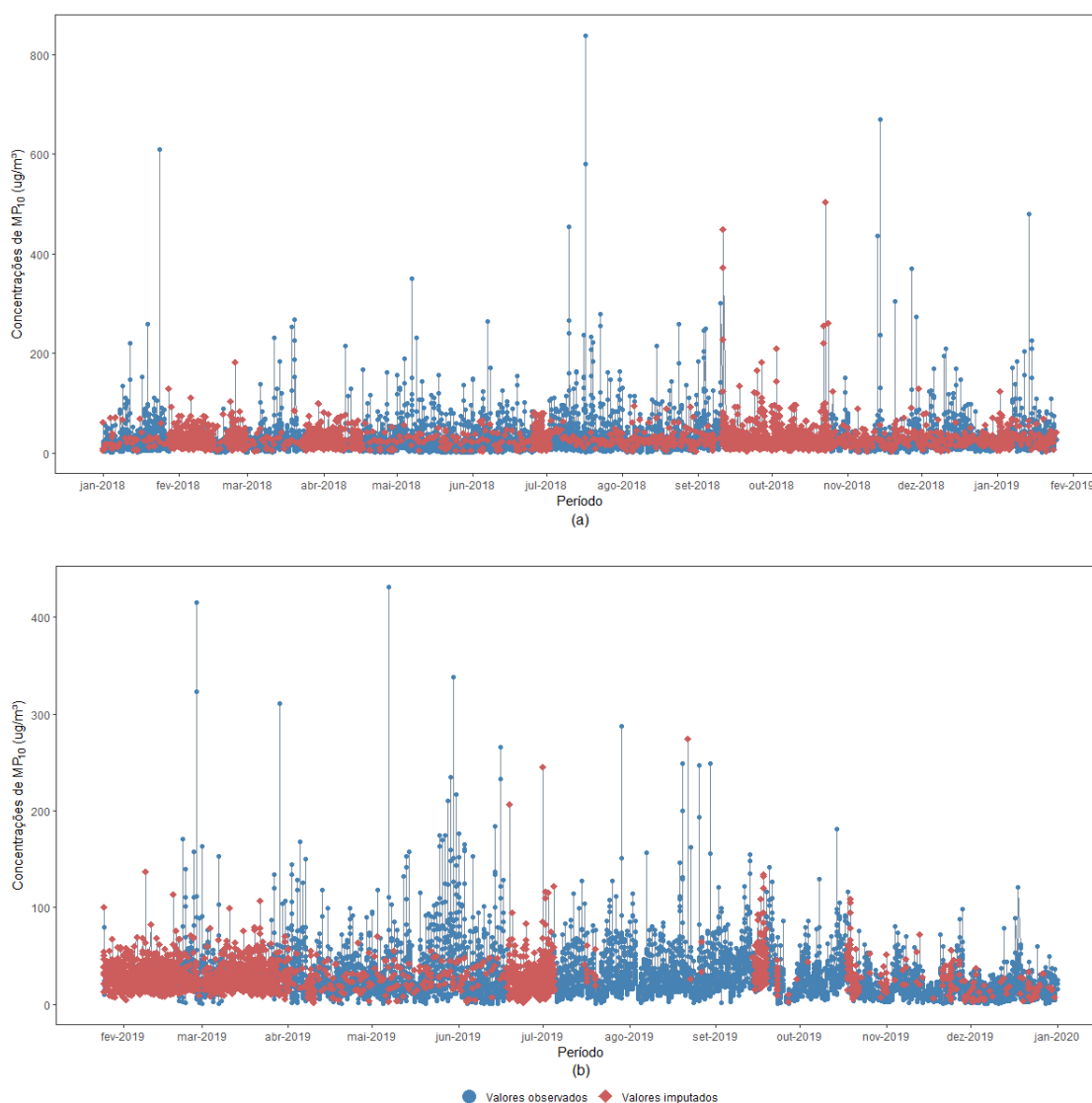


Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Por meio dos gráficos da Figura 5.13, observa-se que tanto durante a semana, como nos finais de semana houve registros de que a concentração de MP_{10} ultrapassou o limite de $50 \mu g/m^3$, em pelo menos alguma hora do dia. Além disso, antes do rompimento da barragem foram identificados maiores concentrações do poluente, principalmente nos primeiros dias da semana. No entanto, em vários horários do final de semana, após o rompimento, houve aumento da concentração de MP_{10} , particularmente no domingo.

Após ser feita uma análise exploratória dos dados de partículas inaláveis, foram então realizadas as imputações utilizando o método PMM. As Figuras 5.14 (a) e (b) representam graficamente as séries imputadas para as concentrações de MP_{10} antes e após o rompimento da barragem, respectivamente.

Figura 5.14 – Gráficos das séries imputadas de concentrações de MP_{10} , em $\mu g/m^3$, antes (a) e após (b) a ruptura da barragem.



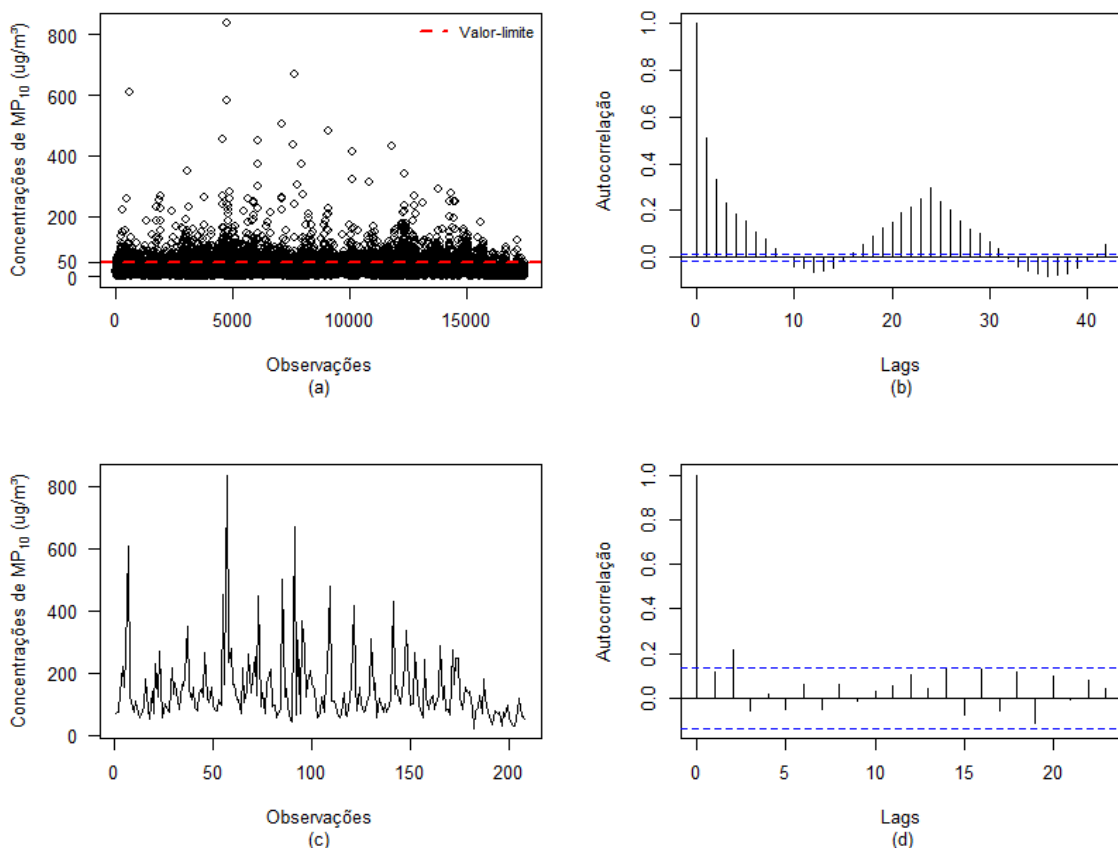
Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

5.2.1 Método de blocos

Após realizadas as imputações, foram obtidos os máximos da série de MP_{10} pelo método dos blocos. Para isso, primeiramente foi feita a junção das séries imputadas antes e após o rompimento, representadas nas Figuras 5.14 (a) e (b), respectivamente. Pela Figura 5.15 (b) observa-se que existe uma forte correlação entre as observações, uma vez que há poucos valores da função de autocorrelação dentro do intervalo de confiança (linha tracejada em azul).

Logo, na busca de uma série independente foram formados blocos de 84 observações a partir da série completa (Figura 5.15 (a)), pois adotando blocos com 24, 48 e 72 observações não foi possível obter uma série de dados não correlacionada. Então, retirou-se a observação máxima de cada bloco contendo 84 observações, obtendo-se a série de máximos de MP_{10} , representada na Figura 5.15 (c). Analisando a função de autocorrelação, Figura 5.15 (d), observa-se que apenas um *lag* da função de autocorrelação se encontra fora do intervalo, podendo ser considerados os dados não correlacionados, ou seja, independentes.

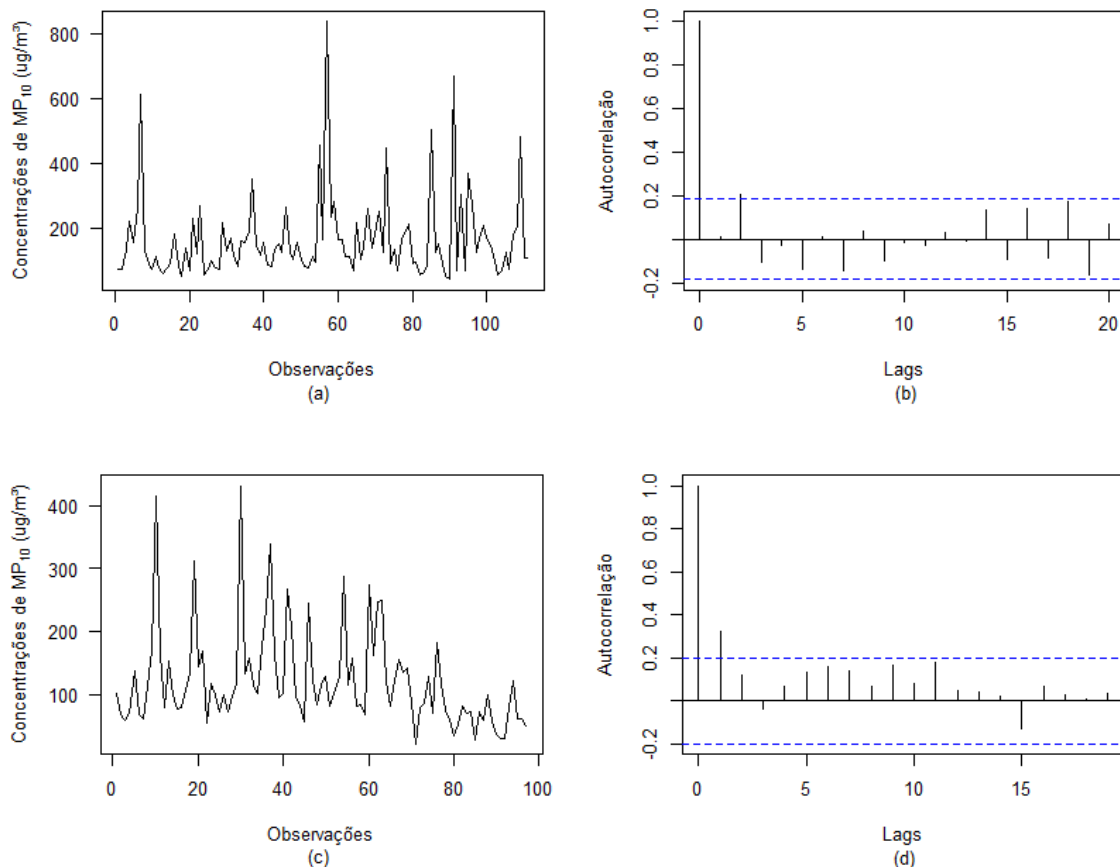
Figura 5.15 – Série completa (observações horárias) da concentração de MP_{10} , em $\mu g/m^3$ (a); Função de autocorrelação para a v.a. concentração de MP_{10} (b); Série de 208 máximos de MP_{10} (blocos de 84 observações) (c); Função de autocorrelação para os máximos de MP_{10} (d).



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Em seguida, foi feito o desdobramento da série de máximos de MP_{10} , apresentada na Figura 5.15 (c). Desse modo, a sequência antes do rompimento ficou composta de 111 observações e após o rompimento da barragem com 97 valores máximos, conforme apresentadas nas Figuras 5.16 (a) e (c), respectivamente. Pelas funções de autocorrelação (Figuras 5.16 (b) e (d)), foi verificado que ambas as séries apresentaram apenas um *lag* da função de autocorrelação fora do intervalo e , portanto, os dados são independentes.

Figura 5.16 – Série de 111 máximos de MP_{10} , em $\mu g/m^3$, antes do rompimento (blocos de 84 observações) (a); Função de autocorrelação para a v.a. 111 máximos de MP_{10} (b); Série de 97 máximos de MP_{10} , em $\mu g/m^3$, após o rompimento (blocos de 84 observações) (c); Função de autocorrelação para a v.a. 97 máximos de MP_{10} (d).



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Além disso, na Figura 5.16 (c) pode-se observar que para o período pós-rompimento a concentração máxima de MP_{10} diminuiu ao longo da série. Após analisar a independência das séries, foi verificada a estacionariedade, ou seja, se as séries de concentrações máximas de MP_{10} possuem tendência e/ou sazonalidade. Para isso, foram realizados os testes de Cox-Stuart e de Fisher, sendo que os resultados obtidos da estatística e seu respectivo valor- p podem ser vistos na Tabela 5.12, respectivamente.

Tabela 5.12 – Resultados do teste de Cox-Stuart e teste de Fisher.

Série	Teste de Cox-Stuart		Teste de Fisher	
	Estatística	valor- <i>p</i>	Estatística	valor- <i>p</i>
Antes do rompimento da barragem	1,1508	0,2498	0,1015	0,1701
Depois do rompimento da barragem	1,5123	0,1305	0,0815	0,8853

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

A partir dos resultados vistos na Tabela 5.12, observa-se que o valor-*p* > 0,05, em ambas as séries, indicando a não presença das componentes tendência e sazonal. Visto que as séries de máximos de MP₁₀, tanto para o conjunto de 111 observações (antes do rompimento) como para o conjunto de 97 observações (depois do rompimento), satisfazem aos pressupostos de independência e estacionaridade foi então possível aplicar a teoria de valores extremos.

Para a estimação dos parâmetros referente aos dados expressos nas Figuras 5.16 (a) e (c) foi utilizado o método da máxima verossimilhança, por meio do procedimento iterativo de Nelder-Mead, já que o sistema de equações de verossimilhança não possui solução analítica. As estimativas e seus respectivos erros padrões, além do intervalo de confiança (IC) de 95% são apresentados na Tabela 5.13.

Tabela 5.13 – Estimativas dos parâmetros, seus respectivos erros padrões e IC_{95%} para a distribuição GEV e Gumbel ajustada a série de máximos de MP₁₀, antes e após o rompimento da barragem.

Série	Modelo	Parâmetro	Estimativa	Erro padrão	IC _{95%}
Antes do rompimento da barragem	GEV	μ	103,3028	5,5715	[92,3829; 114,2228]
		σ	51,0081	5,0809	[41,0498; 60,9664]
		ξ	0,4225	0,0936	[0,2391; 0,6059]
	Gumbel	μ	116,9210	6,7120	[103,7657; 130,0764]
		σ	68,3176	5,5821	[57,3769; 79,2582]
Depois do rompimento da barragem	GEV	μ	84,7064	5,0500	[74,8085; 94,6043]
		σ	44,2616	3,9961	[36,4294; 52,0938]
		ξ	0,2021	0,0790	[0,0474; 0,3569]
	Gumbel	μ	89,8309	5,1776	[79,6830; 99,9789]
		σ	48,8917	4,0954	[40,8649; 56,9186]

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Após obter as estimativas dos parâmetros, foi realizado o teste da razão de verossimilhança para proceder com a escolha dos melhores modelos, ou seja, avaliar se o parâmetro ξ era igual ou diferente de zero. Porém, antes de se realizar o teste, já existe um forte indício de que

$\xi \neq 0$, pois pela Tabela 5.13 é possível ver que o zero não pertence ao intervalo de confiança de 95% para o parâmetro ξ em ambas as séries de material particulado. Para validar o que foi observado, na Tabela 5.14 são apresentados os resultados da TRV de cada série.

Tabela 5.14 – Resultados do TRV.

Série	Teste da razão de verossimilhança	
	Estatística	valor- <i>p</i>
Antes do rompimento da barragem	34,9418	$3,40 \times 10^{-9}$
Depois do rompimento da barragem	8,4421	0,0037

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Conforme os resultados do teste, apresentados na Tabela 5.14, ao nível de significância de 5% deve-se rejeitar a hipótese nula, indicando que o parâmetro ξ é diferente de zero. Sendo assim, tanto para a série anterior, como a posterior ao rompimento da barragem o modelo escolhido, utilizando-se o método dos blocos, foi a distribuição GEV. Desta forma, o modelo para a série de MP_{10} , em $\mu g/m^3$, antes do rompimento pode ser escrito como:

$$F_{GEV}(x) = \exp \left\{ - \left[1 + 0,4225 \left(\frac{x - 103,3028}{51,0081} \right) \right]^{-\frac{1}{0,4225}} \right\}, \quad (5.5)$$

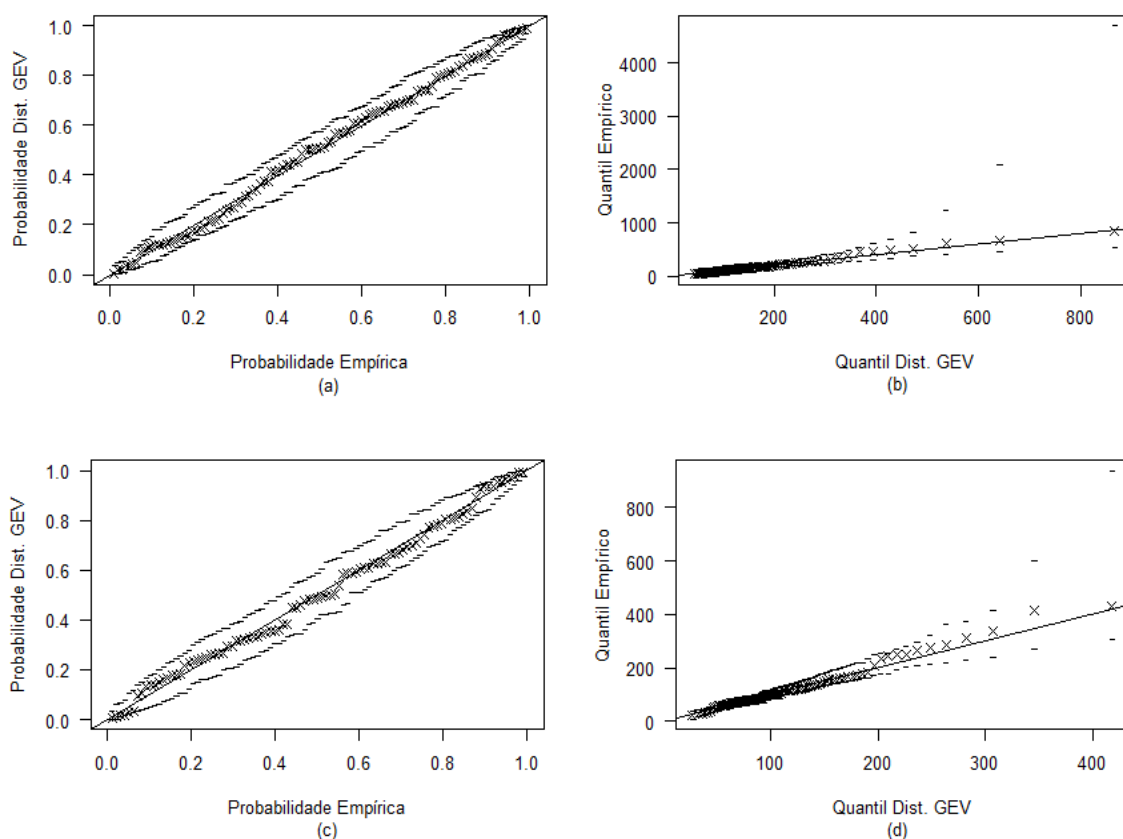
com $x \in [0; +\infty)$. Já para a série de MP_{10} , em $\mu g/m^3$, depois do rompimento, o modelo pode ser expresso da seguinte forma:

$$F_{GEV}(x) = \exp \left\{ - \left[1 + 0,2021 \left(\frac{x - 84,7064}{44,2616} \right) \right]^{-\frac{1}{0,2021}} \right\}, \quad (5.6)$$

com $x \in [0; +\infty)$.

Em seguida, com os parâmetros estimados foi analisado o desempenho de cada modelo por meio dos gráficos de probabilidade-probabilidade (*pp-plot*) e quantil-quantil (*qq-plot*). Através das Figuras 5.17 (a) e (b), observa-se um bom ajuste do modelo para a série antes do rompimento, pois todas as observações estão bem próximas da reta, além de manterem-se dentro do envelope de confiança (ou, intervalo de confiança). O mesmo obteve-se para a série após o rompimento, em que é possível identificar a partir das Figuras 5.17 (c) e (d) que houve um bom ajuste do modelo aos dados observados.

Figura 5.17 – Probabilidades ((a) e (c)) e quantis ((b) e (d)) para a distribuição GEV ajustada as séries de MP_{10} , em $\mu\text{g}/\text{m}^3$, antes e após o rompimento da barragem, respectivamente.



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Entretanto, as conclusões a respeito da qualidade do ajuste não podem ser baseadas somente em uma análise visual. Para isso, foi utilizado o teste de aderência de Anderson-Darling, no qual a estatística do teste é obtida por meio das distribuições teórica e empírica.

A estatística do teste de Anderson Darling para a distribuição GEV ajustada a série de MP_{10} antes do rompimento da barragem foi a seguinte: $A = 2,2782$ e valor- $p = 0,5293$, indicando a não rejeição da hipótese nula, ou seja, $F(x) = \hat{F}(x)$. Quanto que para a série de MP_{10} após o rompimento da barragem, a estatística do teste obtida foi: $A = 1,4388$ e valor- $p = 0,8812$, concluindo também que a série de dados segue a distribuição teórica.

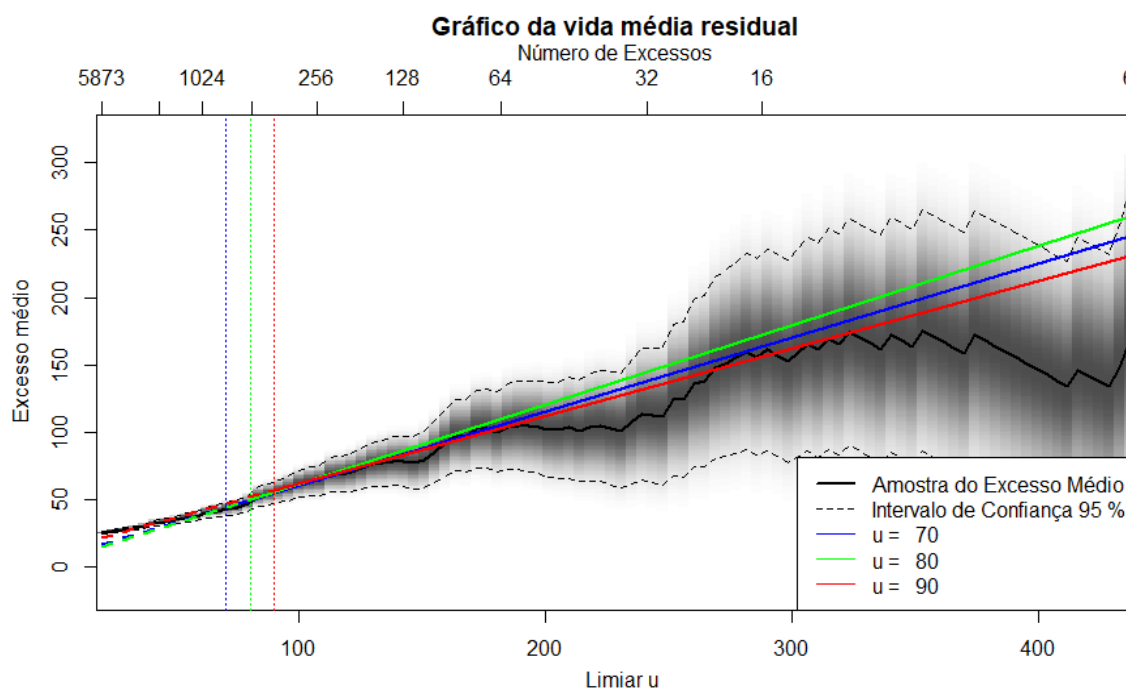
Visto que os modelos se ajustaram aos dados de forma satisfatória, foi então possível calcular as probabilidades das concentrações de MP_{10} ultrapassarem o limite legal de $50 \mu\text{g}/\text{m}^3$, para cada período. Aplicando-se as estimativas dos parâmetros obtidas pela distribuição GEV (Tabela 5.13) na expressão 3.49, obtiveram-se os seguintes resultados: 98,11% para antes do rompimento e 90,45% para após o rompimento. Portanto, a probabilidade da concentração de material particulado ultrapassar o limite foi menor após o rompimento da barragem.

5.2.2 Método dos excessos acima de um limiar u

A utilização deste método foi feito por meio dos valores originais das séries de material particulado, por apresentarem melhores resultados. No entanto, de forma que se tenha o mesmo valor de limiar para a série antes e após o rompimento da barragem, foi necessário realizar a união delas.

Em seguida, foi escolhido o valor do limiar, para assim posteriormente se obter os valores que excedem u . Na Figura 5.18, tem-se o gráfico da média de excessos para os dados de MP_{10} com três possíveis valores para o limiar u . Além disso, os valores propostos para o limiar são próximos dos níveis estabelecidos pela Resolução CONAMA nº 491/2018 como aceitável, apresentados na Figura 1 do apêndice A.

Figura 5.18 – Gráfico de vida média residual para os dados de MP_{10} , em $\mu g/m^3$.

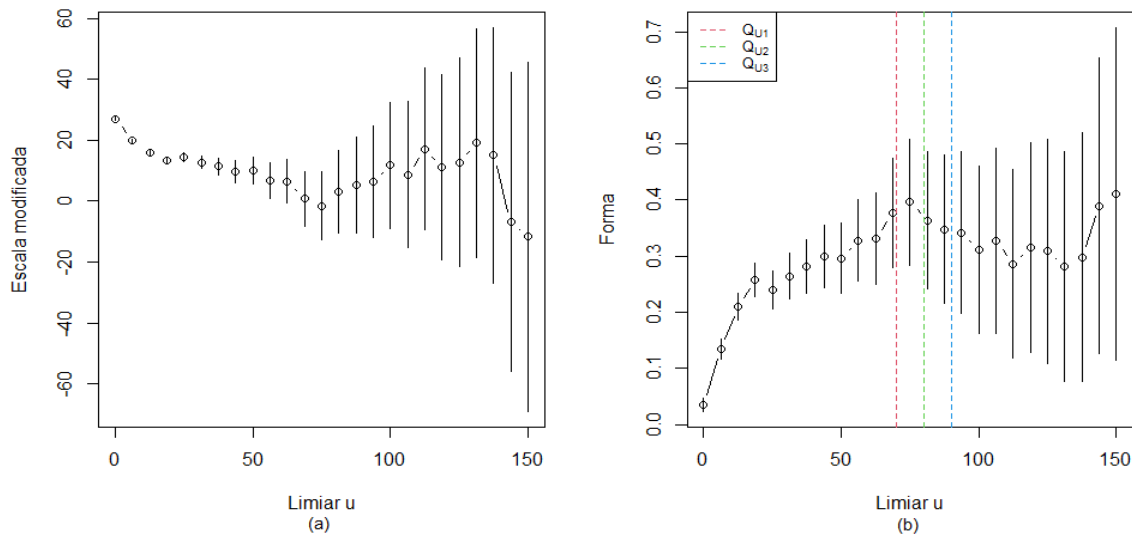


Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Além do gráfico da média residual, foi também utilizado a técnica complementar, em que se obtiveram as estimativas para os parâmetros escala e forma em função de 150 diferentes limiares, que pode ser vista na Figura 5.19.

Na Figura 5.19 (a), pode se observar que a partir de $u = 70$ o comportamento do parâmetro escala é linear, além de que as estimativas para o parâmetro forma, Figura 5.19 (b), é aproximadamente constante a contar deste ponto. Portanto, a partir desse valor foi analisado qual seria o melhor limiar, para assim obter os valores máximos de MP_{10} .

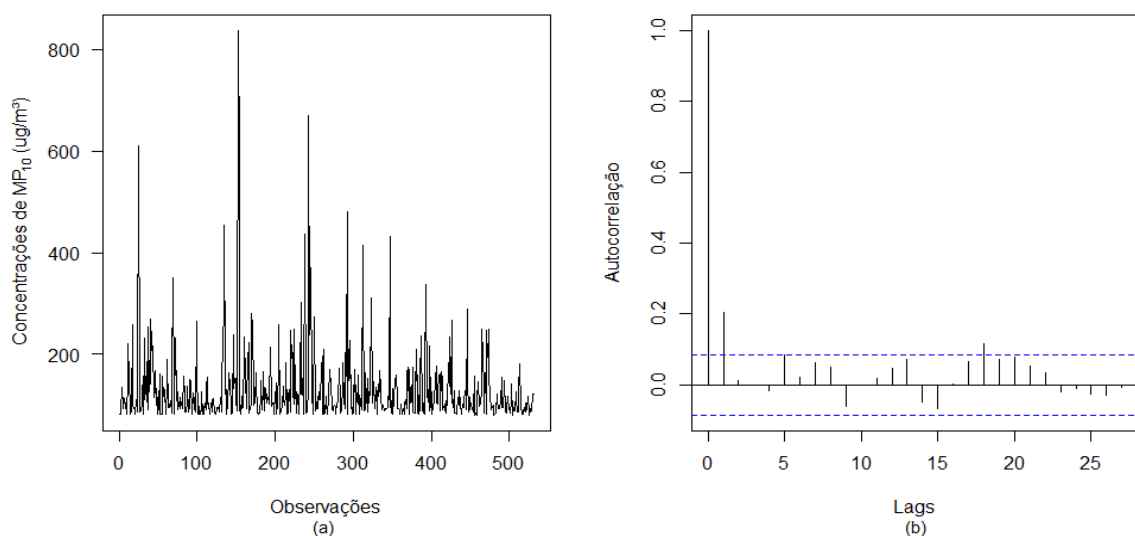
Figura 5.19 – Estimação dos parâmetros escala (a) e forma (b) para 150 limiares diferentes para os dados horários de MP_{10} .



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Ao utilizar como limiar $u = 70$, foi verificado que as observações máximas de MP_{10} não eram independentes. Então, utilizou-se a segunda escolha, que foi $u = 80$. Para verificar a independência da nova série de máximos (Figura 5.20 (a)), composta de 531 observações, utilizou-se o gráfico da função de autocorrelação representado na Figura 5.20 (b). A partir dela foi constatado que os dados são não correlacionados.

Figura 5.20 – Série de 531 máximos de MP_{10} , em $\mu\text{g}/\text{m}^3$, com limiar $u = 90$ (a); Função de autocorrelação para a v.a. 531 máximos de MP_{10} (b).

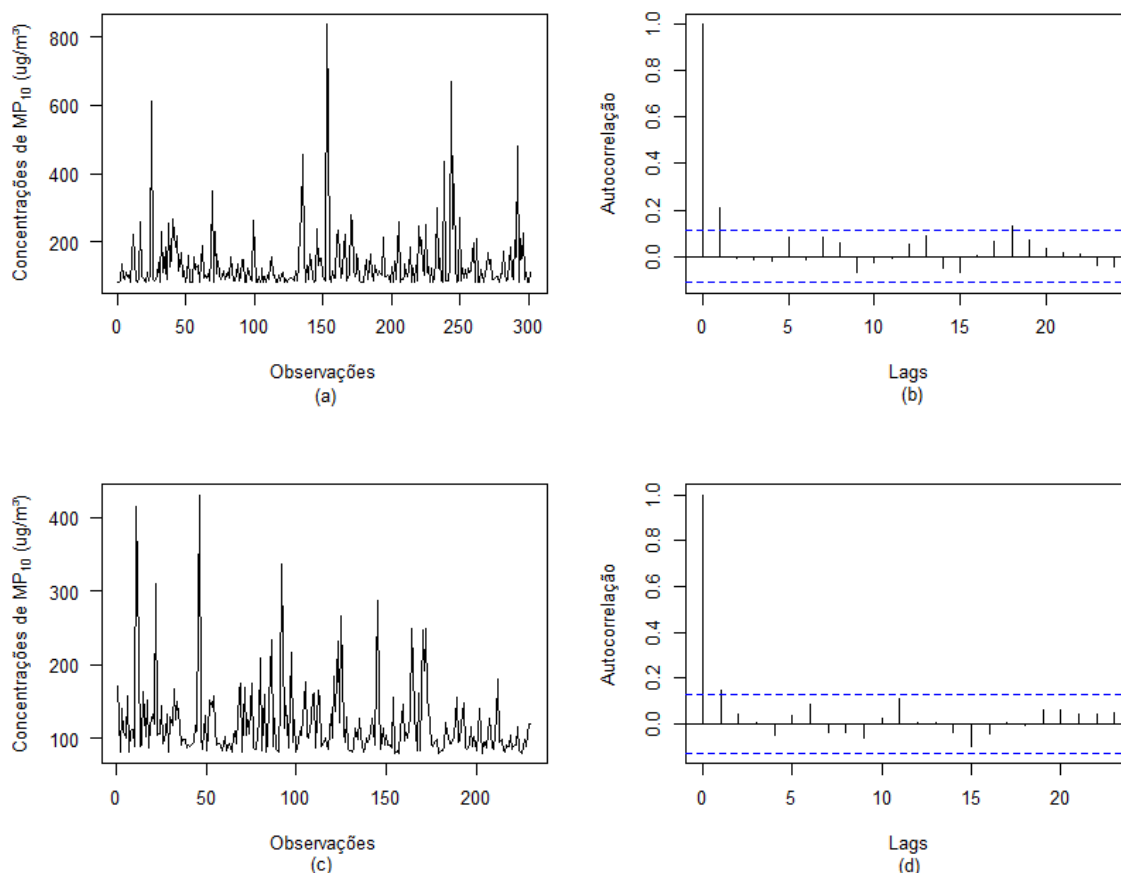


Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Uma vez obtidas as concentrações máximas de MP_{10} , essas foram separadas de acordo com cada período. Dessa forma, a série antes do rompimento apresentou 301 máximos, repre-

sentados na Figura 5.21 (a) e para a série após o rompimento teve-se 230 máximos, como pode ser observado na Figura 5.21 (c). Pelas funções de autocorrelação (Figuras 5.21 (b) e (d)), foi verificado que ambas as séries são independentes.

Figura 5.21 – Série de 301 máximos de MP_{10} , em $\mu g/m^3$, antes do rompimento (com limiar $u = 80$) (a); Função de autocorrelação para a v.a. 301 máximos de MP_{10} (b); Série de 230 máximos de MP_{10} , em $\mu g/m^3$, após o rompimento (com limiar $u = 80$) (c); Função de autocorrelação para a v.a. 230 máximos de MP_{10} (d).



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Após, na Tabela 5.15 encontram-se os valores das estatísticas e os valores- p dos testes de Cox-Stuart e de Fisher, com o objetivo de analisar a estacionariedade.

Tabela 5.15 – Resultados do teste de Cox-Stuart e teste de Fisher.

Série	Teste de Cox-Stuart		Teste de Fisher	
	Estatística	valor- p	Estatística	valor- p
Antes do rompimento da barragem	0,0333	0,9735	0,0432	0,2087
Depois do rompimento da barragem	1,4069	0,0797	0,0334	0,1874

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Pela Tabela 5.15, considerando o nível de 95% de significância, os resultados dos testes indicaram que as séries de máximos de MP_{10} são estacionárias.

Após obter séries de máximos de MP_{10} , antes (Figura 5.21 (a)) e após (Figura 5.21 (c)) o rompimento da barragem, que satisfizessem os pressupostos de independência e estacionariedade, foi então possível obter as estimativas dos parâmetros para cada modelo. Para isso, foi utilizado o método da máxima verossimilhança. Na Tabela 5.16 encontram-se as estimativas, seus respectivos erros padrões e o intervalo de confiança de 95%.

Tabela 5.16 – Estimativas dos parâmetros, seus respectivos erros padrões e $IC_{95\%}$ para a distribuição GP e Exponencial ajustada a série de máximos de MP_{10} , antes e após o rompimento da barragem.

Série	Modelo	Parâmetro	Estimativa	Erro padrão	$IC_{95\%}$
Antes do rompimento da barragem	GPD	σ_u	32,7251	3,2010	[26,4514; 38,9989]
		ξ	0,4399	0,0830	[0,2772; 0,6026]
	Exponencial	σ_u	55,8306	3,2180	[49,5234; 62,1377]
Depois do rompimento da barragem	GPD	σ_u	31,1436	3,2362	[24,8007; 37,4864]
		ξ	0,2417	0,0818	[0,0812; 0,4022]
	Exponencial	σ_u	40,8580	2,6940	[35,5776; 46,1383]

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Uma vez obtidas as estimativas dos parâmetros, foi então realizado o teste da razão de verossimilhança. Os resultados conforme apresentados na Tabela 5.17 indicam que $\xi \neq 0$, pois o valor- p de ambos os períodos foi menor que $\alpha = 0,05$. Sendo assim, a distribuição GP foi a escolhida para descrever o comportamento das concentrações máximas de MP_{10} .

Tabela 5.17 – Resultados do TRV.

Série	Teste da razão de verossimilhança	
	Estatística	valor- p
Antes do rompimento da barragem	56,3331	$6,12 \times 10^{-14}$
Depois do rompimento da barragem	13,5154	0,0002

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Desta forma, o modelo para a série de máximos de MP_{10} , em $\mu g/m^3$, antes do rompimento pode ser expressa pela seguinte função de distribuição:

$$G_{GP}(y) = 1 - \left[1 + 0,4399 \left(\frac{y}{32,7251} \right) \right]^{-0,4399}, \quad (5.7)$$

com $y \in [0; +\infty)$ e $u = 80$. Para a série de máximos de MP_{10} , em $\mu g/m^3$, depois do rompimento, o modelo é dado a seguir:

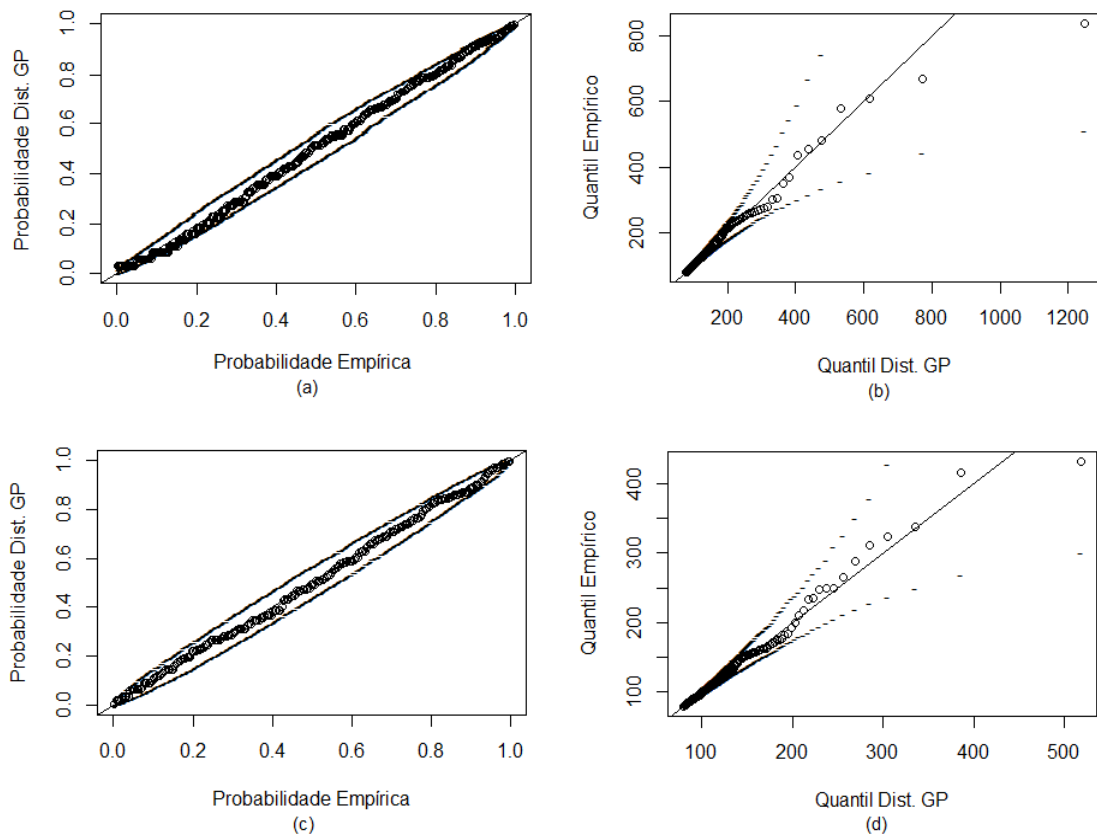
$$G_{GP}(y) = 1 - \left[1 + 0,2417 \left(\frac{y}{31,1436} \right) \right]^{-\frac{1}{0,2417}}, \quad (5.8)$$

com $y \in [0; +\infty)$ e $u = 80$.

Após obter as estimativas dos parâmetros, foi analisado o desempenho de cada modelo utilizando os gráficos *pp-plot* e *qq-plot* que são apresentados na Figura 5.22.

Pelos gráficos *pp-plot* expressos pelas Figuras 5.22 (a) e (c), observa-se que os pontos estão próximos da reta, ou seja, de forma visual se tem um bom ajuste dos modelos aos dados. Para os gráficos *qq-plot* representados pelas Figuras 5.22 (b) e (d) alguns pontos apresentam-se um pouco distantes da reta, mas ainda se encontram dentro do intervalo de confiança.

Figura 5.22 – Probabilidades ((a) e (c)) e quantis ((b) e (d)) para a distribuição GP ajustada as séries de MP_{10} , em $\mu g/m^3$, antes e após o rompimento da barragem, respectivamente.



Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Para confirmar o bom ajuste do modelo aos dados, visto na Figura 5.22, foi realizado o teste de Anderson Darling. A estatística do teste para a distribuição GP ajustada a série de

máximos de MP_{10} antes do rompimento foi $A = 2,9289$ e o seu valor- $p = 0,4072$. Já para a série após o rompimento, tem-se o valor da estatística $A = 2,1381$ e o valor- $p = 0,7046$. Assim, como o valor- $p > 0,05$ para ambas as séries, não se rejeitou H_0 , ou seja, as séries de dados seguem as distribuições teóricas.

A partir dos modelos ajustados e posterior verificação da adequabilidade de cada um deles aos dados de MP_{10} , obteve-se a probabilidade de ter concentrações acima do limite de $120 \mu\text{g}/\text{m}^3$, para que assim fosse possível avaliar antes e após o rompimento. Neste caso, a análise somente tem sentido para concentrações de material particulado superiores aos limiares u , além de que o valor limite considerado é a primeira medida padrão estabelecida pela Resolução CONAMA nº 491/2018 a ser cumprida, conforme Figura 1 do apêndice A.

Portanto, ao aplicar as estimativas dos parâmetros, considerando a distribuição GP, apresentadas na Tabela 5.16 na equação 3.51, a probabilidade de ocorrer uma concentração máxima acima de $120 \mu\text{g}/\text{m}^3$ antes e após o rompimento foi de 37,60% e 32,68%, respectivamente.

5.2.3 Cálculo do VaR

O valor em risco (VaR) para concentração de MP_{10} , utilizando as duas abordagens da teoria de valores extremos, foi obtido a partir das equações 3.69 e 3.74, com os seguintes resultados que podem ser vistos na Tabela 5.18.

Tabela 5.18 – Níveis de risco e valores ao risco para v.a. máximos de MP_{10} , em $\mu\text{g}/\text{m}^3$.

P	Antes do rompimento ($\mu\text{g}/\text{m}^3$)		Depois do rompimento ($\mu\text{g}/\text{m}^3$)	
	VaR pela GEV	VaR pela GPD	VaR pela GEV	VaR pela GPD
0,2	11,6905	27,3482	0,6071	12,9675
0,3	15,4893	28,6634	6,6352	14,9948
0,4	19,5142	30,2810	8,4345	17,4180
0,5	24,1372	32,3413	13,9080	20,4031
0,6	29,8578	35,0985	17,5155	24,2402
0,7	37,6069	39,0769	28,9458	29,5026
0,8	49,6676	45,6120	45,1758	37,5693
0,9	74,6988	59,8737	74,5675	53,3299
0,99	231,1953	155,0344	201,5308	129,4171

Fonte: Elaborado pelo autor a partir dos dados da FEAM, 2022.

Ao observar a Tabela 5.18, verifica-se que os maiores valores em risco, para as duas abordagens de valores extremos, ocorrem no período antes do rompimento da barragem. Sendo

assim, considerando a concentração de $50 \mu\text{g}/\text{m}^3$, as maiores chances de que se alcance o limite legal foi anteriormente ao rompimento.

Para o poluente MP_{10} no período antes do rompimento da barragem, ao utilizar a GEV o nível crítico pertence ao intervalo de probabilidade 0,8 e 0,9. Mais precisamente, a uma probabilidade de 80,21% é o risco de se atingir o nível de $50 \mu\text{g}/\text{m}^3$. A partir da GPD, a Tabela 5.18 indica que o nível crítico deve estar entre uma probabilidade (p) maior que 0,8 e menor que 0,9, ou seja, uma probabilidade de exatamente 84,22% que se alcance o limite legal.

Para a série referente ao período pós-rompimento, por meio da GEV foi obtido uma probabilidade de 82,23%, que é o risco de se atingir o limite legal do poluente MP_{10} . Ao utilizar a GPD, novamente o valor crítico está entre 0,8 e 0,9, ou seja, a uma probabilidade de 88,55%, atingindo, assim, o limite legal de concentrações de partículas inaláveis.

6 DISCUSSÃO

De acordo com os resultados obtidos por meio das análises estatísticas, observou-se um aumento significativo da poluição do ar em Brumadinho, especialmente pelo poluente PTS, após o rompimento da barragem. Além disso, em ambos os períodos (antes e após o rompimento), houve vários registros de concentrações de PTS e MP_{10} que ultrapassaram os limites padrões de qualidade do ar estabelecido pelo CONAMA. Como resultado, a poluição do ar aumentou, colocando em risco a saúde humana, especialmente de crianças e idosos.

Alguns estudos, como de Braga *et al.* (2007), avaliaram a relação entre a poluição atmosférica e doenças respiratórias e cardiovasculares em Itabira, cidade que é conhecida pelas grandes jazidas de minério de ferro. Os resultados comprovaram que com o aumento de $10 \mu\text{g}/\text{m}^3$ de material particulado, gerado pela mineração a céu aberto, houve aumentos de 4% e 12% no número de atendimentos por doenças respiratórias entre menores de 13 anos e adolescentes, respectivamente. Para as doenças cardiovasculares, o aumento foi de 4%, com efeito maior em pacientes com faixa etária de 45 a 64 anos.

Além desse, na pesquisa de Vormittag, Oliveira e Gleriano (2018), os autores identificaram um aumento das doenças respiratórias (asma, enfisema, bronquite e sinusite crônica, entre outras) no município de Barra Longa (MG), após o rompimento da barragem em Mariana (MG). Sendo assim, algumas atitudes devem ser tomadas pelos órgãos responsáveis, com o intuito de evitar que aconteçam outros casos, pois em Brumadinho não deve ter sido diferente.

Os resultados deste estudo também concordam com os resultados encontrados por Galvão, Feroni e Silveira (2020) para o material particulado, que os autores observaram que desde 2016 houve uma melhora significativa na qualidade do ar em Anchieta (ES), com o rompimento de uma barragem de rejeitos de minério de ferro, no final de 2015, no município de Mariana (MG). Com o rompimento da barragem houve a paralização de uma usina de pelotização no município, e de acordo com os autores as médias anuais de MP_{10} diminuíram entre 10 e 30%, dependendo da estação de monitoramento.

Pelas Figuras 5.16 (c) e 5.21 (c) foi observado por volta do último trimestre do ano de 2019 que as concentrações máximas de MP_{10} reduziram. Este fato, provavelmente, se deve à paralização das atividades de mineração na região de Brumadinho, além de que nesse período houve uma diminuição nas buscas por desaparecidos, evitando assim a movimentação de máquinas, como tratores e caminhões.

Outros estudos, como de Vergilio *et al.* (2021) que também identificaram um aumento significativo de material particulado suspenso (PTS) na água, ao longo de todo o curso do Rio Doce, após o rompimento da barragem de Fundão, em Mariana (MG). No trabalho de Alves e Freitas (2021), os autores mostraram que nos dias críticos de poluição na cidade de Itabira ocorrem, predominantemente, anomalias negativas de precipitação, pressão atmosférica, umidade relativa e velocidade do vento. Portanto, como visto, vários são os problemas causados pelo aumento da poluição do ar, que ocorre em muitos casos por falta de cuidado do próprio ser humano, como foi com o caso do rompimento da barragem em Brumadinho.

7 CONCLUSÕES

Ao utilizar a teoria de valores extremos para analisar a concentração máxima de PTS e MP₁₀, na região de Brumadinho (MG), foi verificado que as distribuições GEV e GP se ajustaram de maneira satisfatória à série de dados para ambos os períodos (antes e após o rompimento da barragem). Embora tenham sido aplicadas duas abordagens diferentes para obter as concentrações máximas de cada poluente, em geral, as conclusões a partir dos resultados foram as mesmas.

No caso das séries das concentrações de PTS, em $\mu\text{g}/\text{m}^3$, ao analisar os dois períodos, por meio das distribuições GEV e GP, verificou-se um aumento da probabilidade de ocorrer concentrações máximas que ultrapassem o limite legal ($240 \mu\text{g}/\text{m}^3$) após o rompimento da barragem. Em relação a aplicação do VaR, foram obtidos altos riscos de exposição da população ao particulado, considerando o nível estabelecido pelo CONAMA.

Para as séries das concentrações de MP₁₀, em $\mu\text{g}/\text{m}^3$, ao examinar os dois períodos, utilizando as distribuições GEV e GP, verificou-se uma redução da probabilidade de ocorrer concentrações máximas que excedessem o limite padrão de $50 \mu\text{g}/\text{m}^3$ e o intermediário de $120 \mu\text{g}/\text{m}^3$, depois do rompimento da barragem. Quanto aos resultados obtidos do VaR, o risco de um habitante de Brumadinho à exposição ao material particulado, considerando o limite legal, foi bastante alta, tanto antes como depois do rompimento da barragem, chegando a ultrapassar o nível de risco de 80%.

Portanto, em ambos os períodos (antes e após o rompimento da barragem) houve vários registros de que as concentrações de PTS e MP₁₀ ultrapassaram os limites legais e, portanto, podem afetar a saúde da população e ao meio ambiente. Porém, se a redução das concentrações extremas para o MP₁₀ venham a ocorrer nos anos seguintes, a longo prazo poderá trazer uma qualidade de vida mais adequada para a população.

Uma das limitações deste estudo é o fato de que não há dados disponíveis para os demais anos, pois se houvesse, seria possível determinar se ocorreu ou não a redução das concentrações máximas de MP₁₀. Além disso, poderia ser realizada uma avaliação mais abrangente da qualidade do ar na região atualmente.

Para trabalhos futuros recomenda-se o uso de outros métodos, como por exemplo, o processo pontual de Poisson, ou até mesmo modelos para dados dependentes, evitando assim perda de informação. Além disso, avaliar os índices de poluição dos demais poluentes que foram aqui citados.

REFERÊNCIAS

- ALMEIDA, R. P. d. S. de *et al.* Comparação da qualidade do ar em localidades industrial e urbana. **Revista Internacional de Ciências**, Rio de Janeiro, v. 9, n. 3, p. 47–62, 2019.
- ALVES, H. dos S.; FREITAS, A. C. V. Critical air pollution events analysis in Itabira – Minas Gerais, Brazil. **Research, Society and Development**, Vargem Grande Paulista, v. 10, n. 1, 2021.
- ALVES, I. F. Acerca de testes estatísticos para valores extremos. **Boletim da Sociedade Portuguesa de Estatística**, Lisboa, p. 20–26, 2007.
- AMORIM, N. A.; CONCEIÇÃO, T. L. de A.; FERREIRA, P. D. A. Mineração de agregados e saúde pública: uma perspectiva socioambiental nas comunidades rurais sagui 1 e 2. **Revista E.T.C. Educação, Tecnologia e Cultura**, Salvador, n. 14, p. 1–13, 2016.
- AMPLO ENGENHARIA E GESTÃO DE PROJETOS. **RIMA - Relatório de Impacto Ambiental maio/2019**. Nova Lima, 2019. Disponível em: http://videosvale.intranetvale.com.br/projetos-de-licenciamento-mg/RIMA_LOC_OBRAS_EMERG.pdf. Acesso em: 30 jan. 2022.
- ARAÚJO, C. B. de. **Contribuição ao estudo do comportamento de barragens de rejeito de mineração de ferro**. 2006. 133 p. Dissertação (Mestrado em Engenharia Civil) – Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro, 2006.
- BEHRENS, C. N.; LOPES, H. F.; GAMERMAN, D. Bayesian analysis of extreme events with threshold estimation. **Statistical Modelling**, New York, v. 4, n. 3, p. 227–244, 2004.
- BRAGA, A. L. F. *et al.* Associação entre poluição atmosférica e doenças respiratórias e cardiovasculares na cidade de Itabira, Minas Gerais, Brasil. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 23, n. 4, p. S570–S578, 2007.
- BRASIL. Conselho Nacional do Meio Ambiente. Resolução CONAMA nº 491/2018, de 19 de novembro de 2018. Dispõe sobre qualidade do ar. **Diário Oficial [da] República Federativa do Brasil**, Poder Executivo, Brasília, DF, 21 nov. 2018. Seção 1, p. 155-156. Disponível em: <http://www2.mma.gov.br/port/conama/legiabre.cfm?codlegi=740>. Acesso em: 25 jul. 2021.
- BUUREN, S. V. **Flexible imputation of missing data**. 2. ed. Flórida: CRC press, 2018. 444 p.
- BUUREN, S. V.; GROOTHUIS-OUDSHOORN, K. mice: Multivariate imputation by chained equations in R. **Journal of Statistical Software**, Innsbruck, v. 45, n. 1, p. 1–67, 2011.
- CAEIRO, F.; MATEUS, A. **randtests: Testing randomness in R**. [S.l.], 2014. R package version 1.0. Disponível em: <https://CRAN.R-project.org/package=randtests>.
- CARSLAW, D. C.; ROPKINS, K. Openair — an R package for air quality data analysis. **Environmental Modelling & Software**, Provo, v. 27, p. 52–61, 2012.
- CARVALHO, A. P. M. *et al.* A vigilância em saúde ambiental como resposta ao desastre do rompimento da barragem de rejeitos em brumadinho. **Saúde em Debate**, Rio de Janeiro, v. 44, n. ESPECIAL 2, p. 364–376, 2021.
- CARVALHO, J. R. P. de *et al.* Modelo de imputação múltipla para estimar dados de precipitação diária e preenchimento de falhas. **Revista Brasileira de Meteorologia**, São José dos Campos, v. 32, n. 4, p. 575–583, 2017.

- CASELLA, G.; BERGER, R. L. **Statistical inference**. 2. ed. EUA: Cengage Learning, 2001. 688 p.
- CHAN, S.; NADARAJAH, S.; AFUECHETA, E. An R package for value at risk and expected shortfall. **Communications in Statistics - Simulation and Computation**, New York, v. 45, n. 9, 2016.
- COLES, S. G. **An introduction to statistical modeling of extreme values**. London: Springer, 2001. 226 p.
- COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO. **Qualidade do ar no estado de São Paulo 2018**. São Paulo, 2019. Série Relatórios. Disponível em: <https://cetesb.sp.gov.br/ar/wp-content/uploads/sites/28/2019/07/Relat%C3%B3rio-de-Qualidade-do-Ar-2018.pdf>. Acesso em: 15 jul. 2021.
- COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO. **Dióxido de enxofre**. [S.l.], 2020a. Ficha de Informação Toxicológica. Disponível em: <https://cetesb.sp.gov.br/laboratorios/wp-content/uploads/sites/24/2020/07/Dio%CC%81xido-de-enxofre.pdf>. Acesso em: 12 jul. 2021.
- COMPANHIA AMBIENTAL DO ESTADO DE SÃO PAULO. **Monóxido de carbono**. [S.l.], 2020b. Ficha de Informação Toxicológica. Disponível em: <https://cetesb.sp.gov.br/laboratorios/wp-content/uploads/sites/24/2020/07/Mono%CC%81xido-de-carbono.pdf>. Acesso em: 12 jul. 2021.
- COSTA, D. L. Poluição do Ar. In: KLAASSEN, C. D.; WATKINS, J. B. (Ed.). **Fundamentos em Toxicologia**. 2. ed. Porto Alegre: AMGH Editora, 2012. cap. 28, p. 379–389.
- COSTA, F. M. da; SÁFADI, T. Comparação estatística de duas séries de material particulado (MP10) na cidade de São Paulo. **Revista Brasileira de Biometria**, Lavras, v. 28, n. 3, p. 23–38, 2010.
- DANCEY, C. P.; REIDY, J. **Statistics without maths for psychology**. 7. ed. London: Pearson education, 2017. 640 p.
- DINIZ-EHRHARDT, M. A.; LOPES, V. L. d. R.; PEDROSO, L. G. Métodos sem derivadas para minimização irrestrita. **Notas em Matemática Aplicada**, São Carlos, v. 49, n. 3, p. 87, 2010.
- FARAWAY, J. *et al.* **goftest: Classical goodness-of-fit tests for univariate distributions**. [S.l.], 2021. R package version 1.2-3. Disponível em: <https://CRAN.R-project.org/package=goftest>.
- FISHER, R. A.; TIPPETT, L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. **Mathematical Proceedings of the Cambridge Philosophical Society**, Cambridge, v. 24, n. 2, p. 180–190, 1928.
- FUNDAÇÃO ESTADUAL DO MEIO AMBIENTE. **Monitoramento da Qualidade do Ar na Região Metropolitana de Belo Horizonte - Ano Base de 2013**. Belo Horizonte, 2016. Relatório Técnico. Disponível em: <http://www.feam.br/qualidade-do-ar/relatorios-artigos-e-publicacoes>. Acesso em: 08 ago. 2021.

FUNDAÇÃO ESTADUAL DO MEIO AMBIENTE. **Lista de Barragens 2020**. Belo Horizonte, 2020. Disponível em: <http://www.feam.br/gestao-de-barragens/inventario-de-barragens>. Acesso em: 10 jul. 2021.

FUNDAÇÃO ESTADUAL DO MEIO AMBIENTE. **Institucional**. Belo Horizonte, 2021. Disponível em: <http://feam.br/instituicao>. Acesso em: 10 ago. 2021.

GALVÃO, E. S.; FERONI, R. d. C.; SILVEIRA, A. B. Trends in air quality and hospital admissions due to respiratory diseases since the standstill of an industrial plant in Brazil. **Environmental Science and Pollution Research**, Berlim, v. 27, n. 19, p. 24452–24465, 2020.

GILLELAND, E.; KATZ, R. W. extRemes 2.0: An extreme value analysis package in R. **Journal of Statistical Software**, Innsbruck, v. 72, n. 1, p. 1–39, 2016.

GNEDENKO, B. Sur la distribution limite du terme maximum d'une serie aleatoire. **Annals of Mathematics**, Princeton, v. 44, p. 423–453, 1943.

GOUVEIA, N. *et al.* Hospitalizações por causas respiratórias e cardiovasculares associadas à contaminação atmosférica no Município de São Paulo, Brasil. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 22, p. 2669–2677, 2006.

HADEED, S. J. *et al.* Imputation methods for addressing missing data in short-term monitoring of air pollutants. **Science of The Total Environment**, Amsterdã, v. 730, p. 1–7, 2020.

HARRELL JR., F. E. **Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis**. 2. ed. New York: Springer, 2015. 582 p.

HOSKING, J. R. M. Testing whether the shape parameter is zero in the generalized extreme-value distribution. **Biometrika**, Oxford, v. 71, n. 2, p. 367–374, 1984.

HU, Y.; SCARROTT, C. evmix: An R package for extreme value mixture modeling, threshold estimation and boundary corrected kernel density estimation. **Journal of Statistical Software**, Innsbruck, v. 84, n. 5, p. 1–27, 2018.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Estimativa da População 2021, Brumadinho, Minas Gerais, Brasil**. Rio de Janeiro, 2021. Disponível em: <https://cidades.ibge.gov.br/>. Acesso em: 30 jan. 2022.

INSTITUTO BRASILEIRO DE MINERAÇÃO. **Gestão e Manejo de Rejeitos da Mineração**. Brasília, 2016. Disponível em: <https://ibram.org.br/wp-content/uploads/2021/02/Gestao-e-Manejo-de-Rejeitos-da-Mineracao-2016.pdf>. Acesso em: 08 jul. 2021.

INSTITUTO BRASILEIRO DE MINERAÇÃO. **Informações sobre a economia mineral brasileira 2020 - Ano base 2019**. Brasília, 2020. Disponível em: <https://portaldamineracao.com.br/wp-content/uploads/2021/03/Economia-Mineral-Brasileira-IBRAM-2020.pdf>. Acesso em: 25 jul. 2021.

JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. **Quarterly Journal of the Royal Meteorological Society**, Reading, v. 81, p. 158–171, 1955.

- KLUMPP, A. *et al.* Um novo conceito de monitoramento e comunicação ambiental: a rede européia para a avaliação da qualidade do ar usando plantas bioindicadoras (EuroBionet). **Brazilian Journal of Botany**, São Paulo, v. 24, n. 4, p. 511–518, 2001.
- LACAZ, F. A. d. C.; PORTO, M. F. d. S.; PINHEIRO, T. M. M. Tragédias brasileiras contemporâneas: o caso do rompimento da barragem de rejeitos de Fundão/Samarco. **Revista Brasileira de Saúde Ocupacional**, São Paulo, v. 42, n. 9, p. 1–12, 2017.
- LEMOS, I. P. *et al.* Uma abordagem pela teoria de valores extremos: determinação do tamanho amostral para inspeção de equipamentos de processo utilizando a distribuição generalizada de Pareto e o Peak-Over-Threshold method. **Revista de Matemática**, Ouro Preto, v. 1, p. 29–50, 2020.
- LISKA, G. R.; BEIJO, L. A. Distribuições de probabilidade aplicadas na análise de níveis máximos de MP10 e O3 nas cidades de Cubatão e Paulínia, SP. **Revista de Estudos Ambientais**, Blumenau, v. 14, n. 3, p. 35–47, 2012.
- LIU, P. W. G. *et al.* Establishing multiple regression models for ozone sensitivity analysis to temperature variation in Taiwan. **Atmospheric Environment**, Amsterdã, v. 79, p. 225–235, 2013.
- LOPO, A. B. *et al.* Teoria dos valores extremos aplicada na análise da radiação ultravioleta na cidade de Natal-RN. **Proceeding Series of the Brazilian Society of Computational and Applied Mathematics**, São Carlos, v. 3, n. 1, 2015.
- MARIMOUTOU, V.; RAGGAD, B.; TRABELSI, A. Extreme value theory and value at risk: application to oil market. **Energy Economics**, Brighton, v. 31, n. 4, p. 519–530, 2009.
- MARTINS JÚNIOR, H. **Análise de valores extremos para dados de poluição atmosférica na cidade de São Paulo**. 2010. 88 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) – Universidade Federal de Lavras, UFLA, Lavras, 2010.
- MARTINS JÚNIOR, H.; SÁFADI, T.; CARVALHO, M. Análise de valores de valores extremos para dados de poluição atmosférica. **Revista Sodebras**, Rio de Janeiro, v. 8, n. 85, p. 16–20, 2013.
- MESQUITA, M. E. A. Geografia da Saúde: um estudo sobre o clima e saúde. *In*: ENCONTRO DE GEÓGRAFOS DA AMÉRICA LATINA, 10., 2005, São Paulo. **Anais** [...]. São Paulo: USP, 2005. p. 9398–9408.
- MONTE, E. Z.; ALBUQUERQUE, T. T. d. A.; REISEN, V. A. Inter-relações entre as concentrações de ozônio e de dióxido de nitrogênio na região da Grande Vitória, Espírito Santo, Brasil. **Engenharia Sanitária e Ambiental**, Rio de Janeiro, v. 22, p. 679–690, 2017.
- MORETTIN, P. A.; BUSSAB, W. O. **Estatística básica**. 9. ed. São Paulo: Saraiva, 2017. 568 p.
- MORETTIN, P. A.; TOLOI, C. M. C. **Análise de Séries Temporais**. 2. ed. São Paulo: Edgard Blucher, 2006. 564 p.
- MORITA, R. H.; BUENO, R. D. L. d. S.; PIRES, R. A. Controlando o pânico. **Economia Aplicada**, Ribeirão Preto, v. 12, n. 1, p. 29–54, 2008.

- MORRIS, T. P.; WHITE, I. R.; ROYSTON, P. Tuning multiple imputation by predictive mean matching and local residual draws. **BMC Medical Research Methodology**, London, v. 14, n. 1, p. 1–13, 2014.
- NAGHETTINI, M.; PINTO, É. J. d. A. **Hidrologia Estatística**. Belo Horizonte: CPRM, 2007. 552 p.
- OLEQUES, N. *et al.* Aplicação da técnica de imputação múltipla no preenchimento de falhas em registros atmosféricos de Bagé/RS. **Anais do Salão Internacional de Ensino, Pesquisa e Extensão**, Santana do Livramento, v. 9, n. 2, p. 1–6, 2020.
- OLIVEIRA, D. P. de; KUMMROW, F. Poluentes da Atmosfera. *In*: OGA, S.; CAMARGO, M. M. de A.; BATISTUZZO, J. A. de O. (Ed.). **Fundamentos de Toxicologia**. 4. ed. São Paulo: Atheneu Editora, 2014. cap. 2, p. 143–164.
- PAIVA, D. de A.; SÁFADI, T. Study of tests for trend in time series. **Revista Brasileira de Biometria**, Lavras, v. 39, n. 2, p. 311–333, 2021.
- PICKANDS, J. Statistical inference using extreme order statistics. **Annals of Statistics**, Philadelphia, v. 3, n. 1, p. 119–131, 1975.
- PINTO, W. d. P.; REISEN, V. A.; LIMA, G. B. Imputação de dados faltantes em séries temporais de concentrações de material particulado inalável. *In*: SIMPÓSIO LUSO-BRASILEIRO DE ENGENHARIA SANITÁRIA E AMBIENTAL, 17., 2016, Florianópolis. **Anais [...]**. Florianópolis: ABES, 2016. p. 1–6.
- PINTO, W. d. P.; REISEN, V. A.; MONTE, E. Z. Previsão da concentração de material particulado inalável, na Região da Grande Vitória, ES, Brasil, utilizando o modelo SARIMAX. **Engenharia Sanitária e Ambiental**, Rio de Janeiro, v. 23, n. 2, p. 307–318, 2018.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2021. Disponível em: <https://www.R-project.org/>.
- RAZAK, N. A.; ZUBAIRI, Y. Z.; YUNUS, R. M. Imputing missing values in modelling the PM10 concentrations. **Sains Malaysiana**, Bangi, v. 43, n. 10, p. 1599–1607, 2014.
- RIBATET, M.; DUTANG, C. **POT: Generalized Pareto Distribution and Peaks Over Threshold**. [*S.I.*], 2022. R package version 1.1-10. Disponível em: <https://CRAN.R-project.org/package=POT>.
- RUBIN, D. B. **Multiple imputation for nonresponse in surveys**. New York: John Wiley & Sons, 1987. 285 p.
- RUIZ-CÁRDENAS, R.; KRAINSKI, E. T. Preenchimento de falhas em bancos de dados meteorológicos diários: uma comparação de abordagens. *In*: CONGRESSO BRASILEIRO DE AGROMETEOROLOGIA, 17., 2011, Guarapari. **Anais [...]**. Guarapari: SESC, 2011. p. 1–5.
- SANTOS, J. F. dos; FARIAS, K. G. de P. Agricultura familiar como forma de produção rural para abastecimento de grandes centros urbanos. *In*: SIMPÓSIO MINEIRO DE GEOGRAFIA, 1., 2014, Alfenas. **Anais [...]**. Alfenas: UNIFAL, 2014. p. 692–696.
- SILVA, I. F. da *et al.* Análise da qualidade do ar da região central de Uberlândia: partículas inaláveis (MP10). **Brazilian Journal of Animal and Environmental Research**, Curitiba, v. 3, n. 1, p. 207–217, 2020.

SILVA, K. C. C. *et al.* Avaliação da tendência à formação de ozônio dos gases de escape de um veículo abastecido com combustíveis contendo diferentes teores de etanol. *In: SIMPÓSIO INTERNACIONAL DE ENGENHARIA AUTOMOTIVA*, 22., 2014, São Paulo. *Anais* [...]. São Paulo: Blucher Engineering Proceedings, 2014. p. 1–13.

SILVA, R. R. **A distribuição Generalizada de Pareto e mistura de distribuições de Gumbel no estudo da vazão e da velocidade máxima do vento em Piracicaba, SP.** 2008. 100 p. Dissertação (Mestrado em Agronomia) – Universidade de São Paulo, ESALQ/USP, Piracicaba, 2008.

SMITH, R. L. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, Oxford, v. 72, n. 1, p. 67–90, 1985.

SOARES, L. G. *et al.* Análise da qualidade do ar de Uberlândia: partículas totais em suspensão (PTS). *In: CONGRESSO BRASILEIRO DE ENGENHARIA QUÍMICA EM INICIAÇÃO CIENTÍFICA*, 11., 2015, São Paulo. *Anais* [...]. São Paulo: Blucher, 2015. p. 427–432.

SOARES, L. M. *et al.* Avaliação do impacto na qualidade do ar após implantação do BRS no Rio de Janeiro. *Revista Internacional de Ciências*, Rio de Janeiro, v. 8, n. 2, p. 203–220, 2018.

STEPHENSON, A. G. e.v.d: Extreme value distributions. *R news*, Vienna, Austria, v. 2, n. 2, p. 31–32, 2002. Disponível em: <https://CRAN.R-project.org/doc/Rnews/>.

TEODORO, V. A. **Comparação de método de imputação para dados de precipitação diária.** 2019. 85 p. Tese (Doutorado em Ciências) – Universidade de São Paulo, ESALQ/USP, Piracicaba, 2019.

THOMÉ, R.; PASSINI, M. L. Barragens de rejeitos de mineração: características do método de alteamento para montante que fundamentaram a suspensão de sua utilização em Minas Gerais. *Ciências Sociais Aplicadas em Revista*, Marechal Cândido Rondon, v. 18, n. 34, p. 49–65, 2018.

TSAY, R. S. **Analysis of financial time series.** 3. ed. Chicago: John Wiley & Sons, 2010. 714 p.

VERGÍLIO, C. dos S. *et al.* Immediate and long-term impacts of one of the worst mining tailing dam failure worldwide (Bento Rodrigues, Minas Gerais, Brazil). *Science of The Total Environment*, Amsterdã, v. 756, p. 1–12, 2021.

VORMITAG, E. d. M. P. A. d. A.; OLIVEIRA, M. A. d.; GLERIANO, J. S. Avaliação de saúde da população de Barra Longa afetada pelo desastre de Mariana, Brasil. *Ambiente & Sociedade*, São Paulo, v. 21, p. 1–22, 2018.

WILLMOTT, C. J. *et al.* Statistics for the evaluation and comparison of models. *Journal of Geophysical Research: Oceans*, Hoboken, v. 90, n. C5, p. 8995–9005, 1985.

APÊNDICE A – Padrões de qualidade do ar

Figura 1 – Padrões nacionais de qualidade do ar e diretrizes da Organização Mundial da Saúde (OMS).

				MP _{2,5}	MP ₁₀	PTS	OS	SO ₂	NO ₂	O ₃	CO	Fumaça	Chumbo				
				(µg/m ³)							(ppm)	(µg/m ³)					
Padrão Nacional (CONAMA nº 491/2018)	Exposição	Curta	Padrão	PI-1	60 24h	120 24h	-	-	125 24h	260 1h ²	140 8h ³	-	120 24h	-			
				PI-2	50 24h	100 24h	-	-	50 24h	240 1h ²	130 8h ³	-	100 24h	-			
				PI-3	37 24h	75 24h	-	-	30 24h	220 1h ²	120 8h ³	-	75 24h	-			
				PF	25 24h	50 24h	240 24h	-	20 24h	200 1h ²	100 8h ³	9 8h ³	50 24h	-			
	Longa	Padrão	PI-1	20 Ano ¹	40 Ano ¹	-	-	40 Ano ¹	60 Ano ¹	-	-	40 Ano ¹	-				
			PI-2	17 Ano ¹	35 Ano ¹	-	-	30 Ano ¹	50 Ano ¹	-	-	35 Ano ¹	-				
			PI-3	15 Ano ¹	30 Ano ¹	-	-	20 Ano ¹	45 Ano ¹	-	-	30 Ano ¹	-				
			PF	10 Ano ¹	20 Ano ¹	80 Ano ⁴	-	-	40 Ano ¹	-	-	20 Ano ¹	0,5 Ano ¹				
			Diretriz OMS	Exposição	Curta	Padrão		25 24h	50 24h	-	-	20 24h	200 1h	100 8h	10.000 8h	-	-
											500 10min			30.000 1h			
Longa	Padrão			10 Ano ²	20 Ano ²	-	-	-	40 Ano ²	-	-	-	-				

¹ Não pode ser excedido mais que uma vez por ano.

² Média Aritmética Anual - MAA.

³ Média Geométrica Anual - MGA.

Fonte: Adaptado de relatórios de qualidade do ar, 2020.

<https://iema.es.gov.br/qualidadedoar/relatorios>