Universidade Federal de Ouro Preto
Escola de Minas
Departamento de Engenharia Urbana
Graduação em Engenharia Urbana

Lucas Pereira Golia

CONTROLE DE QUALIDADE E AJUSTE DA DISTRIBUIÇÃO GENERALIZADA DE VALORES EXTREMOS PARA SÉRIES PLUVIOMÉTRICAS DE MINAS GERAIS

Ouro Preto

2026

Lucas Pereira Golia

Controle de qualidade e ajuste da Distribuição Generalizada de Valores Extremos para séries pluviométricas de Minas Gerais

Projeto Final de Curso apresentado como parte dos requisitos para obtenção do Título de Bacharel em Engenharia Urbana na Universidade Federal de Ouro Preto.

Orientador: Prof. D.Sc. Aline de Araújo Nunes - UFOP

Co-orientador: Prof. D.Sc. Éber José de Andrade Pinto

Ouro Preto

2026

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

G626c Golia, Lucas Pereira.
Controle de qualidade e ajuste da distribuição generalizada de valores extremos para séries pluviométricas de Minas Gerais. [manuscrito] / Lucas Pereira Golia. - 2026.
125 f.: il.: color., gráf., tab., mapa.

Orientadora: Profa. Dra. Aline de Araújo Nunes.
Coorientador: Prof. Dr. Éber José de Andrade Pinto.
Monografia (Bacharelado). Universidade Federal de Ouro Preto. Escola de Minas. Graduação em Engenharia Urbana .

1. Análise de séries temporais. 2. Controle de qualidade. 3. Teoria dos valores extremos. I. Nunes, Aline de Araújo. II. Pinto, Éber José de Andrade. III. Universidade Federal de Ouro Preto. IV. Título.

CDU 62:711.4

Bibliotecário(a) Responsável: Soraya Fernanda Ferreira e Souza - CRB6/2322

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE OURO PRETO
REITORIA
ESCOLA DE MINAS
DEPARTAMENTO DE ENGENHARIA URBANA

FOLHA DE APROVAÇÃO

Lucas Pereira Golia

Controle de qualidade e ajuste da Distribuição Generalizada de Valores Extremos para séries pluviométricas de Minas Gerais

Monografia apresentada ao Curso de Engenharia Urbana da Universidade Federal
de Ouro Preto como requisito parcial para obtenção do título de Engenheiro Urbanista

Aprovada em 24 de julho de 2026

Membros da banca

Doutora Aline de Araújo Nunes - Orientadora (Universidade Federal de Ouro Preto)
Doutor Éber José de Andrade Pinto - Coorientador (Serviço Geológico do Brasil/Universidade Federal de Minas Gerais)
Mestre Flávio Henrique Carvalho Cardoso (Agência Nacional de Mineração)
Doutora Bárbara Cristina Mendanha Reis (Universidade Federal de Ouro Preto)

Aline de Araújo Nunes, orientadora do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 04/08/2026

Documento assinado eletronicamente por **Aline de Araujo Nunes**, **PROFESSOR DE MAGISTERIO SUPERIOR**, em 04/08/2026, às 16:12, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0 , informando o código verificador **1152330** e o código CRC **1D955A10**.

R. Diogo de Vasconcelos, 122, - Bairro Pilar Ouro Preto/MG, CEP 35402-163
Telefone: (31)3559-1471 - www.ufop.br

AGRADECIMENTOS

Agradeço primeiramente aos meus pais, Alexandre e Genilda, por sempre me apoiarem e me incentivarem a continuar os estudos, além de me darem forças para seguir até atingir meus objetivos. Agradeço à minha irmã Bia, por sempre estar presente em todas as fases da minha trajetória.

Agradeço, em especial, à minha orientadora Aline Nunes, que fortaleceu meu interesse pelos estudos e despertou meu gosto pela hidrologia, através de suas aulas e orientações impecáveis. Sou grato pela confiança depositada em mim, além da disponibilidade e paciência ao longo das orientações. Agradeço também ao meu co-orientador Éber José, que contribuiu para minha trajetória com seu vasto conhecimento. Agradeço aos demais professores do Departamento de Engenharia Urbana, que tornam o curso acolhedor e de elevado nível técnico, e que me auxiliaram de tantas formas ao longo da minha trajetória.

Agradeço ao CNPq pela concessão da bolsa de iniciação científica, essencial para o desenvolvimento deste trabalho. O incentivo ao desenvolvimento científico e à formação de novos pesquisadores é indispensável para o avanço da ciência no país.

Por fim, agradeço à Universidade Federal de Ouro Preto e à Escola de Minas, pela infraestrutura, oportunidades e pelo ambiente de aprendizado que me proporcionaram uma formação da qual me orgulho.

Resumo

A confiabilidade das análises de frequência de precipitações extremas, que subsidiam estudos hidrológicos e dimensionamentos hidráulicos, depende diretamente da qualidade das séries pluviométricas, cujos registros frequentemente apresentam falhas e inconsistências. O desafio é agravado em Minas Gerais pela expressiva variabilidade climática e topográfica e pela heterogeneidade da rede de monitoramento. Este trabalho objetivou estabelecer um procedimento sistemático de controle de qualidade das séries pluviométricas do estado e, a partir das séries que atenderam aos critérios estabelecidos nesse processo, conduzir a análise regional de frequência das máximas diárias anuais, ajustando a distribuição Generalizada de Valores Extremos (GEV) e caracterizando o comportamento espacial de seu parâmetro de forma. Aplicou-se a metodologia de controle de qualidade de Llabrés-Brustenga et al. (2019), seguida da verificação de condição de Amostra Aleatória Simples e remoção de *outliers* mínimos e máximos, e o ajuste da GEV pelo método dos momentos-L. Complementarmente, avaliou-se o comportamento do parâmetro de forma por região homogênea no estado, por tamanho da série e por duração da chuva. O controle de qualidade absoluta revelou elevada consistência dos dados, com 89,88% das estações apresentando índice superior a 90, enquanto o índice relativo evidenciou a limitação central do método: a baixa densidade da rede mineira, exigindo adaptação. Com o conjunto de séries que atenderam aos critérios de controle de qualidade, verificou-se que todas apresentaram aderência à GEV, com predomínio da família Fréchet, com parâmetro de forma médio de -0,055. O agrupamento separou o estado em duas regiões contrastantes em altitude e precipitação, sem, contudo, distingui-las quanto ao parâmetro de forma. Verificou-se, ainda, tendência a valores mais negativos em séries longas e atenuação do predomínio da família Fréchet com o aumento da duração da chuva. A consistência de séries pluviométricas e o ajuste da GEV subsidiam estudos hidrológicos e projetos hidráulicos, consolidando a relevância deste estudo.

Palavras-chave: Séries de precipitação, Controle de qualidade, Distribuição Generalizada de Valores Extremos.

Abstract

The reliability of frequency analyses of extreme precipitation, which support hydrological studies and the design of hydraulic works, depends directly on the quality of rainfall series, whose records frequently present failures and inconsistencies. This challenge is intensified in Minas Gerais by the marked climatic and topographic variability and by the heterogeneity of the monitoring network. This work aimed to establish a systematic quality control procedure for the state's rainfall series and, from the quality-controlled series, to conduct a regional frequency analysis of the annual daily maxima, fitting the Generalized Extreme Value (GEV) distribution and characterizing the spatial behavior of its shape parameter. The quality control methodology of Llabrés-Brustenga et al. (2019) was applied, followed by a careful filtering through verification of the Simple Random Sample condition and the removal of low and high outliers, and then the GEV fitting by the L-moments method. In addition, the behavior of the shape parameter was assessed by homogeneous region within the state, by series length, and by rainfall duration. The absolute quality control revealed high internal consistency of the data, with 89.88% of the stations presenting an index above 90, whereas the relative index highlighted the method's central limitation given the low density of the Minas Gerais network, requiring adaptation. Once the set of quality-controlled series was established, all of them proved to fit the GEV, with a predominance of the Fréchet family, whose mean shape parameter was −0.055. The clustering separated the state into two regions contrasting in altitude and precipitation, without, however, distinguishing them regarding the shape parameter. A tendency toward more negative values in long series and an attenuation of the predominance of the Fréchet family with increasing rainfall duration were also observed. It is concluded that this study is relevant, since the consistency of rainfall series and the GEV fitting provide support for safer and more reliable hydrological studies and hydraulic projects.

Keywords: Precipitation series, Quality control, Generalized Extreme Value distribution.

Lista de Figuras

Figura 1: Distância mínima a obstáculos para a instalação do pluviômetro..... 18

Figura 2: Pluviômetro convencional..... 18

Figura 3: Proveta graduada..... 19

Figura 4: Modelo de pluviógrafo flutuador..... 20

Figura 5: Modelo de pluviograma, com o tempo registrado no eixo “x” e a altura de chuva no eixo “y”..... 20

Figura 6: Detalhe de uma EMA, com destaque ao pluviômetro de báscula, localizado na esquerda..... 22

Figura 7: Gráfico da curva dupla-massa, relacionando precipitação acumulada medida com precipitação acumulada padrão..... 27

Figura 8: Caudas das distribuições Weibull, Gumbel e Fréchet, em contraste com a distribuição Normal..... 43

Figura 9: Fluxograma das etapas metodológicas do trabalho..... 56

Figura 10: Localização de Minas Gerais no território brasileiro..... 57

Figura 11: Classificação de Thornthwaite em Minas Gerais..... 58

Figura 12: Altimetria de Minas Gerais..... 60

Figura 13: Declividade de Minas Gerais..... 60

Figura 14: Estações pluviométricas de Minas Gerais..... 63

Figura 15: Estrutura das planilhas contendo dados pluviométricos..... 63

Figura 16: Fluxograma do cálculo do índice de qualidade absoluta..... 65

Figura 17: Índice de qualidade absoluta para Minas Gerais..... 84

Figura 18: Índice de qualidade relativa para Minas Gerais considerando um raio de 50 km para estações auxiliares..... 86

Figura 19: Índice de qualidade relativa para Minas Gerais considerando um raio de 75 km para estações auxiliares..... 87

Figura 20: Diagrama de momentos-L para séries completas..... 93

Figura 21: Diagrama de momentos-L para séries sem outliers mínimos..... 93

Figura 22: Gráfico de silhueta..... 95
Figura 23: Distribuição espacial dos clusters..... 97
Figura 25: Distribuição do parâmetro de forma por região homogênea.....100
Figura 26: Histograma do parâmetro de forma com normal ajustada.....101
Figura 27: Histograma do parâmetro de forma com normal ajustada por região homogênea.....102
Figura 28: Intervalo de confiança de 95% da média do parâmetro de forma por tamanho da série.....103
Figura 29: Dispersão da média do parâmetro de forma por tamanho da série.....104
Figura 30: Percentual de estações que se ajustam à Fréchet e Weibull por tamanho da série.....105
Figura 31: Distribuição do parâmetro de forma por estudo.....107
Figura 32: Variação da média do parâmetro de forma por duração de chuva..... 109
Figura 33: Diagrama de momentos-L por duração de chuva..... 110
Figura 34: Distribuição do parâmetro de forma por duração de chuva..... 110

Lista de Tabelas

Tabela 1: Média e mediana dos parâmetros do índice de qualidade absoluta..... 84

Tabela 2: Percentuais de rejeição da hipótese nula nos testes de AAS..... 89

Tabela 3: Resumo dos parâmetros de posição e escala.....92

Tabela 4: Parâmetro de forma por cenário completo e sem outliers mínimos..... 94

Tabela 5: Características médias de cada cluster..... 96

Tabela 6: Estatísticas descritivas do parâmetro de forma..... 98

Tabela 8: Estatísticas descritivas do parâmetro de forma por trabalho.....106

Tabela 9: Estatísticas descritivas do parâmetro de forma por duração de chuva.... 109

Lista de abreviaturas e siglas

AAS - Amostra Aleatória Simples

ANA - Agência Nacional de Águas e Saneamento Básico

DP - Desvio-Padrão

EMA - Estação Meteorológica Automática

ETCCDI - Expert Team on Climate Change Detection and Indices

FDR - False Discovery Rate

GEV - Generalized Extreme Value

IBGE - Instituto Brasileiro de Geografia e Estatística

IDE-SISEMA - Infraestrutura de Dados Espaciais do Sistema Estadual de Meio Ambiente e Recursos Hídricos

IDF - Intensidade-Duração-Frequência

INMET - Instituto Nacional de Meteorologia

IPCC - Intergovernmental Panel on Climate Change

MGBT - Multiple Grubbs-Beck Test

MML - Método dos Momentos-L

NOAA - National Oceanic and Atmospheric Administration

PRCPTOT - Precipitation Total

RHM - Rede Hidrometeorológica Nacional

SGB - Serviço Geológico do Brasil

WMO - World Meteorological Organization

SUMÁRIO

1. Introdução..... 13

1.1 Objetivo..... 15

1.1.1 Objetivos Específicos..... 15

1.2 Justificativa..... 16

2 Revisão Bibliográfica..... 18

2.1 Precipitação e monitoramento hidrometeorológico..... 18

2.2 Erros e consistência em séries pluviométricas..... 24

2.3 Metodologias clássicas de controle de qualidade..... 28

2.4 Metodologia de Llabrés-Brustenga et al. (2019)..... 32

2.5. Aplicações e Importância dos Dados Consistidos..... 36

2.6. Eventos extremos de precipitação e teoria dos valores extremos..... 39

2.7. Distribuição GEV e interpretação do parâmetro de forma..... 41

2.8. Testes estatísticos aplicados a séries hidrológicas..... 44

2.9. Regionalização estatística e análise de agrupamentos..... 49

3 Metodologia..... 55

3.1 Caracterização da área de estudo..... 55

3.2 Seleção das estações e obtenção dos dados..... 60

3.3 Cálculo do índice de qualidade absoluta..... 62

3.4 Cálculo do índice de qualidade relativa..... 66

3.5 Modelagem estatística de valores extremos..... 69

3.5.1 Filtragem inicial e registro dos máximos anuais..... 69

3.5.2 Verificação da condição de Amostra Aleatória Simples (AAS)..... 70

3.5.3 Distribuição Generalizada de Valores Extremos (GEV)..... 74

3.5.4 Análise de cluster..... 76

3.5.5 Estatísticas do parâmetro de forma..... 78

3.5.6 Análise multidiária..... 79
4 Resultados e discussão.....81
4.1 Índice de qualidade absoluta..... 81
4.2 Índice de qualidade relativa..... 83
4.3 Modelagem estatística de valores extremos.....86
4.3.1 Filtragem inicial e registro dos máximos anuais.....86
4.3.2 Verificação da condição de Amostra Aleatória Simples (AAS)..... 87
4.3.3 Distribuição Generalizada de Valores Extremos (GEV)..... 89
4.3.4 Análise de agrupamento..... 92
4.3.5 Estatísticas do parâmetro de forma..... 94
4.3.5.1 Estatísticas globais e por cluster..... 94
4.3.5.2 Verificação da normalidade.....97
4.3.5.3 Análise por faixa de tamanho..... 99
4.3.5.4 Comparação com outros trabalhos.....102
4.3.6 Análise multidiária..... 104
4.3.7 Síntese dos resultados e implicações para os estudos hidrológicos..... 107
5 Conclusão..... 110

1. INTRODUÇÃO

A precipitação constitui uma das variáveis mais relevantes nos estudos hidrológicos e no planejamento e gestão dos recursos hídricos. A confiabilidade de análises estatísticas, da elaboração de curvas Intensidade-Duração-Frequência (IDF), de balanços hídricos e de modelagens hidrológicas depende diretamente da qualidade das séries pluviométricas. Esse vínculo é especialmente crítico em relação aos dados de precipitações máximas, que se concentram nos valores mais extremos das séries, sendo mais sensíveis a pequenas inconsistências nos registros, e cuja análise ampara o dimensionamento de obras de drenagem e o controle de cheias. Contudo, registros observados frequentemente apresentam lacunas, falhas operacionais e inconsistências de medição ou processamento, podendo comprometer as análises e introduzir incertezas em estudos técnicos e projetos de engenharia.

No estado de Minas Gerais, esses desafios assumem maior complexidade em razão da expressiva variabilidade climática e topográfica, associada à heterogeneidade da rede de monitoramento pluviométrico. A representatividade espacial das informações pode ser prejudicada quando não há uma avaliação prévia da consistência e confiabilidade das séries, sobretudo em regiões com baixa densidade de estações ou em aplicações que exigem séries históricas longas e contínuas. Nesse contexto, a ausência de um diagnóstico sistemático de qualidade pode impactar diretamente a tomada de decisão em estudos de drenagem urbana, controle de cheias, segurança hídrica e planejamento territorial.

Diante dessa problemática, metodologias de controle de qualidade de dados pluviométricos vêm sendo aprimoradas com o objetivo de assegurar padronização, reprodutibilidade e automatização dos procedimentos de verificação. Essas abordagens combinam testes internos, como análise de completude, verificação de plausibilidade e identificação de valores anômalos, com avaliações de coerência espacial entre estações vizinhas. Dessa forma, é possível quantificar a qualidade, por meio da interpretação da relação entre o total de registros em uma série e a quantidade de registros que não atendem a padrões de qualidade propostos.

Destaca-se, nesse cenário, a metodologia proposta por Llabrés-Brustenga et al. (2019), que organiza as verificações em uma sequência lógica e sintetiza os resultados por meio de índices de qualidade, permitindo avaliações quantitativas e comparáveis entre diferentes conjuntos de dados.

Esse controle de qualidade torna-se ainda mais relevante quando as séries se destinam à análise de eventos extremos, que se concentra nas caudas da distribuição, onde os valores mais raros, que são os mais influentes sobre as estimativas, são também os mais suscetíveis a erros. Para essa modelagem, a distribuição Generalizada de Valores Extremos (GEV) ocupa posição central por sua flexibilidade para representar diferentes regimes de máximas de precipitação, flexibilidade que decorre do parâmetro de forma, responsável por determinar o comportamento da cauda. Também é o parâmetro que exerce maior influência sobre os quantis associados a diferentes tempos de retorno, insumos para curvas IDF, dimensionamento hidráulico e controle de cheias. A estimativa do parâmetro de forma é fortemente influenciada pelos maiores valores da série, sendo sensível o bastante para que uma única inconsistência comprometa a inferência, justificando a necessidade da aplicação de um criterioso controle de qualidade à base de dados.

Assim, o presente trabalho tem como objetivo aplicar a metodologia de Llabrés-Brustenga et al. (2019) ao contexto do estado de Minas Gerais, buscando quantificar a qualidade das séries pluviométricas disponíveis e identificar possíveis limitações do método frente às especificidades regionais. A partir do conjunto de séries que atenderam aos critérios de controle de qualidade, procede-se ao ajuste da distribuição GEV às séries de precipitação máxima, etapa que permite transformar os dados selecionados em estimativas de eventos extremos.. Espera-se, com isso, contribuir para o aprimoramento da base de dados empregada em estudos hidrológicos no estado e disponibilizar parâmetros ajustados que subsidiem análises futuras com maior robustez técnica e menor incerteza.

1.1 Objetivo

Aplicar um procedimento sistemático de controle de qualidade das séries pluviométricas de Minas Gerais e, a partir das séries que atenderam aos critérios estabelecidos, conduzir a análise regional de frequência das precipitações máximas por ano hidrológico com durações de 1, 2, 3, 5 e 10 dias, ajustando a distribuição GEV, caracterizando o comportamento espacial do parâmetro de forma e identificando regiões homogêneas.

1.1.1 Objetivos Específicos

- Classificar a qualidade das séries de dados pluviométricos diários;
- Avaliar o atendimento das séries de máximas diárias anuais aos pressupostos de uma Amostra Aleatória Simples;
- Avaliar o comportamento da cauda da Distribuição GEV no Estado de Minas Gerais para as precipitações com durações de 1, 2, 3, 5 e 10 dias;
- Analisar o comportamento espacial do parâmetro de forma da distribuição GEV para a duração de 1 dia.

1.2 Justificativa

A qualidade das séries pluviométricas constitui um tema particularmente relevante por abordar um aspecto frequentemente negligenciado no planejamento e na gestão de recursos hídricos: a confiabilidade dos dados utilizados como base para análises e decisões técnicas. Em projetos de engenharia, é comum presumir que os dados disponíveis apresentam qualidade adequada, sem uma verificação crítica sobre sua origem, sua integridade ou os procedimentos de validação aos quais foram submetidos. Entretanto, o dimensionamento de obras hidráulicas, a interpretação do comportamento das chuvas e a calibração e validação de modelos hidrológicos dependem diretamente da representação fidedigna dos eventos observados.

No contexto de Minas Gerais e, de maneira mais ampla, do Brasil, a discussão torna-se ainda mais pertinente em razão da expressiva diversidade climática e topográfica do território, associada à distribuição heterogênea e, em muitos casos, insuficiente da rede de monitoramento meteorológico. Regiões com menor densidade de estações ou com históricos operacionais instáveis tendem a apresentar maiores incertezas quanto à representatividade espacial e temporal das séries. Nesse cenário, a aplicação de uma metodologia estruturada de controle de qualidade permite identificar fragilidades da rede, detectar padrões de inconsistência e levantar hipóteses sobre suas possíveis causas, fornecendo subsídios técnicos para o aprimoramento do monitoramento hidrometeorológico.

Ao aplicar essa metodologia ao estado de Minas Gerais, busca-se não apenas avaliar a qualidade das séries disponíveis, mas também testar a aderência do procedimento ao contexto regional e estabelecer um protocolo sistemático e replicável. Tal abordagem pode ser ampliada e adaptada a outras regiões do país, contribuindo para a promoção de maior transparência, robustez e consistência no uso de dados pluviométricos em estudos hidrológicos, projetos de engenharia e políticas públicas relacionadas à gestão de recursos hídricos.

Para além dessa contribuição metodológica, a aprovação das séries nos critérios de controle de qualidade constitui uma condição necessária, mas não suficiente para

sua aplicação. Verificar a confiabilidade dos dados representa apenas a etapa inicial, pois sua utilidade se concretiza quando eles são empregados na obtenção de informações capazes de subsidiar decisões técnicas, como a estimativa da magnitude e da frequência de precipitações extremas utilizadas em projetos de engenharia.

É nesse ponto que se insere a distribuição Generalizada de Valores Extremos (GEV), um dos modelos mais consolidados para descrever eventos extremos e estimar a magnitude associada a diferentes probabilidades de ocorrência. Aplicá-la de forma sistemática às séries previamente consistidas de Minas Gerais permite caracterizar o regime das precipitações máximas em todo o estado e mapear como ele varia no espaço, informação hoje escassa para um território de grande diversidade climática e topográfica, em que não se espera um comportamento uniforme das chuvas extremas.

Dessa forma, a contribuição do trabalho estende-se para além do diagnóstico da qualidade dos dados. Os parâmetros estimados e o mapeamento de seu comportamento no estado constituem uma base de referência para estudos posteriores, como a elaboração de curvas IDF e a definição de chuvas de projeto, na perspectiva de fornecer recomendações sobre o emprego da GEV em Minas Gerais.

2 Revisão Bibliográfica

Esta revisão bibliográfica apresenta, de forma integrada, os fundamentos teóricos e técnicos que sustentam o estudo. Inicialmente, discute-se a precipitação e o monitoramento hidrometeorológico no Brasil, abordando instrumentos, redes de medição e padronização dos dados. Em seguida, são analisadas as principais fontes de erro em séries pluviométricas e a necessidade de controle de qualidade. Na sequência, revisam-se as metodologias clássicas e recentes de consistência de dados, com destaque para a abordagem de Llabrés-Brustenga et al. (2019), adotada neste trabalho, além das aplicações e importância de séries consistidas para a gestão de recursos hídricos e o planejamento territorial. Então, são apresentados testes estatísticos para análise de eventos extremos de precipitação e sua base teórica, destacando-se a Distribuição Generalizada de Valores Extremos (GEV).

2.1 Precipitação e monitoramento hidrometeorológico

De acordo com a Política Nacional de Recursos Hídricos, instituída pela Lei nº 9.433/1997, a água é reconhecida como um recurso natural limitado, dotado de valor econômico, cujo uso prioritário deve atender ao consumo humano e à dessedentação de animais (Brasil, 1997). Nesse contexto, a compreensão do regime de precipitação assume papel central na gestão dos recursos hídricos, uma vez que a chuva constitui a principal forma de reposição natural dos reservatórios superficiais e subterrâneos responsáveis pelo abastecimento da sociedade e pela manutenção dos ecossistemas.

Segundo a World Meteorological Organization (WMO, 2024), a precipitação corresponde ao produto líquido ou sólido da condensação do vapor d'água presente na atmosfera, que atinge a superfície terrestre sob a forma de chuva, neve, granizo, entre outros. As nuvens, definidas como aglomerados de partículas microscópicas de água líquida ou gelo suspensas no ar, formam-se a partir da evaporação da água superficial e da transpiração das plantas (WMO, 1975). O processo de precipitação ocorre quando as gotículas ou cristais de gelo adquirem massa suficiente para superar a resistência do ar e a sustentação atmosférica, deslocando-se em direção à

superfície (NOAA, 2025).

Ao atingir o solo, a precipitação representa o principal componente de entrada nos balanços hídricos e constitui o primeiro parâmetro considerado na análise dos sistemas hídricos superficiais e subterrâneos (Tucci, 2007). O conhecimento do regime pluviométrico, associado às características geotécnicas, geomorfológicas e de uso e ocupação do solo, possibilita a estimativa de processos como infiltração, escoamento superficial e recarga aquífera. Esses elementos são fundamentais para a elaboração de modelos chuva-vazão empregados no dimensionamento de sistemas de micro e macrodrenagem urbana, bem como para o planejamento de obras hidráulicas (Botto, 2016; El Hachem et al., 2022). Ademais, a análise contínua das séries de precipitação contribui para a previsão e mitigação de desastres hidrológicos, como enxurradas, alagamentos e inundações, além da identificação de tendências associadas a períodos de seca e estiagem.

A caracterização adequada do regime pluviométrico depende da existência de um sistema de monitoramento hidrometeorológico abrangente, contínuo e confiável, baseado em estações distribuídas espacialmente de forma representativa. No Brasil, o monitoramento hidrológico é coordenado pela Agência Nacional de Águas e Saneamento Básico (ANA), enquanto a operação, instalação e manutenção das estações são majoritariamente conduzidas pelo Serviço Geológico do Brasil (SGB), em parceria com outros órgãos federais, estaduais e entidades privadas.

Conforme diretrizes da WMO (2008), a precipitação é expressa pela espessura da lâmina d’água que recobriria uma superfície horizontal, sendo sua medição realizada por meio de pluviômetros devidamente instalados. Para que as medições sejam representativas, a instalação deve minimizar os efeitos do vento e evitar a interferência de obstáculos que possam comprometer a captação das gotas de chuva. No caso do pluviômetro convencional, a distância mínima em relação a obstáculos e a altura adequada de instalação são ilustradas na Figura 1, conforme recomendações da ANA (2016).

O pluviômetro convencional, apresentado na Figura 2, é geralmente fabricado em aço inoxidável, possui área de captação de 400 cm² e reservatório com

capacidade de até 20 litros, sendo destinado à coleta e ao armazenamento da água precipitada para posterior leitura por meio de proveta graduada, conforme ilustrado na Figura 3 (ANA, 2016). A operação desse instrumento é manual, realizada por um observador local que, em geral, efetua a leitura diariamente, registrando a precipitação acumulada nas 24 horas anteriores.

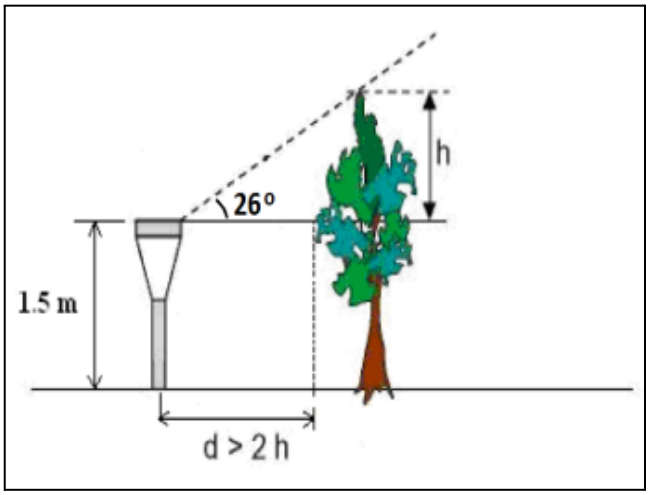


Figura 1: Distância mínima a obstáculos para a instalação do pluviômetro.

Fonte: ANA (2016)



Figura 2: Pluviômetro convencional.

Fonte: ANA (2016).



Figura 3: Proveta graduada.

Fonte: ANA (2016).

Outra forma de monitoramento da precipitação é realizada por meio de pluviógrafos, instrumentos capazes de registrar não apenas o volume total precipitado, mas também a duração e a intensidade da chuva ao longo do tempo (Rocha, 2022). Um exemplo de pluviógrafo do tipo flutuador é apresentado na Figura 4. Diferentemente do pluviômetro convencional, que fornece valores diários acumulados, o pluviógrafo gera registros contínuos representados graficamente em pluviogramas, como o ilustrado na Figura 5, permitindo a análise detalhada da distribuição temporal da precipitação e o cálculo da intensidade de chuva em diferentes intervalos de tempo.



Figura 4: Modelo de pluviógrafo flutuador.

Fonte: Rocha (2022).

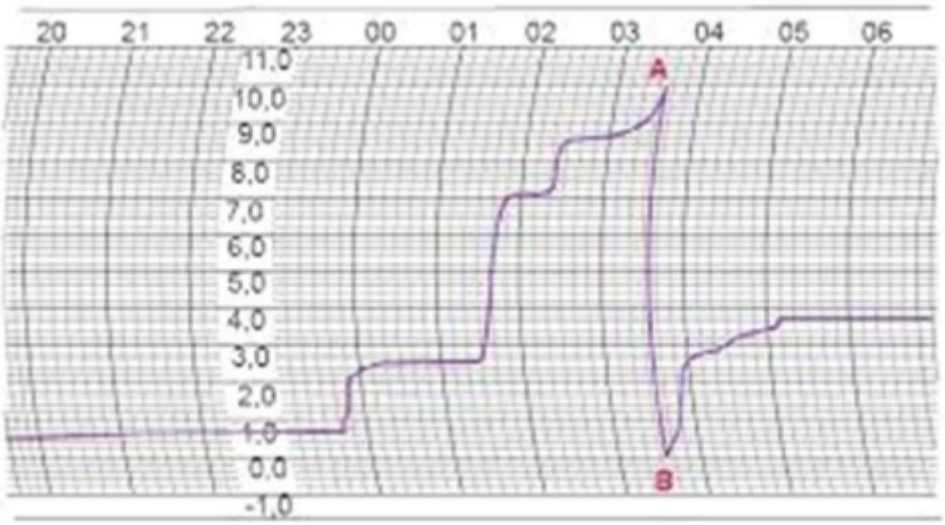


Figura 5: Modelo de pluviograma, com o tempo registrado no eixo “x” e a altura de chuva no eixo “y”.

Fonte: Rocha (2022).

No Brasil, a padronização, organização e disponibilização dos dados hidrometeorológicos são atribuições da ANA, conforme estabelecido pela Lei nº 9.984/2000, com alterações introduzidas pela Lei nº 14.026/2020 (Brasil, 2000). Nesse contexto, o Sistema Nacional de Informações sobre Recursos Hídricos (SNIRH) tem como objetivo integrar, dar consistência e tornar acessíveis dados referentes aos recursos hídricos do país. As informações provenientes da Rede Hidrometeorológica Nacional (RHN) são disponibilizadas por meio da plataforma HidroWeb, que reúne séries históricas de dados pluviométricos e fluviométricos, incluindo registros que remontam ao ano de 1855, com destaque para a estação pluviométrica Morro Velho, em Minas Gerais.

A Rede Hidrometeorológica Nacional é operada pelo SGB, sob coordenação da ANA, e conta com a colaboração de diversas instituições públicas e privadas. Em 2020, a RHN possuía aproximadamente 23.000 estações em operação, das quais cerca de 2.717 eram pluviométricas, conforme dados da ANA (2023).

Além da ANA e do SGB, o Instituto Nacional de Meteorologia (INMET) desempenha papel relevante no monitoramento meteorológico no território brasileiro. Criado em 1909, o INMET é o órgão oficial de meteorologia do país, vinculado ao Ministério da Agricultura e Pecuária (Brasil, 2025). Sua rede é composta por estações meteorológicas convencionais, mas também automáticas (EMA), sendo que as últimas realizam medições em alta resolução temporal, coletando dados a cada minuto e transmitindo-os periodicamente para a central. Um exemplo de EMA, com destaque para o pluviômetro de báscula, é apresentado na Figura 6 (INMET, 2011).



Figura 6: Detalhe de uma EMA, com destaque ao pluviômetro de báscula, localizado na esquerda.

Fonte: INMET (2011).

As EMAs utilizam, em geral, pluviômetros de báscula, que funcionam a partir do basculamento de compartimentos de volume conhecido, registrando a precipitação de forma automática e contínua (INMET, 2011). Esse tipo de instrumento apresenta maior resolução temporal, sendo amplamente empregado em estudos hidrológicos, meteorológicos e de drenagem urbana.

Por fim, ressalta-se que a qualidade, a consistência e a continuidade das séries históricas de precipitação são fundamentais para a confiabilidade das análises hidrológicas e do dimensionamento de estruturas hidráulicas (Botto, 2016; El Hachem et al., 2022). Séries curtas, dados faltantes ou medições inconsistentes aumentam as incertezas associadas aos modelos chuva-vazão e às estimativas de eventos extremos, podendo resultar tanto na subestimação de riscos hidrológicos quanto no superdimensionamento de obras, com implicações técnicas, econômicas e ambientais significativas.

2.2 Erros e consistência em séries pluviométricas

Conforme já destacado, a disponibilidade de séries temporais de precipitação confiáveis é um elemento central para o planejamento e o desenvolvimento das cidades (Botto, 2016). Nesse contexto, a identificação e a análise de erros associados às séries pluviométricas constituem etapas indispensáveis do controle de qualidade dos dados. Tal necessidade decorre do fato de que, mesmo em redes extensas operadas por equipes capacitadas, as estações de medição não estão imunes a falhas de diferentes naturezas.

Erros que comprometem a qualidade das séries podem gerar impactos significativos na gestão dos recursos hídricos. Registros sistematicamente discrepantes em relação à precipitação real podem conduzir à construção de modelos hidrológicos incoerentes com a realidade física, ou até inviabilizar sua elaboração, afetando diretamente o dimensionamento de sistemas de micro e macrodrenagem. Da mesma forma, estudos de frequência de eventos extremos, sistemas de alerta de desastres e o planejamento de medidas mitigatórias podem tornar-se enviesados ou incorretos, dificultando a tomada de decisão por parte do poder público e o planejamento de ações futuras. Deve-se reconhecer que redes de monitoramento de grande porte, como a mantida pela Agência Nacional de Águas e Saneamento Básico (ANA), que conta com aproximadamente 23.000 estações, são naturalmente suscetíveis a falhas, ainda que adequadamente instaladas e operadas. Segundo a ANA (2016), as falhas instrumentais figuram entre as mais recorrentes, mas erros operacionais e humanos também são frequentes, reforçando a importância de sistemas de manutenção e controle eficientes.

As falhas instrumentais podem estar associadas à instalação inadequada dos equipamentos, seja por descumprimento das normas técnicas estabelecidas pela ANA, seja pela execução por pessoal sem a devida capacitação. Incluem-se nesse grupo erros como: instalação do pluviômetro inclinado; posicionamento muito próximo ao solo; localização em áreas excessivamente expostas ao vento; proximidade de obstáculos como árvores e edificações; e perdas por evaporação da água coletada (ANA, 2016). Além disso, mesmo quando a instalação inicial é adequada, a ausência de manutenção periódica pode levar ao surgimento dessas

condições ao longo do tempo, em função do crescimento da vegetação ou da urbanização do entorno. Tais falhas podem resultar em medições incoerentes com a precipitação real, uma vez que ventos e obstáculos tendem a desviar ou obstruir a chuva. De acordo com a World Meteorological Organization (WMO, 2020), a ação do vento pode reduzir em até 10% o volume de chuva registrado, enquanto perdas por evaporação podem alcançar cerca de 4%.

As falhas operacionais, por sua vez, afetam principalmente estações automáticas dotadas de sistemas de telemetria, responsáveis pela transmissão contínua dos dados para centros de armazenamento (CEIVAP, 2025). Essas falhas variam conforme o tipo de estação e podem incluir problemas no sinal de internet ou rádio, danos às antenas, interrupções no fornecimento de energia elétrica, além de falhas de hardware e software. Em muitos casos, tais problemas inviabilizam o registro dos dados até que sejam solucionados, gerando lacunas nas séries temporais e, em situações críticas, atrasando a emissão de alertas de desastres.

Já as falhas humanas estão associadas, predominantemente, às estações de medição manual e tendem a ser mais difíceis de identificar, dada a diversidade de situações que podem ocasioná-las, como desatenção, falta de treinamento ou imprevistos durante a leitura e o registro dos dados (ANA, 2016). Entre os erros mais comuns destacam-se: leitura incorreta do menisco; realização da leitura fora do horário padronizado; ausência de leitura em determinado dia, com posterior anotação de valor estimado ou registro de zero; erros de transcrição, como troca de algarismos ou unidades (mm por cm); nivelamento inadequado do pluviômetro manual; não esvaziamento completo do recipiente antes de nova leitura; respingos de água para fora do equipamento; além de erros na digitalização dos dados, como a inserção de zeros excedentes ou o registro em datas incorretas (ANA, 2016; Hamada et al., 2011; WMO, 2008). Observa-se, portanto, que os erros humanos apresentam elevada variabilidade e, por sua natureza, são mais complexos de prevenir e detectar (Hamada et al., 2011).

Diante da multiplicidade de fontes de erro, torna-se evidente que, além da manutenção contínua realizada por equipes capacitadas, a implementação de processos sistemáticos de detecção de falhas é fundamental para reduzir

inconsistências e aprimorar a qualidade dos dados pluviométricos. Nesse sentido, foram desenvolvidas metodologias específicas para avaliar a qualidade das séries, identificar inconsistências e reconhecer estações com elevado número de falhas, fornecendo subsídios para o aperfeiçoamento dos sistemas de monitoramento.

O primeiro passo para assegurar a qualidade das séries é a garantia da rastreabilidade dos dados. Segundo a WMO (2020), os dados de precipitação são relativamente caros de serem obtidos, insubstituíveis e desempenham papel crucial na análise de eventos extremos. Assim, deve haver rastreabilidade em todas as etapas do processo de coleta, de modo a assegurar a confiabilidade e a integridade das séries. Informações sobre o instrumento utilizado, o operador responsável pela estação e o método de coleta devem ser disponibilizadas, bem como a preservação dos registros originais em cadernetas ou formulários. Ademais, quaisquer alterações realizadas nos dados, como correções de erros evidentes, devem ser devidamente documentadas, permitindo a identificação posterior das modificações e aumentando a confiabilidade das informações.

De forma complementar, a WMO (2020) destaca que um de seus principais objetivos é a promoção da padronização dos métodos de obtenção e divulgação de dados hidrometeorológicos. Entre os elementos fundamentais dessa padronização está a adoção uniforme de unidades e símbolos reconhecidos internacionalmente, conforme suas normas técnicas, como milímetros (mm) para precipitação e metros cúbicos por segundo (m^3/s) para vazão. A instalação e a operação das estações também devem seguir padrões definidos, considerando aspectos como altura do pluviômetro, distância mínima de obstáculos, tipo de instrumento, horário e procedimento de leitura, além da realização de visitas regulares para manutenção.

Nesse contexto, destaca-se que a Agência Nacional de Águas e Saneamento Básico (ANA, 2016) disponibiliza um manual de procedimentos para a instalação, operação e manutenção de estações pluviométricas convencionais, no qual são detalhados critérios técnicos relacionados às dimensões, aos materiais e aos equipamentos empregados. Tal iniciativa reforça o papel institucional da ANA na promoção da rastreabilidade e da padronização dos dados pluviométricos, além de orientar quanto às boas práticas necessárias para a identificação e a correção de

falhas ao longo do processo de coleta.

Entretanto, considerando a extensão das redes de monitoramento, o volume de dados gerados e a diversidade de fontes potenciais de erro, torna-se inviável assegurar a qualidade das séries exclusivamente por meio de procedimentos operacionais e de manutenção em campo. Nesse sentido, as metodologias de análise de consistência de dados pluviométricos surgem como ferramentas complementares e indispensáveis, permitindo verificar, de forma sistemática, automatizada e em larga escala, se as diretrizes técnicas estabelecidas estão sendo efetivamente atendidas, além de possibilitar a detecção de inconsistências, erros pontuais e padrões anômalos nos registros.

Dessa forma, a consolidação de métodos formais de controle de qualidade representa um avanço essencial para garantir a confiabilidade das séries históricas e sua adequada utilização em estudos hidrológicos e aplicações de engenharia. A seguir, são apresentadas e discutidas as principais metodologias clássicas de controle de qualidade de dados pluviométricos, que fundamentaram o desenvolvimento das abordagens mais recentes e subsidiaram a metodologia adotada neste trabalho.

2.3 Metodologias clássicas de controle de qualidade

Dados pluviométricos no Brasil vêm sendo coletados de forma sistemática desde 1855 (ANA, [2025]). Entretanto, as metodologias destinadas à verificação da integridade e da confiabilidade dessas informações foram desenvolvidas de maneira mais recente. Um dos primeiros estudos voltados à análise da consistência de séries pluviométricas foi o de Searcy e Hardison (1960), que introduziu o método da curva de dupla-massa para a comparação de séries anuais de precipitação provenientes de diferentes estações.

Nesse método, a precipitação acumulada ao longo do tempo em uma estação é comparada à precipitação acumulada considerada padrão, obtida a partir de estações próximas, geralmente inseridas na mesma bacia hidrográfica. Essa relação é representada graficamente, conforme ilustrado na Figura 7, por meio da análise do coeficiente angular das retas formadas (Searcy; Hardison, 1960). Espera-se que

estações geograficamente próximas apresentem coeficientes angulares semelhantes; assim, alterações significativas nesse coeficiente podem indicar mudanças no equipamento, no local de instalação ou nos procedimentos de leitura.

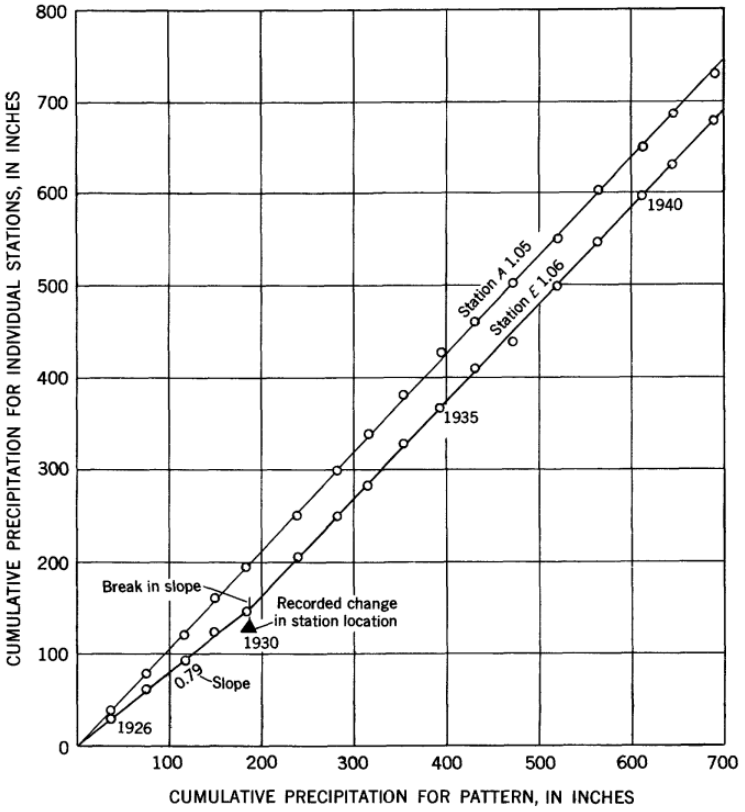


Figura 7: Gráfico da curva dupla-massa, relacionando precipitação acumulada medida com precipitação acumulada padrão.

Fonte: Searcy e Hardison (1960).

A metodologia da curva de dupla-massa (Searcy; Hardison, 1960) mostra-se eficiente na detecção de alterações instrumentais e operacionais em estações pluviométricas, porém apresenta limitações importantes, uma vez que se restringe à análise de dados anuais e não permite a identificação de erros pontuais ou sutis em escalas temporais menores. Em função dessas limitações, métodos baseados em testes estatísticos passaram a ser propostos, com destaque para o trabalho de

Alexandersson (1986). Nesse estudo, é definida a razão q_i, obtida pela divisão da média anual de precipitação de uma estação pela média regional correspondente. Para séries consistentes, espera-se que os valores de q_i permaneçam próximos de 1. Essas razões são posteriormente padronizadas, resultando na série z_i, com média zero e desvio padrão unitário. A série z_i é então submetida a testes estatísticos para verificar a existência de mudanças bruscas na média ao longo do tempo. Valores da estatística de teste superiores aos limites críticos indicam não homogeneidade da série, enquanto valores inferiores sugerem maior confiabilidade para análises climáticas.

Diferentemente da curva de dupla-massa, cuja interpretação é essencialmente subjetiva, uma vez que não há critérios objetivos que definam quando uma variação no coeficiente angular é suficientemente significativa, a metodologia proposta por Alexandersson (1986) introduz maior objetividade ao estabelecer limiares estatísticos claros que subsidiam a tomada de decisão. Ambas as abordagens serviram como base para o desenvolvimento de métodos mais avançados, incorporando novos parâmetros e ampliando o grau de automatização das análises.

Nesse contexto, González-Rouco et al. (2001) propuseram um procedimento sistemático para o tratamento de séries mensais de precipitação, estruturado em um fluxo reprodutível de controle de qualidade e homogeneização, implementado por meio de códigos computacionais. Inicialmente, são aplicados testes absolutos, que avaliam os dados individualmente, sem comparação com estações vizinhas. Esses testes são executados de forma automática e atribuem flags (aceito, suspeito ou rejeitado) com base em critérios como: (i) plausibilidade e formato dos registros, identificando valores negativos, datas inválidas, duplicidades e inconsistências de unidades; (ii) completude temporal, detectando lacunas e meses não informados ou registrados como zero; e (iii) limites climatológicos mensais, utilizados para identificar extremos atípicos incompatíveis com o regime local, geralmente definidos a partir de estatísticas como quartis. O objetivo central é isolar erros grosseiros e anomalias internas, preservando os valores originais e documentando explicitamente

as decisões por meio de flags, sem remoção direta dos dados.

Uma contribuição relevante para a caracterização de erros associados à acumulação de precipitação em múltiplos dias foi apresentada por Viney e Bates (2004). Os autores demonstraram que séries diárias frequentemente apresentam totais acumulados de vários dias registrados em uma única data, configurando o padrão típico de sequências de zeros seguidas por um pico anômalo. Esse comportamento é comum em estações de leitura manual, especialmente quando observações não são realizadas diariamente, como em finais de semana. Tal distorção, identificada inclusive em conjuntos de dados considerados de alta qualidade, compromete a análise de extremos e a comparabilidade entre estações. A caracterização desse tipo de erro subsidiou o aprimoramento de metodologias posteriores, inclusive da abordagem adotada no presente estudo.

As diferentes propostas apresentadas convergem para a metodologia sistematizada por Einfalt e Michaelides (2008), que publicaram um guia abrangente de controle de qualidade para dados de precipitação no livro Quality Control of Precipitation Data. Diferentemente de estudos de caso específicos, esse trabalho configura-se como um manual de referência, definindo quais testes devem ser aplicados, em que ordem e de que forma os dados devem ser classificados por índices de qualidade (aceito, suspeito ou rejeitado), preservando sempre o valor original e a rastreabilidade das decisões, segundo o princípio de “marcar, não apagar”.

No escopo desse processo, os valores analisados correspondem a registros diários ou subdiários de precipitação, expressos em milímetros, provenientes de leituras manuais ou de pluviógrafos, acompanhados de metadados como localização, altitude, tipo de instrumento, horário de leitura e histórico de manutenção. A sequência metodológica envolve: (i) checagens básicas de plausibilidade, visando à identificação de erros grosseiros; (ii) análise da consistência temporal dentro da própria estação, incluindo a detecção de acumulações multi-diárias, sequências improváveis de zeros e deslocamentos de data; (iii) verificação da consistência espacial, por meio da comparação com estações vizinhas, utilizando técnicas como buddy checks e análises estatísticas

robustas; e (iv) verificação multisensorial, quando disponível, integrando informações de radares e pluviômetros. Em todas as etapas, os dados são apenas sinalizados, com registro explícito dos testes aplicados e dos limiares utilizados, os quais demandam calibração regional.

Todas essas abordagens contribuíram de forma decisiva para o desenvolvimento de metodologias cada vez mais abrangentes, reprodutíveis e aplicáveis a grandes bases de dados. A metodologia proposta por Llabrés-Brustenga et al. (2019), adotada no presente estudo, fundamenta-se na integração desses avanços, organizando-os em um esquema sistemático de controle de qualidade. Essa abordagem combina verificações básicas, testes absolutos e testes relativos, aplicados de forma sequencial e passíveis de automatização, permitindo padronizar o tratamento das séries pluviométricas, aumentar a transparência dos critérios adotados e viabilizar sua aplicação em extensos conjuntos de estações, como o analisado neste trabalho.

2.4 Metodologia de Llabrés-Brustenga et al. (2019)

O trabalho de Llabrés-Brustenga et al. (2019) representa um marco recente no desenvolvimento de metodologias de controle de qualidade aplicadas a séries diárias de precipitação, destacando-se pela clara inspiração e integração de métodos consagrados na literatura. Os autores aplicaram e validaram um procedimento inovador sobre um volume expressivo de dados, superior a 100 milhões de registros diários provenientes de 1.726 estações pluviométricas distribuídas pela Catalunha, abrangendo o período de 1855 até o presente (Llabrés-Brustenga et al., 2019). A proposta baseia-se em um controle progressivo da qualidade, estruturado em etapas sucessivas de detecção de erros grosseiros, avaliação da qualidade absoluta das séries e, por fim, análise da qualidade relativa por meio da comparação com estações auxiliares.

Para a avaliação da qualidade absoluta, foi definido um índice global Q, variando de 0 a 100, composto por cinco subíndices: P, que representa o percentual de dados disponíveis; $Qgaps$, relacionado à presença e à distribuição temporal de lacunas; Qm_{zero}, voltado à detecção de meses inteiros com precipitação nula,

indicativos de possíveis falsos zeros; Qw_{zero}, que avalia a existência de padrões semanais anômalos associados a acumulações não identificadas; e $Qoutliers$, responsável pela identificação de valores extremos potencialmente espúrios (Llabrés-Brustenga et al., 2019).

O índice de qualidade absoluta funciona como um resumo numérico abrangente da qualidade da série, ao reunir, em um único valor, diferentes dimensões do controle de qualidade: completude dos dados (P), continuidade temporal ($Qgaps$), identificação de registros fisicamente improváveis (Qm_{zero} e Qw_{zero}) e detecção de extremos atípicos ($Qoutliers$). Ao condensar essas informações em uma escala de 0 a 100, torna-se possível a definição de limiares objetivos de aceitação. Llabrés-Brustenga et al. (2019) estabelecem que valores de Q superiores a 80 indicam séries com qualidade absoluta aceitável, fornecendo embasamento estatístico para a classificação das estações, reduzindo a subjetividade das avaliações e ampliando a transparência e a reprodutibilidade do procedimento metodológico.

Observa-se que o índice de qualidade absoluta sintetiza, em um fluxo unificado, contribuições centrais de estudos anteriores. Estão incorporados os princípios de plausibilidade dos dados, completude temporal, identificação de lacunas e detecção de valores extremos discutidos por González-Rouco et al. (2001), bem como a caracterização de acumulações multi-diárias destacada por Viney e Bates (2004). Soma-se a isso a organização metodológica proposta por Einfalt e Michaelides (2008), que enfatizam a definição clara da ordem dos testes e a rastreabilidade das decisões, reforçando a sistematização e a reprodutibilidade do processo.

Além da avaliação absoluta, Llabrés-Brustenga et al. (2019) incorporam um componente de qualidade relativa, fundamentado na comparação das séries de cada estação com aquelas registradas por estações vizinhas. Para tal, são definidas estações auxiliares localizadas dentro de um raio de até 50 km, sem restrição explícita de diferença altimétrica. A influência da distância entre as estações é considerada por meio de uma penalização exponencial da correlação, de modo que

pares mais próximos geograficamente apresentam maior peso na avaliação, assumindo-se maior similaridade das condições pluviométricas.

O procedimento consiste na comparação diária entre a diferença de precipitação observada em dois postos e um limite dependente do mês. Quando a diferença não excede esse limite, o registro é classificado como “válido”; caso contrário, é considerado “inválido” (Llabrés-Brustenga et al., 2019). Esse processo é repetido sistematicamente ao longo da série, permitindo a identificação de registros discrepantes e incoerentes em relação ao comportamento regional esperado.

A partir dessa etapa, o índice de qualidade relativa é estimado por dia, sendo que a diferença relativa dos registros de precipitação é calculada para a principal e cada uma de suas auxiliares, posteriormente comparando-as com o limite mensal. Após as comparações, os dias são marcados de acordo com a quantidade de comparações válidas, ou seja, quando a diferença relativa é menor que o limite. O índice de qualidade relativa corresponde, assim, à quantidade de dias, para cada estação, que a comparação com suas auxiliares foi válida. (Llabrés-Brustenga et al., 2019).

O índice de qualidade relativa é particularmente relevante por explorar a redundância espacial da rede pluviométrica, permitindo avaliar não apenas a coerência interna dos dados de uma estação, mas também sua consistência em relação ao entorno. Ao confrontar diferenças diárias de precipitação com limites mensais calibrados, o método é capaz de identificar erros sutis que dificilmente seriam detectados por testes absolutos, como acumulações não identificadas, deslocamentos de data ou falhas pontuais de leitura que resultam em valores fisicamente plausíveis, porém espacialmente incoerentes. Dessa forma, torna-se possível classificar as estações segundo sua coerência espacial, destacando aquelas mais adequadas para análises regionais e modelagem hidrológica, bem como aquelas que demandam calibração ou investigação adicional, fortalecendo o controle de qualidade em escala de rede.

Em comparação com metodologias clássicas, a proposta de Llabrés-Brustenga et al. (2019) pode ser interpretada como uma evolução mais abrangente e operacionalmente robusta das ideias introduzidas por Searcy e

Hardison (1960) e Alexandersson (1986). Enquanto a curva de dupla-massa atua em escala anual, com foco na detecção de mudanças estruturais entre séries acumuladas, e o teste de Alexandersson introduz critérios estatísticos de homogeneidade também em base anual, a metodologia de Llabrés-Brustenga avança ao trabalhar com resolução diária. Conceitualmente, mantém-se o princípio da comparação entre a série candidata e um comportamento regional de referência, porém ampliado pela incorporação explícita da representatividade de cada par de estações, o que resulta em maior refinamento e confiabilidade estatística.

A aplicação dessa metodologia, originalmente desenvolvida para a realidade catalã, impõe desafios relevantes quando transposta ao contexto brasileiro, em especial no estado de Minas Gerais, foco deste estudo. O território brasileiro é extenso e apresenta regimes pluviométricos altamente heterogêneos, influenciados por diferentes biomas, variações de altitude e latitude, o que resulta em médias e extremos de precipitação bastante distintos (IBGE, 2002). Ademais, a rede de monitoramento brasileira é significativamente menos densa e mais irregular quando comparada à da Catalunha, o que exige a adequação dos parâmetros da metodologia às condições regionais, refletindo as particularidades climáticas e operacionais do país.

A Catalunha possui área aproximada de 32.108 km² (Catalunya, 2017) e, considerando as 1.726 estações utilizadas por Llabrés-Brustenga et al. (2019), apresenta uma densidade de cerca de 54 estações por 1.000 km². Em contraste, Minas Gerais possui área de 586.513,984 km² (IBGE, 2024) e aproximadamente 652 estações pluviométricas, o que resulta em uma densidade de apenas 1,1 estação por 1.000 km², cerca de 49 vezes menor. Essa diferença é crucial para a aplicabilidade da metodologia, uma vez que a qualidade relativa depende diretamente da quantidade e da distribuição espacial das estações auxiliares disponíveis dentro de um determinado raio de influência.

Além da densidade da rede, a declividade acentuada do relevo mineiro, especialmente na região Sul do estado, deve ser considerada (IDE-Sisema, 2018). O relevo montanhoso exerce forte influência orográfica sobre o regime de chuvas, de modo que estações relativamente próximas, porém localizadas em diferentes

cotas altimétricas ou faces de uma mesma cadeia montanhosa, podem registrar padrões pluviométricos significativamente distintos (Roe, 2005).

Também se destaca a diferença entre os regimes climáticos da Catalunha e de Minas Gerais. Enquanto Minas Gerais apresenta ampla variabilidade de condições de umidade, com regiões superúmidas ao sul e semiáridas ao norte (IDE-Sisema, 2018), a Catalunha também exibe contrastes climáticos (Catalunya, 2014), porém amparados por uma rede de monitoramento muito mais densa, o que aumenta a confiabilidade das comparações espaciais. Assim, os parâmetros utilizados para definir a representatividade geográfica entre estações devem ser cuidadosamente analisados e, quando necessário, recalibrados, de forma a adequar a metodologia às condições específicas do território mineiro.

Mesmo com as possíveis limitações em regiões com diferentes regimes climáticos, a metodologia foi aplicada com sucesso à regiões distintas. Estévez et al. (2022) transferiram o procedimento completo para a região semiárida da Andaluzia, no sul da Espanha, aplicando as três etapas do método a 1.947 estações pluviométricas de três redes de monitoramento distintas, com limiares recalibrados a partir dos dados locais. Os resultados demonstraram que a metodologia é transferível a contextos climáticos distintos do original, desde que seus parâmetros sejam reajustados às características da região de aplicação.

Na América Latina, Terán-chaves, Duarte-carvajalino e Polo-murcia (2021) adotaram o índice de qualidade absoluto de Llabrés-brustenga et al. (2019) para avaliar as séries diárias da rede nacional colombiana do IDEAM, sem implementar o controle relativo. Mesmo em um contexto climático e orográfico marcadamente distinto do original, o índice permitiu classificar a qualidade das estações e validar a maior parte das séries, com desempenho comparável ao da aplicação original, reforçando a portabilidade dos componentes da metodologia. Em conjunto, esses trabalhos evidenciam que a metodologia pode ser aplicada a diferentes regiões, ainda que mediante adaptações às características locais.

2.5. Aplicações e Importância dos Dados Consistidos

Dados confiáveis de precipitação constituem um insumo fundamental para estudos e aplicações em recursos hídricos, uma vez que sustentam a compreensão do regime hidrológico de uma região. A partir dessas informações são desenvolvidas modelagens hidrológicas, estimativas de vazão, dimensionamento de sistemas de drenagem urbana e infraestrutura hidráulica, além de análises voltadas à prevenção e mitigação de eventos extremos (Botto, 2016; El Hachem et al., 2022; FUNASA, 2006). Quando as séries pluviométricas apresentam falhas, lacunas ou inconsistências, as análises subsequentes tornam-se comprometidas, podendo resultar em diagnósticos imprecisos e projetos de engenharia inadequadamente dimensionados. Assim, a garantia de um nível mínimo de qualidade, consistência e representatividade dos dados é condição prévia para qualquer aplicação técnica ou científica.

Segundo Bárdossy et al. (2022), na modelagem hidrológica, a importância de dados consistidos torna-se ainda mais evidente, uma vez que esses registros subsidiam a representação dos processos de transformação da precipitação em vazão em uma bacia hidrográfica. A partir dessa modelagem, são obtidos parâmetros essenciais para o planejamento em recursos hídricos, como vazões de pico, curvas IDF e critérios de dimensionamento de estruturas de controle de cheias (FUNASA, 2006). Erros sistemáticos ou lacunas nos dados de entrada refletem-se diretamente nos resultados dos modelos, aumentando a incerteza e comprometendo a confiabilidade das soluções propostas.

Séries históricas extensas e consistentes também desempenham papel central na previsão e mitigação de desastres hidrológicos (WMO, 2011). O registro contínuo de valores extremos ao longo do tempo possibilita a análise da frequência e da intensidade de eventos críticos, enquanto dados em tempo quase real sustentam o funcionamento de sistemas de alerta precoce (Ghomash et al., 2023; WMO, 2011). Registros faltantes, acumulações multi-diárias não identificadas ou defasagens temporais reduzem a eficácia desses sistemas, elevando o risco tanto de falsos alarmes quanto de eventos não previstos, com consequências diretas para a segurança da população.

Os impactos da má qualidade dos dados pluviométricos extrapolam o campo técnico e atingem a esfera socioeconômica. Walker (2000) destacou que estimativas imprecisas de disponibilidade hídrica afetam diretamente o setor agrícola, podendo resultar em redução da produtividade. De forma mais ampla, o World Bank Group (2023) estimou que o Brasil registre perdas anuais da ordem de R$ 13 bilhões em decorrência de eventos extremos como secas e inundações, com danos significativos à infraestrutura e à economia. Nesse contexto, séries históricas consistidas são fundamentais para a definição de curvas IDF, estimativas de períodos de retorno e estabelecimento de limiares críticos de precipitação utilizados em sistemas de alerta, além de subsidiarem a elaboração de mapas de suscetibilidade a inundações e deslizamentos, essenciais ao planejamento territorial.

Outro aspecto central da gestão de recursos hídricos é a avaliação da disponibilidade hídrica das bacias, necessária para compatibilizar a oferta de água com as demandas humanas, animais, industriais e ambientais. Conforme estabelecido pela Política Nacional de Recursos Hídricos (Brasil, 1997), o abastecimento humano e a dessedentação de animais são usos prioritários em situações de escassez, o que reforça a importância de estudos baseados em dados pluviométricos confiáveis. Associada a esse conceito, a segurança hídrica refere-se à capacidade de uma população garantir acesso sustentável e suficiente à água (UN-WATER, 2013). Séries de precipitação consistidas permitem estimar a recarga de mananciais superficiais e subterrâneos, caracterizar a variabilidade sazonal e interanual das chuvas e identificar períodos críticos de estiagem, subsidiando o dimensionamento de reservatórios, a definição de vazões de referência e o planejamento de usos múltiplos da água.

Como base para essas aplicações, têm-se os métodos estatísticos aplicados aos dados hidrológicos, que servem como base para a determinação de parâmetros que subsidiam o já mencionado planejamento ligado à prevenção de enchentes, estudos ligados à disponibilidade hídrica, dentre outros. Segundo Naghettini e Pinto (2007), a hidrologia estatística busca extrair, com base nas informações contidas em uma amostra, conclusões válidas sobre o comportamento da população da qual ela foi obtida, o que pressupõe a utilização de dados representativos, homogêneos e tão

livres quanto possível de erros de observação e medição. Nesse sentido, a identificação de lacunas, valores anômalos, erros sistemáticos e demais inconsistências reduz o risco de que falhas nos registros sejam interpretadas como características do próprio processo hidrológico, proporcionando maior confiabilidade às análises posteriores. Nota-se a relevância desse controle na análise de eventos extremos, uma vez que a interpretação de dados não consistidos pode resultar em estimativas errôneas de tempos de retorno, prejudicando a gestão de recursos hídricos e prevenção de desastres.

Para que essas análises sejam viáveis em escala regional, a disponibilidade e a organização dos dados consistidos assumem papel estratégico. No contexto brasileiro, observa-se que a extração direta dos registros por meio da plataforma HidroWeb (ANA, [2025]) apresenta limitações para estudos de grande abrangência, sobretudo pela dificuldade de obtenção simultânea de dados de múltiplas estações. Em investigações envolvendo centenas de postos pluviométricos, como no presente estudo, torna-se necessária a utilização de ferramentas externas e procedimentos automatizados. Além disso, informações complementares, como a altitude das estações, disponível apenas para parte dos postos em Minas Gerais, são fundamentais para a correta interpretação dos registros, considerando que diferenças altimétricas significativas podem resultar em regimes pluviométricos distintos mesmo entre estações próximas.

Nesse cenário, a integração entre registros pluviométricos e sistemas de informações geográficas (SIG) revela-se essencial. O georreferenciamento dos dados possibilita a análise da variabilidade espacial da precipitação, a identificação de regiões com maior confiabilidade dos registros e aquelas que demandam aprimoramento da rede de monitoramento (WMO, 2011). O cruzamento com camadas temáticas, como uso e ocupação do solo, sistemas de drenagem, áreas urbanizadas, litologia, pedologia e relevo, amplia o potencial analítico, permitindo avaliar a influência de barreiras topográficas, tipos de solo e substrato geológico sobre os padrões de chuva, o escoamento superficial e a recarga de aquíferos (Amen; Calvache, 2025). Ademais, a associação entre dados pluviométricos e sua localização espacial viabiliza a atualização dinâmica de sistemas de alerta,

permitindo visualizar, em tempo quase real, a distribuição dos acumulados de chuva em relação às áreas de risco e às populações expostas.

Dessa forma, a qualidade e a confiabilidade das séries pluviométricas, aliadas à sua adequada organização e georreferenciamento, ampliam significativamente o potencial de uso dessas informações. Tal integração favorece análises espaciais mais robustas, a aplicação direta em modelos hidrológicos e o suporte a ferramentas de planejamento territorial, contribuindo para a redução de incertezas no processo decisório e para o fortalecimento da gestão integrada de recursos hídricos.

2.6. Eventos extremos de precipitação e teoria dos valores extremos

A precipitação é uma variável altamente irregular no tempo e no espaço. Em uma série histórica diária, há muitos dias sem chuva, muitos eventos fracos e moderados, e poucos eventos muito intensos. Esses eventos intensos podem ser definidos como extremos, dependendo do limiar adotado. Um extremo pode ser classificado por um limiar absoluto, como uma chuva acima de 50 mm/dia, por um limiar relativo, como valores acima do percentil 95, por um máximo em um determinado período, como a maior chuva diária registrada em um ano, ou até por impacto hidrológico, quando o evento gera inundação, alagamento ou danos materiais (WMO, 2009).

Além disso, o *Expert Team on Climate Change Detection and Indices* (ETCCDI) definiu um conjunto de 27 índices de extremos climáticos, sistematizados pela WMO (2009), que descrevem características de valores extremos de temperatura e precipitação em uma série de dados. Esses índices são importantes por padronizarem, em escala mundial, o conceito de extremos climáticos, subsidiando estudos em diferentes regiões geográficas e descrevendo características como a frequência e a intensidade dos eventos. Um índice de destaque é o R95pTOT, que representa a soma acumulada da precipitação registrada nos dias que excederam o percentil 95 (p95 ou Q95), calculado sobre os dias chuvosos de um período de referência (ETCCDI, s.d.). O R95pTOT é utilizado para acompanhar mudanças no regime de chuvas extremas ao longo do tempo, pois

uma tendência de aumento desse índice indica que a precipitação vem se concentrando cada vez mais em eventos extremos, mesmo que o total anual não se altere.

O estudo de Silva Junior et al. (2022) relaciona precipitações intensas e extremas à ocorrência de desastres naturais em Maceió/AL. Os autores classificaram como chuva extrema os registros acima do quantil 99, equivalentes a precipitações superiores a 82,8 mm/dia na cidade, e mostraram que esses eventos estavam associados a decretos de Estado de Calamidade Pública, sobretudo em decorrência de enxurradas. O trabalho ilustra, assim, como a identificação de eventos extremos de precipitação pode subsidiar a antecipação do risco de desastres.

Cabe destacar os relatórios do IPCC (2012, 2021) e do trabalho de Katz et al. (2002), que mostram que os eventos extremos, de modo geral, já se tornaram mais frequentes e intensos desde meados do século XX e indicam a tendência de que essa intensificação prossiga ao longo do século XXI. Desastres causados por chuvas extremas são agravados pelo fato de que boa parte da infraestrutura urbana, sistemas de drenagem, pontes e contenções, foi dimensionada sob a premissa de estacionariedade climática, hipótese cada vez mais incompatível com a intensificação observada dos extremos (Milly et al., 2008; WMO, 2009). Diante disso, observa-se que a demanda por registros confiáveis de extremos, indispensáveis para subsidiar decisões de adaptação às mudanças climáticas, vem aumentando (WMO, 2009). Nesse contexto, caracterizar estatisticamente os eventos extremos de precipitação, estimando sua frequência, intensidade e probabilidade de ocorrência, torna-se fundamental para a prevenção de desastres e o dimensionamento de infraestruturas mais resilientes, papel que cabe à teoria dos valores extremos, adotada neste trabalho.

A maior parte das técnicas estatísticas convencionais descreve o comportamento central de séries de dados, como média e desvio padrão, apoiando-se em distribuições como a Normal. Todavia, em hidrologia, comumente o interesse não se dá pelo comportamento médio, mas sim sobre valores extremos,

utilizados para dimensionar obras hidráulicas e avaliar riscos (Katz et al., 2002). Esses eventos situam-se na cauda da distribuição, região em que as observações são escassas e em que frequentemente é necessário extrapolar para além do registro histórico, estimando, por exemplo, a precipitação esperada a cada cem anos a partir de séries de poucas décadas. Nesse cenário, ajustar uma distribuição ao conjunto completo dos dados tende a priorizar a porção central e a representar mal a cauda, comprometendo justamente as estimativas de maior interesse (Coles, 2001).

Diante dessa limitação, foi desenvolvido um arcabouço estatístico específico para extremos: a Teoria dos Valores Extremos (TVE). Fundamentada pelo Teorema dos Tipos Extremos (Fisher; Tippett, 1928; Gnedenko, 1943), que desempenha, para máximos, papel semelhante ao do Teorema Central do Limite para as médias. Enquanto este mostra que a média de muitas observações tende a seguir uma distribuição Normal, o Teorema dos Tipos Extremos mostra que o maior valor de um grande conjunto de observações, como a maior chuva diária de cada ano, tende a seguir uma entre três distribuições, conhecidas como Gumbel, Fréchet e Weibull, qualquer que seja a forma como os dados originais se distribuem. Esses três padrões são reunidos em uma única formulação, a distribuição Generalizada de Valores Extremos, ou GEV (Jenkinson, 1955). A principal vantagem desse resultado é permitir representar os eventos extremos por meio dessa distribuição sem que seja necessário conhecer a distribuição de probabilidade que descreve todos os dados da série, inclusive valores de precipitação considerados baixos, bastando trabalhar diretamente com os valores extremos (Coles, 2001; Naghettini; Pinto, 2007).

2.7. Distribuição GEV e interpretação do parâmetro de forma

A distribuição GEV descreve a probabilidade de que o valor máximo observado em uma série, por exemplo, a maior chuva diária de um ano, não ultrapasse um determinado valor x. Sua função de distribuição acumulada é dada pela equação 1:

$$F(x) = exp\left\{-\left[1 - \kappa\left(\frac{x-\mu}{\sigma}\right)\right]^{\frac{1}{\kappa}}\right\} \text{ para } 1 - \kappa\left(\frac{x-\mu}{\sigma}\right) > 0 \quad (1)$$

em que μ, σ e κ são, respectivamente, os parâmetros de posição, escala e forma (Jenkinson, 1955). O parâmetro de posição (μ) indica a posição central da distribuição, ou seja, em torno de qual valor as máximas se concentram, e pode assumir qualquer valor real. O parâmetro de escala (σ) mede a dispersão das máximas, o quanto elas variam de um ano para outro, e é sempre positivo ($\sigma > 0$). Por fim, o parâmetro de forma (κ) dita o comportamento da cauda da distribuição, isto é, a probabilidade associada aos eventos mais raros e intensos. É ele que determina qual das três famílias, Gumbel, Fréchet ou Weibull, melhor se ajusta aos dados.

A estimação dos parâmetros da GEV requer um método que relacione as características da amostra aos valores de posição, escala e forma. Um dos métodos mais utilizados para essa estimação é o método dos momentos-L (MML), proposto por Hosking (1990). Os momentos-L são medidas de uma distribuição, análogas aos momentos convencionais, porém construídas a partir de combinações lineares dos valores ordenados da amostra. Essa formulação confere ao método maior robustez frente a valores extremos e menor viés em amostras pequenas ou assimétricas, características típicas das séries de máximas de precipitação. Por esse motivo, o MML é amplamente empregado em estudos hidrológicos (Hosking, 1990; Naghettini; Pinto, 2007).

Quanto à notação, a distribuição GEV admite duas parametrizações com sinais opostos para o parâmetro de forma. A forma adotada neste trabalho é a de Hosking, baseada em Jenkinson (1955), na qual o parâmetro de forma é representado por κ e os valores $\kappa = 0$, $\kappa < 0$ e $\kappa > 0$ correspondem, respectivamente, às famílias Gumbel, Fréchet e Weibull. Contudo, é comum na literatura a adoção da parametrização de Coles (2001), em que o parâmetro de forma tem o sinal invertido e é representado por ξ, sendo $\kappa = -\xi$. Nessa convenção, $\xi > 0$ corresponde à Fréchet e $\xi < 0$ à Weibull. Trata-se apenas de uma diferença de notação, não alterando a natureza da distribuição. Adota-se aqui a convenção de Jenkinson (1955) por ser empregada na formulação dos momentos-L por Hosking (1990).

Conforme antecipado, as famílias Gumbel, Fréchet e Weibull, reunidas na GEV, correspondem a três regimes do parâmetro de forma (κ), que é o de maior interesse na análise de extremos por governar o comportamento da cauda da distribuição. Quando $\kappa = 0$, a GEV assume a forma da distribuição de Gumbel, cuja cauda decai exponencialmente e não possui limite superior, assim eventos cada vez maiores permanecem possíveis, ainda que progressivamente menos prováveis. Quando $\kappa < 0$, corresponde à distribuição de Fréchet, caracterizada por uma cauda pesada, que decai lentamente e admite eventos extremos substancialmente maiores, comportamento frequentemente associado a séries de precipitação mais intensa. Quando $\kappa > 0$, corresponde à distribuição de Weibull, em sua forma reversa (distinta da distribuição de Weibull clássica), cuja cauda é limitada por um valor máximo finito, acima do qual o modelo não prevê ocorrências. A Figura 8 ilustra as caudas de cada distribuição, permitindo observar como a Fréchet admite eventos mais intensos, e contrasta com a distribuição Normal, que nitidamente pode não representar de forma realista probabilidades de ocorrências de chuvas extremas.

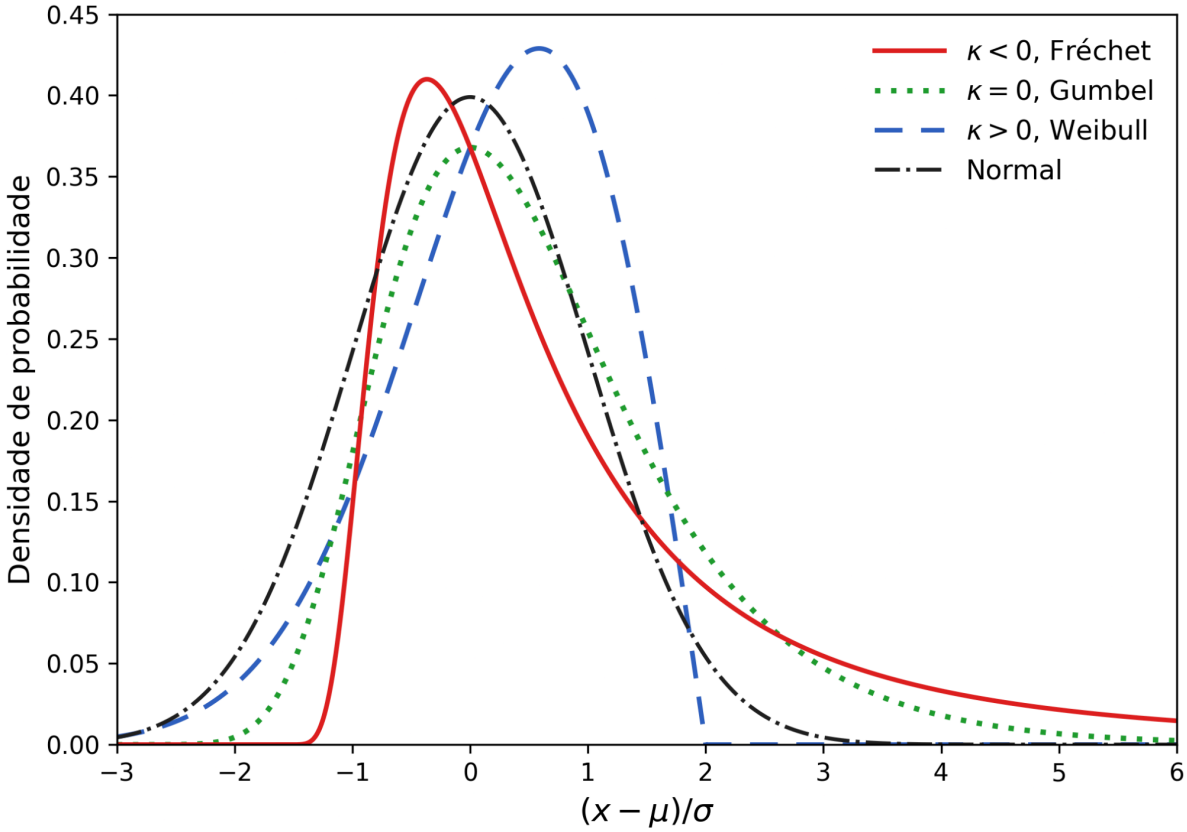


Figura 8: Caudas das distribuições Weibull, Gumbel e Fréchet, em contraste com a distribuição Normal.

Fonte: Adaptado de Naghettini e Pinto (2007)

A interpretação do parâmetro de forma possui, assim, aplicações práticas, pois indica se uma série analisada possui um limite superior para máximas e quão intensos podem ser os eventos raros, refletindo em estimativas de tempo de retorno (Coles, 2001; Naghettini; Pinto, 2007). Cabe ressaltar que o parâmetro de forma κ é o de estimativa mais difícil e o mais sensível ao tamanho da amostra, séries curtas podem comprometer sua determinação precisa, e pequenas alterações em κ podem alterar de modo expressivo a estimação de chuvas para grandes períodos de retorno (Cardoso, 2026). Estudos de larga escala corroboram a relevância desse parâmetro: ao analisarem 15.137 séries de máximas diárias anuais ao redor do mundo, Papalexiou e Koutsoyiannis (2013) verificaram o predomínio de distribuições do tipo

Fréchet sobre a Gumbel, o que, na convenção de Hosking aqui adotada, corresponde a valores de κ predominantemente negativos e, portanto, a caudas sem limite superior para as máximas.

2.8. Testes estatísticos aplicados a séries hidrológicas

A aplicação de modelos probabilísticos em análises de frequência hidrológica pressupõe que a série analisada represente adequadamente toda sua população estatística. Para séries de máximos anuais, é necessário avaliar se os dados apresentam características de uma amostra independente, aleatória, homogênea e sem tendência monotônica, caracterizando-a como uma Amostra Aleatória Simples, ou AAS (Khaliq et al., 2006; Naghettini; Pinto, 2007). Essas características são avaliadas em testes de hipótese, em que a hipótese nula (H0) pressupõe tais características a série, e a hipótese alternativa (H1) assume o contrário. Essa análise se justifica, pois, a presença de mudança de regime ou tendência temporal pode comprometer a interpretação dos parâmetros estimados, prejudicando a capacidade do modelo de representar o cenário real da série em projeções de eventos extremos. Assim, antes do ajuste da GEV, a aplicação de testes estatísticos permite verificar se as séries atendem aos pressupostos da análise de frequência hidrológica.

Para a verificação dos pressupostos que caracterizam uma AAS, a literatura recomenda preferencialmente o emprego de testes não-paramétricos. A distinção entre as duas classes de testes reside nas premissas impostas aos dados: os testes paramétricos pressupõem que a amostra provém de uma população com distribuição de probabilidades conhecida, usualmente a Normal, e baseiam suas estatísticas em grandezas como a média e a variância amostrais. Os testes não-paramétricos, ditos "livres de distribuição", dispensam essa premissa e fundamentam suas estatísticas em características derivadas dos dados da série, como postos, sinais e contagens (Naghettini; Pinto, 2007). Essa propriedade é apontada como particularmente vantajosa em hidrologia, pois as variáveis de interesse, em especial as séries de máximos anuais, tendem a ter forte assimetria

positiva e ser limitadas inferiormente, condições sob as quais a suposição de normalidade exigida pelos testes paramétricos não é considerada adequada (Helsel et al., 2020). Argumenta-se ainda que, por a verificação dos pressupostos anteceder o ajuste de qualquer modelo probabilístico, assumir a distribuição Normal nessa etapa seria prematuro. Por não dependerem de uma distribuição, os procedimentos não-paramétricos têm sido recomendados como abordagem preferencial na detecção de mudanças e na verificação de hipóteses em registros hidrológicos (Kundzewicz; Robson, 2004).

A verificação dos pressupostos da AAS apoia-se em um conjunto de testes não-paramétricos, cada um destinado a uma propriedade esperada da amostra. A aleatoriedade é comumente avaliada pelo *Turning Point Test* (NERC, 1975 apud Naghettini; Pinto, 2007), que se baseia na contagem de inflexões (picos e vales) observadas ao longo do tempo. Um número excessivamente grande ou pequeno de inflexões pode indicar uma não aleatoriedade. A independência, por sua vez, costuma ser examinada pelo teste de Wald-Wolfowitz (1943), que avalia se cada observação influencia a ocorrência da seguinte, isto é, se há correlação serial entre elementos consecutivos da série. Nota-se que uma série pode ser aleatória e, ainda assim, apresentar observações dependentes no tempo, razão pela qual as hipóteses são verificadas separadamente (Naghettini; Pinto, 2007).

As próximas duas propriedades verificadas se referem à invariância das características da série ao longo do tempo. A homogeneidade é frequentemente avaliada pelo teste de Mann-Whitney, que compara, em geral, duas metades da série, rejeitando-se a homogeneidade caso os valores de uma sejam sistematicamente superiores ou inferiores a da outra (Mann; Whitney, 1947). Já a ausência de tendência monotônica é verificada pelo teste de Mann-Kendall, fundamentado na correlação entre os valores da série e o índice temporal, capaz de detectar tendências crescentes ou decrescentes. Como séries hidrológicas podem apresentar autocorrelação serial, a literatura registra variantes desses testes que incorporam etapas de pré-branqueamento (*pre-whitening*), procedimento que remove a autocorrelação da série antes da aplicação do teste, evitando que a

dependência temporal seja confundida com tendência (Yue et al., 2002; Kundzewicz; Robson, 2004).

Quando um mesmo teste é aplicado simultaneamente a um grande número de estações, surge o problema da multiplicidade, em que uma parte das séries rejeita a hipótese nula por acaso, em proporção próxima ao nível de significância adotado (usualmente de 5%). Para corrigir esse efeito, Benjamini e Hochberg (1995) propuseram o controle *False Discovery Rate* (FDR), que limita a proporção esperada de rejeições incorretas entre o total de hipóteses rejeitadas. Sua aplicação em estudos hidrológicos foi consolidada por Wilks (2006), e, dessa forma, a quantidade de estações que rejeitam H0 incorretamente é muito reduzida.

Após a verificação dos pressupostos das amostras, é importante uma etapa preparatória das séries, em especial com o tratamento dos *outliers* mínimos, registros geralmente muito abaixo do restante, que podem exercer influência desproporcional nos ajustes de distribuições estatísticas. Assim, Cohn et al. (2013) destacam que os chamados *low outliers* não são informativos para a estimativa de quantis superiores, foco da análise de frequência de extremos. A literatura dispõe de diferentes alternativas para a identificação desses *outliers* mínimos. A mais simples sendo baseada na amplitude interquartil (AIQ) que classifica como atípico todo valor inferior a Q1 − 1,5AIQ, porém, trata-se de um procedimento simplificado, sem nível de significância associado, que não constitui, do ponto de vista estatístico, um teste de hipótese (Naghettini; Pinto, 2007).

Como alternativa formal, *Grubbs-Beck test* (Grubbs; Beck, 1972 apud Cohn et al., 2013) compara o menor valor da série a um limite crítico, construído a partir da média e do desvio padrão amostrais e de um fator crítico que depende do tamanho da amostra e do nível de significância adotado, então, o valor é classificado como outlier quando esse desvio é estatisticamente significativo. A limitação do método reside na identificação de apenas um valor por vez, levando ao mascaramento quando há vários *outliers* próximos. Para superar essa restrição, Cohn et al. (2013) propuseram o *Multiple Grubbs-Beck Test* (MGBT), uma generalização iterativa que

detecta sucessivamente múltiplos pontos atípicos inferiores, repetindo o teste até que nenhum novo seja identificado.

Uma vez verificados os pressupostos da amostra e tratados os pontos atípicos inferiores, resta avaliar se a distribuição de probabilidades candidata representa adequadamente os dados, papel dos testes de aderência, que confrontam a distribuição empírica da amostra com a distribuição teórica hipotetizada. A construção dessa distribuição empírica requer que se associe a cada valor observado uma estimativa de sua probabilidade de não superação. Como tal probabilidade não é conhecida, ela é estimada a partir da posição que o valor ocupa na série ordenada, por meio das chamadas posições de plotagem (*plotting positions*). Ordenados os dados, atribui-se ao valor de ordem *i*, em uma série de n observações, uma probabilidade segundo a forma geral $Pi=(i-a)/(n+1-2a)$, na qual a constante *a* distingue as diversas fórmulas propostas na literatura, como as de Weibull, Blom, Hazen e Gringorten (Cunnane, 1978 apud Naghettini; Pinto, 2007). Tais fórmulas produzem resultados semelhantes no centro da distribuição, mas divergem nas caudas, razão pela qual a escolha deve considerar a distribuição em análise. A posição de plotagem de Gringorten, correspondente a $a = 0,44$, é a recomendada para os quantis das distribuições de Gumbel e GEV, o que a torna apropriada à análise de máximos (Naghettini; Pinto, 2007).

Entre os mais difundidos testes de aderência está o de Kolmogorov-Smirnov, de natureza não-paramétrica, cuja estatística corresponde à máxima diferença absoluta entre a função de distribuição acumulada empírica e a teórica (Naghettini; Pinto, 2007). Apesar de sua simplicidade e ampla utilização, o teste apresenta limitações relevantes. É considerado conservador, com baixo poder de rejeição, e tem sensibilidade reduzida nas caudas da distribuição, que é justamente a região de maior interesse na análise de valores extremos (Naghettini; Pinto, 2007). Além disso, quando os parâmetros da distribuição teórica são estimados a partir da própria amostra, situação usual na prática, os valores críticos tabelados deixam de ser válidos, tornando necessário obtê-los por simulação, de forma específica para cada distribuição. Por essas razões, em estudos de extremos costuma-se recorrer a

testes que atribuem maior peso às caudas, como o de Anderson-Darling (Laio, 2004).

Quando se deseja verificar a aderência à distribuição Normal, os testes de aderência passam a se chamar testes de normalidade. Ressalta-se que na análise de eventos extremos, a normalidade não é esperada para os dados de máximas, então os testes geralmente são feitos sobre outras características da série, como o parâmetro de forma da GEV. Um dos procedimentos empregados é o teste de Anderson-Darling que, semelhante ao de Kolmogorov-Smirnov, fundamenta-se na diferença entre as funções de distribuição acumuladas empírica e teórica, mas atribui maior peso às caudas, dividindo a diferença ao quadrado por um fator que é máximo no centro da distribuição e tente a 0 nas caudas, conferindo ao teste maior sensibilidade às caudas (Anderson; Darling, 1952).

Outro teste amplamente utilizado para essa finalidade é o de Shapiro-Wilk, considerado um dos mais potentes na detecção de afastamentos da normalidade, sobretudo em amostras de tamanho pequeno a moderado (Shapiro; Wilk, 1965). Nele, é avaliado o grau de linearidade que os dados exibem em um gráfico de probabilidade Normal. Nesse gráfico, cada valor da amostra, depois de ordenado, é confrontado com o valor que se esperaria observar naquela posição caso a amostra proviesse de uma distribuição Normal. Quando os dados são compatíveis com a Normal, esses pares tendem a se dispor sobre uma reta, e conjuntos de dados com assimetria ou caudas pesadas geram pontos que se distanciam dessa reta. A estatística de teste corresponde ao quadrado da correlação entre os valores ordenados e as estatísticas normais esperadas, assumindo valores próximos de 1 quando o alinhamento é forte. Dessa forma, a hipótese de normalidade é rejeitada quando a estatística de teste for suficientemente baixa, quando os pontos se afastam da reta de referência.

A aplicação dos testes de normalidade justifica-se pela posição importante ocupada pela Normal na inferência estatística. Boa parte dos procedimentos paramétricos usuais, entre eles os testes de comparação de médias, a análise de variância e os modelos de regressão, dependem da hipótese de que os dados, ou

uma grandeza deles derivada, sigam distribuição Normal para gerar resultados mais precisos (Naghettini; Pinto, 2007). Verificar essa hipótese permite decidir entre empregar esses métodos mais potentes ou recorrer a alternativas não paramétricas, que dispensam suposições sobre a forma da distribuição. Dessa forma, a verificação da normalidade serve como um critério que orienta a escolha entre as abordagens paramétrica e não-paramétrica, algumas das quais foram discutidas nesse trabalho.

Os procedimentos aqui citados compartilham da lógica dos testes de hipótese, partindo da hipótese nula que assume uma característica desejável à série, como aleatoriedade, ausência de tendência, aderência a um modelo, e confrontando essa premissa com a evidência amostral, decidindo entre rejeitá-la ou não a um nível de significância previamente fixado. Embora cada um seja diferente quanto à estatística de teste e a distribuição de referência, todos cumprem o propósito de traduzir as características das amostras em critérios de decisão, reduzindo a subjetividade no processo de análise (Naghettini; Pinto, 2007). Em conjunto, os testes estatísticos servem como base para sustentar a confiabilidade de análises de frequência, verificando a qualidade dos dados e a adequação dos modelos adotados.

2.9. Regionalização estatística e análise de agrupamentos

A análise de frequência conduzida posto a posto esbarra na extensão muitas vezes reduzida das séries observadas, o que compromete a confiabilidade das estimativas de quantis associados a períodos de retorno elevados. Para atenuar essa restrição, a literatura recomenda a análise regional de frequência, cujo princípio é reunir as informações de postos que apresentam comportamento estatístico semelhante, estimando os parâmetros da distribuição adotada de maneira mais robusta do que seria possível a partir de uma única amostra local (Hosking; Wallis, 1997). Esse enfoque organiza-se em torno do conceito de região homogênea: um conjunto de postos cujas séries seguem o mesmo tipo de distribuição de frequências, diferindo apenas por um fator de escala que reflete as condições locais de cada posto (Hosking; Wallis, 1997).

A própria existência de regiões homogêneas decorre da natureza do processo que se busca modelar. A precipitação resulta da interação de múltiplos fatores físicos como o relevo, a atuação das massas de ar e a disponibilidade de umidade, que variam no espaço e escapam ao controle do analista (Hosking; Wallis, 1997). Como esses fatores atuam de maneira semelhante sobre postos sujeitos a condições geográficas e climáticas próximas, é razoável esperar que tais postos compartilhem um mesmo comportamento estatístico, constituindo uma região homogênea. Essa mesma complexidade, contudo, faz com que as fronteiras entre regiões raramente sejam nítidas, razão pela qual a identificação das regiões homogêneas é reconhecida como a etapa mais difícil e a mais sujeita a subjetividades de toda a análise regional de frequência (Naghettini; Pinto, 2007).

Parte dessa dificuldade se dá pois, embora o termo região possa remeter a um conjunto de postos vizinhos no espaço, a proximidade geográfica não constitui, por si só, garantia de similaridade entre as distribuições de frequências (Hosking; Wallis, 1997). O que delimita uma região homogênea é a semelhança entre as características que condicionam o comportamento das chuvas em cada posto, de natureza geográfica, como latitude e longitude, ou física, como altitude e precipitação média anual, e não a mera contiguidade no mapa.

A reunião dos postos semelhantes é feita por meio da análise de agrupamento (*cluster analysis*), técnica da estatística que separa os indivíduos (aqui, estações pluviométricas), em grupos, de modo que postos parecidos fiquem juntos e os diferentes em grupos distintos. A semelhança entre dois postos é medida pela distância entre suas características, em geral a euclidiana, em que quanto menor a distância, maior a semelhança (Melo Júnior et al., 2006). Como essa distância é sensível à escala das variáveis, aquelas de maior magnitude, como a altitude em metros frente à latitude e à longitude em graus, dominariam o cálculo caso fossem usadas em seus valores brutos. Por isso, antes do agrupamento é usual padronizar as variáveis, de modo que todas contribuam com pesos equivalentes (Hosking; Wallis, 1997).

Medida a semelhança, resta agrupar os postos, e a literatura consolida duas grandes formas: os métodos hierárquicos e os não hierárquicos, ou particionais. Nos métodos hierárquicos, cada posto começa sozinho e os mais próximos vão sendo unidos sucessivamente, até formar uma árvore de agrupamentos, chamada dendrograma, na qual se escolhe onde cortar para definir os grupos, sendo o método de Ward o mais conhecido (Melo Júnior et al., 2006).

Já nos métodos não hierárquicos, decide-se antes quantos grupos se quer formar, e os postos vão sendo trocados de grupo até se chegar à melhor divisão possível, sem construir uma árvore. O mais usado dessa família é o método k-means, no qual se escolhe um número k de grupos e se definem, de início, k pontos centrais, os centroides, que funcionam como centros provisórios em torno dos quais os grupos se organizam. A partir desses centros, cada posto é colocado no grupo cujo centroide lhe é mais próximo; em seguida, cada centroide é reposicionado no centro dos postos que passaram a compô-lo, e essa alternância entre redistribuir os postos e reposicionar os centroides se repete até que os postos de um mesmo grupo fiquem o mais parecidos possível entre si e os grupos, os mais distintos uns dos outros (James et al., 2021; Naghettini; Pinto, 2007). Diferentemente dos métodos hierárquicos, o k-means precisa que o número de grupos seja definido de antemão, o que exige um critério para decidir quantas regiões homogêneas adotar.

Esse critério pode ser fornecido pelo índice de silhueta. Cada posto é descrito por um conjunto de características, como a latitude e a longitude, e a distância entre dois postos nada mais é do que o quanto esses valores diferem entre si, medida pela distância euclidiana já apresentada: quanto mais parecidas as características, menor a distância e maior a semelhança. Assim, é decidido em qual grupo cada posto fica. Para cada posto, o índice compara a distância média até os postos do próprio grupo com a distância até os do grupo vizinho mais próximo, resumindo esse contraste em um valor de −1 a 1, no qual valores mais altos indicam grupos mais coesos e mais bem separados (Rousseeuw, 1987). Adota-se, então, o número de grupos que maximiza a silhueta média.

Em síntese, o valor da análise regional está em compensar a escassez temporal das séries individuais com a informação espacial do conjunto de postos. Ao tratar estações semelhantes como parte de uma mesma população, torna-se possível estimar o comportamento das chuvas extremas de forma mais robusta e menos sensível à variabilidade amostral de cada série isolada, ganho especialmente relevante para os tempos de retorno mais elevados, em que as séries curtas são pouco informativas (Hosking; Wallis, 1997).

2.10. Estudos sobre o parâmetro de forma da distribuição GEV

A escolha da família de valores extremos mais adequada às máximas diárias de precipitação foi investigada em escala global por Papalexiou e Koutsoyiannis (2013). Até então, conforme destacam os autores, os estudos sobre o tema possuíam caráter predominantemente local, restritos a áreas específicas, o que limitava conclusões de alcance mais amplo. A principal inovação do trabalho reside na escala da análise: foram examinadas 15.137 séries de máximas diárias anuais obtidas da base Global Historical Climatology Network-Daily, distribuídas por todo o globo e com extensões entre 40 e 163 anos.

Os autores ajustaram a distribuição GEV pelo método dos momentos-L, justificando a escolha pelo menor viés e pela menor variância desses estimadores em amostras de tamanho reduzido. Na notação por eles adotada, em que $\xi = -\kappa$, obtiveram média de 0,092 e mediana de 0,093 para o parâmetro de forma, indicando predominância da família Fréchet e, portanto, de caudas pesadas. Constataram, ainda, que a extensão da série afeta sistematicamente a estimativa: a média do parâmetro de forma cresceu de 0,077, nas séries de 40 a 50 anos, para 0,116, naquelas com mais de 121 anos, enquanto o desvio-padrão reduziu-se de 0,141 para 0,088. Diante disso, propuseram um estimador corrigido, livre do efeito do tamanho amostral, cuja distribuição limite seria normal com média 0,114 e desvio-padrão 0,045. A hipótese de normalidade foi verificada por meio de simulação de Monte Carlo, técnica que consiste em gerar séries sintéticas a partir de uma distribuição de parâmetros conhecidos e reestimá-los em cada uma delas,

isolando a parcela de variabilidade atribuível ao acaso. Quanto à variabilidade espacial, os autores verificaram que extensas áreas do globo compartilham valores semelhantes do parâmetro, ainda que regiões distintas apresentem comportamentos diversos dos extremos.

Enquanto os levantamentos de larga escala estabelecem o predomínio de caudas pesadas, outros estudos voltaram-se ao comportamento espacial e à estimabilidade do parâmetro. Ragulina e Reitan (2017) dedicaram-se especificamente à caracterização do parâmetro de forma da GEV para precipitações extremas, a partir de séries longas e de uma abordagem bayesiana. Os autores constataram que a magnitude do parâmetro varia com a localização geográfica, mas que essa variação é limitada o bastante para que, dentro de uma mesma região, o parâmetro de forma possa ser tratado como aproximadamente constante, mediante a adoção de um fator de correção regional. Tal resultado sugere que, apesar da reconhecida dificuldade de estimar o parâmetro de forma a partir de séries individuais, seu comportamento espacial relativamente homogêneo pode favorecer abordagens regionais para a sua determinação.

Em âmbito estadual, Cardoso (2026) analisou o parâmetro de forma da distribuição GEV ajustada a séries de precipitação diária máxima por ano hidrológico em Minas Gerais, motivado pela concentração, no estado, de estruturas hidráulicas de elevado potencial de dano, como as barragens de mineração. Foram analisadas 496 estações, todas com no mínimo 30 anos hidrológicos válidos, e o ajuste foi realizado pelos métodos dos momentos-L e máxima verossimilhança, em cenários com e sem os *outliers* mínimos identificados pelo teste MGBT. O autor adotou as estimativas por momentos-L nas análises subsequentes do parâmetro de forma, por se mostrarem mais robustas: o algoritmo de máxima verossimilhança não convergiu em cinco estações nas séries completas e revelou-se mais sensível à variabilidade amostral local, gerando padrões espaciais artificiais nos mapas de quantis.

Ao aplicar o método, o autor obteve média e mediana nulas até a segunda casa decimal para ambos os métodos no cenário sem *outliers* mínimos, resultado compatível com a família Gumbel e com a predominância de caudas leves no

estado, que é um cenário mais favorável, sob a ótica da segurança de estruturas hidráulicas, do que o indicado pelo estudo global. Diferentemente de Papalexiou e Koutsoyiannis (2013), não identificou influência sistemática do tamanho da série sobre os valores médios do parâmetro, razão pela qual ponderou que a aplicação direta da correção proposta por aqueles autores deve ser avaliada com cautela no contexto regional. Ao agrupar as estações em dois clusters definidos por latitude, longitude, altitude e precipitação média anual, um na porção norte/nordeste e outro na porção sul/sudoeste do estado, tampouco identificou diferença entre os grupos, cujas médias coincidiram até a segunda casa decimal, sugerindo que tais variáveis não explicam a variabilidade das estimativas.

Ainda no contexto brasileiro, Correia Filho, Lucio e Spyrides (2016) caracterizaram os extremos de precipitação diária no Nordeste do Brasil por meio do ajuste das distribuições Generalizada de Valores Extremos e Generalizada de Pareto às séries de estações da região. Os autores investigaram a distribuição espacial dos parâmetros ajustados, associando os padrões de posição e escala à influência da costa leste, do litoral norte e da topografia do planalto da Borborema e da chapada do Araripe. Quanto ao parâmetro de forma, os resultados indicaram comportamento do tipo Weibull nas regiões analisadas, isto é, caudas com limite superior para as máximas. Trata-se de um resultado digno de nota, uma vez que séries de precipitação máxima tendem, em geral, a apresentar comportamento do tipo Gumbel ou Fréchet, (Papalexiou; Koutsoyiannis, 2013), de modo que a predominância do tipo Weibull pode estar relacionada às características semiáridas do Nordeste.

Em conjunto, esses trabalhos delimitam as escalas de referência desta pesquisa. Papalexiou e Koutsoyiannis (2013) fornecem o arcabouço metodológico adotado, aplicando o ajuste por momentos-L e a análise do parâmetro de forma segundo o tamanho da série e a localização geográfica, e o parâmetro de comparação global, ao passo que Cardoso (2026), por tratar da mesma variável, do mesmo estado e do mesmo recorte de ano hidrológico, constitui o antecedente mais próximo e o principal referencial de confronto dos resultados aqui obtidos. Cabe destacar que os estudos aqui apresentados chegaram a conclusões distintas quanto

à família predominante, o que evidencia a necessidade de avaliar o comportamento do parâmetro de forma em âmbito regional, em vez de transpor diretamente conclusões de escala global.

3 METODOLOGIA

A organização das etapas metodológicas desenvolvidas neste trabalho é apresentada no fluxograma da Figura 9.

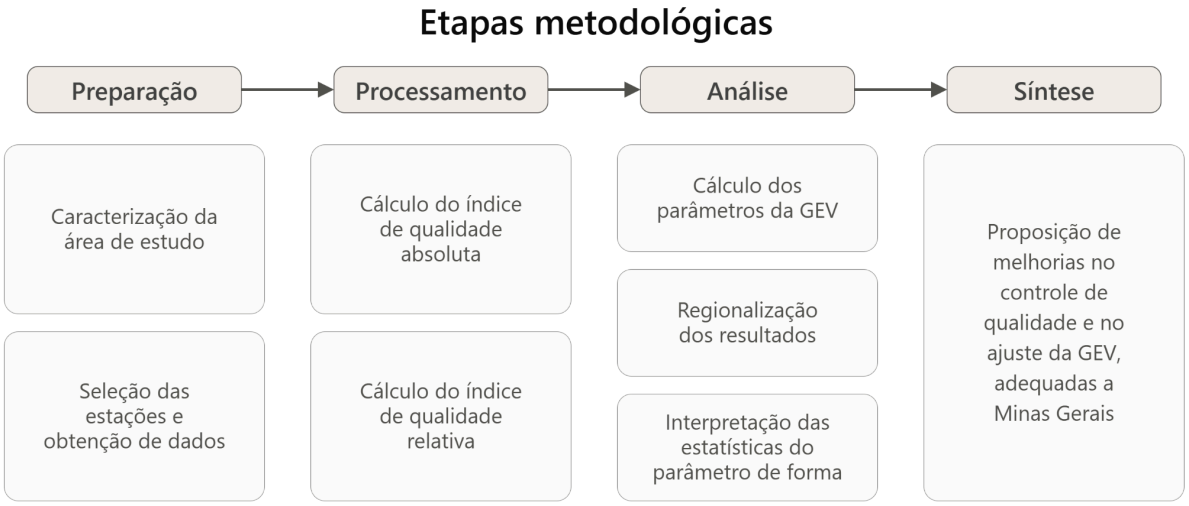


Figura 9: Fluxograma das etapas metodológicas do trabalho.

Fonte: Autoral (2026).

3.1 Caracterização da área de estudo

Minas Gerais é um estado localizado na região Sudeste do Brasil (Figura 10). De acordo com o Censo Demográfico de 2022, possui uma população de 20.539.989 habitantes distribuídos em uma área de 586.513,984 km², resultando em uma densidade demográfica aproximada de 35 hab/km². Trata-se de um dos maiores estados brasileiros em extensão territorial, caracterizado por significativa diversidade física, climática e ambiental.

Aspectos topográficos exercem influência direta sobre a representatividade espacial das estações pluviométricas, uma vez que pontos de medição geograficamente próximos podem registrar regimes de precipitação bastante distintos quando separados por barreiras orográficas, como cadeias montanhosas. De modo semelhante, as condições climáticas regionais interferem na interpretação

dos resultados do controle de qualidade: um evento considerado extremo em áreas mais áridas pode corresponder a uma ocorrência relativamente comum em regiões úmidas. Assim, para interpretar adequadamente os índices de qualidade e ajustar os métodos às particularidades regionais, é fundamental compreender as características topográficas e climáticas da área de estudo.

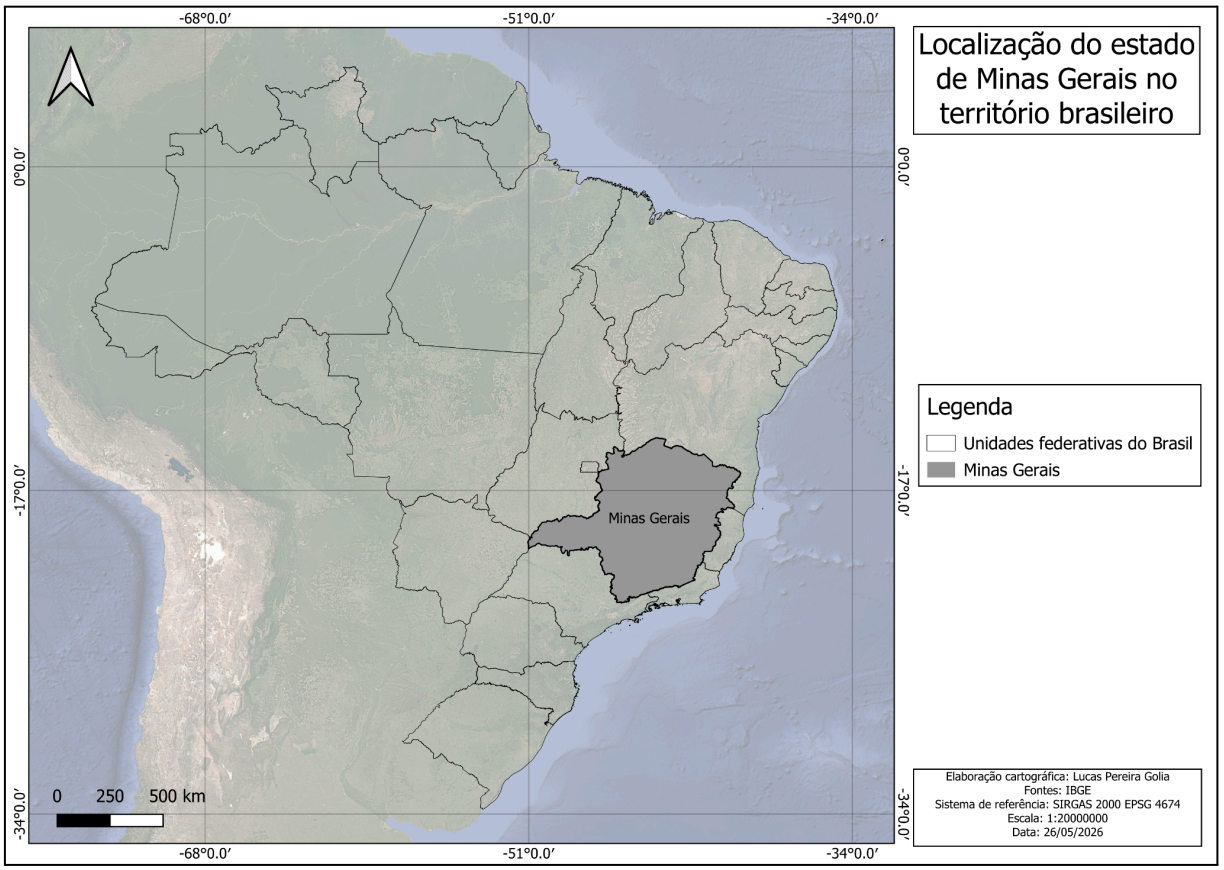


Figura 10: Localização de Minas Gerais no território brasileiro.

Fonte: Autoral (2026).

O mapa climático de Minas Gerais, baseado na classificação de Thornthwaite (Figura 11), evidencia a variabilidade espacial das condições de umidade no estado. Essa classificação não se restringe aos totais anuais de precipitação, pois deriva da relação entre precipitação e evapotranspiração potencial (ETP), representando o balanço entre a oferta e a demanda hídrica do ambiente. Dessa forma, regiões com

totais pluviométricos semelhantes podem enquadrar-se em classes climáticas distintas, a depender da intensidade da evapotranspiração. Observam-se, no estado, áreas superúmidas e úmidas (classes A e B), predominantes no Sul e na Zona da Mata; áreas subúmidas (classe C), que abrangem grande parte da região central e do Cerrado mineiro; e áreas semiáridas (classe D), no Norte e Nordeste, onde baixos totais de chuva combinados a elevadas taxas de evaporação resultam em déficits hídricos mais acentuados.

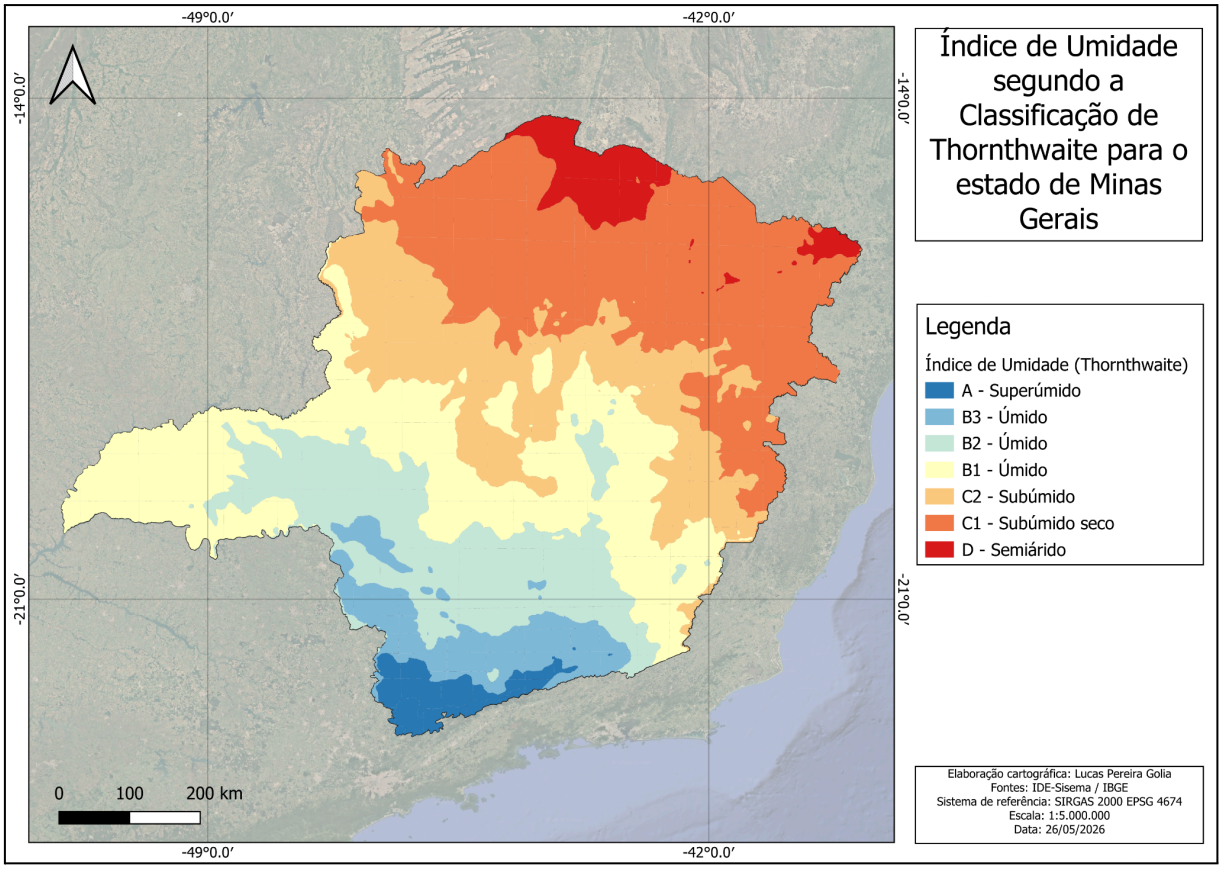


Figura 11: Classificação de Thornthwaite em Minas Gerais.

Fonte: Autoral (2026).

Essa variabilidade climática influencia diretamente os padrões estatísticos das séries pluviométricas. Em regiões mais secas, a maior frequência de dias sem chuva pode afetar índices relacionados à ocorrência de zeros, como Qm_{zero} e Qw_{zero}. Já

em áreas mais úmidas, eventos intensos podem ser identificados como *outliers*, exigindo a distinção entre ocorrências hidrologicamente plausíveis e possíveis erros de medição ou processamento. Assim, a distribuição climática deve ser considerada na interpretação dos índices de qualidade.

O mapa de altimetria de Minas Gerais (Figura 12) é relevante porque a altitude influencia a circulação atmosférica e os padrões de precipitação. Em áreas montanhosas, variações altimétricas podem alterar a incidência de ventos úmidos, a formação de nuvens e a distribuição espacial das chuvas. Desse modo, a topografia é importante tanto para compreender a distribuição das precipitações quanto para interpretar os índices de qualidade. O mapa de declividade (Figura 13) também contribui para avaliar a aplicabilidade da metodologia de Llabrés-Brustenga et al. (2019) em Minas Gerais. Áreas de relevo mais acidentado apresentam maior influência topográfica e maior variabilidade espacial da precipitação, podendo gerar grandes diferenças entre estações próximas. Essas características afetam principalmente os índices de qualidade relativa, baseados na correlação entre séries vizinhas e na representatividade espacial calculada a partir da distância e da altitude. Portanto, a declividade reforça a importância dos fatores topográficos considerados na Equação 6, utilizada para calcular a representatividade R entre estações.

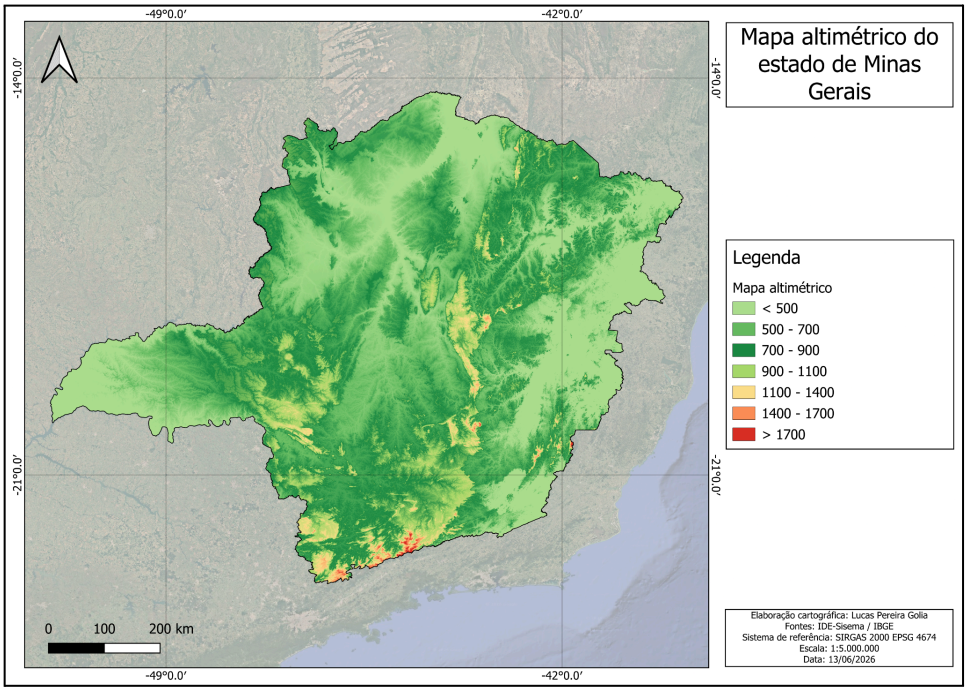


Figura 12: Altimetria de Minas Gerais.

Fonte: Autoral (2026).

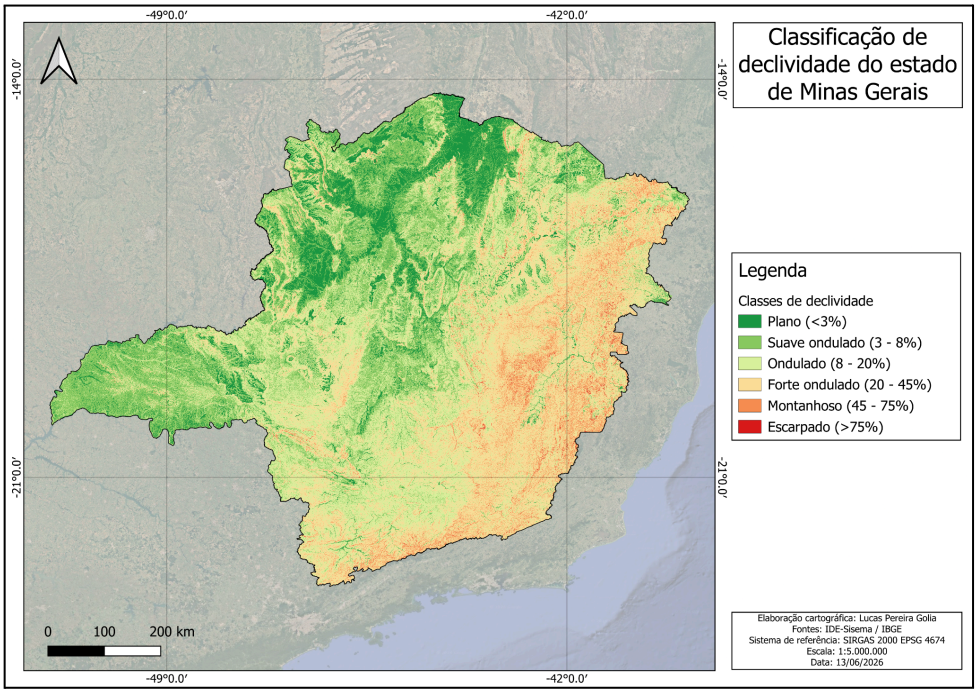


Figura 13: Declividade de Minas Gerais.

Fonte: IDE-Sisema (2025).

A metodologia de controle de qualidade proposta por Llabrés-Brustenga et al. (2019) foi originalmente desenvolvida e validada na Catalunha (Espanha), uma região territorialmente mais compacta, onde a variabilidade espacial da precipitação tende a ser menos acentuada, o que facilita a identificação de falhas e inconsistências em séries históricas. Em Minas Gerais, entretanto, sua aplicação exige maior cautela interpretativa, considerando as dimensões territoriais mais amplas e a elevada heterogeneidade climática e topográfica do estado. A forte influência do relevo, a transição entre diferentes biomas e a coexistência de áreas tropicais úmidas e zonas semiáridas no norte conferem maior complexidade ao padrão espacial das chuvas.

Assim, ao aplicar essa metodologia em Minas Gerais, o presente estudo amplia seu campo de empregabilidade e testa a robustez do procedimento em um contexto tropical de elevada complexidade espacial, contribuindo para a análise da consistência de séries pluviométricas em territórios extensos e climaticamente diversos.

3.2 Seleção das estações e obtenção dos dados

A Agência Nacional de Águas e Saneamento Básico (ANA) é o órgão federal responsável pela implementação da Política Nacional de Recursos Hídricos e pela coordenação da Rede Hidrometeorológica Nacional (RHN), atuando na coleta, gestão e disponibilização de dados hidrológicos e meteorológicos em âmbito nacional. A plataforma Hidroweb constitui a principal base pública de dados da ANA, reunindo séries históricas de precipitação, vazão, níveis fluviométricos e qualidade da água em todo o território brasileiro. No presente estudo, essa base foi utilizada como fonte primária para obtenção das séries pluviométricas analisadas.

Complementarmente, foi consultada a Infraestrutura de Dados Espaciais do Sistema Estadual de Meio Ambiente e Recursos Hídricos de Minas Gerais (IDE-Sisema), plataforma oficial do Governo do Estado que organiza e disponibiliza dados geoespaciais, ambientais e hidrológicos. O IDE-Sisema integra informações provenientes de diferentes instituições estaduais, oferecendo acesso a mapas

temáticos como uso e ocupação do solo, cobertura vegetal, hidrografia, geologia e modelos digitais de elevação (MDE), sendo essencial para a caracterização espacial da área de estudo.

Apesar de sua relevância para a transparência e disseminação de informações hidrometeorológicas, a plataforma Hidroweb apresenta limitações operacionais, como a ausência de opção para *download* simultâneo de dados de múltiplas estações, exigindo que o procedimento seja realizado individualmente para cada posto. Considerando que este estudo envolve 652 estações pluviométricas (Figura 14), a coleta manual seria excessivamente onerosa. Para viabilizar o processo, foi utilizado um *script* em linguagem R, disponibilizado pelo Serviço Geológico do Brasil (SGB), que permite agilizar e sistematizar o *download* das planilhas. Outra limitação refere-se à disponibilidade da informação de altitude, presente no Hidroweb para apenas 322 estações em Minas Gerais. Assim, torna-se necessário complementar esses dados por meio do Modelo Digital de Elevação do IDE-Sisema. Embora essa fonte apresente precisão inferior à de levantamentos altimétricos oficiais específicos de cada estação, sua utilização foi fundamental para possibilitar a aplicação da metodologia de controle de qualidade adotada.

As planilhas obtidas por meio do código automatizado contêm, como principais informações, as colunas de precipitação diária, identificadas como Chuva01 a Chuva31, que representam os valores registrados em milímetros para cada dia do mês. Além disso, incluem campos de identificação e metadados, como o EstacaoCodigo, o NivelConsistencia (sendo 1 para dados brutos e 2 para dados consistidos) e a coluna DataHora, que indica o mês, ano e horário de referência das medições. Essas informações constituem a base para as etapas subsequentes de organização, tratamento e aplicação dos testes de qualidade e para análises estatísticas, e estão presentes no Apêndice A[1]. A Figura 15 exemplifica a estrutura das planilhas supracitadas.

[1] Os apêndices deste trabalho estão disponíveis em: https://drive.google.com/drive/folders/1-RkAcHNwBmP7pBs09nKVmO-5M1gqvk2-?usp=sharing

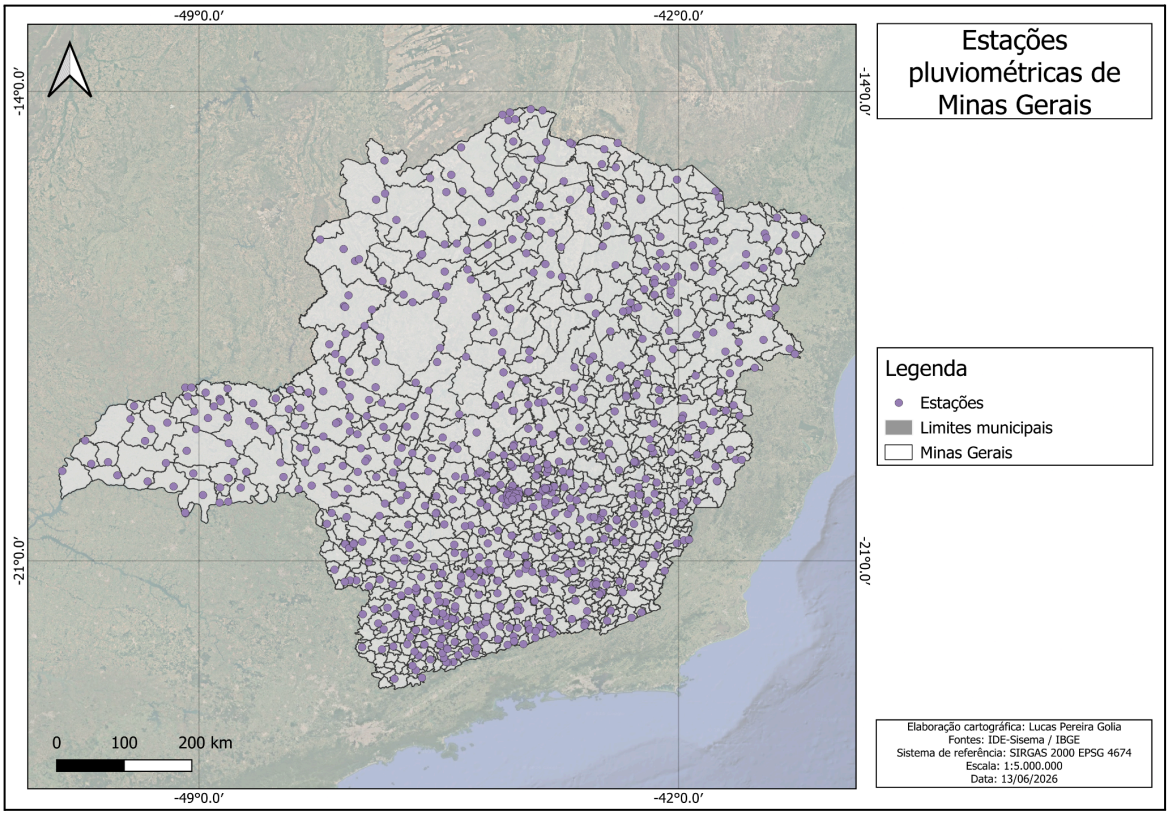


Figura 14: Estações pluviométricas de Minas Gerais.

Fonte: Autoral (2026).

EstacaoCodigo	NivelConsistencia	DataHora	Chuva01	Chuva02	Chuva03	Chuva04
1443001	1	2025-02-01 00:00:00	6,1	0	0	0
1443001	1	2025-01-01 00:00:00	0	1,8	0	1,5
1443001	1	2024-12-01 00:00:00	6,5	2,2	0	0
1443001	1	2024-11-01 00:00:00	0	0	0	0
1443001	1	2024-10-01 00:00:00	0	0	0	0
1443001	1	2024-09-01 00:00:00	0	0	0	0
1443001	1	2024-08-01 00:00:00	0	0	0	0
1443001	1	2024-07-01 00:00:00	0	0	0	0
1443001	1	2024-06-01 00:00:00	0	0	0	0
1443001	1	2024-05-01 00:00:00	0	0	0	0
1443001	1	2024-04-01 00:00:00	0	77,1	3,8	0
1443001	1	2024-03-01 00:00:00	0	0	0	2,1
1443001	1	2024-02-01 00:00:00	0	0	2,3	0
1443001	1	2024-01-01 00:00:00	0	39,3	3,6	17,7

Figura 15: Estrutura das planilhas contendo dados pluviométricos.

Fonte: Hidroweb (2025).

Para a obtenção da localização geográfica das estações pluviométricas, foi utilizado um *shapefile* disponibilizado pelo Hidroweb, contendo a posição de todas as estações cadastradas, a partir do qual foram extraídas as coordenadas de latitude e longitude, já inerentes à camada. Em relação à altitude, foi utilizado inicialmente um arquivo em formato XML, também fornecido pelo Hidroweb, que contém essa informação para apenas 322 estações. Para complementar os valores faltantes, foi empregada a camada de altimetria do IDE-Sisema, em formato *geopackage*, a qual, após ser adicionada ao QGIS, permitiu realizar uma junção espacial entre a grade altimétrica e os pontos das estações, possibilitando assim a obtenção das altitudes ausentes. Tais informações estão presentes no Apêndice B.

3.3 Cálculo do índice de qualidade absoluta

Após a etapa de coleta e organização dos dados, iniciou-se a implementação das equações propostas por Llabrés-Brustenga et al. (2019) para o cálculo dos índices de qualidade absoluta e relativa, utilizando a linguagem de programação Python. Dessa forma, a compreensão detalhada da formulação matemática e da lógica subjacente às equações constitui o primeiro passo para sua correta operacionalização e posterior análise dos resultados.

O índice de qualidade absoluta Q é dividido em 5 parâmetros diferentes, e visa determinar uma pontuação para a confiabilidade dos dados de cada estação, sem compará-los com os de nenhuma outra. O valor de Q varia de 0 a 100, com os valores acima de 80 indicando uma qualidade absoluta dos dados considerada aceitável. Consiste de uma média entre todos os 5 parâmetros calculados, onde cada um também varia de 0 a 100. Nota-se que cada índice é calculado por ano, para depois ser compilado por estação através da média dos valores anuais pela quantidade de anos. A Figura 16 sintetiza de forma simplificada a composição do índice de qualidade absoluta Q, evidenciando como o valor final resulta da média de seus parâmetros e a equação associada a cada um deles. Todas as etapas metodológicas foram implementadas em um único *script* em Python, que realiza a leitura dos registros diários de precipitação e aplica automaticamente as equações

dos índices de qualidade, as quais foram codificadas com apoio de inteligência artificial generativa, sem a utilização de pacotes específicos para o cálculo desses indicadores. O uso de pacotes auxiliares, como *datetime* e *calendar*, restringiu-se ao tratamento das datas, possibilitando a distinção entre lacunas reais na série e dias inexistentes no calendário, como 30 de fevereiro.

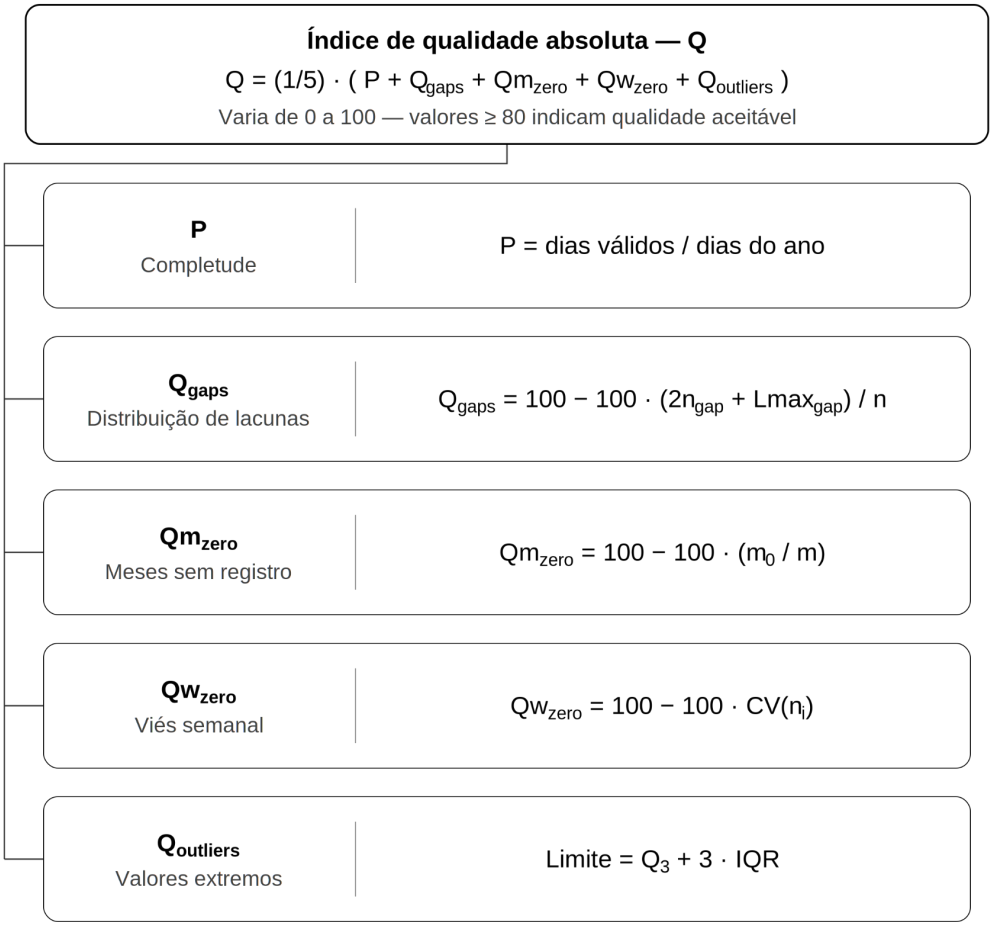


Figura 16: Fluxograma do cálculo do índice de qualidade absoluta.

Fonte: Autoral (2026).

O parâmetro P corresponde à proporção de dados anuais válidos, calculada pela razão entre os dias que contam com medições válidas (não nulas) e o número

total de dias do ano. O componente $Qgaps$ (Equação 2) mede a distribuição das lacunas na série, penalizando de forma mais intensa quando os dias sem registro se encontram intercalados entre dias com medições válidas (padrão "cheio–vazio–cheio"), e de maneira menos rigorosa quando as ausências se concentram em blocos contínuos. Ressalta-se que o *script* classifica como lacuna um dia em que não houve nenhum registro, ou seja, P = N/A, assim, precipitação igual a 0 não é lacuna. Nessa equação, n_{gap} representa o número de dias sem dados, $Lmax_{gap}$ corresponde ao maior intervalo contínuo sem registros, e n indica o total de dias no ano.

$$Qgaps = 100 - 100\frac{2n_{gap}+Lmax_{gap}}{n} \quad (2)$$

Valores elevados de $Qgaps$ indicam poucas lacunas ou ausências concentradas em períodos contínuos, possivelmente associados à manutenções ou interrupções operacionais, enquanto valores reduzidos sinalizam vazios distribuídos de forma dispersa, prejudicando a consistência da série.

O componente Qm_{zero} (Equação 3) penaliza anos com meses inteiros sem registros de precipitação, situação que pode refletir preenchimento artificial de valores iguais a zero em períodos de baixa pluviosidade ou ausência de medições. Nesse índice, $m0$ representa o número de meses cuja precipitação total foi igual a 0 e m o total de meses no ano.

$$Qm_{zero} = 100 - 100\frac{m0}{m} \quad (4)$$

O parâmetro Qw_{zero} (Equação 4) avalia a existência de viés semanal nos registros de precipitação, associando a ocorrência de eventos chuvosos (1, se $P \geq 1$ mm) ou secos (0) a cada dia da semana, com n_i representando o dia da semana, de 1 a 7. A distribuição esperada é uniforme, uma vez que a precipitação não deveria estar vinculada a dias específicos. Desvios acentuados dessa uniformidade, como a

concentração sistemática de registros em determinados dias, indicam a presença de acumulações multidiárias não registradas corretamente. O índice é calculado a partir do coeficiente de variação CV, definido pela razão entre o desvio-padrão e a média da frequência de dias chuvosos por semana.

$$Qw_{zero} = 100 - 100CV\left(n_i\right) \quad (4)$$

Por fim, o componente $Qoutliers$ avalia a proporção de dias cuja precipitação ultrapassa limites climatológicos superiores definidos para cada mês (Equação 5). O limite é determinado pelo terceiro quartil $Q3$ acrescido de três vezes o intervalo interquartílico IQR, considerando apenas dias com precipitação ≥ 1 mm.

$$Limite = Q3 + 3IQR \quad (5)$$

3.4 Cálculo do índice de qualidade relativa

A segunda etapa da metodologia corresponde ao cálculo do índice de qualidade relativa, destinado a avaliar a consistência dos registros de uma estação em comparação com postos vizinhos. Ressalta-se que algumas alterações foram feitas na metodologia original, para melhor se adequar à realidade mineira. Para tal, a simulação foi feita duas vezes, na primeira selecionando estações auxiliares em um raio máximo de 50 km e, na segunda, de 75 km, com penalização crescente associada à diferença altimétrica entre os pontos. O procedimento foi aplicado em ambas as situações para avaliar se o aumento da abrangência espacial resulta em maior representatividade. A coerência entre as séries foi estimada pela correlação das precipitações diárias, atribuindo valores mais elevados de qualidade relativa às estações que apresentaram maior similaridade espacial. O índice varia de 0 a 100, sendo que valores próximos de 100 indicam alta representatividade, enquanto valores reduzidos sugerem inconsistências ou baixa coerência regional. Todas as etapas foram aplicadas por um *script* em Python, desenvolvido para ler as planilhas

com registro de precipitação e aplicar as equações aqui citadas, codificadas com apoio de inteligência artificial generativa.

O primeiro parâmetro utilizado nessa etapa é o índice R (Equação 6), que expressa a representatividade espacial entre pares de estações. Nessa equação, da corresponde à distância máxima admitida (50 ou 75 km), d representa a distância real entre as estações, calculada no próprio *script* por meio da fórmula de Haversine, implementada manualmente em Python com o auxílio das funções trigonométricas do pacote *math*, a partir das coordenadas geográficas de latitude e longitude de cada estação. h representa a diferença de altitude em metros, e $Ccorr$ o coeficiente de correlação de Pearson calculado para os registros diários do par. O coeficiente de Pearson foi calculado com a biblioteca *NumPy* utilizando apenas os dias com registros simultâneos e válidos entre as duas séries, de modo a assegurar a comparabilidade entre os pares analisados. Esse coeficiente varia entre –1 e 1, indicando desde forte correlação negativa até forte correlação positiva. No presente estudo, o cálculo será restrito aos dias com registros simultâneos, assegurando comparabilidade. Valores elevados refletem maior semelhança nos padrões diários de precipitação, enquanto valores reduzidos indicam divergência. Apenas pares com $R \geq 70$ serão considerados válidos para as etapas subsequentes.

$$R = \frac{100}{3}\left(\frac{da-d}{da} + 0,5^{\frac{h}{500}} + Ccorr\right) \quad (6)$$

É calculada, então, a diferença diária relativa Dif (Equação 7) entre os registros das estações, comparada a um limite de tolerância T (Equação 8), ajustado para cada mês de forma a considerar a variabilidade sazonal da precipitação.

$$Dif = 100\frac{|PPTc-PPTaux|}{Av(PPTc,PPTaux)}Max(PPTc, PPTaux) \quad (7)$$

$$T = Cm \ln ln\,(101 - R) \quad (8)$$

Nessas equações, $PPTc$ representa a precipitação da estação central, $PPTaux$ a da estação auxiliar, Av a média aritmética dos dois valores, Max o maior

entre eles, e Cm um coeficiente de ajuste sazonal dependente da variabilidade espacial típica da precipitação em cada mês. Valores mais altos de Cm são associados a meses chuvosos e convectivos, enquanto valores mais baixos refletem a estação seca, caracterizada por maior homogeneidade espacial.

Na metodologia original, o índice Wm, representando a qualidade relativa, é calculado diariamente para cada estação, a partir da comparação entre ela e suas auxiliares, por meio da equação 9. Nela, $Rmin$ corresponde ao limiar mínimo de representatividade adotado (70 neste estudo), e L assume valor 1 para comparações entre Dif e T válidas e 0 para inválidas. O índice varia de 0 a 100, sendo valores acima de 50 considerados aceitáveis.

$$Wm = 100\frac{\sum Aux(R-Rmin)^2 L}{\sum Aux\,(R-Rmin)^2} \quad (9)$$

Os dias são classificados como insuficientes quando há menos de três estações auxiliares disponíveis, inválidos quando Wm < 20, dúbios para 20 < Wm < 50 e válidos quando Wm > 50. Entretanto, a aplicação direta desse critério em Minas Gerais mostrou-se inviável, devido à baixa densidade da rede pluviométrica. Aproximadamente 79% dos dias com medições foram classificados como insuficientes, uma vez que não havia o mínimo de três estações auxiliares para a realização das comparações, mesmo já ajustando a distância máxima admitida para pares de 75 km.

Diante dessa limitação, o critério mínimo de três estações auxiliares foi flexibilizado para viabilizar a aplicação da metodologia. Assim, neste estudo, o Wm passou a representar a porcentagem de dias em que Dif não superou o limite T, para cada par de estações. Posteriormente, os valores de Wm obtidos nas comparações entre a estação principal e cada uma de suas auxiliares foram agregados por meio de uma média ponderada pela representatividade R , resultando em um único valor de Wm para cada estação. Então, primeiramente, aplica-se a

Equação 10, calculando o Wm por par. Nela, n_{diasL} representa o total de dias entre as duas estações em que L foi 1, e n_{diasC} o total de dias comparáveis entre elas.

$$Wm_{par} = 100\frac{n_{diasL}}{n_{diasC}} \quad (10)$$

Por fim, para representar a qualidade relativa de uma estação, o $Wm_{estação}$ (Equação 11) é calculado a partir da média ponderada dos valores de Wm_{par} obtidos em todos os pares formados por essa estação. A ponderação é feita pelo termo $(R - Rmin)^2$, sendo R a representatividade do par e $Rmin$ o limiar mínimo adotado. Com isso, pares menos representativos, normalmente associados à maior distância, à maior diferença de altitude ou à menor correlação entre as séries, contribuem menos para o valor final.

$$Wm_{estação} = 100\frac{\sum(R-R_{min})^2 Wm_{par}}{\sum(R-R_{min})^2} \quad (11)$$

3.5 Modelagem estatística de valores extremos

3.5.1 Filtragem inicial e registro dos máximos anuais

Essa etapa iniciou-se com a exclusão das séries pluviométricas que não contam com um índice de qualidade absoluta Q igual ou superior a 80, e disponibilidade de, no mínimo, 30 anos de dados. A filtragem foi realizada por meio de um *script* em Python, responsável por percorrer as planilhas individuais de cada estação pluviométrica e, a partir da análise da coluna contendo as datas dos registros, contabilizar o número de anos com dados disponíveis. Paralelamente, o código da estação foi confrontado com a planilha contendo os respectivos índices de qualidade absoluta, permitindo identificar as séries que atendem simultaneamente

aos dois critérios. Como resultado, foi gerada uma nova planilha contendo apenas as estações selecionadas para as etapas posteriores da análise.

Em seguida, um segundo *script* em Python extraiu os maiores registros diários de precipitação de cada ano hidrológico, aplicando uma filtragem mais restritiva dos dados e gerando, paralelamente, uma planilha com a data (dia e mês) de ocorrência de cada máxima. Com o auxílio das bibliotecas *datetime* e *calendar* para a contabilização precisa dos dias de cada mês, o código percorreu os registros diários e contabilizou a quantidade de dias ausentes no período chuvoso de cada ano hidrológico, então os anos com 30 ou mais dias de falha nesse período foram excluídos.

Com o apoio da planilha de datas, os registros diários mais elevados foram examinados individualmente, verificando se o valor poderia corresponder a uma precipitação acumulada ao longo de vários dias e atribuída indevidamente a uma única data. Complementarmente, foi realizada a pesquisa por notícias de desastres associados ao dia em questão e os valores foram comparados com os das estações vizinhas. Essa análise fundamentou a adoção do limiar físico de 300 mm, valor com baixa probabilidade de ocorrência para a área de estudo, acima do qual os hidrológicos foram excluídos. Após essa etapa, as estações que passaram a dispor de menos de 30 anos de dados válidos também foram removidas da planilha de máximas gerada. Para fins deste estudo, o ano hidrológico foi definido como o intervalo entre o primeiro dia de outubro de um ano e o último dia de setembro do ano seguinte, e o período chuvoso, como os meses de outubro a março do respectivo ano hidrológico.

A aplicação desses filtros é essencial, pois ao mesmo tempo em que séries curtas tornam a base amostral menos robusta,registros incompletos ou inconsistentes também comprometem a representatividade dos eventos extremos (Naghettini; Pinto, 2007).

3.5.2 Verificação da condição de Amostra Aleatória Simples (AAS)

Para avaliar se as séries de máximas anuais podem ser tratadas como amostras estatisticamente adequadas, ou seja, se atendem às condições de Amostra Aleatória Simples, foram aplicados três testes que verificaram as hipóteses nulas H_0 de aleatoriedade, homogeneidade e tendência dos dados de máximas. Os testes foram aplicados para as séries através de um *script* em R. Para avaliar a aleatoriedade dos dados, foi utilizado o teste estatístico *Turning Point Test*, através do pacote *spgs*. Nele, conta-se o número p de pontos de inversão (observações maiores ou menores que ambos os vizinhos), e aplica-se as seguintes equações (NERC, 1975 apud Naghettini; Pinto, 2007):

$$E[p] = \frac{2(n-2)}{3} \tag{12}$$

$$Var[p] = \frac{16n-29}{90} \tag{13}$$

$$Z = \frac{p - E[p]}{\sqrt{Var[p]}} \tag{14}$$

Em que n é o número de observações da série; p, o número de pontos de inversão da série; $E[p]$ é o valor esperado do número de pontos de inversão sob H_0; $Var[p]$ a variância do número de pontos de inversão sob H_0 e Z, a estatística padronizada de teste.

Para verificar a homogeneidade da série, foi aplicado o Teste de Wilcoxon-Mann-Whitney entre a primeira e a segunda metade da série, pelo pacote *stats*, de acordo com as seguintes equações (Mann; Whitney, 1947):

$$U = R_1 - \frac{n_1(n_1+1)}{2} \tag{15}$$

$$E[U] = \frac{n_1 n_2}{2} \tag{16}$$

$$Var[U] = \frac{n_1 n_2 \left(n_1 + n_2 + 1\right)}{12} \quad (17)$$

$$Z = \frac{U - E[U]}{\sqrt{Var[U]}} \quad (18)$$

Nas equações, U representa a estatística de teste, R_1 a soma dos postos atribuídos às observações da primeira metade da série, após ordenar de forma crescente o conjunto das observações n, em que n_1 e n_2 são os tamanhos da primeira e segunda metade, respectivamente. $E[U]$ e $Var[U]$ correspondem ao valor esperado e à variância de U sob a hipótese nula, e Z é a estatística padronizada, que segue aproximadamente uma distribuição normal. Cabe observar que as Equações 15 a 18 correspondem à forma assintótica do teste, válida para amostras grandes. Como as metades das séries mais curtas contêm cerca de 15 observações, o pacote stats do R calcula o p-valor a partir da distribuição exata da estatística, recorrendo à aproximação Normal apenas na presença de empates ou de amostras maiores.

Para verificar a tendência da série, foi utilizado o teste de Mann-Kendall com *prewhitening*, representado pelas Equações 19 e 20, utilizando-se o pacote modifiedmk (Mann, 1945; Kendall, 1975; Yue et al., 2002).

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sinal(x_j - x_i) \quad (19)$$

$$sinal(x_j - x_i) = \begin{cases} +1, & se\ x_j > x_i \\ 0, & se\ x_j = x_i \\ -1, & se\ x_j < x_i \end{cases} \quad (20)$$

Na estatística S, a série de máximas anuais é entendida como uma sequência de valores ordenados no tempo, com um valor para cada ano hidrológico, sendo que o primeiro corresponde ao ano mais antigo e o último, ao mais recente. Cada posição dessa sequência representa um instante de tempo, e os índices i e j identificam duas dessas posições, sempre com j posterior a i, ou seja, $1 \leq i < j \leq n$

, sendo n o número de anos válidos da estação. A estatística S soma esses sinais para todos os pares possíveis de anos da série: quando os aumentos predominam, S assume valores positivos, indicando tendência de crescimento ao longo do tempo; quando predominam as reduções, torna-se negativo, sinalizando tendência de queda; e valores próximos de zero sugerem ausência de tendência.

Devido à aplicação simultânea dos testes em diversas estações, adotou-se o controle da False Discovery Rate (FDR), conforme proposto por Benjamini e Hochberg (1995). Esse procedimento reduz a probabilidade de identificar como significativos resultados que, na realidade, correspondem a falsas rejeições da hipótese nula. Assim, o FDR permite uma avaliação mais criteriosa dos p-valores obtidos, especialmente em análises envolvendo muitas séries pluviométricas (Cardoso, 2026). O controle da FDR foi aplicado separadamente aos p-valores de cada teste (aleatoriedade, homogeneidade e tendência), para todas as estações, seguindo a Equação 21. Para um conjunto de *m* testes, o método de Benjamini e Hochberg ordena os p-valores de forma crescente, e associa a cada posição *i* um p-valor ajustado, definido pela Equação 21. A hipótese nula é rejeitada para toda estação cujo p-valor ajustado seja inferior ao nível de significância regional α_R, adotado como 5%. Na implementação, o ajuste foi realizado pela função *p.adjust* do R, com o método BH (Benjamini; Hochberg, 1995).

$$p_{(i)}^{BH} = min_{j \geq i}\left[min\left(\frac{m}{j} p_{(j)}, 1\right)\right] \quad (21)$$

Nessa equação, p representa os p-valores das m estações ordenados de forma crescente, e o índice (i) indica a posição do p-valor nessa ordenação, de modo que $p_{(1)}$ é o menor e $p_{(m)}$ o maior. O m representa a quantidade de vezes em que o teste foi reproduzido, a mesma quantidade de estações no conjunto. O índice *j* é apenas um índice auxiliar de cálculo. Enquanto *i* identifica a estação cujo p-valor ajustado está sendo determinado, *j* percorre todas as posições a partir de *i* até a última, *m*, calculando o termo $\frac{m}{j} p_{(j)}$ em cada uma delas e retendo o menor valor

encontrado. Esse mecanismo assegura que os p-valores ajustados preservem a mesma ordem crescente dos p-valores originais, impedindo que uma estação com p-valor menor receba um ajuste maior que o de outra com p-valor maior.

Com a planilha de máximas anuais gerada e as séries consideradas amostras aleatórias simples selecionadas, um *script* na linguagem de programação R aplica o teste MGBT (*Multiple Grubbs–Beck Test*), baseado nas Equações 22 e 23, proposto por Cohn et al. (2013), através do pacote MGBT disponível no CRAN (*Comprehensive R Archive Network*), a fim de excluir anos cujo registro da máxima anual é considerado um *outlier* mínimo.

$$G = \frac{\underline{Y} - y_{(1)}}{S} \tag{22}$$

$$\omega_r = \frac{x_{(r)} - \underline{x}_{(r+1:n)}}{S_{(r+1:n)}} \tag{23}$$

A Equação 22 trata-se do caso mais simples, testando-se apenas o menor valor da série, em que $y_{(1)}$ é o menor valor, $\underline{Y}$ a média dos logaritmos dos dados e S o desvio padrão amostral desses logaritmos, repetindo-se o procedimento a cada valor removido. A forma generalizada, efetivamente aplicada neste trabalho, avalia cada posição candidata r pela Equação 23, na qual $x_{(r)}$ é o r-ésimo menor valor e $\underline{x}_{(r+1:n)}$ e $S_{(r+1:n)}$ são a média e o desvio padrão calculados apenas com as observações acima dele. Ao excluir os próprios candidatos a outlier do cálculo, essa formulação evita o efeito de mascaramento, em que vários valores baixos contaminam as estatísticas da amostra e impedem sua detecção. A significância de cada ω decorre da distribuição da r-ésima estatística de ordem de uma amostra normal, e um procedimento de varredura (*sweep out*) define quantos outliers mínimos excluir ao nível de significância (neste trabalho, de 5%).

A remoção desses outliers mínimos nas séries de precipitação diária máxima anual justifica-se pela necessidade de assegurar que os valores utilizados no ajuste estatístico representem adequadamente os eventos extremos observados em cada

ano hidrológico. Em séries de máximas anuais, valores excepcionalmente baixos podem estar associados não necessariamente a anos hidrologicamente pouco chuvosos, mas a falhas de registro, ausência de dados no período chuvoso ou inconsistências que impedem a correta identificação do verdadeiro máximo anual (Cardoso, 2026).

3.5.3 Distribuição Generalizada de Valores Extremos (GEV)

Com as estações selecionadas, iniciam-se os cálculos dos parâmetros da GEV, através do método dos momentos-L (MML). No ambiente R, os momentos-L para cada série são calculados pelo pacote *lmom*, através da função *samlmu*, configurada para retornar os quatro primeiros momentos. A partir desses momentos, a função *pelgev* estima os parâmetros de posição, escala e forma, sendo o parâmetro de forma expresso na convenção de Hosking (κ) adotada ao longo do trabalho.

Os quatro momentos-L calculados pelo *script* correspondem, respectivamente, à média da série L_1, à dispersão L_2, à assimetria L_3 e à curtose L_4. A partir deles, definem-se a assimetria-L ($\tau_3 = L_3/L_2$) e a curtose-L ($\tau_4 = L_4/L_2$). A estimação dos parâmetros consiste em igualar esses momentos-L amostrais às suas expressões teóricas para a GEV, apresentadas nas Equações 24 a 26 (Hosking, 1990):

$$L_1 = \mu + \frac{\sigma}{\kappa}[1 - \Gamma(1 + \kappa)] \quad (24)$$

$$L_2 = \frac{\sigma}{\kappa}\left(1 - 2^{-\kappa}\right)\Gamma(1 + \kappa) \quad (25)$$

$$\tau_3 = \frac{2\left(1-3^{-\kappa}\right)}{1-2^{-4}} - 3 \quad (26)$$

Nas equações, Γ é a função gama e κ é o parâmetro de forma. A estimação dos parâmetros se inicia pela Equação 26, a partir da assimetria-L (τ_3) da série, obtendo-se o parâmetro de forma. Com ele, obtém-se a escala (σ) a partir da Equação 25, e, em sequência, a posição (μ), pela Equação 24. Dessa forma, ficam atribuídos os três parâmetros da GEV para uma série de precipitação, viabilizando análises subsequentes relacionadas à frequência de chuvas extremas.

Com os parâmetros estimados, verifica-se o quanto a GEV ajustada descreve cada série por meio do teste de aderência de Kolmogorov-Smirnov, que compara a função de distribuição empírica dos dados observados com a função de distribuição acumulada da GEV ajustada, tomando como estatística a maior discrepância vertical entre ambas. O teste foi aplicado em ambiente R pela função *ks.test* (pacote *stats*), utilizando a função de distribuição acumulada *cdfgev* (pacote *lmom*) com os parâmetros estimados para cada estação. Adota-se o nível de significância de 5%, de modo que as séries com p-valor superior a 0,05 são consideradas aderentes à GEV. Ressalva-se que, por serem os parâmetros estimados a partir da própria amostra, o teste tende a ser conservador, produzindo p-valores superiores aos que se obteriam com parâmetros conhecidos a priori, por isso, a aderência é avaliada também por evidência gráfica.

Para complementar a avaliação da aderência, elaborou-se um diagrama de razões de momentos-L, confrontando os pares de assimetria-L (τ_3) e curtose-L (τ_4) de cada estação com a relação teórica da GEV. Diferentemente do teste de Kolmogorov-Smirnov, que avalia cada série isoladamente e tende a não rejeitar a hipótese de aderência, o diagrama examina o conjunto das estações de forma simultânea, sendo ferramenta consolidada para a discriminação entre distribuições em análise de frequência.

3.5.4 Análise de cluster

Com o objetivo de investigar se a variabilidade espacial do parâmetro de forma se organiza em regiões geograficamente coerentes e de subsidiar as análises

conduzidas por grupo, delimitaram-se regiões homogêneas por análise de agrupamento pelo método *k-means*, aplicado às estações que passaram pelo controle de qualidade e pelos testes de AAS e tiveram a GEV ajustada. Cada estação foi caracterizada por quatro variáveis: latitude; longitude; altitude e PRCPTOT, padronizadas por z-score (Equação 27) para assegurar contribuição equivalente ao cálculo das distâncias. O método particiona as estações em K grupos minimizando a soma dos quadrados das distâncias euclidianas de cada estação ao centroide de seu grupo (Equação 28).

Dentre as quatro variáveis, apenas o PRCPTOT não está disponível diretamente pelas fontes descritas no item 3.2, exigindo cálculo específico. Trata-se do índice que expressa a precipitação total acumulada nos dias úmidos do ano, considerando-se úmido todo dia com precipitação igual ou superior a 1 mm. O índice foi obtido por um *script* em Python a partir das planilhas de precipitação, dando preferência ao registro de chuva consistido em detrimento do bruto, quando existem ambos. A validade de cada ano civil foi avaliada segundo os critérios de consistência do RClimDex, descartando-se os anos com mais de 15 dias de falha no total ou com mais de 3 dias de falha em qualquer um de seus meses. Sobre a série de valores anuais válidos aplicou-se ainda um filtro pela amplitude interquartil, excluindo-se os anos cujo PRCPTOT fosse inferior a Q1 - 1,5AIQ, de modo a remover totais anuais anomalamente baixos, associados a registros incompletos. O PRCPTOT médio de cada estação corresponde à média dos valores anuais remanescentes, sendo consideradas apenas as estações com pelo menos 30 anos válidos após essa filtragem.

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \tag{27}$$

Na equação, x_{ij} é o valor da variável j para a estação i, e $\bar{x}_j$ e s_j são, respectivamente, a média e o desvio padrão amostral da variável j entre todas as estações.

$$min_{C_1,\ldots,C_k} \sum_{k=1}^{K} \sum_{x_1 \in C_k} \left|\left|x_i - \mu_k\right|\right|^2 \Big|\Big| \quad , \quad \mu_k = \frac{1}{|C_k|} \sum_{x_1 \in C_k} x_1 \tag{28}$$

Nas equações, C_k é o conjunto das estações atribuídas ao cluster K; x_1 o vetor padronizado das quatro variáveis da estação i; μ_k é o centroide do grupo; K é o número de grupos; e $\left|\left|x_i - \mu_k\right|\right|^2 = \sqrt{\sum_{j=1}^{p} (x_{ij} - \mu_{kj})^2}$ é a distância euclidiana no espaço das quatro variáveis. Minimizar essa soma equivale, portanto, a tornar cada estação o mais próxima possível do centro do seu próprio grupo, resultando em clusters internamente compactos.

O número de grupos, exigido a priori pelo método, foi determinado pelo índice de silhueta (Equação 29), calculado para valores de k de 2 a 10 com a função *fviz_nbclust* do pacote *factoextra*, adotando-se o k que maximizou a silhueta média. O agrupamento foi obtido pela função *kmeans* do pacote *stats*, com 25 inicializações aleatórias independentes (*nstart* = 25), retendo-se a solução de menor soma de quadrados, e fixou-se a semente do gerador de números aleatórios para garantir a reprodutibilidade. Por fim, os grupos foram rotulados pela latitude média, e constituíram a base para as análises subsequentes, conduzidas tanto para o conjunto global quanto para cada região homogênea.

$$s(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{29}$$

Aqui, $a(i)$ é a distância média da estação i às demais do seu grupo e $b(i)$ a distância média às estações do grupo vizinho mais próximo; adota-se o número de grupos que maximiza a média de $s(i)$.

3.5.5 Estatísticas do parâmetro de forma

A partir dos valores de κ estimados para as estações, calcularam-se as estatísticas descritivas, sendo elas o mínimo, máximo, primeiro e terceiro quartis,

mediana, média, desvio padrão, coeficiente de variação e assimetria, obtidas pelas respectivas funções do R, à exceção da assimetria, calculada pela função skewness do pacote e1071 (type = 2). Essas medidas foram obtidas para o conjunto global e, separadamente, para cada região homogênea, com a distribuição dos valores de κ visualizada por boxplots e histogramas com as normais ajustadas.

Partiu-se então para a verificação da normalidade do conjunto dos valores do parâmetro de forma. Para tal, aplicou-se, em ambiente R, o teste de Shapiro-Wilk por meio da função *shapiro.test* (pacote stats), adotando-se como hipótese nula a de que esses valores provêm de uma população com distribuição normal. A estatística de teste W quantifica o grau de aderência dos dados ordenados a uma distribuição normal e é definida pela Equação 30:

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n} (x_i - \underline{x})^2} \tag{30}$$

Em que $x_{(i)}$ é o i-ésimo valor da amostra ordenada de forma crescente, x_i é a i-ésima observação da amostra, $\underline{x}$ é a média amostral, n é o número de observações e a_i é um coeficiente de ponderação predefinido pelo teste, que depende apenas do tamanho da amostra e das propriedades de uma distribuição normal. Quanto mais próximo o W for de 1, maior é a compatibilidade da amostra com a normalidade. Adota-se o nível de significância de 5%, de modo que um valor-p superior a 0,05 leva à não rejeição da hipótese de normalidade.

Aplicou-se o teste em dois níveis: sobre o conjunto global dos valores de κ e, separadamente, sobre cada uma das regiões homogêneas delimitadas na análise de agrupamento, com o objetivo de caracterizar essas distribuições e subsidiar as comparações posteriores. A verificação da normalidade fundamenta, ainda, a comparação com outros trabalhos, uma vez que estes reportam apenas a média e o desvio padrão do parâmetro de forma, grandezas suficientes para caracterizar uma distribuição normal.

Após a avaliação das estatísticas gerais, partiu-se para a avaliação da influência do comprimento das séries sobre a qualidade da estimativa, em que as estações foram agrupadas por faixa de anos hidrológicos válidos (30-40, 41-50, 51-60, 61-70, 71-80 e 81-90), calculando-se, em cada faixa, a média de κ e seu intervalo de confiança de 95%, o desvio padrão e a proporção de estações de cada família da GEV (Fréchet, $\kappa < 0$; Weibull, $\kappa > 0$). A divisão visou verificar se séries mais curtas produzem estimativas menos precisas e eventualmente enviesadas, refletidas em maior dispersão de κ.

Por fim, as estatísticas do conjunto global foram comparadas às de Papalexiou e Koutsoyiannis (2013) e Cardoso (2026), por meio de tabela comparativa e de histograma com as normais ajustadas de cada conjunto. Como esses trabalhos adotam a parametrização de Coles (2001), de sinal oposto à convenção de Hosking, suas estatísticas foram convertidas ($\kappa = -\xi$): inverte-se o sinal da média, mediana, assimetria e coeficiente de variação; mantém-se o desvio padrão; e trocam-se entre si, com sinal invertido, o mínimo e o máximo e o primeiro e terceiro quartis.

3.5.6 Análise multidiária

Para avaliar se o parâmetro de forma se mantém estável quando a precipitação é agregada em diferentes durações, comportamento teoricamente esperado, cuja confirmação reforçaria κ como um descritor regional robusto e coerente entre durações, as análises foram estendidas a somas móveis de 2, 3, 5 e 10 dias, comparadas em conjunto com a série de máximas diárias. Para cada estação, obteve-se a máxima soma de cada ano hidrológico, considerando-se válidas apenas janelas de dias consecutivos e sem falhas, de forma que a diferença entre as datas final e inicial deve se igualar a duração (para 5 dias, de 06/10 a 10/10) e nenhum registro diário pode ser nulo. As quatro séries resultantes foram submetidas aos mesmos critérios de validade e ao mesmo procedimento metodológico das máximas diárias, incluindo a análise detalhada dos anos com

máxima diária acima de 300 mm, o controle de qualidade, a verificação de AAS, a remoção de outliers mínimos e o ajuste da GEV por momentos-L.

Sobre cada série, repetiram-se as análises do parâmetro de forma. Foram levantadas e comparadas as estatísticas descritivas de κ e sua média por duração, para avaliar a tendência com o aumento da janela. Também foram feitos o teste de normalidade de Shapiro-Wilk, com as respectivas normais ajustadas; o diagrama de razões de momentos-L (τ_3, τ_4) por duração; e a proporção de estações de cada família da GEV (Fréchet, $\kappa < 0$; Weibull, $\kappa > 0$), dividida por faixa de tamanho de série.

4 Resultados e discussão

4.1 Índice de qualidade absoluta

O processo do controle de qualidade absoluta gerou uma planilha com 651 estações (Apêndice C), após excluir a estação 2146084, classificada como *data logger*, que conta com registros que não contribuem para este estudo. A planilha inclui o índice de cada estação, além do valor de cada parâmetro utilizado para calculá-lo. O índice indica qualidade elevada, com 89,88% das estações pontuando acima de 90, com o restante, excluindo uma, conseguindo pontuação acima de 80, limite considerado aceitável.

A Figura 17 ilustra a distribuição espacial do índice em Minas Gerais, revelando uma queda na qualidade na porção Noroeste do estado. Nota-se que a queda na qualidade também coincide com uma menor densidade de estações, indicando que a qualidade pode estar relacionada à disposição de infraestrutura na região. Além disso, índices mais baixos se encontram em regiões mais áridas segundo a classificação de Thornthwaite, podendo ser reflexo de uma maior quantidade de meses com precipitação igual a 0, que impacta negativamente o índice. Salienta-se que a maior qualidade dos dados está concentrada na região central e na Zona da Mata, onde estão municípios que frequentemente enfrentam problemas de inundação, como Belo Horizonte, Juiz de Fora e Ubá.

Ressalta-se que o mapa não revela quais fatores mais prejudicaram isoladamente a qualidade, então, a Tabela 1 contém as médias e medianas de cada parâmetro separado. A tabela evidencia que o Qm_zero e o Qw_zero, responsáveis por penalizar, respectivamente, anos com meses cujos registros totalizam 0 mm e séries com padrões de registro semanal, ficaram visivelmente inferiores aos demais componentes do índice. Esse comportamento indica que uma parcela importante da perda de qualidade das séries está associada a possíveis inconsistências no padrão de registro diário da precipitação, especialmente pela ocorrência de meses inteiramente zerados e pela presença de viés semanal nos dados. Essa análise é de grande importância, pois revela quais aspectos mais prejudicam a qualidade dos

dados pluviométricos do estado, podendo contribuir para diagnósticos direcionados à operação, manutenção e melhoria na rede.

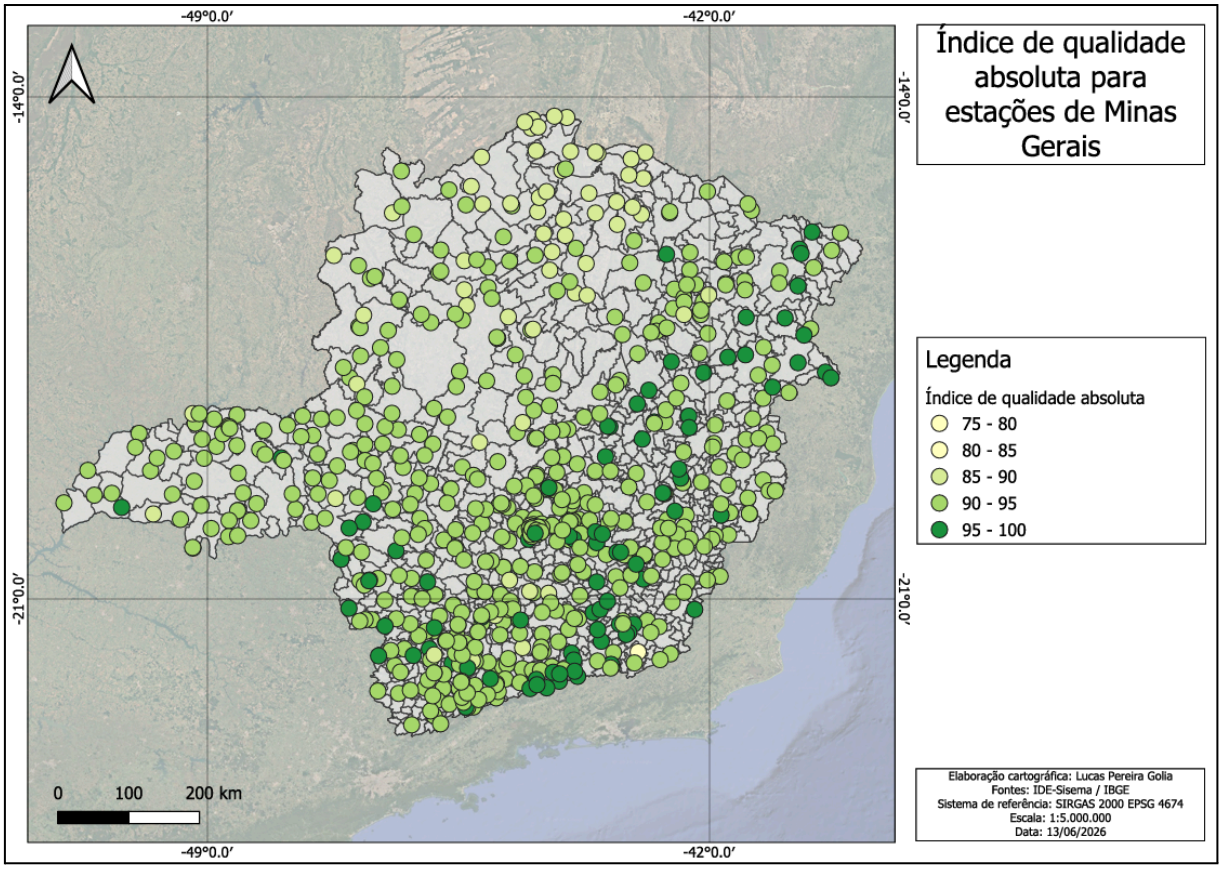


Figura 17: Índice de qualidade absoluta para Minas Gerais.

Fonte: Autoral (2026)

Tabela 1: Média e mediana dos parâmetros do índice de qualidade absoluta.

Parâmetro	P (%)	Qgaps (%)	Qm_zero (%)	Qw_zero (%)	Qoutliers (%)
Média	98,80	97,06	88,35	80,59	99,60
Mediana	99,41	98,33	90,20	81,14	99,65

Fonte: Autoral (2026).

É importante destacar que o índice de qualidade absoluta representa bem a consistência interna das séries, porém não se traduz em uma evidência direta de adequada cobertura espacial da rede. O índice avalia características internas, mas não considera a densidade de estações, sua representatividade espacial e nem sua capacidade de captar a variabilidade pluviométrica local. Assim, verifica-se a importância da etapa de qualidade relativa, que visa quantificar se uma estação representa bem o regime pluviométrico local, contribuindo para a análise da qualidade da rede.

4.2 Índice de qualidade relativa

A aplicação do processo de cálculo do índice de qualidade relativa revelou algumas limitações do método na aplicação em Minas Gerais, principalmente derivados da menor densidade de estações, se comparada a região onde a metodologia foi desenvolvida. Primeiramente, as Figuras 18 e 19 representam a qualidade relativa associada a cada estação do estado que obteve ao menos um par elegível para comparação ($R \geq 70$). Essa exigência se constitui como limitação, especialmente em regiões com menor densidade de estações, pois observa-se que, ao considerar-se uma distância máxima de 50 km (Figura 18), apenas 292, ou 44,72% do total de estações possuíam par elegível e obtiveram um valor de Wm. Ao elevar-se a distância máxima para 75 km (Figura 19), ampliando a abrangência da análise, a quantidade de estações aumentou para 421, ou 64,47%. No entanto, a adoção de um limite de distância mais permissivo deve ser realizada com cautela, uma vez que o aumento da separação espacial entre as estações tende a reduzir a similaridade entre os regimes pluviométricos por elas representados, comprometendo a representatividade e a confiabilidade das comparações. O Apêndice D detalha o cálculo da correlação entre entre pares considerando a distância de 75 km, e os Apêndices E e F contém os resultados do cálculo do índice de qualidade relativa por estação, considerando 50 e 75 km, respectivamente.

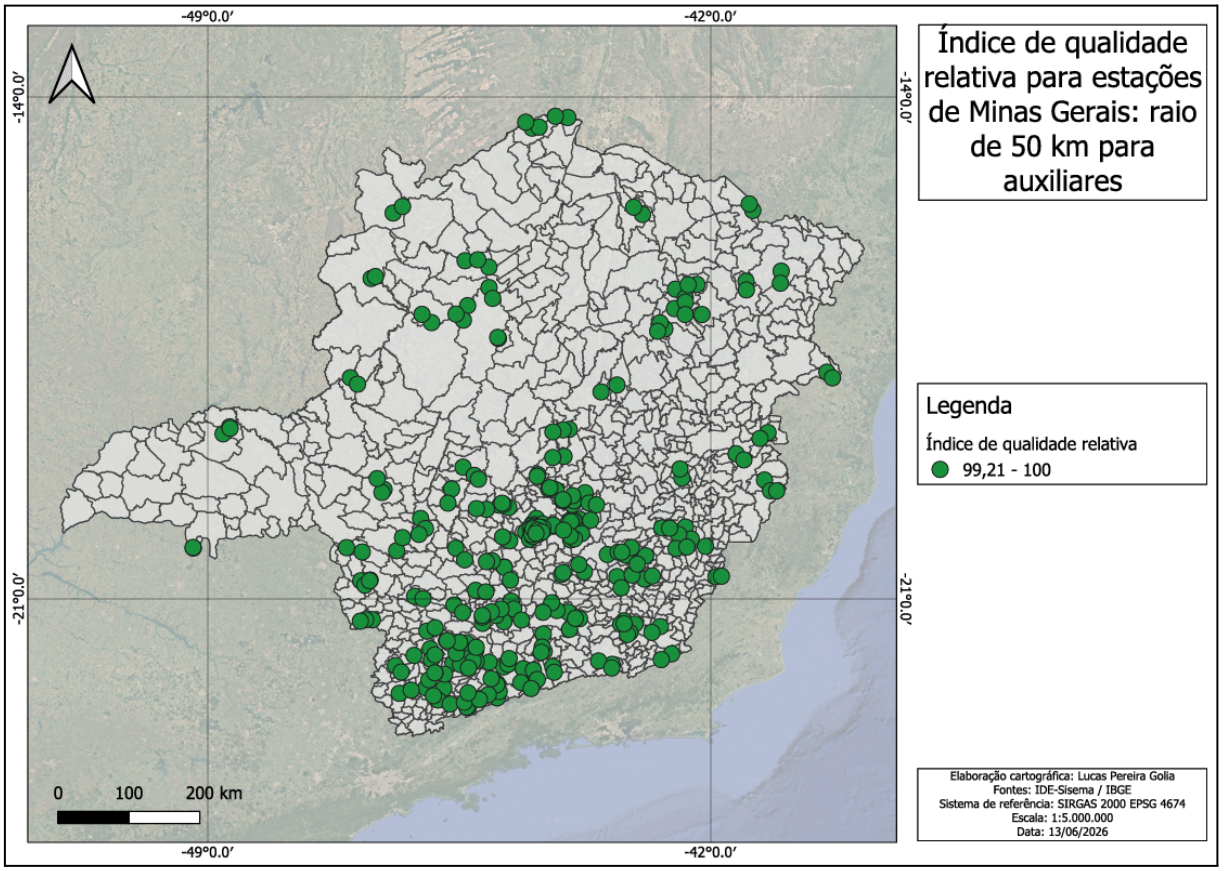


Figura 18: Índice de qualidade relativa para Minas Gerais considerando um raio de 50 km para estações auxiliares.

Fonte: Autoral (2026).

A análise do índice relativo revela uma qualidade muito elevada. Para as distâncias de 50 e 75 km, a quantidade de estações com Wm igual à 100 foi de 39,7% e 33,72%, respectivamente. Isso revela que, em grande maioria, os valores diários de Dif não ultrapassaram o limite T, ou seja, a diferença entre um registro diário de precipitação entre uma estação principal e sua auxiliar não foi grande o suficiente para ultrapassar o limite estabelecido, indicando que o par representa bem o regime da região geográfica próxima. É importante destacar que o valores do limite T poderiam ser revisados com profundidade, pois foram obtidas diferenças diárias na ordem de 80 mm entre as estações de um par, que ainda foram consideradas

aceitáveis, podendo indicar que o limite está muito permissível. O Apêndice G registra a comparação entre o Dif e o T por par, considerando 75 km.

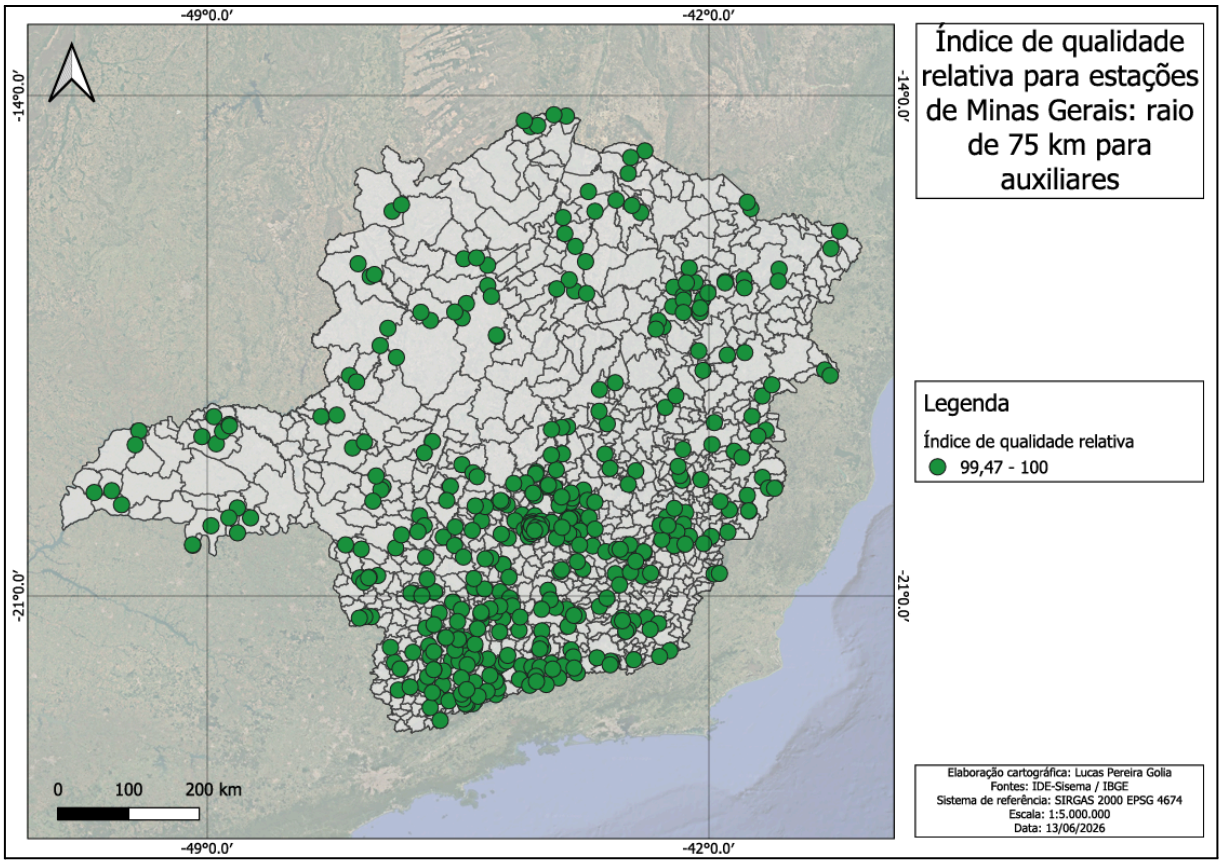


Figura 19: Índice de qualidade relativa para Minas Gerais considerando um raio de 75 km para estações auxiliares.

Fonte: Autoral (2026).

Na metodologia original, é exigido que uma estação principal tenha no mínimo três auxiliares com dados para que o Wm seja calculado por dia. Essa exigência foi a maior limitação encontrada para a aplicação em Minas Gerais. No artigo original, adotando-se uma distância máxima de 50 km, 25% dos dias, para todas as estações, foram classificados como insuficientes, ou seja, não houve três ou mais auxiliares. Neste estudo, mesmo já ampliando a distância máxima para 75 km, 79% dos dias ficaram classificados como insuficientes, resultado da baixa densidade de estações no estado. Assim, a metodologia foi reajustada: originalmente, uma

estação principal é comparada simultaneamente com várias auxiliares, e como o índice depende da concordância com a maioria, uma auxiliar defeituosa não reduz a qualidade. Neste trabalho, uma estação principal é comparada com suas auxiliares individualmente, par a par, e como o índice resulta da média entre a pontuação por par, uma auxiliar defeituosa poderia jogar o índice de uma boa estação para baixo. Nessa formulação, a média é ponderada pela representatividade do par, que pode atenuar o efeito de uma auxiliar defeituosa. Essa adaptação possibilitou avaliar estações que ficariam sem estimativa de qualidade relativa, ao custo de menor robustez.

4.3 Modelagem estatística de valores extremos

4.3.1 Filtragem inicial e registro dos máximos anuais

A filtragem foi conduzida em duas etapas, por meio de dois scripts em Python. Na primeira, das 652 estações inicialmente consideradas, selecionaram-se as 449 que atenderam simultaneamente ao índice de qualidade absoluta Q igual ou superior a 80 e à disponibilidade mínima de 30 anos de dados, compiladas no Apêndice H. Na segunda, a partir dessas 449 estações, registraram-se as máximas por ano hidrológico e excluíram-se os anos com 30 ou mais falhas no período chuvoso. Também foram avaliadas detalhadamente registros diários superiores a 300 mm, valor utilizado como limiar de referência. Em seguida, as estações que passaram a dispor de menos de 30 anos válidos foram removidas, restando 404 estações.

O produto desta etapa é a planilha de máximas diárias anuais, com uma coluna por estação e uma linha por ano hidrológico, abrangendo o período de 1912 a 2024. O mesmo script gerou ainda uma planilha auxiliar com a data (dia e mês) de cada máxima, útil para a conferência posterior dos registros. Os Apêndices I e J registram as máximas anuais de cada estação e suas datas, respectivamente.

A comparação entre as etapas evidencia que a redução do número de estações decorreu quase inteiramente da exigência de séries longas, pois das 652 estações iniciais, apenas uma foi excluída por índice de qualidade inferior a 80, de modo que praticamente todas as 204 exclusões da primeira filtragem se deveram à

disponibilidade de menos de 30 anos de dados. Esse predomínio revela que a principal limitação do conjunto não é a qualidade dos registros, mas a extensão temporal das séries, o que pode estar associado à instalação recente de parte das estações ou à operação descontínua da rede pluviométrica. A opção por reter apenas séries com pelo menos 30 anos torna as estimativas dos parâmetros da GEV mais robustas, sobretudo a do parâmetro de forma, cuja estimação é particularmente sensível à extensão da amostra, porém, em contrapartida, reduz a densidade espacial da rede analisada, o que pode comprometer a representatividade de sub-regiões menos monitoradas e deve ser considerado na interpretação das regiões homogêneas delimitadas adiante.

4.3.2 Verificação da condição de Amostra Aleatória Simples (AAS)

Aplicados os três testes às 404 estações aprovadas na filtragem inicial, avaliou-se em que medida as séries de máximas anuais não rejeitam as condições de amostra aleatória simples. A Tabela 2 apresenta o número de rejeições da hipótese nula em cada teste, antes e após o controle da *False Discovery Rate*. Sem a correção, 87 estações seriam reprovadas em pelo menos um teste, e com o controle da FDR, esse total reduziu-se a 35 (8,7% do conjunto), restando 369 séries para as etapas subsequentes. Como essas séries rejeitaram a hipótese de pertencer a uma amostra aleatória simples, foram removidas do conjunto, uma vez que o ajuste da GEV pressupõe observações independentes e provenientes de uma mesma população. Nota-se que uma mesma estação pode ser reprovada em mais de um teste, de modo que a última coluna da tabela não corresponde à soma das anteriores. O Apêndice K registra, para cada estação, os p-valores associados a cada teste aplicado, além da correção feita pelo controle FDR e as rejeições da hipótese nula.

Tabela 2: Percentuais de rejeição da hipótese nula nos testes de AAS.

Tipo de controle	Aleatoriedade (TP)	Homogeneidade (WMW)	Tendência (MK-TFPW)	Ao menos um teste

p-valor	4,5%	13,9%	15,6%	21,5%
FDR	0,0%	5,2%	7,2%	8,7%

Fonte: Autoral (2026).

A magnitude dessa redução evidencia o efeito da multiplicidade: ao submeter centenas de séries a três testes simultâneos, parte das rejeições ao nível de 5% decorre do acaso. O teste de aleatoriedade é ilustrativo, pois nenhuma das 18 rejeições iniciais resistiu à correção, e sua proporção de 4,5% aproxima-se da taxa de falsos positivos esperada sob a hipótese nula, corroborando a independência das máximas anuais. Já a homogeneidade e a tendência preservaram, respectivamente, 21 e 29 rejeições após o ajuste, indicando que, nessas séries, as diferenças detectadas são acentuadas o bastante para não serem atribuídas ao acaso. Cabe ressaltar que, ao controlar a proporção esperada de falsas descobertas, o procedimento reduz a sensibilidade a sinais fracos, de modo que tendências de menor magnitude podem não ser detectadas.

Além do efeito da correção, a baixa proporção de estações reprovadas pode estar associada às características climáticas da área de estudo. Minas Gerais situa-se em uma zona de transição, abrangendo desde o Norte, de clima semiárido e chuvas irregulares, até o Sul, com os maiores totais pluviométricos do estado. Nesse contexto, os sinais de tendência tendem a ser heterogêneos e espacialmente contraditórios, dificultando que alcancem significância estatística de forma generalizada.

A remoção de outliers mínimos foi aplicada exclusivamente às 369 estações aprovadas nos testes de AAS. O MGBT identificou outliers em 105 delas (28,5%), enquanto 264 (71,5%) não apresentaram nenhum, totalizando 903 anos removidos, 4,66% do total. Após a remoção, 17 estações passaram a dispor de menos de 30 anos válidos e foram igualmente excluídas, restando 352 séries para o ajuste da GEV. O Apêndice L contém todos os registros excluídos por estação.

A distribuição das remoções mostrou-se bastante assimétrica. Em 29 estações, apenas um único ano foi excluído, e metade das estações afetadas perdeu até cinco anos, o que sugere ocorrências pontuais. Em contrapartida, seis estações tiveram removida exatamente a metade de seus registros, chegando a 41 anos excluídos em uma delas. Esse número corresponde justamente ao limite máximo admitido pelo procedimento de varredura do MGBT, que testa candidatos apenas até a metade da amostra, tratando-se, portanto, de uma saturação do teste, e não da identificação de um conjunto genuíno de valores atípicos.

A inspeção dessas séries afasta as explicações mais imediatas, pois os valores removidos distribuem-se ao longo de todo o registro, e não em períodos contínuos, e as estações não apresentam máximas atipicamente elevadas, sendo suas medianas situadas em torno de 77 mm, próximas à mediana do conjunto. O comportamento decorre da própria formulação do teste, que pressupõe log-normalidade e supõe que os valores atípicos inferiores sejam poucos em relação ao conjunto, e quando a metade inferior da série se afasta desse pressuposto, o procedimento sinaliza sucessivamente os candidatos até atingir o teto. O efeito é favorecido pelo nível de significância de 5% adotado, dez vezes superior ao padrão de 0,005 da implementação do teste. Como consequência, essas seis estações passaram a dispor de menos de 30 anos válidos e foram excluídas, o que evidencia uma limitação metodológica, o MGBT requer atenção ao número de valores sinalizados, sobretudo quando este se aproxima do limite do procedimento.

4.3.3 Distribuição Generalizada de Valores Extremos (GEV)

A distribuição GEV foi ajustada pelo método dos momentos-L às 369 estações aprovadas na verificação das condições de AAS e às 352 remanescentes após a remoção dos outliers mínimos, obtendo-se estimativas dos três parâmetros para a totalidade das séries. A Tabela 3 resume os parâmetros de posição e escala estimados, contidos no Apêndice M.

Tabela 3: Resumo dos parâmetros de posição e escala.

Parâmetro	Cenário	Mínimo	Mediana	Máximo
Posição (mm)	Completas	50,0	70,1	90,3
	Sem outliers	53,6	71,1	95,0
Escala (mm)	Completas	7,5	18,4	33,1
	Sem outliers	7,5	17,3	27,0

Fonte: Autoral (2026).

A posição mediana, próxima de 70 mm, indica o valor mediano da máxima diária anual típica das estações mineiras, enquanto a escala, com mediana em torno de 18 mm, expressa a variabilidade interanual dessas máximas. A remoção dos outliers mínimos elevou a posição e reduziu a escala, comportamento coerente com a supressão dos menores registros de cada série, que desloca o centro e comprime a dispersão.

A aderência foi avaliada pelo teste de Kolmogorov-Smirnov, e para as 352 séries, ao nível de 5%, não foram encontradas evidências para rejeitar o ajuste da GEV (D mediano de 0,074; p-valor mínimo de 0,107; resultados completos no Apêndice M). Contudo, o teste é conservador quando os parâmetros são estimados a partir da própria amostra, de modo que a ausência de rejeições indica ausência de evidência contrária ao ajuste, e não sua confirmação definitiva. Por isso, a aderência foi examinada também pelo diagrama de razões de momentos-L (Figuras 20 e 21), no qual os pares (τ_3, τ_4) distribuem-se em torno da curva teórica da GEV, com o ponto médio praticamente sobre ela em ambos os cenários, evidência independente do teste de aderência da família. A Tabela 4 complementa a comparação entre os dois cenários.

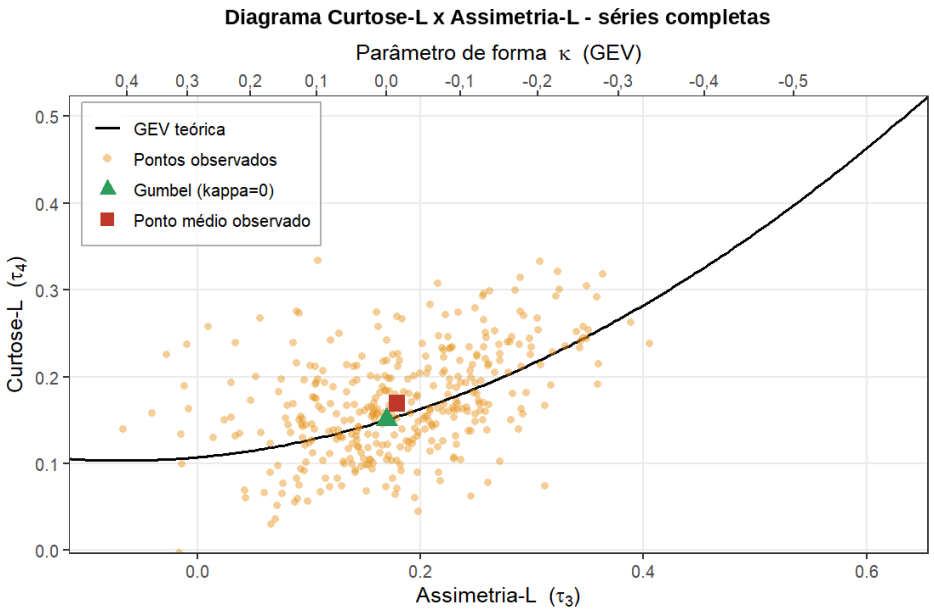


Figura 20: Diagrama de momentos-L para séries completas.

Fonte: Autoral (2026).

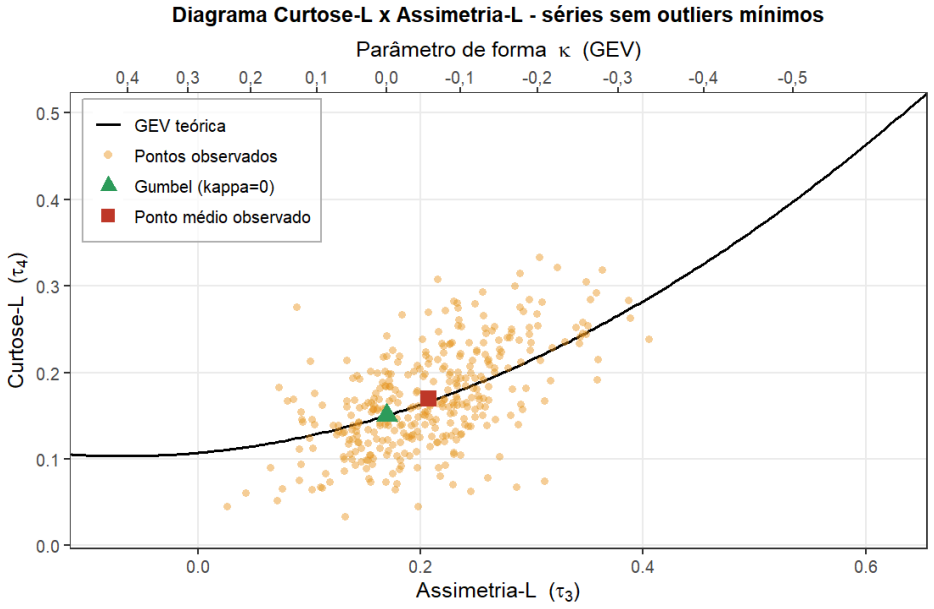


Figura 21: Diagrama de momentos-L para séries sem outliers mínimos.

Fonte: Autoral (2026).

Tabela 4: Parâmetro de forma por cenário completo e sem outliers mínimos.

	Completas (369)	Sem outliers (352)
κ médio	-0,010	-0,055
κ mediano	-0,011	-0,057
DP	0,129	0,098
Fréchet (%)	53,4	69,3
Weibull (%)	46,6	30,7

Fonte: Autoral (2026).

A remoção dos outliers mínimos reduziu a dispersão das estimativas, com o desvio-padrão de κ caindo de 0,129 para 0,098, e deslocou o ponto médio do diagrama de razões de momentos-L de uma posição praticamente coincidente com o ponto de Gumbel para o domínio da família Fréchet, elevando a proporção de estações com $\kappa < 0$ de 53,4% para 69,3%. Ao todo, 49 estações (13,9%) trocaram de família.

É importante registrar a exclusão das dez estações com assimetria-L negativa após a remoção de outliers mínimos, condição fisicamente implausível para máximas anuais de precipitação, inerentemente assimétricas à direita.. Essa escolha, contudo, não é neutra. As 17 estações excluídas apresentavam, sem exceção, parâmetro de forma positivo, com média de +0,198, pertencendo integralmente à família Weibull, de modo que sua saída contribui para o deslocamento observado em direção à Fréchet.

4.3.4 Análise de agrupamento

A análise de agrupamento buscou investigar se a variabilidade espacial do parâmetro de forma pode ser associada a fatores geográficos e climáticos, exigindo, para cada estação, a disponibilidade simultânea das quatro variáveis empregadas. Essa condição reduziu o conjunto de trabalho, pois a obtenção de um PRCPTOT

médio válido mostrou-se mais restritivo que as filtragens anteriores. Das 652 estações iniciais, apenas 378 (57,9%) alcançaram 30 anos válidos após os critérios do RClimDex e o filtro pela amplitude interquartil, e, quando confrontadas com as 352 estações com a GEV ajustada, apenas 331 apareceram nos dois conjuntos simultaneamente, consolidando o grupo submetido ao agrupamento.

A Figura 22 apresenta a largura média da silhueta, cujo máximo ocorre em k = 2 (0,41), com estabilização a partir de k = 3. Resultado semelhante foi obtido por Cardoso (2026), que identificou valor de silhueta próximo de 0,42 para (k = 2), além de uma curva com comportamento semelhante ao observado neste estudo. Valores dessa ordem caracterizam, contudo, uma estrutura de agrupamento fraca: os grupos são distinguíveis, mas não fortemente separados. A convergência entre os dois estudos sugere que essa característica não decorre da filtragem adotada, mas do comportamento das próprias variáveis em Minas Gerais, estado situado em zona de transição, onde as fronteiras entre regiões pluviométricas tendem a ser graduais.

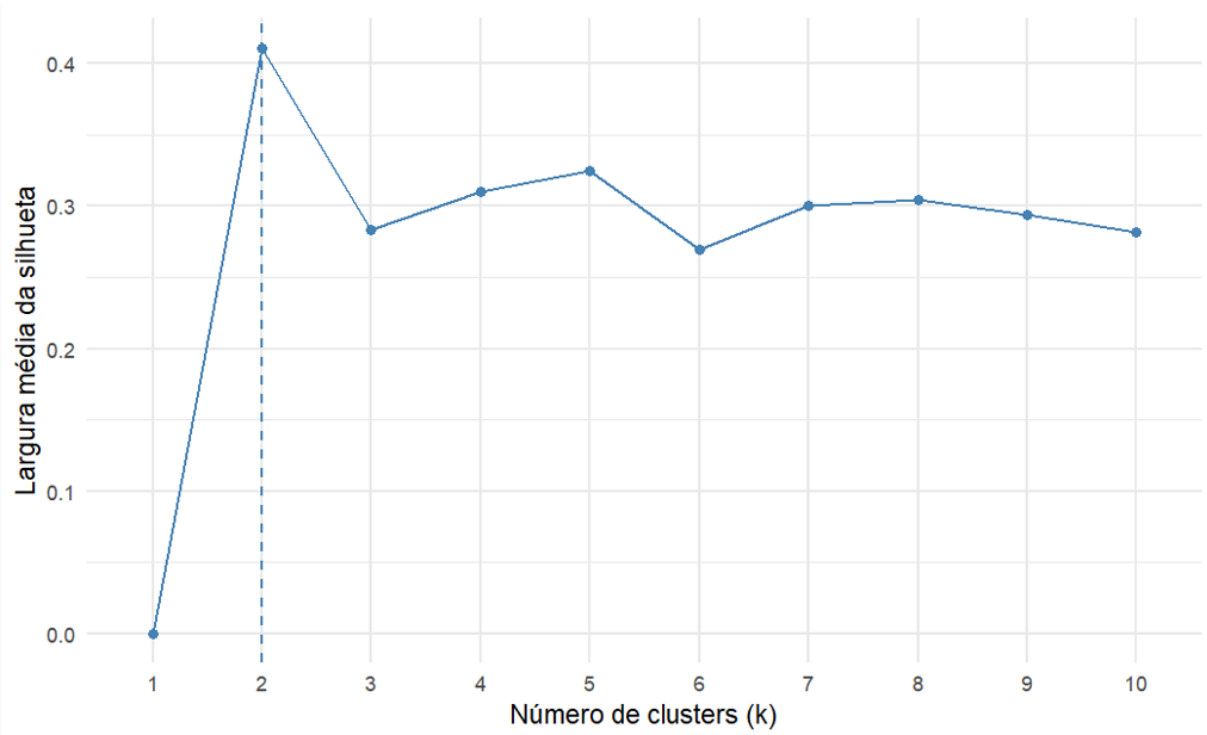


Figura 22: Gráfico de silhueta.

Fonte: Autoral (2026).

A Tabela 5 apresenta as características médias de cada grupo e a Figura 23, sua distribuição espacial. O Cluster 1 (Norte) reúne 118 estações e o Cluster 2 (Sul), 213, com as atribuições por cluster registradas no Apêndice N. A distinção entre eles não se reduz à posição latitudinal: o Cluster 2 apresenta altitude média quase duas vezes superior (818 m contra 447 m) e precipitação média anual cerca de 40% maior (1.431 mm contra 1.023 mm). A partição segue, assim, o gradiente de umidade de Thonthwaite, descrito na caracterização da área de estudo, item 3.2, separando a porção norte, mais baixa e seca, da porção sul, mais elevada e úmida. Registra-se que os valores médios do parâmetro de forma resultaram muito próximos entre os grupos (−0,047 e −0,059), a despeito do contraste geográfico e climático observado, comparação detalhada no item seguinte (4.3.5).

Tabela 5: Características médias de cada cluster.

	Cluster 1 (Norte)	Cluster 2 (Sul)
Estações	118	213
Latitude média	-17,72°	-20,37°
Longitude média	-43,11°	-45,44°
Altitude média	447 m	818 m
PRCPTOT médio	1.023 mm	1.431 mm
κ médio	-0,047	-0,059

Fonte: Autoral (2026).

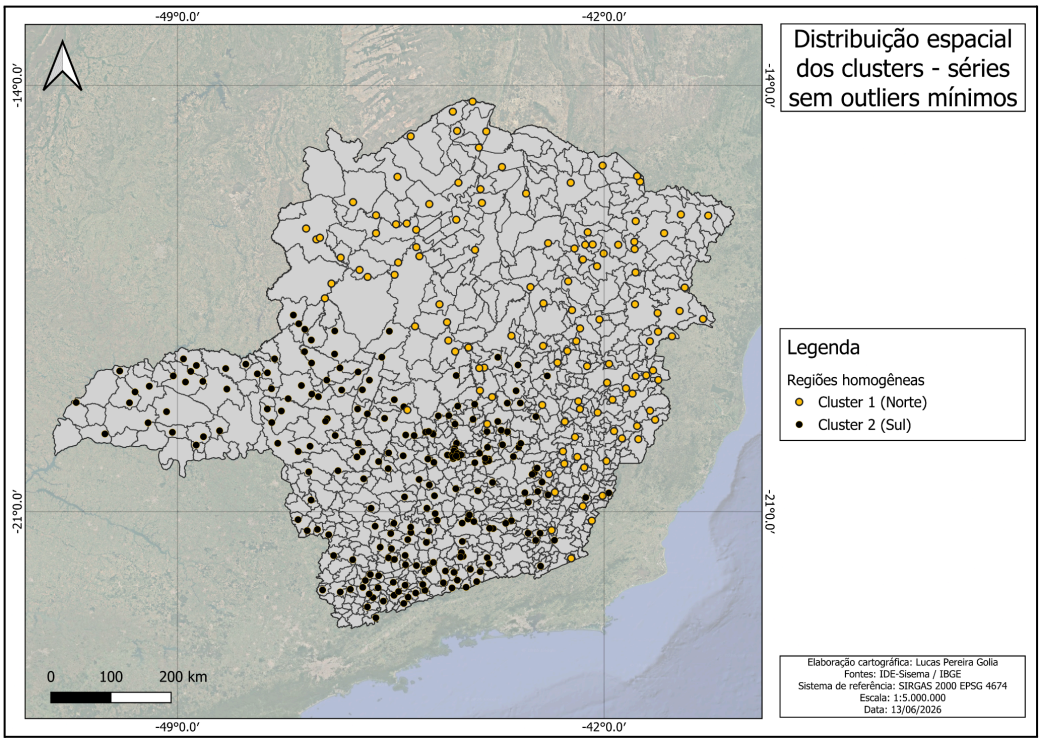


Figura 23: Distribuição espacial dos clusters.

Fonte: Autoral (2026).

4.3.5 Estatísticas do parâmetro de forma

4.3.5.1 Estatísticas globais e por cluster

Caracterizado o ajuste da GEV e definidas as regiões homogêneas, descreveu-se o comportamento do parâmetro de forma no conjunto das 352 estações sem outliers mínimos, cujas estatísticas descritivas são apresentadas na Tabela 6, com os valores individuais disponíveis no Apêndice M. A média e a mediana, ambas próximas de -0,055, situam-se no domínio da família Fréchet, e a predominância de valores negativos indica o prevalecimento de caudas sem limite superior para as máximas. O desvio-padrão de 0,098, contudo, é elevado frente a essa média, revelando acentuada heterogeneidade entre as estações, então, embora o predomínio da Fréchet seja consistente para o conjunto, o parâmetro de forma de estações individuais estende-se por ambas as famílias, o que recomenda cautela na classificação de cada estação isoladamente. A Figura 24 apresenta esses

valores em função do tamanho da série, com a nuvem de pontos revelando ampla dispersão, mais acentuada entre as séries curtas, e uma leve tendência de valores mais negativos à medida que o registro se alonga, relação explorada mais adiante.

Tabela 6: Estatísticas descritivas do parâmetro de forma.

Estatística	Valor
Mínimo	-0,336
1º quartil (Q1)	-0,117
Mediana	-0,057
Média	-0,055
3º quartil (Q3)	0,014
Máximo	0,237
Desvio-padrão	0,098
Assimetria	-0,104

Fonte: Autoral (2026).

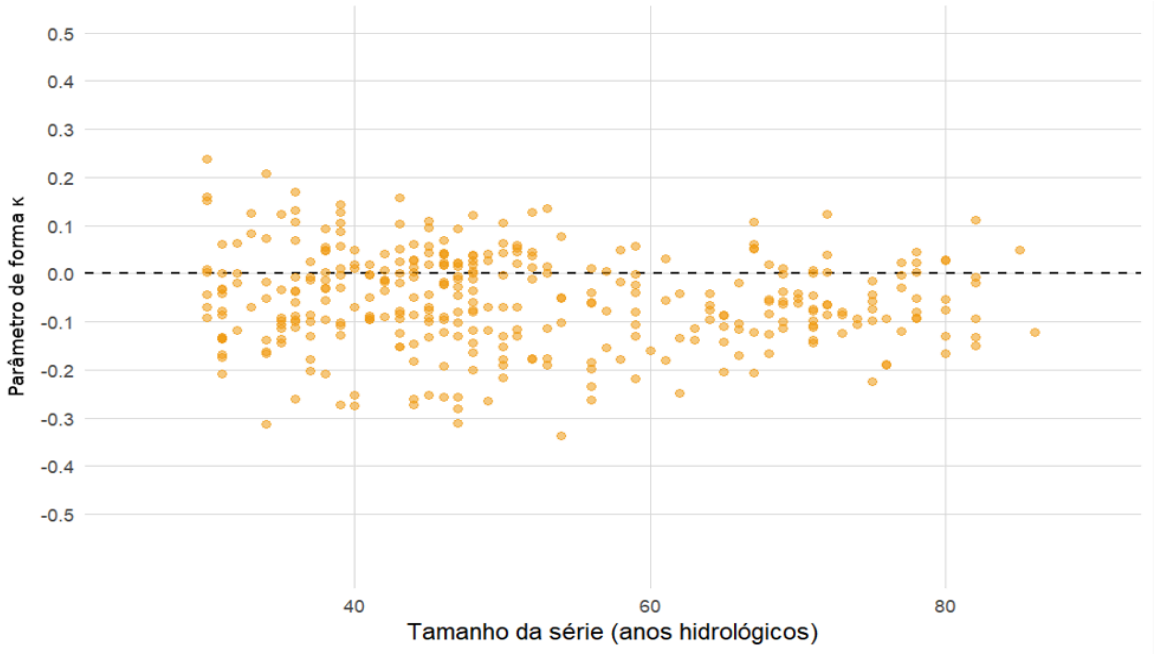


Figura 24: Gráfico de dispersão do parâmetro de forma por tamanho da série.

Fonte: Autoral (2026).

A Tabela 7 e a Figura 25 comparam a distribuição do parâmetro de forma entre as duas regiões homogêneas. As estatísticas são próximas, ainda que o Cluster 2 (Sul) apresente valores ligeiramente mais negativos, com mediana de -0,063, contra -0,039 do Cluster 1 (Norte). As médias seguem o mesmo padrão (-0,059 e -0,047) e os desvios-padrão são praticamente iguais (0,094 e 0,100). Nos *boxplots*, as duas regiões são bem similares, ambas com mediana no domínio da família Fréchet. O Cluster 2 exibe dois valores atípicos, um em cada extremo (aproximadamente -0,33 e 0,24), o que, somado à sua maior quantidade de estações, revela uma amplitude maior, sem alterar muito a posição central. No conjunto, embora o agrupamento tenha separado regiões geográfica e climaticamente contrastantes, essa distinção não se traduz em uma separação clara das distribuições do parâmetro de forma entre as regiões, resultado convergente com Cardoso (2026) para Minas Gerais

Tabela 7: Estatísticas descritivas do parâmetro de forma por região homogênea..

Estatística	Cluster 1 (Norte)	Cluster 2 (Sul)
n	118	213
Mínimo	-0,311	-0,336
1º quartil (Q1)	-0,118	-0,114
Mediana	-0,039	-0,063
Média	-0,047	-0,059
3º quartil (Q3)	0,026	0,006
Máximo	0,207	0,237
Desvio-padrão	0,100	0,094

Assimetria	-0,165	-0,094

Fonte: Autoral (2026).

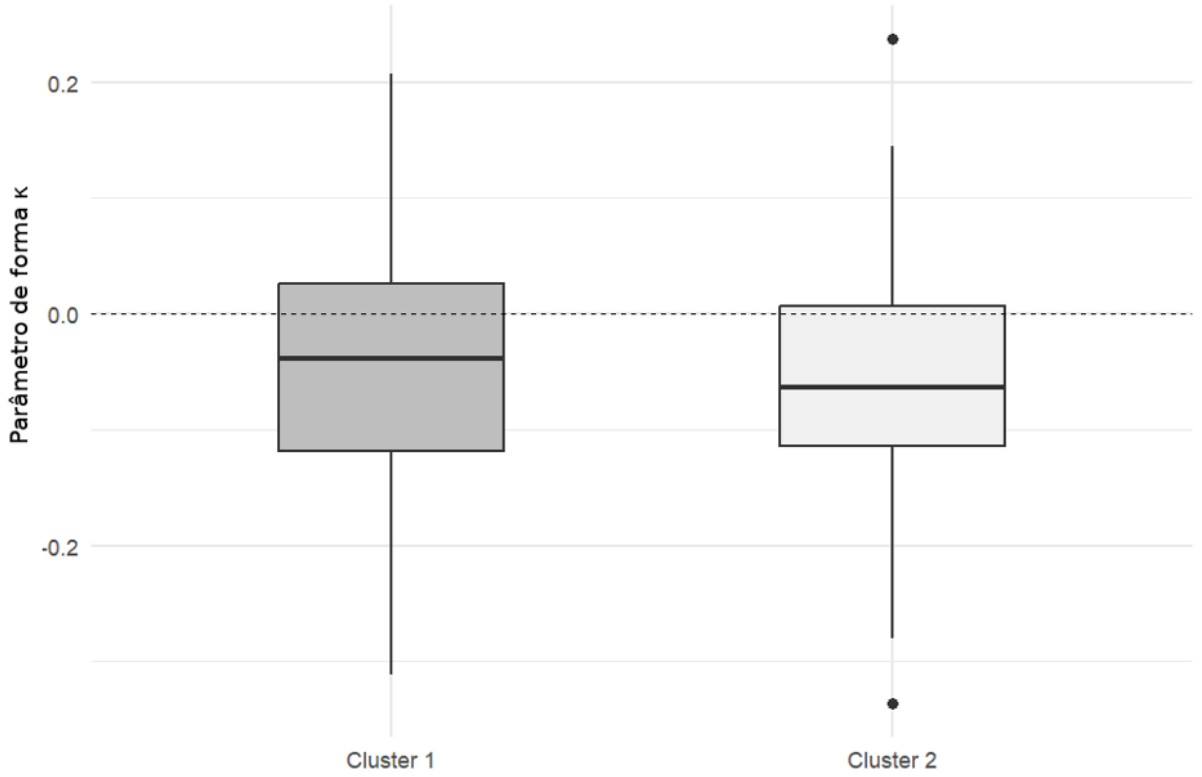


Figura 25: Distribuição do parâmetro de forma por região homogênea.

Fonte: Autoral (2026).

4.3.5.2 Verificação da normalidade

A hipótese de normalidade da distribuição do parâmetro de forma foi avaliada pelo teste de Shapiro-Wilk, no conjunto global e em cada região homogênea. A Figura 26 apresenta o histograma dos valores de κ para as 352 estações, com a curva normal ajustada sobreposta, e a Figura 27 traz os histogramas correspondentes aos Clusters 1 e 2. Em nenhum dos três casos a hipótese de normalidade foi rejeitada ao nível de 5%: o p-valor foi de 0,560 para o conjunto global e de 0,750 e 0,497 para os Clusters 1 (Norte) e 2 (Sul), respectivamente. Não há, portanto, evidência de afastamento da normalidade na distribuição do parâmetro

de forma, o que é reforçado visualmente pela boa aderência dos histogramas às curvas normais e pela baixa assimetria já observada nas estatísticas descritivas.

Esse resultado tem função metodológica, pois uma distribuição aproximadamente normal fica inteiramente caracterizada por sua média e seu desvio-padrão, o que fundamenta a comparação com outros trabalhos apresentada na seção seguinte, admitindo-se que também verificaram a normalidade e que reportaram justamente essas duas estatísticas. Cabe ainda notar a semelhança da distribuição do parâmetro de forma entre as regiões homogêneas, que se mostram aproximadamente normal e de dispersão e assimetria comparáveis em ambas, reforçando a conclusão de que os grupos pouco se distinguem quanto ao parâmetro de forma.

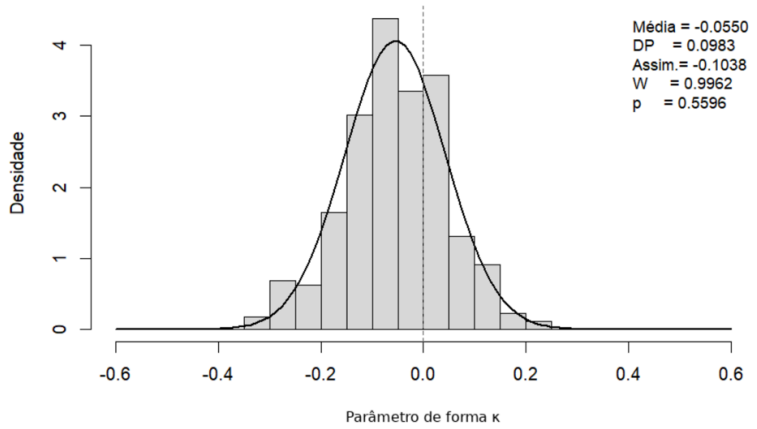


Figura 26: Histograma do parâmetro de forma com normal ajustada.

Fonte: Autoral (2026).

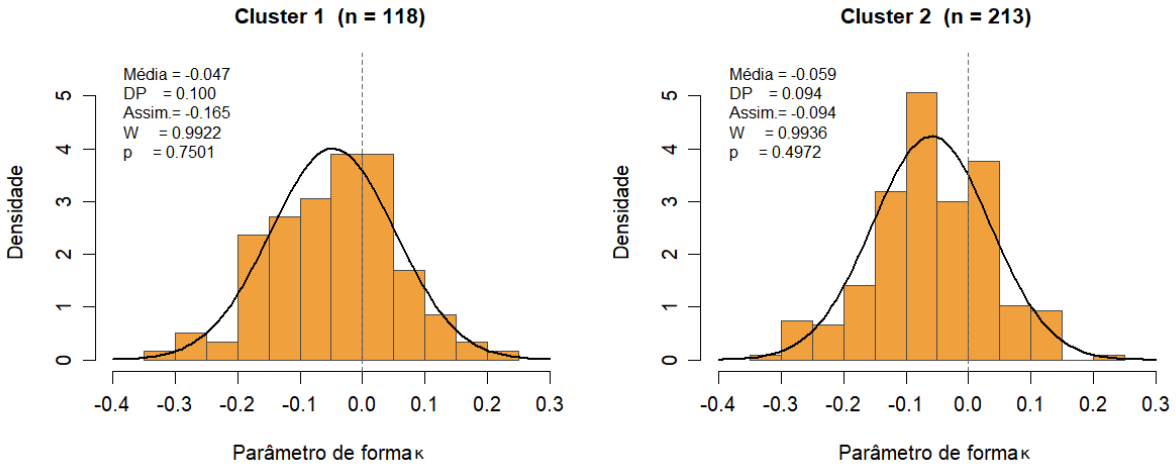


Figura 27: Histograma do parâmetro de forma com normal ajustada por região homogênea.

Fonte: Autoral (2026).

4.3.5.3 Análise por faixa de tamanho

Em seguida, avaliou-se a influência do comprimento das séries sobre a estimativa, agrupando-se as estações em faixas de anos hidrológicos válidos (Figura 28). Nota-se uma distribuição desigual, em que as duas primeiras faixas concentram 206 das 352 estações e a última apenas 8, de modo que a média dessa última, de intervalo de confiança muito amplo, é pouco confiável, e as tendências se apoiam nas faixas anteriores. Nestas, a média torna-se progressivamente mais negativa com o alongamento das séries, de −0,041 a −0,075, indicando que séries curtas produzem estimativas enviesadas em direção a valores menos negativos, próximos da Gumbel. O padrão coincide com Papalexiou e Koutsoyiannis (2013), porém contrasta com Cardoso (2026), que não identificou tal influência na mesma região.

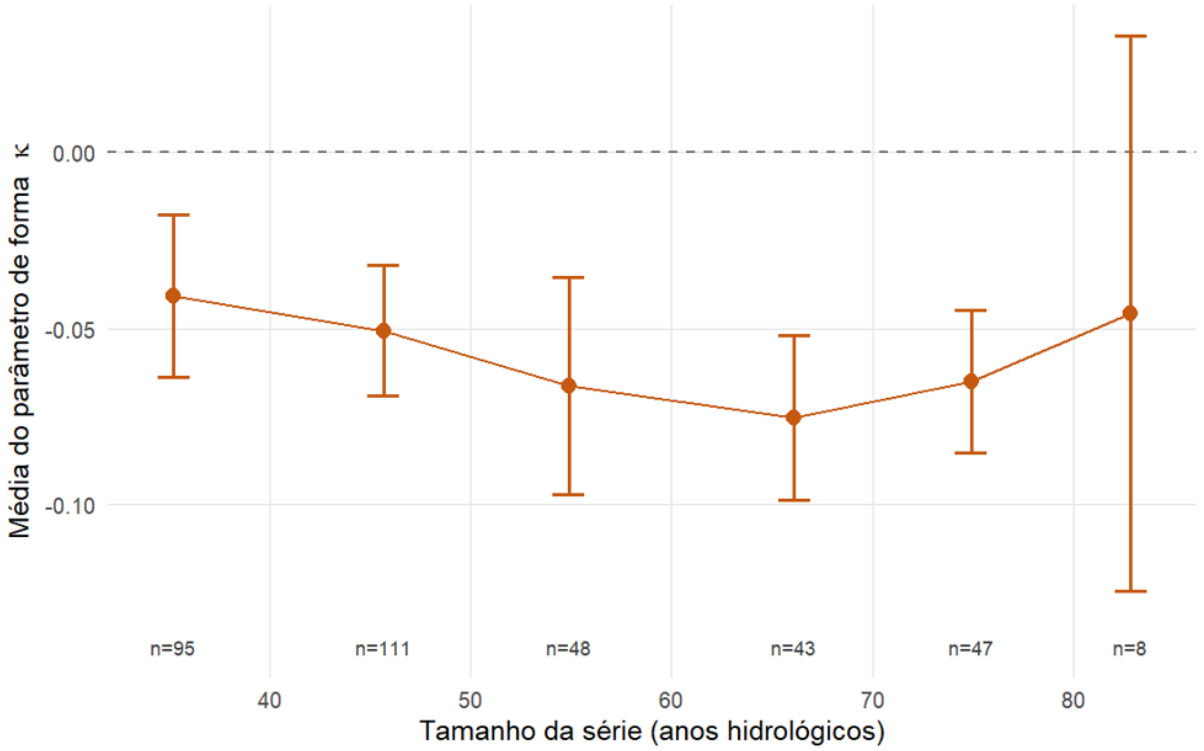


Figura 28: Intervalo de confiança de 95% da média do parâmetro de forma por tamanho da série.

Fonte: Autoral (2026).

A Figura 29 contribui para essa leitura apresentando a dispersão, em que a faixa da média mais ou menos um desvio-padrão estreita-se à medida que as séries se alongam, refletindo a menor variabilidade das estimativas. Aqui há convergência com Cardoso (2026), cujo desvio-padrão cai de 0,12 a 0,08, então os dois estudos concordam quanto ao ganho de precisão com séries longas, divergindo apenas quanto ao viés na média.

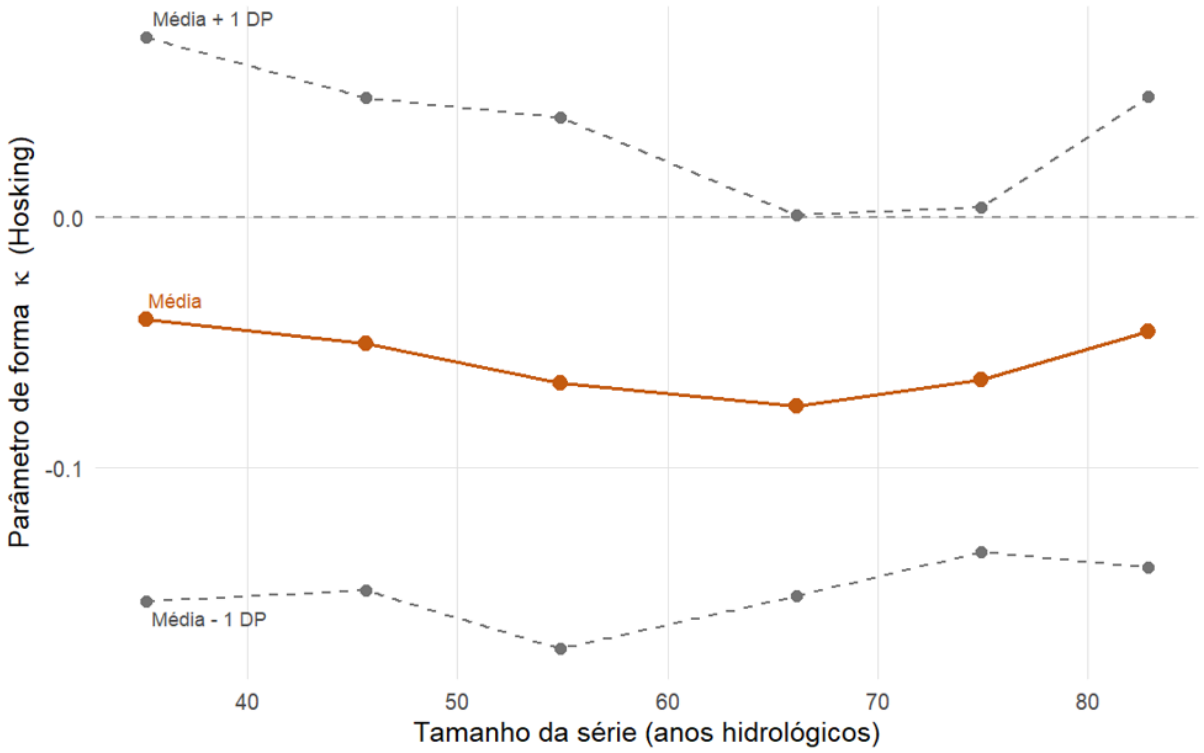


Figura 29: Dispersão da média do parâmetro de forma por tamanho da série.

Fonte: Autoral (2026).

Coerentemente, a Figura 30 mostra que a fração de estações que indicam ajuste à Fréchet sobe de 67% a 84% com o comprimento das séries, enquanto a fração Weibull diminui. Isso indica que a predominância da cauda Fréchet se torna mais nítida com registros longos, sugerindo que parte das estações Weibull nas séries curtas pode decorrer da imprecisão da estimativa, e não de comportamento físico distinto. O Apêndice O contém todas as estatísticas descritivas por faixa de tamanho, servindo como base de dados para todas as análises feitas nesta seção.

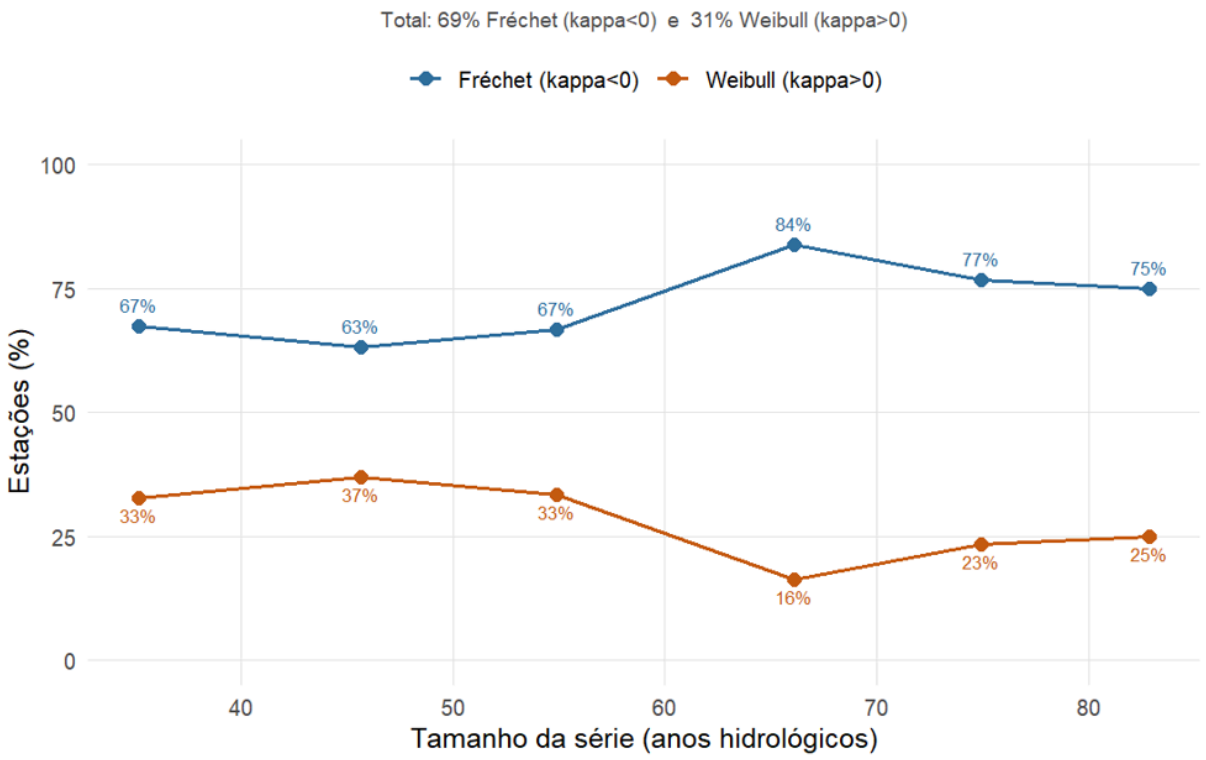


Figura 30: Percentual de estações que se ajustam à Fréchet e Weibull por tamanho da série.

Fonte: Autoral (2026).

4.3.5.4 Comparação com outros trabalhos

A Tabela 8 apresenta as estatísticas descritivas do parâmetro de forma nos estudos de Papalexiou e Koutsoyiannis (2013) e Cardoso (2026), com os valores dos dois trabalhos de referência convertidos para a convenção de Hosking. A média obtida no presente estudo (−0,055) situa-se entre a de Cardoso (2026), praticamente nula, e a de Papalexiou e Koutsoyiannis (2013), mais negativa (−0,092). Os três estudos concordam quanto ao sinal predominante, todos com mediana igual ou inferior a zero, e quanto à ordem de grandeza da dispersão, com desvios-padrão entre 0,098 e 0,120.

Tabela 8: Estatísticas descritivas do parâmetro de forma por trabalho.

Estudo	n	Média	Mediana	DP	Assimetria
Papalexiou; Koutsoyiannis (2013)	15.137	-0,092	-0,093	0,120	0,130
Cardoso (2026)	496	0,000	0,000	0,110	-0,200
Presente estudo	352	-0,055	-0,057	0,098	-0,104

Estudo	Mínimo	Máximo	Q1	Q3
Papalexiou; Koutsoyiannis (2013)	-0,760	0,587	-0,169	-0,020
Cardoso (2026)	-0,520	0,320	-0,070	0,070
Presente estudo	-0,336	0,237	-0,117	0,014

Fonte: Autoral, com dados de Papalexiou e Koutsoyiannis (2013), Cardoso (2026) e do presente estudo (2026).

Cabe considerar que os dois trabalhos de referência dispõem de conjuntos maiores e operam em escalas geográficas distintas. Enquanto o presente estudo e Cardoso (2026), com 496 estações, restringem-se a Minas Gerais, Papalexiou e Koutsoyiannis (2013) analisam 15.137 estações de todo o mundo, contra as 352 aqui utilizadas. Cardoso (2026) constitui, assim, o termo de comparação mais direto, por compartilhar a região e a base de dados, porém decorre de uma filtragem distinta e a incorporação da base de dados do INMET, conferindo maior precisão às estatísticas globais, devido ao maior tamanho do conjunto. Papalexiou e Koutsoyiannis (2013), por sua vez, reúnem número muito superior de estações e

séries substancialmente mais extensas (40 a 163 anos), o que ajuda a explicar sua média mais negativa (−0,092), coerente com a tendência observada nas faixas de maior tamanho no presente estudo, ainda que a diferença entre a escala global e regional possa contribuir com essa diminuição.

A Figura 31 sobrepõe ao histograma dos valores de κ do presente estudo as distribuições normais ajustadas a cada conjunto, procedimento viabilizado pela normalidade verificada anteriormente. A curva do presente estudo posiciona-se entre a de Cardoso (2026), centrada em zero, e as de Papalexiou e Koutsoyiannis (2013), deslocadas para valores mais negativos. Papalexiou e Koutsoyiannis (2013 apresentam uma curva ajustada às suas estimativas brutas e outra obtida com o estimador corrigido do efeito do tamanho de série, proposto pelos próprios autores, mais estreita e centrada em −0,114. Visualmente, confirma-se que os três estudos identificam o predomínio da família Fréchet, porém com esse predomínio variando por trabalho.

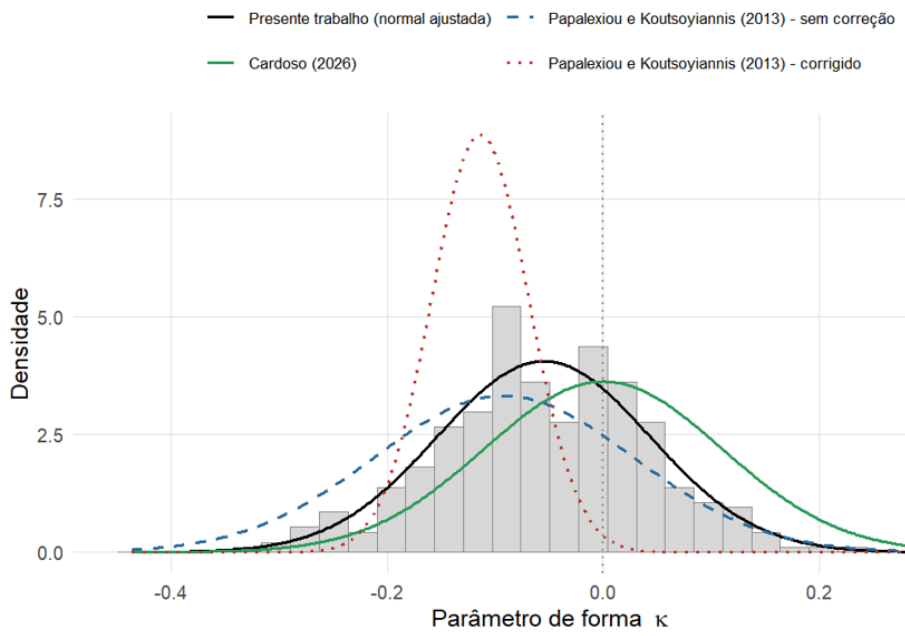


Figura 31: Distribuição do parâmetro de forma por estudo.

Fonte: Autoral, com dados de Papalexiou e Koutsoyiannis (2013), Cardoso (2026) e do presente estudo (2026).

4.3.6 Análise multidiária

Estendeu-se a análise do parâmetro de forma às precipitações máximas acumuladas em diferentes durações de 1, 2, 3, 5 e 10 dias, com o objetivo de verificar se o comportamento da cauda das distribuições se mantém à medida que se agregam dias consecutivos de chuva. O número de estações com ajuste válido variou entre as durações, entre 352 para 1 dia e 371 para 10 dias, em razão de os filtros de validade das somas móveis, das condições de AAS e da remoção de outliers mínimos produzirem resultados ligeiramente distintos em cada duração.

A Figura 40 apresenta a média do parâmetro de forma para cada duração. Observa-se uma tendência consistente de aproximação de zero à medida que a duração aumenta, passando de -0,055, para 1 dia, a -0,027, para 10 dias. A Tabela 9 detalha essa evolução e mostra que a mediana acompanha o mesmo movimento (de -0,057 a -0,018), enquanto o desvio-padrão permanece praticamente estável (entre 0,091 e 0,098). Coerentemente, a Figura 32 também mostra a leve redução do percentual de estações se ajustando à Fréchet. O deslocamento é gradual, mas nítido, indicando que a agregação em durações maiores atenua a tendência à família Fréchet observada nas máximas diárias.

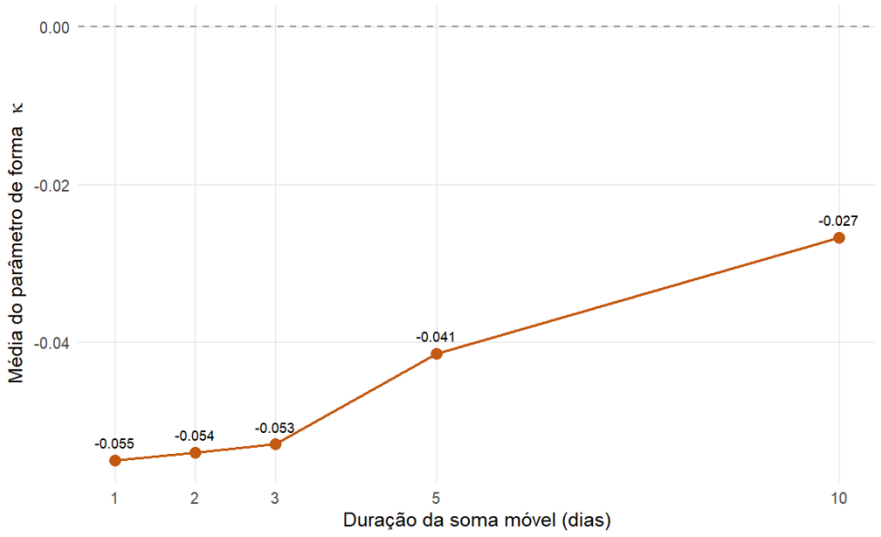


Figura 32: Variação da média do parâmetro de forma por duração de chuva.

Fonte: Autoral (2026).

Tabela 9: Estatísticas descritivas do parâmetro de forma por duração de chuva.

Duração	n	Média	Mediana	DP	Fréchet (%)
1	352	-0,055	-0,057	0,098	69
2	357	-0,054	-0,050	0,093	69
3	352	-0,053	-0,049	0,094	71
5	365	-0,041	-0,042	0,094	68
10	371	-0,027	-0,018	0,091	60

Fonte: Autoral (2026).

A figura 33 apresenta os diagramas de momentos-L por duração, revelando que os pares (τ_3, τ_4) distribuem-se em torno da curva teórica da GEV, e sua sobreposição revela distribuições semelhantes, sem nenhum afastamento da curva teórica.

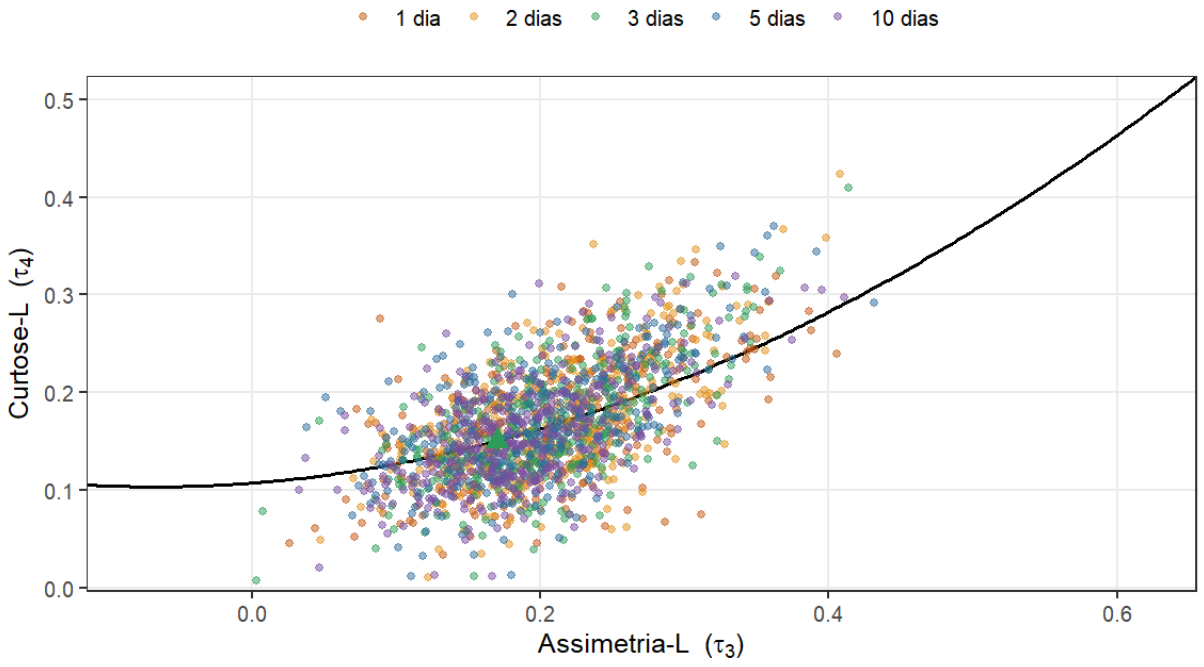


Figura 33: Diagrama de momentos-L por duração de chuva.

Fonte: Autoral (2026).

Diferentemente do que se poderia esperar sob a hipótese de preservação da forma da distribuição entre durações, os resultados evidenciados pela figura 34 revelam uma variação sistemática do parâmetro de forma com a duração de agregação, com atenuação progressiva da tendência à família Fréchet. Esse comportamento sugere que a cauda das distribuições de precipitação acumulada torna-se relativamente mais leve à medida que se somam dias consecutivos, aproximando-se do modelo de Gumbel.

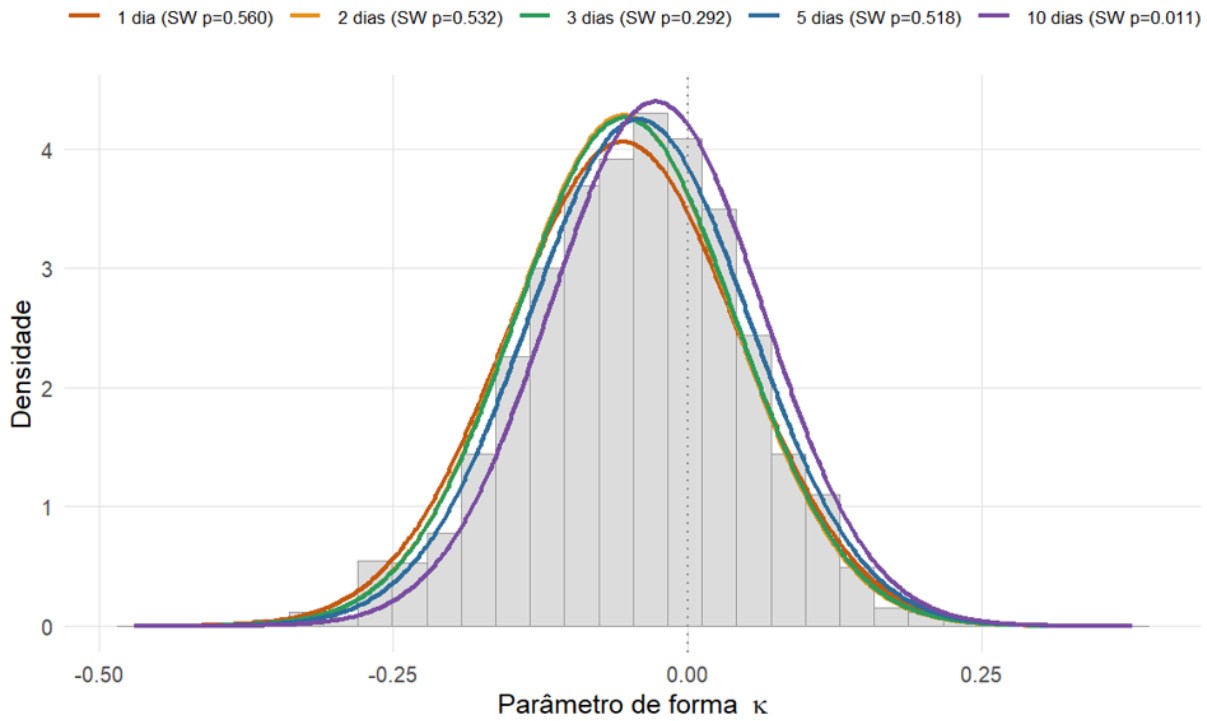


Figura 34: Distribuição do parâmetro de forma por duração de chuva.

Fonte: Autoral (2026).

Destaca-se que o histograma presente na Figura 34 agrega o parâmetro de forma de todas as durações, servindo apenas como referência visual. A hipótese de

normalidade não foi rejeitada para as durações de 1 a 5 dias (p-valor entre 0,29 e 0,56), mas o foi para a de 10 dias (p = 0,011), indicando que, nesta última, a distribuição do parâmetro de forma afasta-se da normal, o que recomenda cautela ao caracterizá-la apenas por média e desvio-padrão.

4.3.7 Síntese dos resultados e implicações para os estudos hidrológicos

De forma integrada, os resultados demonstram que a estimativa confiável das precipitações extremas depende não apenas da escolha de uma distribuição probabilística adequada, mas também de um processo rigoroso de avaliação, filtragem e consistência das séries pluviométricas. Embora a maior parte das estações tenha apresentado elevada qualidade interna, verificaram-se limitações relacionadas à extensão temporal dos registros, à distribuição espacial da rede e à disponibilidade de estações auxiliares para a análise de qualidade relativa. A aplicação da metodologia evidenciou, portanto, que a qualidade pontual dos dados não garante, isoladamente, uma representação espacial adequada da variabilidade pluviométrica, especialmente em regiões menos monitoradas do estado.

Após as diferentes etapas de controle de qualidade e verificação dos pressupostos estatísticos, a distribuição GEV mostrou-se adequada à representação das séries de precipitações máximas, com predominância da família Fréchet e variabilidade espacial do parâmetro de forma. A tendência de obtenção de parâmetros mais negativos e menos dispersos nas séries mais longas reforça a importância da continuidade do monitoramento, uma vez que registros de pequena extensão podem produzir estimativas mais incertas e próximas da distribuição de Gumbel. Esse aspecto tem consequência prática direta, pois como as diferentes famílias produzem quantis cada vez mais discrepantes à medida que se avança à direita da cauda, a adoção de Gumbel ou Weibull pode subestimar de forma expressiva as precipitações de projeto associadas a grandes períodos de retorno. Além disso, a variação do parâmetro de forma com a duração da precipitação indica que a adoção de um único comportamento distributivo para diferentes tempos de acumulação deve ser realizada com cautela.

Esses resultados possuem implicações diretas para os estudos hidrológicos e

para o planejamento das infraestruturas hídricas. A estimativa das chuvas associadas a diferentes períodos de retorno constitui uma das principais entradas para o dimensionamento de sistemas de drenagem urbana, bacias de detenção, canais, galerias pluviais, barragens, vertedouros, pontes, dispositivos de controle de cheias e demais estruturas sujeitas aos efeitos de eventos extremos. Inconsistências nos registros ou escolhas inadequadas da distribuição de frequência podem resultar tanto no subdimensionamento das estruturas, elevando os riscos de falhas, inundações e prejuízos socioeconômicos, quanto no superdimensionamento, aumentando desnecessariamente os custos de implantação.

Assim, a principal contribuição deste trabalho consiste no estabelecimento de um procedimento sistemático e replicável para avaliar a qualidade das séries, selecionar registros adequados e caracterizar estatisticamente as precipitações máximas em Minas Gerais. Ao explicitar as limitações dos dados e a sensibilidade dos resultados às etapas de filtragem, ao comprimento das séries e à duração das chuvas, o estudo fornece subsídios para uma utilização mais criteriosa das informações pluviométricas. Assim, os resultados não devem ser compreendidos apenas como parâmetros estatísticos isolados, mas como uma base técnica para aprimorar estudos de frequência, regionalizações hidrológicas e critérios de projeto, contribuindo para o planejamento de infraestruturas mais seguras, economicamente eficientes e resilientes diante da ocorrência de eventos hidrológicos extremos.

5 Conclusão

Este trabalho compilou e organizou as séries pluviométricas de Minas Gerais e as submeteu a um procedimento sistemático de controle de qualidade, contemplando a avaliação da qualidade interna dos registros e de sua coerência espacial. A partir das séries selecionadas, ajustou-se a distribuição Generalizada de Valores Extremos às precipitações máximas anuais e caracterizou-se o comportamento do parâmetro de forma em diferentes regiões, extensões de séries e durações de precipitação. Dessa forma, os objetivos propostos foram atingidos.

O controle de qualidade mostrou-se uma etapa essencial para estudos hidrológicos e projetos hidráulicos, uma vez que falhas, lacunas e inconsistências nos registros podem afetar a estimativa de eventos extremos e, consequentemente, o dimensionamento e a segurança de estruturas hidráulicas. Diante da inexistência de um procedimento único e amplamente adotado no Brasil, este trabalho contribui para tornar mais transparente, sistemática e criteriosa a avaliação das séries pluviométricas.

O índice de qualidade absoluta apresentou boa aplicabilidade e indicou elevada qualidade interna para a maior parte das séries, além de permitir a identificação dos componentes que mais contribuíram para a redução das pontuações. Já o índice de qualidade relativa apresentou limitações decorrentes da baixa e desigual densidade da rede pluviométrica de Minas Gerais. A exigência de estações auxiliares próximas restringiu sua aplicação, tornando necessária uma adaptação metodológica que ampliou a cobertura da análise, embora com redução da robustez originalmente proporcionada pela comparação simultânea entre várias estações.

A distribuição GEV mostrou-se adequada à representação das precipitações máximas anuais, com predominância da família Fréchet. Esse resultado possui relevância para a engenharia hidrológica, pois distribuições com cauda do tipo Fréchet não apresentam limite superior e podem produzir estimativas mais elevadas para grandes períodos de retorno do que modelos de cauda mais leve. Assim, a

adoção indiscriminada da distribuição de Gumbel pode levar à subestimação das precipitações de projeto. Verificou-se também que a predominância da Fréchet se torna mais evidente em séries longas, reforçando a necessidade de cautela na utilização de registros curtos.

Embora a análise de agrupamento tenha separado o estado segundo gradientes de altitude e umidade, as regiões delimitadas apresentaram comportamento semelhante quanto ao parâmetro de forma. Portanto, as diferenças climáticas e geográficas identificadas não se refletiram em uma regionalização clara desse parâmetro, resultado convergente com Cardoso (2026). A análise multidiária mostrou ainda que a predominância da família Fréchet diminui com o aumento da duração da precipitação acumulada, indicando que a escolha da distribuição deve considerar também a duração analisada.

Os resultados obtidos constituem uma base para aplicações futuras, como a estimativa de precipitações associadas a diferentes períodos de retorno, a elaboração de mapas de chuvas de projeto e a construção de curvas intensidade-duração-frequência. Essas informações são fundamentais para o dimensionamento de sistemas de drenagem, estruturas hidráulicas, medidas de controle de cheias e ações de planejamento territorial.

Para estudos futuros, recomenda-se:

- Investigar a relação entre o índice de qualidade absoluta e variáveis físicas, climáticas e socioeconômicas, bem como sua evolução ao longo do tempo, visando identificar fragilidades e possibilidades de melhoria da rede de monitoramento;
- aprimorar o índice de qualidade relativa, incorporando não apenas a ocorrência de diferenças superiores a determinado limite, mas também a magnitude dessas diferenças;
- estimar precipitações de projeto, elaborar mapas regionais e construir curvas intensidade-duração-frequência a partir das séries avaliadas, ampliando a aplicação prática dos resultados no planejamento e dimensionamento de infraestruturas hídricas.

Referências

AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO (ANA). **HidroWeb: Sistema de Informações Hidrológicas.** Brasília: ANA, [2025]. Disponível em: https://www.snirh.gov.br/hidroweb/apresentacao. Acesso em: 1 dez. 2025.

AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO (ANA). **Manual de procedimentos para instalação, operação e manutenção de estações pluviométricas convencionais.** Brasília, 2016. Disponível em: https://www.gov.br/ana/pt-br/assuntos/monitoramento-e-eventos-criticos/monitoramento-hidrologico/monitoramento-hidrologico-do-setor-eletrico/resolucao-conjunta-ana-aneel-127-2022/ManualInstalaoEstaesPluviomtricas.pdf. Acesso em: 10 nov. 2025.

AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO (ANA). **Rede Hidrometeorológica Nacional.** Brasília: ANA, 2023. Disponível em: https://metadados.snirh.gov.br/geonetwork/srv/api/records/f85dbf06-a869-414c-afc5-bb01869e9156. Acesso em: 30 nov. 2025.

AGÊNCIA NACIONAL DE ÁGUAS E SANEAMENTO BÁSICO (ANA). **Sistema Nacional de Informações sobre Recursos Hídricos.** Brasília, DF, s.d. Disponível em:<https://www.gov.br/ana/pt-br/assuntos/gestao-das-aguas/politica-nacional-de-recursos-hidricos/sistema-de-informacoes-sobre-recursos-hidricos\>. Acesso em: 11 nov. 2025.

AGÊNCIA NACIONAL DE ÁGUAS (ANA); SERVIÇO GEOLÓGICO DO BRASIL (CPRM). **Rede Hidrometeorológica Nacional de Referência – RHNR: relato do planejamento da RHNR e definição das estratégias de implementação para os próximos anos (5 anos).** Brasília, DF: ANA, 2017. Disponível em: https://rigeo.sgb.gov.br/server/api/core/bitstreams/b6ea5258-e80d-4542-83df-eb5e797d75a3/content Acesso em: 19 dez. 2025.

ALEXANDERSSON, H. **A homogeneity test applied to precipitation data.** Journal of Climatology, v. 6, n. 6, p. 661–675, 1986. Disponível em: <https://rmets.onlinelibrary.wiley.com/doi/10.1002/joc.3370060607>. Acesso em: 9 dez. 2025.

AMEN, Ehab Mohammad; CALVACHE, María Luisa. **Hydrological model and integrated GIS techniques for simulation and analysis of the long-term water balance in a coastal aquifer, Spain.** Scientific Reports, v. 15, art. 27148, 2025. DOI: 10.1038/s41598-025-07543-z.

ANDERSON, T. W.; DARLING, D. A. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. **The Annals of Mathematical Statistics**, v. 23, n. 2, p. 193-212, 1952. DOI: https://doi.org/10.1214/aoms/1177729437.

BÁRDOSSY, A. et al. **Is Precipitation Responsible for the Most Hydrological Model Uncertainty?** Frontiers in Water, v. 4, 2022. DOI: 10.3389/frwa.2022.836554.

BOTTO, Anna et al. **Technical note: Design flood under hydrological uncertainty.** Hydrology and Earth System Sciences, v. 21, p. 3353–3358, 2017. DOI: 10.5194/hess-21-3353-2017.

BENJAMINI, Y.; HOCHBERG, Y. **Controlling the false discovery rate: a practical and powerful approach to multiple testing**. Journal of the Royal Statistical Society: Series B (Methodological), v. 57, n. 1, p. 289-300, 1995. DOI: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

BRASIL. Lei nº 9.433, de 8 de janeiro de 1997. **Institui a Política Nacional de Recursos Hídricos, cria o Sistema Nacional de Gerenciamento de Recursos Hídricos e dá outras providências.** Disponível em: https://www.planalto.gov.br/ccivil_03/leis/l9433.htm. Acesso em: 27 nov. 2025.

BRASIL. Lei nº 9.984, de 17 de julho de 2000. **Dispõe sobre a criação da Agência Nacional de Águas e Saneamento Básico (ANA).** Diário Oficial da União: seção 1, Brasília, DF, 18 jul. 2000. Disponível em: <https://www.planalto.gov.br/ccivil_03/leis/l9984.htm\>. Acesso em: 11 nov. 2025.

CARDOSO, F. H. C. **Avaliação do comportamento do parâmetro de forma da distribuição generalizada de valores extremos e de índices de extremos de precipitação em Minas Gerais**. 2026. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) – Escola de Engenharia, Universidade Federal de Minas Gerais, Belo Horizonte, 2026.

CATALUNYA. Agència per a la Competitivitat de l'Empresa. **Mission to China and Hong Kong 2017: CataloniaConnects, building bridges between Barcelona, Shanghai and Hong Kong.** Barcelona: ACCIÓ, 2017. Disponível em: https://www.accio.gencat.cat/web/.content/bancconeixement/documents/5d3428e.pdf . Acesso em: 10 dez. 2024.

COHN, T. A.; ENGLAND, J. F.; BERENBROCK, C. E.; MASON, R. R.; STEDINGER, J. R.; LAMONTAGNE, J. R. A generalized Grubbs-Beck test statistic for detecting multiple potentially influential low outliers in flood series. **Water Resources Research**, v. 49, n. 8, p. 5047-5058, 2013. DOI: https://doi.org/10.1002/wrcr.20392.

COLES, S. **An introduction to statistical modeling of extreme values**. London: Springer, 2001. (Springer Series in Statistics). DOI: 10.1007/978-1-4471-3675-0. Disponível em: https://link.springer.com/book/10.1007/978-1-4471-3675-0. Acesso em: 06/06/2026.

COMITÊ DA BACIA HIDROGRÁFICA DO RIO PARAÍBA DO SUL (CEIVAP). **Operação e manutenção de estação telemétrica.** 2025. Disponível em: https://sigaceivap.org.br/publicacoesArquivos/ceivap/arq_pubMidia_23006-ATV5-P3-0-00-00.pdf. Acesso em: 19 dez. 2025.

Correia Filho, W. L. F.; Lucio, P. S.; Spyrides, M. H. C. **Caracterização dos extremos de precipitação diária no Nordeste do Brasil.** Boletim Goiano de Geografia, v. 36, n. 3, p. 539–554, 2016. DOI: 10.5216/bgg.v36i3.44557

EINFALT, T.; MICHAELIDES, S. **Quality control of precipitation data.** In: MICHAELIDES, S. (ed.). Precipitation: Advances in Measurement, Estimation and Prediction. Berlin/Heidelberg: Springer, 2008. p. 101–126. DOI: 10.1007/978-3-540-77655-0_5.

EL HACHEM, Abbas et al. **Technical Note: Space–time statistical quality control of extreme precipitation observations.** Hydrology and Earth System Sciences, v. 26, p. 6137–6146, 2022. DOI: 10.5194/hess-26-6137-2022.

Estévez, J.; Llabrés-Brustenga, A.; Casas-Castillo, M. C.; García-Marín, A. P.; Kirchner, R.; Rodríguez-Solà, R. **A quality control procedure for long-term series**

of daily precipitation data in a semiarid environment. Theoretical and Applied Climatology, v. 149, n. 3–4, p. 1029–1041, 2022. DOI: 10.1007/s00704-022-04089-2

EXPERT TEAM ON CLIMATE CHANGE DETECTION AND INDICES (ETCCDI). **Climate Change Indices: definitions of the 27 core indices**. [*s.l.*], [s.d.]. Disponível em: https://etccdi.pacificclimate.org/list_27_indices.shtml. Acesso em: 2 jun. 2026.

FISHER, R. A.; TIPPETT, L. H. C. **Limiting forms of the frequency distribution of the largest or smallest member of a sample.** Proceedings of the Cambridge Philosophical Society, v. 24, n. 2, p. 180-190, 1928. DOI: 10.1017/S0305004100015681.

FUNDAÇÃO NACIONAL DE SAÚDE (FUNASA). **Manual de saneamento.** 3. ed. rev. Brasília: FUNASA, 2006. Cap. 5: Drenagem. Disponível em: https://bvsms.saude.gov.br/bvs/publicacoes/funasa/capitulo_5.pdf Acesso em: 19 dez. 2025.

GNEDENKO, B. **Sur la distribution limite du terme maximum d'une série aléatoire.** Annals of Mathematics, v. 44, n. 3, p. 423-453, 1943. DOI: 10.2307/1968974.

GONZÁLEZ-ROUCO, J. F. et al. **Quality Control and Homogeneity of Precipitation Data in the Southwest of Europe**. Journal of Climate, v. 14, p. 964–978, 2001. DOI: https://doi.org/10.1175/1520-0442(2001)014<0964:QCAHOP>2.0.CO;2

HAMADA, Atsushi; ARAKAWA, Osamu; YATAGAI, Akiyo. **An Automated Quality Control Method for Daily Rain-gauge Data. Global Environmental Research,** v. 15, p. 183–192, 2011. Disponível em: https://airies.wikiplus.net/attach.php/6a6f75726e616c5f31352d32656e67/save/0/0/15_2-12.pdf. Acesso em: 19 dez. 2025.

HELSEL, D. R.; HIRSCH, R. M.; RYBERG, K. R.; ARCHFIELD, S. A.; GILROY, E. J. **Statistical methods in water resources**. Reston: U.S. Geological Survey, 2020. 458 p. (Techniques and Methods, book 4, chapter A3). DOI: https://doi.org/10.3133/tm4a3.

HOSKING, J. R. M. L-moments: analysis and estimation of distributions using linear combinations of order statistics. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 52, n. 1, p. 105-124, 1990. DOI: https://doi.org/10.1111/j.2517-6161.1990.tb01775.x

HOSKING, J. R. M.; WALLIS, J. R. **Regional frequency analysis: an approach based on L-moments.** Cambridge: Cambridge University Press, 1997.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Clima – 1:5 000 000 (2002): O que é.** Rio de Janeiro: IBGE, 2002. Disponível em: https://www.ibge.gov.br/geociencias/informacoes-ambientais/climatologia/15817-clima.html. Acesso em: 19 dez. 2025.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). Minas Gerais. Rio de Janeiro: IBGE, 2024. Disponível em: https://www.ibge.gov.br/cidades-e-estados/mg.html. Acesso em: 10 dez. 2024.

INSTITUT CARTOGRÀFIC DE CATALUNYA (ICC). **Tipus de clima segons l'índex d'humitat de Thornthwaite.** Catalunya: Institut Cartogràfic de Catalunya, 2014. Disponível em: https://cartotecadigital.icgc.cat/digital/collection/catalunya/id/3232/. Acesso em: 10 dez. 2025.

INSTITUTO NACIONAL DE METEOROLOGIA (INMET). **Rede de Estações Meteorológicas Automáticas do INMET.** Nota Técnica n. 001/2011/SEGER/LAIME/CSC/INMET. Brasília, DF, 2011. Acesso em: 11 dez. 2025.

INSTITUTO NACIONAL DE METEOROLOGIA (INMET). **Sobre o INMET.** Brasília, DF: INMET, [2025]. Disponível em: https://portal.inmet.gov.br/sobre. Acesso em: 11 dez. 2025.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (IPCC). **Managing the risks of extreme events and disasters to advance climate change adaptation**: a special report of Working Groups I and II of the Intergovernmental Panel on Climate Change. Edição de C. B. Field *et al.* Cambridge: Cambridge University Press, 2012. 582 p.

INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE (IPCC). **Climate Change 2021: the physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change.** Masson-Delmotte, V. *et al.* (eds.). Cambridge; New York: Cambridge University Press, 2021. Link: https://www.ipcc.ch/report/ar6/wg1/ .

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. **An Introduction to Statistical Learning: with Applications in R**. 2. ed. New York: Springer, 2021. Disponível em: https://www.statlearning.com. Acesso em: 6 jul. 2026.

JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. **Quarterly Journal of the Royal Meteorological Society,** v. 81, n. 348, p. 158-171, 1955. DOI: 10.1002/qj.49708134804.

KATZ, R. W.; PARLANGE, M. B.; NAVEAU, P. Statistics of extremes in hydrology. ***Advances in Water Resources*,** v. 25, n. 8-12, p. 1287-1304, 2002. DOI: 10.1016/S0309-1708(02)00056-8.

KENDALL, M. G. **A new measure of rank correlation.** Biometrika, v. 30, n. 1/2, p. 81–93, 1938. DOI: https://doi.org/10.1093/biomet/30.1-2.81

KHALIQ, M. N.; OUARDA, T. B. M. J.; ONDO, J. C.; GACHON, P.; BOBÉE, B. Frequency analysis of a sequence of dependent and/or non-stationary hydro-meteorological observations: a review. ***Journal of Hydrology*,** v. 329, n. 3–4, p. 534–552, 2006. DOI: 10.1016/j.jhydrol.2006.03.004.

KHOSH BIN GHOMASH, Shahin; APEL, Heiko; SCHRÖTER, Kai; STEINHAUSEN, Max. Rapid high-resolution impact-based flood early warning is possible with RIM2D: a showcase for the 2023 pluvial flood in Braunschweig. **Natural Hazards and Earth System Sciences,** v. 25, p. 1737–1749, 2025. DOI: 10.5194/nhess-25-1737-2025.

KUNDZEWICZ, Z. W.; ROBSON, A. J. **C**hange detection in hydrological records: a review of the methodology. **Hydrological Sciences Journal**, v. 49, n. 1, p. 7-19, 2004. DOI: https://doi.org/10.1623/hysj.49.1.7.53993.

LAIO, F. Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters. **Water Resources Research**, v. 40, n. 9, W09308, 2004. DOI: https://doi.org/10.1029/2004WR003204.

LLABRÉS-BRUSTENGA, A.; RIUS, A.; RODRÍGUEZ-SOLÀ, R. et al. **Quality control process of the daily rainfall series available in Catalonia from 1855 to the present. Theoretical and Applied Climatology,** v. 137, p. 2715–2729, 2019. Disponível em: https://link.springer.com/article/10.1007/s00704-019-02772-5. Acesso em: 10 dez. 2024.

MANN, H. B. **Nonparametric tests against trend.** Econometrica, v. 13, n. 3, p. 245–259, 1945. DOI: https://doi.org/10.2307/1907187

MANN, H. B.; WHITNEY, D. R. **On a test of whether one of two random variables is stochastically larger than the other.** The Annals of Mathematical Statistics, v. 18, n. 1, p. 50–60, 1947. DOI: 10.1214/aoms/1177730491.

MELO JÚNIOR, J. C. F.; SEDIYAMA, G. C.; FERREIRA, P. A.; LEAL, B. G. Determinação de regiões homogêneas quanto à distribuição de freqüência de chuvas no leste do Estado de Minas Gerais. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 10, n. 2, p. 408-416, 2006. DOI: https://doi.org/10.1590/S1415-43662006000200023.

MILLY, P. C. D.; BETANCOURT, Julio; FALKENMARK, Malin; HIRSCH, Robert M.; KUNDZEWICZ, Zbigniew W.; LETTENMAIER, Dennis P.; STOUFFER, Ronald J. **Stationarity is dead: whither water management?** Science, v. 319, n. 5863, p. 573-574, 2008. DOI: 10.1126/science.1151915.

MINAS GERAIS. Sistema Estadual de Meio Ambiente e Recursos Hídricos (SISEMA). **Infraestrutura de Dados Espaciais do Sisema – IDE-Sisema.** Belo Horizonte, 2018. Disponível em: <https://visualizador.idesisema.meioambiente.mg.gov.br\>. Acesso em: 10 dez. 2025.

NAGHETTINI, M.; PINTO, É. J. A. **Hidrologia estatística**. Belo Horizonte: CPRM, 2007.

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (NOAA). **How clouds form.** JetStream: Online School for Weather, 2025. Disponível em: https://www.noaa.gov/jetstream/clouds/how-clouds-form. Acesso em: 27 nov. 2025.

PAPALEXIOU, S. M.; KOUTSOYIANNIS, D. **Battle of extreme value distributions: a global survey on extreme daily rainfall**. Water Resources Research, v. 49, n. 1, p. 187–201, 2013. DOI: https://doi.org/10.1029/2012WR012557.

Ragulina, G.; Reitan, T. **Generalized extreme value shape parameter and its nature for extreme precipitation using long time series and the Bayesian approach.** Hydrological Sciences Journal, v. 62, n. 6, p. 863–879, 2017. DOI: https://doi.org/10.1080/02626667.2016.1260134

ROCHA, Eider Gomes de Azevedo. **Conceitos básicos de hidrologia e drenagem para projetos rodoviários.** Rio de Janeiro: IPR; DNIT, 2022. Disponível em: https://bibliotecadigital.enap.gov.br/bitstream/1/7814/2/M%C3%B3dulo%202%20-%20Pluviomtria.pdf. Acesso em: 11 dez. 2025.

ROE, Gerard H. **Orographic precipitation.** Annual Review of Earth and Planetary Sciences, v. 33, p. 645–671, 2005. DOI: 10.1146/annurev.earth.33.092203.122541.

Rousseeuw, P. J. **Silhouettes: a graphical aid to the interpretation and validation of cluster analysis.** Journal of Computational and Applied Mathematics, v. 20, p. 53–65, 1987. https://doi.org/10.1016/0377-0427(87)90125-7

SEARCY, J. K.; HARDISON, C. H. **Double-mass curves.** Washington, D.C.: United States Government Printing Office, 1960. 66 p. (U.S. Geological Survey Water-Supply Paper, 1541-B). Disponível em: <https://pubs.usgs.gov/wsp/1541b/report.pdf>. Acesso em: 09 dez. 2025.

SERVEI METEOROLÒGIC DE CATALUNYA (Meteocat). **Atles climàtics de Catalunya: precipitación/precipitació mitjana anual.** Barcelona: Meteocat, [s.d.]. Disponível em: https://www.meteo.cat/climatologia/atles_climatic/. Acesso em: 25 set. 2025.

SERVIÇO GEOLÓGICO DO BRASIL (SGB). **Serviço Geológico do Brasil avançará em 2025 com ações para operar a Rede Hidrometeorológica**

Nacional. Brasília: SGB, 5 dez. 2024. Disponível em: https://sgb.gov.br/w/servico-geologico-do-brasil-avancara-em-2025-com-acoes-para-operar-a-rede-hidrometeorologica-nacional. Acesso em: 1 dez. 2025.

SHAPIRO, S. S.; WILK, M. B. **An analysis of variance test for normality (complete samples).** Biometrika, v. 52, n. 3-4, p. 591-611, 1965. Disponível em: https://sci2s.ugr.es/keel/pdf/algorithm/articulo/shapiro1965.pdf. Acesso em: 21 jun. 2026.

SILVA JUNIOR, R. S. da; GAMA, M. C. C. da; SILVA, E. H. de L.; MARIANO, G. L.; OLIVEIRA JÚNIOR, J. F. de; SILVA, L. S. de O.; CARDOSO, K. R. A. **Avaliação de eventos extremos de precipitação, associados a desastres naturais**. Revista Brasileira de Geografia Física, v. 15, n. 6, p. 2755-2767, 2022. DOI: 10.26848/rbgf.v15.6.p2755-2767.

Terán-Chaves, C. A.; Duarte-Carvajalino, J. M.; Polo-Murcia, S. M. **Quality control and filling of daily temperature and precipitation time series in Colombia.** Meteorologische Zeitschrift, v. 30, n. 6, p. 489–501, 2021. DOI: 10.1127/metz/2021/1077.

TUCCI, Carlos E. M. (org.). **Hidrologia: ciência e aplicação.** 4. ed. Porto Alegre: Editora da UFRGS; ABRH, 2007. 943 p. ISBN 978-85-7025-924-0.

UN-WATER. **Water Security.** [S. l.]: UN-Water, Oct. 2013. Disponível em: https://www.unwater.org/sites/default/files/app/uploads/2017/05/unwater_poster_Oct2013.pdf. Acesso em: 11 dez. 2025.

VINEY, Neil R.; BATES, Bryson C. **It never rains on Sunday: the prevalence and implications of untagged multi-day rainfall accumulations in the Australian high quality data set.** International Journal of Climatology, v. 24, n. 9, p. 1171–1192, 2004. DOI: 10.1002/joc.1053.

WALD, A.; WOLFOWITZ, J. **An exact test for randomness in the non-parametric case based on serial correlation.** The Annals of Mathematical Statistics, v. 14, n. 4, p. 378-388, 1943. DOI: https://doi.org/10.1214/aoms/1177731358.

WALKER, S. **The value of hydrometric information in water resources management and flood control.** Meteorological Applications, v. 7, n. 4, p. 387-397, 2000.

WILKS, D. S. **On "field significance" and the false discovery rate**. Journal of Applied Meteorology and Climatology, v. 45, n. 9, p. 1181-1189, 2006. DOI: https://doi.org/10.1175/JAM2404.1.

WORLD BANK GROUP (WBG). **Brazil Country Climate and Development Report.** CCDR Series. Washington, DC: World Bank, 2023. Disponível em: https://openknowledge.worldbank.org/handle/10986/39782. Acesso em: 11 dez. 2025.

WORLD METEOROLOGICAL ORGANIZATION (WMO). **Manual on Flood Forecasting and Warning.** 2011 ed. Geneva: WMO, 2011. (WMO-No. 1072). ISBN 978-92-63-11072-5.

WORLD METEOROLOGICAL ORGANIZATION (WMO). **Manual on the Observation of Clouds and Other Meteors (International Cloud Atlas, WMO-No. 407).** Revised edition 1975. Geneva: Secretariat of the World Meteorological Organization, 1975. Disponível em: https://cloudatlas.wmo.int/docs/wmo_407_en-v1.pdf. Acesso em: 27 nov. 2025.

WORLD METEOROLOGICAL ORGANIZATION. **Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation**. Geneva: WMO, 2009. (Climate Data and Monitoring, WCDMP-No. 72; WMO-TD No. 1500). Disponível em: <https://www.ecad.eu/documents/WCDMP_72_TD_1500_en_1.pdf>. Acesso em: 02 jun. 2026.

WORLD METEOROLOGICAL ORGANIZATION (WMO). **Guide to Hydrological Practice. Volume I: Hydrology – From Measurement to Hydrological Information.** 2008 ed., updated in 2020. Geneva: World Meteorological Organization, 2020. (WMO-No. 168). Disponível em: <https://www.gov.br/ana/pt-br/assuntos/monitoramento-e-eventos-criticos/monitoram

ento-hidrologico/orientacoes-manuais/entidades/hydrology-2013-from-measurement-to_2020.pdf\>. Acesso em: 10 nov. 2025.

WORLD METEOROLOGICAL ORGANIZATION (WMO). **Guide to instruments and methods of observation. Volume I.** Geneva: World Meteorological Organization, 2024. (WMO-No. 8). Disponível em: <https://library.wmo.int/viewer/68695/>. Acesso em: 10 nov. 2025.

YUE, S.; PILON, P.; PHINNEY, B.; CAVADIAS, G. The influence of autocorrelation on the ability to detect trend in hydrological series. **Hydrological Processes**, v. 16, n. 9, p. 1807–1829, 2002. DOI: https://doi.org/10.1002/hyp.1095