UNIVERSIDADE DE SÃO PAULO
ESCOLA DE ENGENHARIA DE SÃO CARLOS
PROGRAMA DE PÓS GRADUAÇÃO EM HIDRÁULICA E SANEAMENTO

ANDRÉ SIMÕES BALLARIN

Caracterização de eventos extremos de precipitação e suas
consequências em um contexto de mudanças climáticas

São Carlos

2021

ANDRÉ SIMÕES BALLARIN

Caracterização de eventos extremos de precipitação e suas
consequências em um contexto de mudanças climáticas

Dissertação apresentada à Escola de Engenharia
de São Carlos da Universidade de São Paulo,
como parte das exigências do Programa de Pós-
Graduação em Hidráulica e Saneamento para a
obtenção do título de Mestre em Ciências:
Engenharia Hidráulica e Saneamento.

Orientador: Prof. Tit. Edson Wendland

VERSÃO CORRIGIDA

São Carlos

2021

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Dr. Sérgio Rodrigues Fontes da
EESC/USP com os dados inseridos pelo(a) autor(a).

B189c Ballarin, André Simões
 Caracterização de eventos extremos de
precipitação e suas consequências em um contexto de
mudanças climáticas / André Simões Ballarin; orientador
Edson Wendland. São Carlos, 2021.

 Dissertação (Mestrado) - Programa de
Pós-Graduação em Engenharia Hidráulica e Saneamento e
Área de Concentração em Hidráulica e Saneamento --
Escola de Engenharia de São Carlos da Universidade de
São Paulo, 2021.

 1. eventos hidrológicos extremos. 2.
distribuições de probabilidade. 3. mudanças climáticas.
4. precipitação. 5. São Paulo. 6. teoria de valores
extremos. I. Título.

Eduardo Graziosi Silva - CRB - 8/8907

FOLHA DE AVALIAÇÃO

FOLHA DE JULGAMENTO

Candidato: Engenheiro **ANDRÉ SIMÕES BALLARIN**.

Título da dissertação: "Caracterização de eventos extremos de precipitação e suas consequências em um contexto de mudanças climáticas".

Data da defesa: 02/03/2021.

Comissão Julgadora

Resultado

Prof. Titular **Edson Cezar Wendland**
(Orientador)
(Escola de Engenharia de São Carlos/EESC-USP)

Aprovado

Prof. Dr. **Wilson Santos Fernandes**
(Universidade Federal de Minas Gerais/UFMG)

Aprovado

Prof. Dr. **Carlos Henrique Ribeiro Lima**
(Universidade do Brasil/UnB)

Aprovado

Coordenador do Programa de Pós-Graduação em Engenharia Hidráulica e Saneamento:

Prof. Dr. **Eduardo Mario Mendiola**

Presidente da Comissão de Pós-Graduação:

Prof. Titular **Murilo Araujo Romero**

AGRADECIMENTOS

À minha família, razão do meu viver: Adriano, Luisa, Raquel, Caio e Mariana, por todo o apoio diário, carinho, amor, companhia, suporte e compreensão. Obrigado por traçarem esse caminho comigo, tornando-o muito mais leve.

Ao professor Edson, por todo apoio, sugestões e conselhos que com certeza contribuíram muito para minha formação pessoal e profissional.

Aos colegas do Laboratório de Hidráulica Computacional, pela boa convivência, companheirismo e troca mútua de conhecimentos e aprendizados.

Aos funcionários do Departamento de Hidráulica e Saneamento, em especial à Sá e Priscila, por todo suporte com questões administrativas e logísticas.

Aos demais amigos de Botucatu e São Carlos, que me acompanharam ao longo desse processo. Agradeço pela companhia e apoio ao longo desses últimos anos.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (processo nº 131238/2019-1) pelo suporte financeiro.

RESUMO

BALLARIN, A. S. **Caracterização de eventos extremos de precipitação e suas consequências em um contexto de mudanças climáticas.** 2021. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2021.

A caracterização dos eventos extremos de precipitação é fundamental para o planejamento e dimensionamento de sistemas hidráulicos. Nesse processo, existem diversas fontes de incerteza que podem afetar significativamente as estimativas realizadas. Entretanto, não há ainda um consenso sobre como contabilizá-las nas diversas etapas de modelagem. Além disso, o contexto atual de mudanças climáticas acaba dificultando ainda mais o entendimento das propriedades desses eventos. Em vista disso, o presente trabalho buscou analisar as diferentes etapas do processo de caracterização de eventos extremos e suas consequências nas estimativas realizadas. A partir de um estudo de caso realizado no estado de São Paulo, foi proposta uma metodologia para a escolha de distribuições de probabilidade utilizadas nesta caracterização, sintetizando informações fornecidas por diferentes critérios de escolha, comumente utilizados nas análises hidrológicas. Considerando a potencial intensificação dos eventos extremos, avaliou-se a evolução temporal dos quantis de precipitação estimados, considerando também, o impacto da representatividade dos dados observados e da redução da série temporal nas estimativas realizadas. Por fim, o estudo avaliou o desempenho de uma metodologia estatística alternativa à teoria clássica de valores extremos para caracterizar chuvas intensas, chamada de teoria Metaestatística de Valores Extremos (MEV). Os resultados mostraram que a associação de diferentes critérios na escolha das distribuições de probabilidade aperfeiçoa a seleção da distribuição mais adequada e que, embora a utilização de um modelo de probabilidade regionalmente facilite o planejamento hidráulico, essa prática pode aumentar significativamente os erros nas estimativas. A distribuição de Gumbel, amplamente utilizada nos estudos hidrológicos do país, subestimou em muitas situações os eventos hidrológicos. Em contrapartida, a distribuição Generalizada de Valores Extremos se mostrou adequada para a região de estudo. Os quantis de precipitação estimados estão se alterando ao longo do tempo, sendo, necessário que os estudos hidrológicos sejam atualizados constantemente para contabilizar suas propriedades atuais na estimativa de eventos futuros. A análise da representatividade dos dados observados se mostrou como uma alternativa para o processo de caracterização, buscando evitar subestimativas de eventos futuros. Por fim, a distribuição MEV se mostrou mais adequada que a metodologia clássica para caracterizar eventos extremos em um contexto de mudanças climáticas, apresentando menores incertezas em suas estimativas.

Palavras-chave: eventos hidrológicos extremos; mudanças climáticas; distribuições de probabilidade; incertezas.

ABSTRACT

BALLARIN, A. S. **Characterization of extreme rainfall events and its consequences in a changing climate context.** 2021. Dissertação (Mestrado) – Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos, 2021.

The characterization of extreme rainfall events is a fundamental task for the design and management of hydraulic systems. There are, in this process, several sources of uncertainty that may significantly affect the extreme rainfall estimation. However, there is still no general agreement on how to account for them in the modeling steps. In addition, changing climate context makes it even more difficult to understand the properties of these events. Thus, the present work aimed at analysing the different steps of the characterization process and their consequences on the evaluated quantiles. Based on a case study carried out in the state of São Paulo, we proposed a methodology for the selection of probability distributions used in characterization, summarizing information provided by different statistical metrics commonly used in hydrological analyzes. Considering the potential intensification of extreme events, the temporal evolution of the estimated precipitation quantiles was evaluated. Due to this possible intensification, the impact of the representativeness of the observed data and the reduction of the time series in the modeling process were also analyzed. Finally, we evaluated the performance of an alternative statistical methodology to the classic extreme value theory for extreme rainfall characterization, called Metastatistical Extreme Value theory (MEV). The results showed that the association of different selection approaches improves the choice of the most appropriate distribution and that, despite the use of a regional probability model facilitates hydrological studies, this practice may significantly increase estimation errors. The Gumbel distribution, widely used in hydrological studies in the country, underestimated hydrological events in many situations. In contrast, the Generalized Extreme Value distribution proved to be adequate for characterizing extreme events in the state. Estimated precipitation quantiles have been shown to change over time, highlighting the need for constant update of hydrological studies to account for current characteristics of rainfall events in future estimations. The analysis of the representativeness of the observed data figured out as an alternative for the characterization process, avoiding possible underestimation of future events. Finally, the MEV distribution figured out as more appropriate than the classic methodology for characterizing extreme events in a changing climate context, presenting a lower degree of uncertainty in its process.

Keywords: extreme hydrological extremes; climate change; probability distributions; uncertainties.

Lista de Figuras

Figura 1 - Etapas do processo de escolha do modelo de probabilidade.	33
Figura 2 - Etapas do processo de realização dos testes preditivos.	37
Figura 3 – (a) Processo de determinação dos testes descritivos e (b) Processo de determinação dos testes preditivos. As letras P, t, f, s, e N, nos eixos dos gráficos, referem-se a altura de precipitação, tempo, frequência, valor da estatística de estudo e modelos avaliados, respectivamente.	38
Figura 4 – Proposta de avaliação dos desempenhos preditivos dos modelos. (a) Modelos cujo melhor desempenho é expresso por menores valores das métricas (X). (b) Modelos cujo melhor desempenho é expresso por maiores valores das métricas (Z). (c) Esquema do diagrama box-plot das métricas X	39
Figura 5 – Área de estudo e estações meteorológicas utilizadas: (a) América do Sul, com o Brasil e o estado de São Paulo destacados em cinza claro e escuro, respectivamente. (b) Distribuição das 699 estações meteorológicas utilizadas no estudo. (c) Número de estações com dados válidos em cada um dos anos avaliados. (d) Tamanho das séries presentes no banco de dados. Foram removidas 3 estações, com extensa série amostral (com dados a partir de 1890) dos gráficos (c) e (d), para facilitar a visualização.	41
Figura 6 – Distribuição da pontuação (ranking) obtida por cada modelo nas 699 estações consideradas no estudo de caso, considerando as diferentes métricas descritivas e preditivas avaliadas.	45
Figura 7 – Box plot: distribuição das pontuações preditiva, descritiva e final para cada modelo avaliado. Pontuação igual a 1 indica o modelo com melhor desempenho. Linhas central, inferior (da caixa) e superior (da caixa) representam a mediana e percentil 25 e 75, respectivamente. O triângulo central (cinza) indica a pontuação média. Tabela: número de estações onde cada um dos modelos foi escolhido como o mais adequado, de acordo com as pontuações descritiva, preditiva e final.	46
Figura 8 – Distribuição espacial das distribuições de probabilidade escolhidas pela metodologia proposta no estado de São Paulo.	49
Figura 9 – Comparação dos quantis extremos de precipitação estimados pela metodologia proposta com aqueles estimados a partir de uma única distribuição de probabilidade adequada para o estado inteiro. Esquerda: diferenças dos quantis associados aos períodos de retorno de 10 (a) e 100 anos (c) usando a distribuição GEV para todo o estado. Direita:	

diferenças entre os quantis associados aos períodos de retorno de 10 anos (b) e 100 anos (d) utilizando a distribuição GUM para todo o estado. Valores negativos representam subestimativas e valores positivos, superestimativas.....	51
Figura 10 - Estimativa dos eventos extremos de precipitação do estado de SP associados ao período de retorno de 10 anos (a) e 25 anos (b).....	53
Figura 11 - Estimativa dos eventos extremos de precipitação do estado de SP associados ao período de retorno de 50 anos (a) e 100 anos (b).....	54
Figura 12 – Área de estudo e estações meteorológicas utilizadas: (a) Brasil, com o estado de São Paulo e América do Sul destacados em cinza escuro e claro, respectivamente. (b) Estações pluviométricas utilizadas no estudo. (c) Tamanho amostral das séries avaliadas. (d) Número de estações com dados válidos em cada um dos anos avaliados. Foram removidas 3 estações, com extensa série amostral (com dados a partir de 1890) dos gráficos (c) e (d), para facilitar a visualização.....	61
Figura 13 – Etapas para a construção das séries crescente - <i>Si</i> (a) e móvel - <i>Sf</i> (b). <i>Qji</i> e <i>Qjf</i> representam os quantis de precipitação estimados que compõem as séries crescente e móvel, respectivamente.....	63
Figura 14 – Tendências dos quantis de precipitação das séries <i>Si</i> e <i>Sf</i> , considerando os períodos de retorno de 5, 10, 25, 50, 100 e 200 anos. A distribuição espacial das tendências encontradas foi espacialmente interpolada pelo método do inverso quadrado das distâncias (IDW).....	71
Figura 15 – Série temporal de precipitação observada (AMS) para as estações São Paulo-1 e Campinas-2, divididas no ponto de mudança e na metade, respectivamente. Distribuição dos dados da subsérie antiga (azul/verde), recente (amarelo/vermelho) e da série completa estão representados no diagrama box-plot (com/sem mudança abrupta).....	73
Figura 16 – Séries de precipitação observadas (AMS) das estações avaliadas. Esquerda: estações com mudanças abruptas. Direita: estações sem mudanças abruptas.....	74
Figura 17 - Estimativa dos quantis de precipitação associados a diferentes períodos de retorno para as estações São Paulo-1 (a-b) e Campinas-2 (c-d). Esquerda: comparação dos quantis estimados usando as subséries antiga (azul//verde), recente (amarelo/vermelho) e completa (preto) junto ao $CI_{95\%}$, calculado pelo método de reamostragem bootstrap. Direita: detalhes das diferenças nos períodos de retorno do quantis estimados.....	75
Figura 18 – Diagramas de período de retorno para todas as estações avaliadas, comparando os quantis estimados pelas séries antiga, recente e completa. Esquerda e direita: estações com	

séries com e sem ponto de mudança significativo, respectivamente. Todos os gráficos apresentam o intervalo de confiança $CI_{95\%}$	77
Figura 19 – Distribuição dos quantis de precipitação estimados pelas séries reduzida e completa, para os períodos de retorno de 10, 25, 50, 100 e 200 anos. Linhas tracejadas representam a diferença entre as amplitudes do $CI_{95\%}$ considerando as séries reduzidas e completas.	79
Figura 20 – Esquema conceitual da construção da distribuição MEV aplicado à máxima precipitação diária. Parâmetros de Weibull (em azul e vermelho) e o número de dias chuvosos (verde) são determinados para cada ano, definindo uma distribuição acumulada anual de máxima precipitação diária. Distribuições acumuladas anuais (cinza) e suas projeções (preto) definem a variabilidade dos valores de C, w, e N. A distribuição MEV é obtida pela média das distribuições acumuladas de cada ano.	88
Figura 21 - Distribuição dos parâmetros da distribuição de probabilidade de Weibull, considerando todos os anos válidos de todas as estações avaliadas.	90
Figura 22 - Adequabilidade da distribuição de Weibull para descrever os eventos ordinários de precipitação, segundo os testes de aderência de Kolmogorov-Smirnov (KS), Anderson-Darling (AD) e Cramer Von Mises (CVM). A linha tracejada escura indica o limite onde rejeita-se a hipótese nula (valor-p = 0,05).	90
Figura 23 - Esquema da construção das (A) séries com mudanças abruptas e das (B) séries com tendências lineares em seus parâmetros. No primeiro caso, os blocos vermelho e verde representam o intervalo percentil de onde foram escolhidos os parâmetros de geração da primeira e segunda metade da série completa. No segundo caso, o bloco amarelo indica a janela móvel, que representa o intervalo percentil de onde foram escolhidos os parâmetros de geração dos valores de precipitação que compuseram a série completa.	96
Figura 24 – Quantis estimados pelas distribuições MEV, GEV (técnica AMS) e GPD (técnica POT) considerando diferentes tamanhos amostrais (L) utilizados no processo de modelagem (10 a 200 anos). Os pontos pretos indicam os quantis empíricos extraídos da série de 1000 anos, calculados pela fórmula de posição de plotagem de Weibull. O sombreado colorido representa o intervalo de confiança ao nível de significância de 95% ($CI_{95\%}$) e as linhas sólidas indicam a média dos quantis estimados por cada um dos modelos.	99
Figura 25 - Quantis estimados pelas distribuições MEV, GEV (técnica AMS) e GPD (técnica POT) considerando diferentes tamanhos amostrais (L) utilizados no processo de modelagem (10 a 200 anos) para séries com mudanças abruptas. Os pontos pretos indicam os quantis empíricos extraídos da série de 1000 anos, calculados pela fórmula de posição de plotagem de Weibull. O sombreado colorido representa o intervalo de confiança ao nível de	

significância de 95% ($CI_{95\%}$) e as linhas sólidas indicam a média dos quantis estimados por cada um dos modelos. 101

Figura 26 - Razão entre eventos extremos de precipitação (período de retorno de 100 anos) estimados e observados para todas as séries sintéticas com mudanças abruptas (CP1 a CP6). MEV, GEV e POT foram modelados considerando tamanhos de amostra de 20 (topo) e 100 (base) anos..... 102

Figura 27 – Distribuição das estatísticas RMSE (*violin plot*), obtida para os três modelos avaliados, considerando os quatro tamanhos amostrais. Caixas pretas centrais indicam a distância inter-quartil do diagrama box-plot. Pontos brancos representam o valor médio da estatística RMSE. 104

Figura 28 – Quantis estimados pelas distribuições MEV, GEV (técnica AMS) e GPD (técnica POT) considerando diferentes tamanhos amostrais (L) utilizados no processo de modelagem (10 a 200 anos) para as séries com tendência linear nos parâmetros. Os pontos pretos indicam os quantis empíricos extraídos da série de 1000 anos, calculados pela fórmula de posição de plotagem de Weibull. O sombreado colorido representa o intervalo de confiança ao nível de significância de 95% ($CI_{95\%}$) e as linhas sólidas indicam a média dos quantis estimados por cada um dos modelos..... 106

Figura 29 - Razão entre eventos extremos de precipitação (período de retorno de 100 anos) estimados e observados para todas as séries sintéticas com tendências lineares em seus parâmetros (T1 a T6). MEV, AMS e POT foram modelados considerando tamanhos de amostra de 20 (topo) e 100 (base) anos..... 107

Figura 30 - Distribuição das estatísticas RMSE (*violin plot*) obtida para os três modelos avaliados, considerando os quatro tamanhos amostrais. Caixas pretas indicam a distância inter-quartil do diagrama box-plot. Pontos brancos representam o valor médio da estatística RMSE..... 108

Figura 31 – Desempenho (via RMSE) das distribuições MEV e GEV para caracterizar os eventos extremos de precipitação da área de estudo, considerando 48 estações com tamanho amostral maior ou igual a 70 anos. Áreas sombreadas representam o intervalo de confiança do RMSE ao nível de significância de 95% ($CI_{95\%}$). A linha tracejada preta representa o valor do RMSE relativo ao desempenho ideal ($RMSE = 0$ mm)..... 111

Figura 32 – Diferenças entre os quantis de precipitação associados ao período de retorno de 100 anos, estimados pelas distribuições MEV e GEV. (a) América do Sul, com o Brasil e o estado de São Paulo destacados em cinza claro e escuro, respectivamente. (b) Distribuição

especial das diferenças (MEV – GEV) entre os quantis de precipitação estimados pelos dois modelos. Dados pontuais de cada estação (pontos pretos) interpolados espacialmente pelo método IDW. Valores positivos (negativos) representam subestimativa (superestimativa) dos quantis estimados pela GEV em relação aos quantis estimados pela MEV. 112

Lista de tabelas

Tabela 1 - Distribuições de probabilidade avaliadas (funções densidade de probabilidade e parâmetros).....	43
Tabela 2 - Estatísticas descritivas dos quantis de precipitação estimados para o estado de SP....	52
Tabela 3 – Parcela das estações avaliadas (%) que apresentaram tendência significativa (nível de significância de 5%) considerando a média das tendências encontradas para as séries crescente e móvel.	70
Tabela 4 - Dados das estações avaliadas em relação à presença de ponto de mudança. Linhas destacadas em cinza representam estações com ponto de mudança significativo.	72
Tabela 5 - Intervalo de percentis definido para seleção dos parâmetros das subséries inicial e final, que compõem a série sintética completa. Intervalo P1-P2 foi utilizado para a construção da subsérie inicial e intervalo P3-P4 foi utilizado para a construção da subsérie final.....	93
Tabela 6 – Tamanho da janela móvel, parâmetro variável (com tendência linear positiva) e intervalo inicial e final dos percentis utilizados para a construção das séries sintéticas completas com tendência linear positiva em seus parâmetros.	95

SUMÁRIO

CAPÍTULO I – Introdução Geral	25
1.1 OBJETIVOS	29
1.2 ESTRUTURA DO TRABALHO	29
CAPÍTULO II – Sistematização da seleção de distribuição de probabilidade para estimativa de eventos extremos de precipitação: combinando testes descritivos e preditivos.....	30
2.1 INTRODUÇÃO.....	31
2.2 MATERIAL E MÉTODOS.....	32
2.2.1 Testes descritivos	33
2.2.2 Testes preditivos	36
2.2.3 Critério de escolha da distribuição de probabilidade proposto	40
2.3 Estudo de caso.....	40
2.3.1 Área de Estudo	40
2.3.2 Distribuições de probabilidade candidatas.....	41
2.3.3 Estimativa dos parâmetros das distribuições de probabilidade.....	43
2.4 RESULTADOS E DISCUSSÃO.....	44
2.4.1 Desempenho dos modelos avaliados	44
2.4.2 Caracterização das chuvas intensas na área de estudo.....	51
2.5 CONCLUSÃO.....	55
CAPÍTULO III – Avaliação da representatividade dos dados observados e suas influências nas estimativas de precipitação.....	57
3.1 INTRODUÇÃO.....	58
3.2 MATERIAIS E MÉTODOS.....	60
3.2.1 Área de estudo e obtenção das séries de quantis de precipitação	60
3.2.2 Análise de tendência temporal	64
3.2.3 Análise da representatividade dos dados	66
3.3 RESULTADOS E DISCUSSÃO.....	69
3.3.1 Análise de tendência temporal	69

3.3.2	Análise da representatividade dos dados	71
3.4	<i>CONCLUSÃO</i>	79

CAPÍTULO IV – Avaliação da distribuição metaestatística de valores extremos para caracterização da precipitação em um contexto de mudanças climáticas **81**

4.1	<i>INTRODUÇÃO</i>	82
4.2	<i>ESTRUTURA TEÓRICA</i>	84
4.3	<i>MATERIAIS E MÉTODOS</i>	88
4.3.1	Área de estudo	88
4.3.2	Estimativa dos parâmetros das distribuições de probabilidade	91
4.3.3	Séries sintéticas.....	91
4.3.4	Séries sintéticas com mudanças abruptas	91
4.3.5	Séries sintéticas com tendência linear em seus parâmetros.....	93
4.3.6	Validação da MEV para a área de estudo	97
4.4	<i>RESULTADOS E DISCUSSÃO</i>	98
4.4.1	Séries Sintéticas	98
4.4.2	Séries sintéticas com mudanças abruptas	100
4.4.3	Séries sintéticas com tendência linear em seus parâmetros.....	105
4.4.4	Caracterização da precipitação na área de estudo.....	109
4.5	<i>CONCLUSÃO</i>	113

CAPÍTULO V – Considerações Finais..... **115**

5.1	<i>CONSIDERAÇÕES FINAIS</i>	116
-----	-----------------------------------	-----

REFERÊNCIAS **120**

APÊNDICE A **131**

APÊNDICE B **151**

APÊNDICE C **154**

APÊNDICE D **156**

APÊNDICE E	162
-------------------------	------------

CAPÍTULO I

Introdução geral

Os eventos extremos de precipitação possuem grande importância nos estudos hidrológicos, como no dimensionamento de estruturas hidráulicas e na gestão de recursos hídricos. O correto entendimento sobre eles permite uma antecipação mais precisa de possíveis danos aos mais diversos sistemas naturais e socioeconômicos (ALAM et al., 2018; NGUYEN; NGUYEN, 2016). A análise estatística de frequência, metodologia comumente aplicada na caracterização dos eventos de máxima precipitação (SILVA et al., 2012), consiste em, basicamente, ajustar um modelo de probabilidade à série temporal do local de interesse. A partir disso, são construídas curvas de Intensidade-Duração-Frequência (IDF), que associam uma intensidade de precipitação a um determinado período de retorno, para um evento com duração estipulada. Essas curvas, derivadas de séries de precipitação observadas, acabam sendo utilizadas como uma das principais ferramentas no dimensionamento e planejamento de estruturas hidráulicas (SIMONOVIC et al., 2016).

Embora simples, esse procedimento ainda representa um grande desafio na prática de Engenharia. Diversos fatores, associados às diferentes etapas do processo de caracterização, são responsáveis pelas variações e incertezas existentes nas estimativas de precipitação (GAUME, 2018; VIGLIONE et al., 2013). Inicialmente, existem as incertezas associadas à qualidade e disponibilidade dos dados observados, que interferem diretamente na precisão das estimativas realizadas (OKOLI et al., 2018). Numa segunda etapa, na definição da série de dados observados a serem utilizados, existem os erros associados à representatividade da série amostral (MERZ; BLÖSCHL, 2008; MONTANARI; KOUTSOYIANNIS, 2014). Isto é, considerando o contexto recente de alterações dos eventos extremos de precipitação (ASADIEH; KRAKAUER, 2015), deve-se avaliar com cautela as consequências da utilização de dados antigos com características não mais observadas no processo de caracterização. Há também a influência dos diferentes métodos de construção das séries: séries de máxima anual, que selecionam o evento de maior intensidade em cada ano dos dados observados; séries de duração parcial, que selecionam todos eventos superiores a um limite pré-estabelecido; e as séries de precipitação ordinárias, que consideram todos os eventos observados no período (MADSEN; RASMUSSEN; ROSBJERG, 1997; MARANI; IGNACCOLO, 2015).

Considerando, posteriormente, a etapa de ajuste de uma distribuição de probabilidade, encontram-se incertezas associadas a diferentes aspectos. A escolha da posição de plotagem dos dados observados (EWEMOJE; EWEMOOJE, 2011; MAKKONEN, 2006) e as diferentes distribuições de probabilidade usadas nas análises hidrológicas (NGUYEN; NGUYEN, 2019; OKOLI et al., 2018; SALINAS et al., 2014a, 2014b) alteram

significativamente as estimativas realizadas. Existem, ainda, diferentes técnicas de determinação dos parâmetros da distribuição de probabilidade (HAKTANIR; COBANER; GORKEMLI, 2013; HOSKING, 1990) e critérios de escolha do modelo mais adequado (LAIO; DI BALDASSARRE; MONTANARI, 2009), não havendo um consenso, em ambos os casos, sobre a técnica mais indicada. O desempenho dos estimadores dos parâmetros acaba variando de acordo com o tamanho e as características da série de dados observados (KATZ; PARLANGE; NAVEAU, 2002).

Em relação aos critérios de seleção de distribuições de probabilidade, as metodologias comumente utilizadas são baseadas na capacidade do modelo de descrever os dados observados, comparando-se os dados amostrais com os dados obtidos através do modelo (NGUYEN et al., 2017; ÖZTEKIN, 2007). Com isso, não avalia-se a capacidade do modelo de caracterizar eventos ainda não observados, isto é, presentes na faixa de extrapolação (VOLPI, 2019). Dessa maneira, ocorrem inadequações relacionadas à distribuição de probabilidade escolhida, pelo fato das propriedades usualmente avaliadas no processo de escolha não serem, em muitas situações, as mais adequadas.

Devido a todas essas dificuldades, muitos países descrevem, em suas diretrizes, critérios e técnicas de caracterização a serem utilizadas, em toda sua área, na estimativa dos quantis de precipitação (SALINAS et al., 2014a, 2014b). Entretanto, as diferentes propriedades apresentadas por esses eventos dificultam a adequação de um único modelo a todos os eventos da região. Isto é, a variabilidade espacial e temporal apresentada pela precipitação dificulta a definição de um único modelo como o mais adequado para todas estações meteorológicas de uma região.

No caso do Brasil, ainda não há uma recomendação sobre o procedimento de caracterização de eventos de precipitação nas diretrizes nacionais, embora encontrem-se diversos trabalhos na literatura, que avaliam a adequabilidade de distribuições de probabilidade aos dados de máxima precipitação (BATISTA; COELHO, 2013; BESKOW et al., 2015; BLAIN, 2011; CALDEIRA et al., 2015; PEDRON et al., 2017; SANTOS et al., 2016; SILVA et al., 2002). Entretanto, esses trabalhos têm como foco a avaliação da adequabilidade de diferentes modelos de probabilidade aos dados observados, não abordando questões sobre as diferentes fontes de incertezas e suas consequências nas estimativas realizadas. Além disso, a avaliação da adequabilidade é feita através de testes de aderência, cuja utilização para comparação de desempenhos de diferentes distribuições e seleção não é a mais recomendada,

visto que eles avaliam somente se um determinado modelo é adequado ou não, impossibilitando a comparação entre diferentes distribuições candidatas (NAGHETTINI; PINTO, 2007). Aspectos quanto à representatividade dos dados observados; avaliação de mudanças nas séries, por meio da análise de padrões temporais e espaciais; e metodologias de caracterização alternativas; não são abordados.

Considerando o contexto atual de mudanças climáticas e de projeção de mudanças nos padrões dos eventos extremos, há a necessidade de serem aprimoradas as técnicas de caracterização dos eventos de precipitação. Muitas dessas questões associadas às incertezas, nas diversas etapas da caracterização dos eventos extremos, ainda não foram resolvidas nos estudos hidrológicos. Em uma iniciativa para sintetizar as questões ainda não solucionadas na Hidrologia, Blöschl et al. (2019) apontaram a variabilidade dos eventos extremos como um tema de destaque dentre os problemas em aberto. A dificuldade de serem precisamente compreendidos temporal e espacialmente, além da magnitude de seus impactos sociais e ambientais, os caracterizam como ponto vital para o avanço dos estudos hidrológicos. Nesse sentido, têm sido buscadas alternativas às práticas comumente aplicadas, com o intuito de aprimorar o entendimento desses eventos, para que seja possível um correto planejamento de obras de infraestrutura hidráulica, evitando possíveis sub ou superdimensionamentos (PAPALEXIOU; KOUTSOYIANNIS, 2013).

Deve-se buscar uma abordagem sistemática que consiga avaliar os diferentes aspectos referentes às incertezas presentes nas estimativas dos quantis de precipitação (MONTANARI; KOUTSOYIANNIS, 2014). Para tal, uma avaliação holística, que consiga considerar, entre outros aspectos, a capacidade preditiva dos modelos, a representatividade dos dados observados e as diferentes técnicas de escolha de distribuições de probabilidade, se faz necessária. Essa visão mais abrangente facilitará as práticas de engenharia hidrológica, auxiliando tomadores de decisões na gestão e dimensionamento de sistemas hidráulicos, devido à possibilidade de caracterizar os eventos extremos de maneira mais precisa, considerando suas dinâmicas temporais e possíveis alterações de suas propriedades futuras.

1.1 OBJETIVOS

Este trabalho tem como objetivo, sistematizar o processo de caracterização dos eventos extremos de precipitação, buscando aprimorar o entendimento quanto a suas propriedades estatísticas e contabilizar as possíveis incertezas nas diversas etapas de modelagem. Para isso, foram delineadas as seguintes metas específicas:

- I. Avaliar as metodologias atuais de escolha de distribuições de probabilidade, analisando tanto a capacidade descritiva como a preditiva dos modelos candidatos;
- II. Investigar um novo critério de escolha de distribuições de probabilidade, baseado no ranqueamento entre modelos, de acordo com seus desempenhos relacionados às capacidades descritivas e preditivas;
- III. Avaliar a existência de tendências temporais nos quantis de precipitação estimados na área de estudo, considerando o contexto atual de mudanças climáticas;
- IV. Analisar os processos usuais de construções das séries de máxima precipitação e de escolha das distribuições de probabilidade, examinando as possíveis consequências da falta de representatividade de dados nos quantis de precipitação estimados.
- V. Analisar o desempenho da teoria Metaestatística de Valores Extremos (MEV) e da tradicional teoria de Valores Extremos (EVT) para caracterizar eventos extremos de precipitação de séries temporais em um contexto de intensificações climáticas.

1.2 ESTRUTURA DO TRABALHO

O Capítulo II busca responder os objetivos propostos I e II, abordando temas relacionados à caracterização dos modelos de probabilidade, sistematização do procedimento de escolha de distribuições de probabilidade em uma nova metodologia e avaliação do desempenho da proposta por meio de um estudo de caso, caracterizando os eventos de máxima precipitação da região de estudo. O Capítulo III inclui a avaliação das mudanças climáticas por meio da análise de tendência, e suas influências no processo de caracterização dos eventos extremos de precipitação, abordando, principalmente, o processo de construção das séries temporais e seus impactos nas estimativas realizadas (objetivos III e IV). Por fim, no Capítulo IV, é apresentada uma comparação entre a análise de frequência clássica (objetivo V), que considera eventos extremos de máximo para a modelagem das distribuições, com uma metodologia alternativa, chamada de Metaestatística de Valores Extremos, que utiliza todos os eventos ordinários de precipitação observados no processo de modelagem.

CAPÍTULO II

**Sistematização da seleção de distribuição de probabilidade para estimativa
de eventos extremos de precipitação: combinando testes descritivos e
preditivos**

2.1 INTRODUÇÃO

As catástrofes associadas aos eventos extremos de máxima precipitação, como as inundações, podem gerar grandes danos à sociedade (SLATER; VILLARINI, 2016). Apesar dos significativos avanços científicos no entendimento desses processos, o contexto atual de mudanças climáticas evidencia a necessidade de aperfeiçoar o conhecimento sobre os eventos extremos, visando uma correta antecipação de seus possíveis impactos (ALAM et al., 2018; MONTANARI; KOUTSOYIANNIS, 2014; NGUYEN; NGUYEN, 2016). Geralmente, a caracterização de eventos extremos é baseada na análise estatística de frequência (NGUYEN; NGUYEN, 2019). Este processo envolve o ajuste de uma distribuição de probabilidade que seja adequada para descrever as variáveis de interesse observadas (BARATTI et al., 2012; OKOLI et al., 2019a). O modelo ajustado é, então, extrapolado para associar eventos de magnitude específica a um período de retorno, que expressa a raridade de um evento, de acordo com o seu tempo médio de recorrência (VOLPI, 2019).

Embora pareça uma tarefa bem estabelecida, esse procedimento possui um alto grau de incerteza (SALINAS et al., 2014b). Muitos são os métodos de caracterização de eventos extremos, variando na forma de se obter as séries amostrais, nos critérios de escolha de modelos de probabilidade adequados e na estimativa de seus parâmetros (GAUME, 2018; OKOLI et al., 2018). Em relação à seleção de distribuições de probabilidade, há uma série de modelos que podem ser aplicados para representar eventos hidrológicos extremos, sem haver, entretanto, um consenso sobre o procedimento adequado para escolha de distribuições de probabilidade (NGUYEN et al., 2017). Entretanto, a adoção de um modelo que não seja o mais adequado pode levar a sérias sub ou superestimativas do período de retorno associado ao evento de interesse (PAPALEXIOU; KOUTSOYIANNIS, 2013). Dessa forma, os modelos são usualmente selecionados de acordo com seu desempenho em diferentes testes estatísticos (OKOLI et al., 2018).

O procedimento comumente aplicado nessa seleção é baseado na avaliação da capacidade descritiva dos modelos candidatos, por meio de testes de aderência e estatísticas de erro que verificam a qualidade do modelo para descrever os dados observados (NGUYEN et al., 2017). No entanto, essa abordagem possui alguns obstáculos que acabam influenciando o processo de caracterização. Primeiro, existem diferentes métricas de erro que podem ser usadas para avaliar o desempenho descritivo do modelo. A seleção de uma métrica apropriada não é uma tarefa simples, pelo fato de cada métrica apresentar pontos fortes e fracos associados aos diferentes tipos de erro e de viés, não contabilizando o real desempenho dos modelos avaliados

a partir de uma perspectiva abrangente (JACKSON et al., 2019; KRAUSE; BOYLE; BÄSE, 2005).

Além disso, os testes descritivos baseiam-se na comparação dos dados observados com os simulados pelo modelo. Assim, sua eficácia é limitada pelo tamanho da série observada. Caso o evento de interesse esteja na faixa de extrapolação, isto é, eventos com período de retorno maior do que o tamanho da série temporal (VOLPI, 2019), o modelo estará sendo utilizado para extrapolação de dados, cujo desempenho não é avaliado pelos testes descritivos. Essa propriedade, no entanto, é essencial para selecionar distribuições de probabilidade, especialmente em locais com baixa disponibilidade de dados, onde o evento de interesse frequentemente pertence à faixa de extrapolação. Embora importante, poucos estudos avaliam a capacidade preditiva de modelos utilizados na caracterização de eventos extremos (LUKE et al., 2017; NGUYEN et al., 2017; ÖZTEKIN, 2007).

Com base no exposto, neste capítulo II, foi desenvolvido um procedimento para auxiliar o processo de escolha de distribuições de probabilidade, sintetizando as informações sobre os desempenhos descritivo e preditivo dos modelos de probabilidade em uma pontuação final. A abordagem proposta busca sistematizar a metodologia de seleção de distribuições de probabilidade, auxiliando tomadores de decisões nas práticas de gerenciamento e dimensionamento de sistemas hidráulicos. A metodologia proposta foi aplicada em um estudo de caso, no estado de São Paulo, buscando (1) comparar seu desempenho com a metodologia tradicional de seleção de distribuições de probabilidade (baseada apenas em testes descritivos) e (2) encontrar a distribuição de probabilidade mais adequada para caracterizar os eventos extremos de chuva da região. Dessa maneira, este estudo tem como objetivo principal aprimorar o processo de caracterização de eventos extremos de precipitação, refletindo principalmente na escolha de distribuições de probabilidade adequadas para descreverem os dados observados.

2.2 MATERIAL E MÉTODOS

O procedimento desenvolvido para definir a metodologia de caracterização de eventos de máxima precipitação está sintetizado na Figura 1. A partir da série de precipitação, modelam-se as distribuições de probabilidade a serem avaliadas. Aplicam-se então os testes descritivos e preditivos, ranqueando, para ambos os critérios, as distribuições de probabilidade de acordo com seus desempenhos. Com base nesse ranqueamento, uma pontuação final é

atribuída para cada distribuição, sintetizando as informações relativas aos seus desempenhos nos testes descritivos e preditivos avaliados. Com isso, define-se a distribuição de probabilidade mais adequada para a série avaliada, comparando as pontuações finais obtidas por cada uma das distribuições candidatas. Nos itens 2.2.1 e 2.2.2, são apresentados os testes descritivos e preditivos avaliados na metodologia, respectivamente. No item 2.2.3, é apresentado o critério de pontuação definido para a escolha da distribuição mais adequada, considerando seu desempenho em ambos os testes realizados.

Figura 1 - Etapas do processo de escolha do modelo de probabilidade.

Fonte: Autor (2021).

2.2.1 Testes descritivos

Para a aplicação dos testes descritivos deve-se, inicialmente, construir o diagrama de frequências relativas acumuladas, também chamado de posição de plotagem, que associa, a cada um dos dados observados, uma probabilidade empírica de não superação. Neste trabalho, optamos pela fórmula de Weibull (Equação 1), por apresentar melhor desempenho quando comparada com demais formulações encontradas na literatura (EWEMOJE; EWEMOOJE, 2011; MAKKONEN, 2006), não enviesando as probabilidades de superação para todas as distribuições (NAGHETTINI; PINTO, 2007).

$$q_i = \frac{i}{N+1}, i = 1, 2, \dots, m \quad (1)$$

onde q_i é a probabilidade empírica de não superação associada ao evento, N é o tamanho da amostra e m , a posição do evento, classificado em ordem crescente.

Os testes descritivos comumente utilizados na avaliação da adequação de distribuições de probabilidade são: Kolmogorov-Smirnov, Chi-Quadrado e Anderson-Darling (NAGHETTINI; PINTO, 2007). Esses testes, chamados de testes de aderência, baseiam-se no conceito dos testes de hipóteses, avaliando se a amostra é ou não retirada de uma distribuição específica. Entretanto, possuem um procedimento de cálculo mais complexo e específico. Além disso, avaliam apenas a adequabilidade das distribuições de probabilidade, não sendo utilizadas para comparar desempenho entre elas (NAGHETTINI; PINTO, 2007). Portanto, neste estudo, optou-se por não utilizar os testes de aderência na avaliação do desempenho descritivo dos modelos. Alternativamente, seis estatísticas de uso corrente na avaliação da qualidade de modelos hidrológicos foram utilizadas: raiz do erro quadrado médio (RMSE), coeficiente de correlação (CC), Critério de Informação de Akaike Modificado (AICc), Eficiência de Kling-Gupta (KGE), Critério de Informação Bayesiano (BIC) e o erro absoluto médio (MAE). Foram utilizados mais de um teste, pelo fato de não haver uma única métrica que consiga avaliar o desempenho de um modelo por completo. Cada uma delas possui sua particularidade (JACKSON et al., 2019).

O RMSE (Equação 2) é uma métrica comumente utilizada na avaliação de modelos meteorológicos e climáticos, apresentando os erros na mesma unidade dos dados observados (CHAI; DRAXLER, 2014). Avalia a diferença entre os valores simulados e observados, atribuindo maior relevância a maiores erros absolutos. Consequentemente, o RMSE é influenciado pela presença de *outliers* nos valores observados (NGUYEN et al., 2017).

$$RMSE = \left(\sum \frac{(x_i - y_i)^2}{n} \right)^{1/2} \quad (2)$$

onde $(x_i - y_i)$, $i = 1, 2, \dots, n$ representam os erros, calculados pela diferença entre os valores observados (x_i) e os respectivos valores estimados (y_i). Menores valores do RMSE indicam melhores desempenhos. O CC (Equação 3) é um coeficiente estatístico que avalia o grau e a direção da correlação entre duas variáveis. Valores de CC com módulo próximo à unidade, indicam que há uma relação linear muito forte - positiva ou negativa - entre as variáveis comparadas.

$$CC = \frac{\sum[(x_i - \bar{x}) \cdot (y_i - \bar{y})]}{\sqrt{\sum(x_i - \bar{x})^2 \cdot \sum(y_i - \bar{y})^2}} \quad (3)$$

sendo que x_i e y_i , $i = 1, 2, \dots, n$ representam os valores observados e estimados respectivamente, e $\bar{x}$ e $\bar{y}$ representam suas médias. O AICc é uma correção do Critério de Informação de Akaike (AIC), que adapta seu uso para amostras pequenas, uma vez que o teste original é recomendado para amostras com $n > 40$ (BURNHAM; ANDERSON, 2004). O AIC computa a quantidade de informação perdida no processo de modelagem, calculando-a com base na maximização das funções de log-verossimilhança. Além disso, o AIC avalia a complexidade do modelo, adicionando um termo de penalidade calculado com base no número de parâmetros do modelo (SADEGH; RAGNO; AGHAKOUCHAK, 2017). De acordo com Nguyen et al., (2017) e Liu et al., (2020), outra maneira de avaliar o AIC é através da utilização da soma quadrática dos resíduos (SQR), de acordo com o conjunto de Equações 4. O critério de informação Bayesiano (BIC) é semelhante ao AIC. Seu cálculo também depende da maximização das funções de log-verossimilhança e ele também penaliza modelos com maior número de parâmetros (Equação 4).

$$\begin{aligned} SQR &= \sum (x_i - y_i)^2 \\ AIC &= 2 \cdot k + n \cdot \ln(SQR) \\ AICc &= AIC + \frac{2k(k+1)}{n-k-1} \\ BIC &= n \cdot \ln(SQR) + \ln(n) \cdot k \end{aligned} \quad \dots(4)$$

sendo que n denota o tamanho da amostra e k , o número de parâmetros da distribuição analisada.

A eficiência de Kling-Gupta (KGE), definida por Gupta et al. (2009), foi introduzida como uma melhoria da eficiência de Nash-Sutcliffe (NSE) - amplamente utilizada em análises hidrológicas (NASH; SUTCLIFFE, 1970). Sua base está alinhada com a ideia de usar múltiplos objetivos nos processos de modelagem e calibração, reduzindo incertezas e garantindo previsões mais confiáveis, ao serem considerados diferentes aspectos de erros presentes na modelagem (POOL; VIS; SEIBERT, 2018). Pode-se definir a KGE, de acordo com o conjunto de Equações 5, representando uma métrica de três componentes: viés (β), correlação (r) e variabilidade relativa (α).

$$\begin{aligned}
KGE &= 1 - ED \\
ED &= \sqrt{(r - 1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \\
r &= CV_y / CV_x \\
\alpha &= \sigma_y / \sigma_x \\
\beta &= \bar{y} / \bar{x}
\end{aligned}
\tag{5}$$

onde $\bar{x}$ e $\bar{y}$ representam as médias e σ_x e σ_y os desvios padrões, com os índices x e y representando os valores observados e simulados, respectivamente. CV representa o coeficiente de variação.

O MAE (Equação 6), amplamente utilizado na avaliação de desempenho de modelos, atribui o mesmo efeito para os erros entre os valores observados e simulados pelo modelo, diferentemente do RMSE, que atribui efeito diferente aos maiores erros (CHAI; DRAXLER, 2014).

$$MAE = \frac{\sum |x_i - y_i|}{n} \tag{6}$$

sendo que $|x_i - y_i|$, $i = 1, 2, \dots, n$ representam os erros absolutos, calculados pelo módulo da diferença entre os valores observados e os respectivos valores estimados e n , o tamanho da amostra.

2.2.2 Testes preditivos

Os testes preditivos buscam avaliar o desempenho dos modelos quando utilizados para prever informações sobre alguns eventos localizados na faixa de extrapolação. São essenciais no processo de escolha de distribuições de probabilidade por avaliarem propriedades complementares àquelas avaliadas pelos testes descritivos. Neste trabalho, foi proposta uma metodologia com base na técnica descrita por Nguyen et al. (2017), que se baseia na reamostragem das séries observadas pelo método *bootstrap* de reamostragem com reposição (EFRON; TIBSHIRANI, 1986), para simular os maiores eventos de precipitação observados (Figura 2).

A proposta baseia-se na geração de amostras sintéticas com tamanho amostral igual a metade do tamanho da amostra original. As distribuições de probabilidade são, então,

modeladas considerando essas amostras sintéticas reduzidas, com o intuito de estimar todos os eventos observados na amostra original, em um esquema de validação cruzada. Com isso, calcula-se as seis estatísticas de erro (RMSE, CC, AICc, KGE, BIC e MAE), comparando os dados simulados por meio das distribuições e das séries sintéticas reduzidas e os dados observados, presentes na série original. Esse procedimento é repetido mil vezes, permitindo uma análise visual da distribuição das estatísticas calculadas em cada uma das repetições por meio de diagramas box-plot. Na Figura 3, são indicadas as etapas de determinação dos testes descritivos e preditivos, considerando N como o número de distribuições de probabilidade avaliadas.

Figura 2 - Etapas do processo de realização dos testes preditivos.

Fonte: Autor (2021).

Figura 3 – (a) Processo de determinação dos testes descritivos e (b) Processo de determinação dos testes preditivos. As letras P, t, f, s, e N, nos eixos dos gráficos, referem-se a altura de precipitação, tempo, frequência, valor da estatística de estudo e modelos avaliados, respectivamente.

Fonte: Autor (2021).

Essa prática possibilita a avaliação da capacidade preditiva dos modelos pelo fato deles serem utilizados para estimar eventos presentes na faixa de extrapolação, com período de retorno maior do que o tamanho amostral utilizado na modelagem, visto que se utiliza uma série reduzida para caracterizar todos eventos presentes na série completa. Pelo fato de a análise visual dos diagramas ser subjetiva e de difícil sistematização e comparação, optou-se por avaliar numericamente o desempenho das distribuições candidatas. Para isso, foram utilizados os valores médios e os limites interquartil dos modelos avaliados, obtidos no diagrama box-plot a partir das 1000 repetições, para ranqueá-los.

Para as métricas em que o maior valor representa o melhor desempenho (CC e KGE), utilizamos a Equação 8 para pontuar cada um dos modelos. Melhores modelos apresentam maiores valores de Z_{final} . Para as métricas cujos menores valores são associados a melhores desempenhos (RMSE, AICc, BIC e MAE), utilizamos a Equação 9. Nesse caso, melhores modelos apresentam menores valores de X_{final} . Por fim, ranqueamos os modelos candidatos com base nas pontuações preditivas finais (X_{final} e Z_{final}). Na Figura 4, encontra-se ilustrada a proposta de comparação de desempenhos preditivos dos modelos. Essas equações

garantem a escolha do modelo com valor médio mais próximo ao ótimo, penalizando-o pela variabilidade de desempenhos indicada pela amplitude dos diagramas *box-plot* nas 1000 repetições. Finalmente,

$$Z_{final} = \bar{Z}_l + Z_{25p} + Z_{75p} \quad (8)$$

$$X_{final} = \bar{X}_l + X_{25p} + X_{75p} \quad (9)$$

sendo Z_{final} , $\bar{Z}_l$, Z_{25p} e Z_{75p} a pontuação preditiva final, valor médio e percentis 25 e 75 da métrica avaliada, obtidos através do diagrama *box-plot*. O mesmo vale para as métricas do tipo X.

Figura 4 – Proposta de avaliação dos desempenhos preditivos dos modelos. (a) Modelos cujo melhor desempenho é expresso por menores valores das métricas (X). (b) Modelos cujo melhor desempenho é expresso por maiores valores das métricas (Z). (c) Esquema do diagrama *box-plot* das métricas X.

Fonte: Autor (2021).

2.2.3 Critério de escolha da distribuição de probabilidade proposto

Para classificar os modelos de probabilidade quanto ao seu desempenho nos testes descritivos e preditivos, foi proposta uma metodologia de pontuação dos modelos candidatos, com base no ranqueamento dos modelos de acordo com os testes avaliados. Para cada um dos testes (preditivos e descritivos), atribuiu-se uma pontuação para as distribuições candidatas relativa à sua posição no ranking de desempenho. Sintetizando a performance das distribuições, calcula-se a pontuação média obtida nos testes descritivos e preditivos (P_D e P_P , respectivamente). Por fim, obtém-se a pontuação final (P_F) a partir da média de P_D e P_P (Equação 10). Comparando as pontuações finais obtidas por cada uma das distribuições candidatas, pode-se optar pela melhor distribuição, considerando que menores valores de P_F representam melhores desempenhos.

$$P_F = \left(\frac{P_D + P_P}{2} \right) \quad (10)$$

2.3 Estudo de caso

2.3.1 Área de Estudo

Com o intuito de aplicar a metodologia desenvolvida para escolha de distribuições de probabilidade foi conduzido um estudo de caso com os dados de precipitação do estado de São Paulo – Brasil, acessados no banco de dados do Departamento de Água e Energia Elétrica (disponível em <http://www.hidrologia.daee.sp.gov.br/>). Consideramos somente as séries que tivessem, ao menos, 30 anos válidos de dados para construção das séries de máxima precipitação diária anual, totalizando 699 estações avaliadas. Informações complementares sobre as estações avaliadas encontram-se no Apêndice A. Um ano foi considerado válido se: (1) não possuísse nenhuma falha (dias sem dados) nos meses chuvosos (Outubro – Março) e (2) se o número total de falhas, no ano inteiro, fosse menor ou igual a 3. Na Figura 5, são ilustradas a distribuição espacial das séries selecionadas, o histograma do tamanho das séries e a série temporal de anos válidos.

Figura 5 – Área de estudo e estações meteorológicas utilizadas: (a) América do Sul, com o Brasil e o estado de São Paulo destacados em cinza claro e escuro, respectivamente. (b) Distribuição das 699 estações meteorológicas utilizadas no estudo. (c) Número de estações com dados válidos em cada um dos anos avaliados. (d) Tamanho das séries presentes no banco de dados. Foram removidas 3 estações, com extensa série amostral (com dados a partir de 1890) dos gráficos (c) e (d), para facilitar a visualização.

Fonte: Autor (2021).

2.3.2 Distribuições de probabilidade candidatas

Foram utilizadas 7 distribuições de probabilidade, comumente aplicadas na análise de eventos hidrológicos extremos em diferentes países: Generalizada de Valores Extremos (GEV), Gumbel (GUM), Pearson III (PE3), Log Pearson III (LP3), Log Normal a 2 parâmetros (LNO), Generalizada Logística (GLO) e Burr Tipo 3 (BUR). As distribuições GEV e GUM são da família da distribuição assintótica de valores extremos. A primeira, introduzida por Jenkinson (1955), incorpora o tipo 1 (GUM), tipo 2 (Fréchet) e tipo 3 (Weibull), em uma única expressão (CHENG et al., 2014). Já a segunda, é uma forma específica da primeira, com a redução do número de parâmetros ao assumir o parâmetro de forma igual a 0. Ambas distribuições são recomendadas em diferentes países, como Espanha, Áustria e Alemanha para a distribuição GEV e Finlândia e Espanha, para GUM, que é também comumente utilizada para caracterizar eventos hidrológicos extremos no Brasil (SALINAS et al., 2014b).

A distribuição PE3, também chamada de Gama a 3 parâmetros, é um caso especial da distribuição de probabilidade de Gama, onde é assumido parâmetro de posição não nulo (ALAM et al., 2018; NAGHETTINI; PINTO, 2007). Já a distribuição LP3, é utilizada caso a transformada logarítmica da variável aleatória (precipitação) seja distribuída de acordo com PE3. Ambas distribuições são também amplamente utilizadas em estudos hidrológicos (KAO; GOVINDARAJU, 2007; LIU et al., 2020; OKOLI et al., 2019b), sendo recomendadas como método base para a análise de frequência nos Estados Unidos (NGUYEN et al., 2017).

Analogamente ao caso da distribuição LP3, diz-se que uma variável segue uma distribuição LNO caso sua transformada logarítmica seja normalmente distribuída. Poucas variáveis hidrológicas podem ser representadas pela distribuição normal, por apresentarem, usualmente, um certo grau de assimetria. Entretanto, quando a transformada logarítmica é aplicada, esse efeito é reduzido, podendo-se representar os dados através da LNO (SANGAL; BISWAS, 1970). A distribuição GLO, introduzida por Hosking e Wallis (1997) é baseada na distribuição logística, assumindo parâmetro de forma não nulo. Apresenta comportamento semelhante à GEV para modelar eventos extremos presentes na cauda superior da distribuição. Por fim, avaliamos a distribuição BUR, utilizada por Papalexiou (2018) e Zaghloul et al. (2020) como uma alternativa às distribuições de valores extremos para caracterização de eventos hidrológicos extremos. Embora não se caracterize como uma distribuição de valores extremos, a BUR consegue descrever bem a cauda superior, apresentando comportamento assintótico semelhante à distribuição GEV. Maiores detalhes das distribuições de probabilidade avaliadas e seus respectivos parâmetros são apresentados na Tabela 1.

Tabela 1 - Distribuições de probabilidade avaliadas (funções densidade de probabilidade e parâmetros).

Distribuição	Função densidade de probabilidade (f) ou função acumulada de probabilidade (F)	Parâmetros
GEV	$f_x(x) = \frac{1}{\alpha} \left[1 - \kappa \left(\frac{x-\beta}{\alpha} \right) \right]^{1/\kappa-1} \exp \left\{ - \left[1 - \kappa \left(\frac{x-\beta}{\alpha} \right) \right]^{\frac{1}{\kappa}} \right\}$	α : escala β : posição κ : forma
GUM	$f_x(x) = \frac{1}{\alpha} \exp \left[\frac{x-\beta}{\alpha} - \exp \left(\frac{x-\beta}{\alpha} \right) \right]$	α : escala β : posição
PE3	$f_x(x) = \frac{1}{\alpha \cdot \Gamma(\beta)} \left(\frac{x-\gamma}{\alpha} \right)^{\beta-1} \exp \left(-\frac{x-\gamma}{\alpha} \right)$	α : escala γ : posição β : forma
LP3¹	$f_x(x) = \frac{1}{\alpha \cdot \Gamma(\beta)} \left(\frac{y-\gamma}{\alpha} \right)^{\beta-1} \exp \left(-\frac{y-\gamma}{\alpha} \right)$	α : escala γ : posição β : forma
LNO¹	$f_x(x) = \frac{1}{x \cdot \sigma \sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{y-\mu}{\sigma} \right)^2 \right]$	μ : posição σ : escala
GLO	$f_x(x) = \frac{\alpha^{-1} \exp \left[-(1-\kappa) \cdot \kappa^{-1} \cdot \log \left(1 - \frac{\kappa(x-\theta)}{\alpha} \right) \right]}{\left\{ 1 + \exp \left[-\kappa^{-1} \cdot \log \left(1 - \frac{\kappa(x-\theta)}{\alpha} \right) \right] \right\}^2}$	θ : posição α : escala κ : forma
BUR	$F_x(x) = \left(1 + \frac{1}{\kappa_1} \cdot \left(\frac{x}{\alpha} \right)^{\frac{-1}{\kappa_2}} \right)^{-\kappa_1 \cdot \kappa_2}$	α : escala κ_1 : forma κ_2 : forma

¹ y representa a transformada logarítmica da variável aleatória x; ($y = \ln x$)

Equações obtidas de Nguyen et al. (2017), Ye et al. (2018) e Zaghloul et al. (2020)

Fonte: Autor (2021).

2.3.3 Estimativa dos parâmetros das distribuições de probabilidade

De acordo com Koutsoyiannis (2008), os métodos clássicos de estimativa de parâmetros de funções de distribuição na estatística são: o método de máxima verossimilhança (MV), o método dos momentos (MM) e o método dos momentos-L (MM-L). O MV, embora apresente boa eficiência, pode produzir estimadores de qualidade inferior aos outros métodos para amostras pequenas, gerando parâmetros com valores absurdos em algumas situações (KATZ; PARLANGE; NAVEAU, 2002; MARTINS; STEDINGER, 2000). O MM-L apresenta eficiência semelhante ao MV, entretanto, exige menor esforço computacional e é menos sensível a variabilidade de amostragem, erros de medição e a presença de *outliers* (HOSKING, 1990). Sendo assim, neste estudo, optou-se pelo MM-L para a estimativa dos parâmetros das

distribuições de probabilidade avaliadas, com exceção da distribuição BUR, cujos parâmetros foram estimados de acordo com o método MV, devido à inexistência de bibliotecas *Python* (utilizado em todas as análises estatísticas) que façam a estimativa via MM-L para essa distribuição.

2.4 RESULTADOS E DISCUSSÃO

2.4.1 Desempenho dos modelos avaliados

A distribuição das pontuações (*ranking*) obtidas por cada um dos modelos avaliados é apresentada na Figura 6. De maneira geral, os resultados indicaram que a avaliação do desempenho dos modelos depende da métrica escolhida. Isto é, bons desempenhos obtidos pelos modelos em uma determinada métrica não foram necessariamente reproduzidos quando analisadas métricas diferentes. Confirma-se, dessa maneira, a necessidade de serem considerados diferentes testes na escolha de distribuições de probabilidade, pelo fato de avaliarem diferentes aspectos e, conseqüentemente, selecionarem diferentes modelos. A combinação de métricas diferentes busca selecionar modelos que possuam um bom desempenho geral. Em contrapartida, a utilização de uma única métrica seleciona modelos com base em um único aspecto.

Os modelos apresentaram, ainda, uma grande diferença entre suas capacidades descritivas e preditivas. Como casos extremos, pode-se destacar as distribuições GUM e PE3. A primeira apresentou um bom desempenho preditivo em todas as métricas avaliadas, mas um baixo desempenho descritivo. A situação contrária ocorre para a distribuição PE3, que apresentou, em geral, um bom desempenho descritivo e um baixo desempenho preditivo em todas as métricas avaliadas. Os resultados confirmam a necessidade de contabilizar a capacidade preditiva dos modelos, em adição à análise descritiva. Modelos podem apresentar uma boa capacidade de reprodução da série observada (capacidade descritiva), mas não serem adequados para caracterizar eventos na faixa de extrapolação (capacidade preditiva).

Esses resultados são confirmados pelos gráficos de frequência acumulada dos desempenhos dos modelos nos diferentes testes avaliados (Apêndice B). Com exceção do modelo GEV, que apresentou uma menor variabilidade entre os desempenhos nos diferentes testes avaliados, os demais modelos apresentaram grande variabilidade tanto entre métricas como entre os tipos de testes avaliados (descritivos e preditivos). Considerando os testes preditivos, a distribuição de GUM apresentou o melhor desempenho em todas as métricas avaliadas, ficando entre as 2 melhores distribuições de probabilidade em aproximadamente

50% das estações. Para os testes descritivos, os modelos apresentaram desempenho mais similares. Em geral, as distribuições GEV, LP3, BUR e GLO apresentaram melhores desempenhos descritivos. A distribuição LNO se destacou negativamente, apresentando um desempenho descritivo inferior em relação aos demais modelos em todas as métricas avaliadas.

Figura 6 – Distribuição da pontuação (*ranking*) obtida por cada modelo nas 699 estações consideradas no estudo de caso, considerando as diferentes métricas descritivas e preditivas avaliadas.

Fonte: Autor (2021).

Os desempenhos dos modelos nos testes avaliados foram sumarizados nas pontuações descritivas, preditivas e finais, de acordo com a metodologia proposta (Figura 7). Essa abordagem visa sintetizar as diferentes análises de desempenhos no processo de escolha de

modelos. É importante ressaltar que, embora a metodologia proposta considere o desempenho dos modelos em diferentes testes, pode haver situações em que duas ou mais distribuições de probabilidade sejam indicadas como a mais adequada, por terem obtido pontuações finais iguais.

Dentre os modelos avaliados, GUM e BUR destacaram-se como os mais indicados em um maior número de estações da região, sendo escolhidos em 160 e 123 estações, respectivamente. Ambos modelos apresentaram, em geral, um bom desempenho preditivo. A GUM, entretanto, apresentou o pior desempenho descritivo dentre os modelos avaliados, mesmo sendo a distribuição mais escolhida de acordo com a pontuação final. Esses resultados sugerem que a capacidade preditiva do modelo é um diferencial para sua escolha de acordo com a metodologia proposta. Isso se dá, provavelmente, pelo fato de todos os modelos apresentarem desempenhos descritivos semelhantes nas diferentes estatísticas avaliadas, sendo, portanto, o desempenho preditivo, essencial para a escolha de acordo com a pontuação final.

Figura 7 – Box plot: distribuição das pontuações preditiva, descritiva e final para cada modelo avaliado. Pontuação igual a 1 indica o modelo com melhor desempenho. Linhas central, inferior (da caixa) e superior (da caixa) representam a mediana e percentil 25 e 75, respectivamente. O triângulo central (cinza) indica a pontuação média. Tabela: número de estações onde cada um dos modelos foi escolhido como o mais adequado, de acordo com as pontuações descritiva, preditiva e final.

Fonte: Autor (2021).

Para avaliar a disparidade entre os modelos selecionados através das pontuações descritivas e preditivas, aplicou-se o coeficiente de concordância de Cohen Kappa (CK), que tem como objetivo quantificar a similaridade entre dois ou mais critérios de escolha (COHEN, 1960). Valores de CK próximo a 1 revelam uma alta concordância entre os critérios, enquanto que valores próximos a 0 indicam que a concordância entre os critérios é próxima àquela esperada pelo acaso. No caso, quando comparadas ambas pontuações (descritiva e preditiva) como critério de seleção, obteve-se um CK de 0,125, indicando uma baixíssima concordância. Esse resultado confirma a necessidade de considerar ambos critérios no processo de escolha de distribuições de probabilidade, visto que avaliam propriedades diferentes e essenciais em seus critérios de escolha.

Considerando o desempenho das distribuições a partir de um ponto de vista regional, a distribuição GEV se destacou como a mais indicada. O modelo apresentou uma pequena amplitude interquartil (IQR) em ambas características avaliadas, e pontuações médias mais próximas do melhor desempenho (*ranking* próximo à 1). Quando consideramos o desempenho dos modelos por um ponto de vista pontual, isto é, considerando o número de estações em que a distribuição foi indicada como a mais adequada, a GEV não se destacou, pelo fato de outros modelos, como BUR e GUM, terem apresentado um ótimo desempenho em um maior número de estações. Isto é, foram selecionadas como as distribuições mais adequadas com maior frequência. Entretanto, essas distribuições apresentaram maior variabilidade em seus desempenhos, representada pelos maiores IQR e grandes amplitudes entre os limites inferior e superior, impossibilitando a indicação dessas distribuições como adequadas regionalmente, pelo baixo desempenho apresentado em algumas estações.

Resultados semelhantes foram encontrados por Nguyen et al., (2017), em um estudo de caso aplicado na região de Ontario (Canadá), que avaliou diferentes métricas de seleção de distribuições de probabilidade para descrever eventos extremos de precipitação, como a RMSE, MAE, AICc, CC, entre outras. Seus resultados indicaram uma grande variabilidade entre as métricas descritivas na seleção de modelos mais adequados. Essas diferenças também foram encontradas quando comparados os desempenhos descritivos e preditivos dos modelos. Isto é, de forma geral, modelos que apresentaram bom desempenho descritivo não necessariamente obtiveram um bom desempenho preditivo. No entanto, o estudo não resumiu ambas perspectivas preditivas e descritivas em uma pontuação final, conforme a metodologia proposta, auxiliando o processo de seleção de modelos a partir da comparação de diferentes distribuições candidatas. A proposta de pontuação permitiu ainda a avaliação do desempenho dos modelos a

partir de uma perspectiva regional, através de uma visão mais abrangente dos desempenhos dos modelos. Além disso, o acesso às capacidades preditivas do estudo foi feito de acordo com uma análise visual, considerando apenas a capacidade do modelo em descrever os maiores eventos observados, diferentemente da proposta apresentada, que avalia a capacidade preditiva de maneira sistemática, considerando diversas estatísticas para avaliar o desempenho dos modelos em caracterizar todos os eventos observados.

A distribuição espacial da adequabilidade dos modelos de probabilidade (Figura 8), selecionados a partir da metodologia proposta, indicou a ausência de concentração espacial dos modelos no estado de São Paulo. Mesmo considerando escalas sub-regionais, é difícil definir locais onde há a predominância de algum modelo sobre os demais, devido à grande dispersão espacial de suas adequabilidades.

Conforme mencionado, diferentes aspectos podem afetar o processo de caracterização, como a disponibilidade de dados e a infinidade de métricas de desempenho. Mesmo considerando diferentes testes descritivos e preditivos, a proposta de pontuação pode selecionar modelos diferentes para estações vizinhas, que apresentam comportamento de eventos de precipitação extremamente semelhantes, pelo fato de não conseguir captar todos os fatores que podem interferir no processo de seleção. Além disso, em muitos casos, diferentes modelos podem apresentar desempenho extremamente satisfatório, sendo o processo de seleção definido por pequenos detalhes nas métricas avaliadas. Nestes casos, a análise regional proposta neste estudo aparece como uma alternativa para o processo de seleção, indicando um modelo capaz de caracterizar bem todos os eventos extremos de precipitação de uma certa região, pelo fato de apresentar menor variabilidade e um bom desempenho médio geral a partir de um ponto de vista regional. Além disso, a abordagem regional é menos sensível às variações na escolha de distribuições definidas por pequenos detalhes, visto que considera um desempenho geral para a seleção do modelo mais adequado.

Figura 8 – Distribuição espacial das distribuições de probabilidade escolhidas pela metodologia proposta no estado de São Paulo.

Fonte: Autor (2021).

Em suma, pode-se destacar os modelos GEV, BUR e GUM como os mais adequados para caracterizar as chuvas intensas da região. O primeiro, por ter apresentado um bom desempenho geral, é o mais indicado para ser usado para região. Embora não tenha sido selecionado com tanta frequência, apresentou desempenho próximo do melhor modelo na grande maioria das estações. Os dois últimos (GUM e BUR) se destacaram pelo ótimo desempenho pontual. Isto é, têm um grande potencial para caracterizar os eventos da região. Entretanto, pela maior variabilidade encontrada em seus desempenhos, deve-se atentar para suas utilizações sem um correto procedimento de análise.

Com o intuito de avaliar a validade dos resultados obtidos, foram comparados os quantis de precipitação estimados pela metodologia proposta, pontualmente para cada estação, com os quantis estimados considerando um único modelo regionalmente adequado (Figura 9). Para isso, as distribuições GEV e GUM foram utilizadas como modelos adequados para a região. A primeira por ter obtido um bom desempenho de acordo com a perspectiva regional e a segunda, pelo melhor desempenho pontual, escolhida como a mais adequada em um grande número de estações (Figura 6). Foram considerados os períodos de retorno de 10 e 100 anos.

A distribuição GEV apresentou bom desempenho em ambos períodos de retorno avaliados. Para todo o estado, foram encontradas diferenças médias de ± 2 mm entre os quantis estimados pela GEV e pela metodologia proposta, em relação ao período de retorno de 10 anos. Para o período de retorno de 100 anos, as diferenças encontradas ficaram na faixa de aproximadamente ± 8 mm. Sendo assim, mesmo que com algumas diferenças pontuais maiores, a distribuição GEV apresentou um bom desempenho geral para todo estado, podendo ser considerada uma distribuição adequada para caracterizar os eventos de precipitação da região.

A distribuição GUM apresentou comportamento semelhante para os quantis associados ao período de retorno de 10 anos, com diferenças de aproximadamente ± 2 mm. Entretanto, para períodos de retorno mais altos, a GUM apresentou maiores diferenças em relação aos quantis estimados pela distribuição mais indicada de acordo com a pontuação final, superestimando os eventos de precipitação em grande parte do estado, com diferenças maiores que 16 mm em muitas estações. Dessa maneira, embora a distribuição GUM seja amplamente utilizada, deve-se atentar para seu uso sem estudos prévios sobre a adequabilidade da distribuição, pois há a possibilidade de superestimativa dos quantis. Esse fato, somado à ótima performance da distribuição GEV, confirma que a abordagem regional supera a abordagem ‘pontual’ para selecionar um único modelo de probabilidade para uma região inteira, especialmente quando há interesse em caracterizar eventos com alto período de retorno.

Figura 9 – Comparação dos quantis extremos de precipitação estimados pela metodologia proposta com aqueles estimados a partir de uma única distribuição de probabilidade adequada para o estado inteiro. Esquerda: diferenças dos quantis associados aos períodos de retorno de 10 (a) e 100 anos (c) usando a distribuição GEV para todo o estado. Direita: diferenças entre os quantis associados aos períodos de retorno de 10 anos (b) e 100 anos (d) utilizando a distribuição GUM para todo o estado. Valores negativos representam subestimativas e valores positivos, superestimativas.

Fonte: Autor (2021).

2.4.2 Caracterização das chuvas intensas na área de estudo

A partir dos modelos definidos para cada estação, foram estimados os quantis de precipitação associados aos períodos de retorno de 10, 25, 50 e 100 anos. As estatísticas descritivas são apresentadas na Tabela 2. O desvio padrão, entendido como uma representação da variabilidade da precipitação na região, aumenta de acordo com o período de retorno, devido, principalmente, às maiores incertezas associadas à estimativa desses eventos. Observa-se

também uma grande amplitude dos quantis estimados para um mesmo período de retorno, representando a heterogeneidade da região, onde, pontualmente, eventos com período de 10 anos chegam a superar, em intensidade, eventos com período de retorno de 100 anos de outro local na região.

Tabela 2 - Estatísticas descritivas dos quantis de precipitação estimados para o estado de SP.

Período de retorno - Tr (anos)	Média (mm)	Desvio Padrão (mm)	CV (%)	Mínimo (mm)	Máximo (mm)
10	117,63	31,04	26%	83,40	340,54
25	139,69	41,91	30%	92,84	393,29
50	157,86	52,05	33%	98,86	459,75
100	177,80	64,77	36%	104,13	573,02

Fonte: Autor (2021).

As estimativas realizadas foram interpoladas espacialmente, pelo método do inverso do quadrado da distância (IDW), que apresenta geralmente um bom desempenho para o mapeamento de eventos de precipitação (GARDIMAN JUNIOR et al., 2012; VIOLA et al., 2010). Para mais detalhes sobre metodologias de interpolação espacial comumente utilizadas em eventos de precipitação, sugere-se consultar Ly, Charles e Degré (2011). A distribuição espacial dos quantis de precipitação associados aos períodos de retorno de 10 e 25 anos (Figura 10) e 50 e 100 anos (Figura 11), sugerem uma concentração dos maiores eventos na região litorânea, como esperado. Nas demais regiões do estado, foram encontrados comportamento semelhante dos quantis de precipitação, para todos os períodos de retorno avaliados, apresentando pequena variação espacial. Nesses locais, pode-se recomendar a utilização de séries mais longas, localizadas em cidades próximas ao local de interesse, para a caracterização dos eventos de precipitação, visto que a região apresenta, no geral, baixa variabilidade em todos os períodos de retorno avaliados.

Figura 10 - Estimativa dos eventos extremos de precipitação do estado de SP associados ao período de retorno de 10 anos (a) e 25 anos (b).

Fonte: Autor (2021).

Figura 11 - Estimativa dos eventos extremos de precipitação do estado de SP associados ao período de retorno de 50 anos (a) e 100 anos (b).

Fonte: Autor (2021).

2.5 CONCLUSÃO

Neste trabalho, buscou-se propor um método de escolha de distribuições de probabilidade, a partir de diferentes critérios de seleção, que avaliam tanto a capacidade descritiva quanto a preditiva de cada modelo. Foi proposta uma associação desses diferentes critérios com o intuito de otimizar a seleção do modelo mais adequado. Além disso, determinaram-se as diferenças encontradas nos quantis de precipitação estimados ao serem comparadas as estimativas feitas por modelos regionalmente selecionados ou por modelos selecionados pontualmente para cada estação.

O método de escolha proposto e as comparações de desempenho das diferentes distribuições foram avaliados em um estudo de caso, com os dados de 699 estações pluviométricas localizadas no estado de São Paulo. A distribuição espacial dos modelos de probabilidade selecionados para cada estação não apresentou um padrão, confirmando a necessidade de serem investigados diferentes modelos de probabilidade para a estação de interesse, sempre que possível. Dentre as 7 distribuições de probabilidade avaliadas, GEV apresentou um bom desempenho geral, permanecendo entre as 3 melhores distribuições na grande maioria dos casos. Avaliando o desempenho pontualmente para cada estação, as distribuições GUM e BUR se destacaram, sendo escolhidas em 160 e 123 estações, respectivamente. Entretanto, apresentaram uma maior variabilidade de desempenho na região de estudo, sendo necessário, portanto, se atentar ao uso dessas duas distribuições sem uma correta análise de adequabilidade dos modelos para descrever os eventos do local de interesse.

A consideração dos testes preditivos na escolha de distribuições de probabilidade acrescenta informações ao processo de caracterização de eventos extremos. Como visto, modelos com boa capacidade descritiva podem apresentar baixo desempenho preditivo, não sendo adequados para caracterizar eventos presentes na faixa de extrapolação. Além disso, a utilização de diferentes métricas na seleção dos modelos garante que modelos que apresentem bons desempenhos sejam escolhidos considerando diferentes aspectos. A sintetização das capacidades preditivas e descritivas em uma pontuação final facilita a comparação entre modelos, sem deixar de considerar as diferentes propriedades descritivas e preditivas. Ainda, a metodologia proposta permite uma melhor visualização da adequabilidade das distribuições de probabilidade em escala regional.

A comparação dos quantis de precipitação estimados pela metodologia proposta com os quantis estimados por uma única distribuição regional (GUM ou GEV) ressaltou o

problema da utilização de um modelo regional sem um estudo adequado. Popularmente utilizada, a distribuição de GUM apresentou grandes diferenças, em grande parte do estado, nas estimativas dos eventos associadas ao período de retorno de 100 anos. Já a distribuição GEV, apresentou um bom desempenho regional, com menores diferenças nos quantis estimados para os períodos de retorno de 10 e 100 anos.

Em suma, foi apresentada no trabalho uma análise sobre o processo de escolha de distribuições de probabilidade e suas influências nos quantis de precipitação estimados. O critério de escolha proposto pode ser considerado como uma alternativa a abordagem tradicional, que considera somente a capacidade preditiva do modelo, para uma caracterização mais precisa dos eventos de precipitação, permitindo considerar diferentes propriedades no processo de escolha. A avaliação espacial da adequabilidade dos modelos confirmou a ausência de padrões no estado, indicando que diferentes modelos devem ser avaliados para um aprimoramento do processo de caracterização. Caso seja necessário indicar um único modelo para todo o estado, uma análise regional, considerando a variabilidade do desempenho de diferentes modelos em toda região, pode ser mais adequada.

CAPÍTULO III

**Avaliação da representatividade dos dados observados e suas influências
nas estimativas de precipitação**

3.1 INTRODUÇÃO

A caracterização espacial e temporal de eventos extremos é indispensável para as análises hidrológicas (BRUNSELL, 2010). Tem como objetivo fundamental auxiliar o dimensionamento e planejamento hídrico com a previsão e caracterização de eventos futuros, protegendo a sociedade contra os impactos de eventos extremos, como secas e inundações (SALAS; OBEYSEKERA; VOGEL, 2018). Esse procedimento, comumente realizado por meio da análise estatística de frequências (NGUYEN; NGUYEN, 2019), assume que as características de eventos futuros se assemelham às características de eventos passados. Admite-se, nesse sentido, o conceito de estacionariedade, isto é, que não ocorrerá nenhuma mudança significativa nas propriedades estatísticas das variáveis observadas (LINS; COHN, 2011; RAGNO et al., 2018). Sendo assim, através de observações passadas do mesmo processo físico, consegue-se definir modelos com a capacidade de inferir propriedades de eventos futuros (LINS; COHN, 2011).

Entretanto, o atual contexto de mudanças climáticas com possíveis mudanças nos padrões dos eventos extremos (FISCHER; KNUTTI, 2015) exige uma nova abordagem de caracterização desses eventos. Nos últimos anos, pesquisadores buscaram verificar a tendência nas séries de precipitação, constatando a alteração no comportamento dos eventos extremos devido às mudanças climáticas (ASADIEH; KRAKAUER, 2015; CHENG; AGHAKOUCHAK, 2014; FISCHER; KNUTTI, 2016; MUKHERJEE et al., 2018). Ademais, eventos extremos de precipitação podem apresentar flutuações de longo prazo, decorrentes de diferentes padrões de circulação atmosférica, que muitas vezes não são captados pelas séries observadas por apresentarem tamanho amostral inferior a periodicidade dessas oscilações (MARANI; ZANETTI, 2015). Com isso, questiona-se a representatividade dos eventos passados nas propriedades de eventos presentes e futuros (MERZ; BLÖSCHL, 2008; MONTANARI; KOUTSOYIANNIS, 2014). Isto é, qual o impacto de considerar informações passadas, com características e regimes diferentes dos observados atualmente, como representativas para caracterizar eventos futuros?

Uma diferente abordagem do problema consiste em considerar modelos não estacionários na representação dos dados observados, assumindo que as características das variáveis que compõem o processo variam ao longo do tempo. Milly et al., (2008) ao afirmar que a “estacionariedade está morta”, sugeriram que as análises hidrológicas, afetadas direta e indiretamente por ações antrópicas, devem considerar as variáveis ambientais e suas evoluções temporais para além das séries históricas, através de modelos não-estacionários. No entanto,

esse procedimento exige uma melhor compreensão das mudanças temporais das variáveis envolvidas no fenômeno de estudo, definindo deterministicamente as relações entre os parâmetros dos modelos e suas covariáveis, como o tempo, população ou outras variáveis climáticas (SERINALDI; KILSBY, 2015).

Em situações onde essas relações não são bem compreendidas, sem uma função determinística que explique as mudanças observadas, a suposição de não estacionariedade está apenas adicionando outra fonte de incerteza à análise (KOUTSOYIANNIS; MONTANARI, 2015; RAGNO et al., 2019). Sendo assim, para os casos em que não seja possível definir as relações que regem a não estacionariedade, indica-se a adoção de modelos estacionários, por serem mais adequados e teoricamente consistentes, não subestimando riscos e incertezas presentes no processo (LUKE et al., 2017). Entretanto, na estacionariedade, o período de retorno associado a um determinado evento, considerado como uma informação vital no dimensionamento e na gestão de sistemas hidráulicos, é assumido como invariável ao longo do tempo (CHENG et al., 2014; RAGNO et al., 2018). Dessa maneira, as recentes mudanças climáticas e as flutuações de longo prazo associadas à padrões de oscilação atmosférica são, muitas vezes, desconsideradas. Além do mais, muitas vezes são considerados eventos com características e regimes muito diferentes dos eventos atuais no processo de modelagem, podendo apresentar grande impacto no processo de caracterização (MARANI; ZANETTI, 2015; VILLARINI et al., 2009).

Diante do exposto, investigou-se, nesse estudo, a representatividade de dados de precipitação históricos observados para inferir condições futuras, avaliando sua influência na caracterização de eventos extremos. Para isso, foi conduzido um estudo de caso para o estado de São Paulo, utilizando dados de 323 estações pluviométricas, com pelo menos 50 anos completos registrados. Com base na metodologia proposta por DeGaetano (2009), investigou-se as mudanças temporais nos quantis de precipitação associados aos períodos de retorno de 5 a 200 anos. Com base nisso, as consequências da não representatividade dos dados nas incertezas e precisão das estimativas foram avaliadas, por meio de uma análise de ponto de mudança associada a uma estrutura de simulação *bootstrap*. O principal objetivo deste estudo foi avaliar o procedimento de inferência estacionária em um contexto de mudanças climáticas, fornecendo suporte aos tomadores de decisões na avaliação das consequências das mudanças climáticas, sejam elas graduais ou abruptas, no processo de caracterização.

3.2 MATERIAIS E MÉTODOS

Nesta seção, apresenta-se a metodologia definida para avaliar o efeito das mudanças nas séries observadas no procedimento de caracterização dos eventos extremos de precipitação. O item 3.2.1 refere-se ao método utilizado para seleção dos dados e construção das séries temporais. No item 3.2.2, introduz-se a análise de tendência aplicada às séries de quantis de precipitação e, no item 3.2.3, é apresentada a análise da representatividade das séries temporais utilizadas nos procedimentos de caracterização.

3.2.1 Área de estudo e obtenção das séries de quantis de precipitação

Assim como no capítulo II, foram utilizados os dados de precipitação disponibilizados pelo Departamento de Água e Energia Elétrica em seu banco de dados (disponível em <http://www.hidrologia.dae.sp.gov.br/>). Para a construção das séries de máxima precipitação diária anual (AMS), selecionamos apenas os anos que não possuísem nenhuma falha nos meses chuvosos (Outubro – Março) e, no máximo, 3 falhas nos meses menos chuvosos (Abril – Setembro). Optou-se por selecionar, para o estudo, somente as estações cujas séries possuísem no mínimo 50 anos de dados completos, resultando em um total de 323 estações avaliadas (Figura 12). Para todas as estações, utilizamos a GEV para estimar os quantis de precipitação, por ter apresentado um bom desempenho na caracterização dos eventos da região (Capítulo II).

Figura 12 – Área de estudo e estações meteorológicas utilizadas: (a) Brasil, com o estado de São Paulo e América do Sul destacados em cinza escuro e claro, respectivamente. (b) Estações pluviométricas utilizadas no estudo. (c) Tamanho amostral das séries avaliadas. (d) Número de estações com dados válidos em cada um dos anos avaliados. Foram removidas 3 estações, com extensa série amostral (com dados a partir de 1890) dos gráficos (c) e (d), para facilitar a visualização.

Fonte: Autor (2021).

A partir de uma adaptação da metodologia descrita por DeGaetano (2009), utilizou-se a série anual de máxima precipitação diária (AMS) de cada estação para construir duas séries de quantis de precipitação associados a diferentes períodos de retorno (5, 10, 25, 50, 100 e 200 anos). A primeira (Figura 13), aqui chamada de série crescente (S_i), foi construída considerando quantis de precipitação estimados pela distribuição GEV modelada a partir de séries AMS com tamanho amostral crescente. O primeiro quantil da S_i (Q_1^i) foi estimado considerando a primeira metade da AMS original para modelar a distribuição GEV, enquanto que o último quantil foi estimado considerando a AMS completa (Q_M^i , sendo M o tamanho da S_i e igual à metade da série AMS original). Isto é, a série AMS utilizada para estimativa dos quantis da série S_i teve seu tamanho amostral gradualmente aumentado. A série $S_i = \{Q_1^i, Q_2^i, \dots, Q_M^i\}$ foi construída para avaliar a tendência dos quantis de precipitação considerando o acréscimo gradual de dados observados, simulando a situação real, em que, com o passar do tempo, a disponibilidade de dados observados aumenta.

A segunda série (Figura 13), denominada de série móvel (S_f), foi construída considerando séries AMS com tamanho fixo e igual à metade do tamanho da série completa para a modelagem da GEV. Nesse caso, o primeiro quantil da $S_f(Q_1^f)$ foi estimado utilizando a primeira metade da AMS original e o último (Q_M^f), considerando a segunda metade da AMS original. Ou seja, a AMS utilizada para a estimativa dos quantis da série S_f foi determinada com base em uma janela móvel, de tamanho fixo, que percorreu os dados de precipitação observados. A $S_f = \{Q_1^f, Q_2^f, \dots, Q_M^f\}$ não considera o efeito do acréscimo de dados observados na estimativa dos quantis, variando somente o período dos dados observados utilizados na modelagem. Ambas as séries (S_i e S_f) foram construídas para os períodos de retorno de 5, 10, 25, 50, 100 e 200 anos.

Figura 13 – Etapas para a construção das séries crescente - S_i (a) e móvel - S_f (b). Q_j^i e Q_j^f representam os quantis de precipitação estimados que compõem as séries crescente e móvel, respectivamente.

Fonte: Autor (2021).

3.2.2 Análise de tendência temporal

A avaliação da existência de tendências significativas nas séries crescente e móvel foi feita de acordo com o teste de Mann-Kendall modificado e a verificação de suas magnitudes com o estimador de Theil-Sen. Essa análise foi realizada com o intuito de identificar possíveis padrões de intensificação ou redução nos eventos extremos de precipitação associados a diferentes períodos de retorno. Assim como no Capítulo II, todas as análises estatísticas foram feitas com auxílio da linguagem de programação *Python*.

3.2.2.1 Existência de tendência - teste de Mann-Kendall

O teste de Mann-Kendall (MK), muito utilizado em análises de séries hidrológicas foi escolhido por ser não paramétrico e por ser pouco influenciado por *outliers* (BARTIKO et al., 2019; KUMAR et al., 2013; MORBIDELLI et al., 2018; VILLARINI et al., 2009, 2011a). A hipótese nula (H_0) assume que os dados da amostra são independentes e identicamente distribuídos. A hipótese alternativa (H_1) assume a presença de certa tendência monotônica na amostra. Considerando uma série temporal $(x_1, x_2, \dots, x_n)$, calcula-se o teste MK de acordo com o conjunto de Equações 11.

$$S = \sum_{j=1}^{n-1} \sum_{i=j+1}^n \text{sgn}(x_i - x_j)$$

$$\text{sgn}(x_i - x_j) = \begin{cases} 1, & \text{se } x_i > x_j \\ 0, & \text{se } x_i = x_j \\ -1, & \text{se } x_i < x_j \end{cases} \quad \dots(11)$$

$$\text{Var}(S) = \frac{[n(n-1)(2n-5)] - \sum_{i=1}^m t_i(t_i-1)(2t_i+5)}{18}$$

sendo n o número de observações presentes na amostra; x o quantil de precipitação; m , o número de grupos com observações iguais na série de dados e t_i , o número de observações com valores iguais dentro do grupo i . Com o intuito de considerar a influência de autocorrelação entre os dados observados, que aumentam a variância de S (YUE et al., 2002), adotou-se a modificação proposta por Hamed e Rao (1998) que considera um fator de correção para o cálculo da variância de S (Equações 12 e 13):

$$Var^*(S) = Var(S) \cdot \frac{n}{n_s} \quad (12)$$

$$\frac{n}{n_s} = 1 + \frac{2}{n(n-1)(n-2)} \cdot \sum_{i=1}^{n-1} (n-1)(n-i-1)(n-i-2)p_s(i) \quad (13)$$

sendo que n é o número de observações e $p_s(i)$ os coeficientes de autocorrelação referentes à classificação dos dados observados. Os resultados obtidos são utilizados para determinar o valor da estatística Z (Equação 14), que avalia a presença de tendência.

$$Z = \begin{cases} \frac{S-1}{\sqrt{Var^*(S)}}, & se S > 0 \\ 0, & se S = 0 \\ \frac{S-1}{\sqrt{Var^*(S)}}, & se S < 0 \end{cases} \quad (14)$$

A hipótese nula de não tendência é rejeitada caso o valor de Z seja maior que o valor crítico da estatística referente ao nível de significância. Por ser normalmente distribuída, os valores críticos da estatística Z podem ser obtidos em tabelas de distribuições normais. Valores positivos de Z , indicam tendência crescente e valores negativos indicam tendência decrescente.

3.2.2.2 Magnitude da tendência - estimador Theil-Sen

Segundo Yue et al. (2002), o estimador Theil-Sen (TS) apresenta estimativas robustas da magnitude da tendência existente em uma série temporal. Além disso, é pouco influenciado pela presença de *outliers* (SAYEMUZZAMAN; JHA, 2014). Sendo assim, neste trabalho, optou-se por utilizar o estimador TS (Equação 15) para calcular a intensidade das tendências significativas indicadas pelo teste MK.

$$b = Mediana\left(\frac{x_i - x_j}{t_i - t_j}\right) \quad (15)$$

sendo x_i e x_j ($i > j$) as observações de precipitação relativas às posições t_i e t_j da série temporal.

3.2.3 Análise da representatividade dos dados

3.2.3.1 Análise de ponto de mudança

A análise de ponto de mudança tem como objetivo avaliar se há, na série estudada, pontos de mudança abrupta de comportamento, separando a série em subséries com comportamento e distribuições diferentes. Dessa forma, os pontos de mudança são momentos de descontinuidade nos processos estocásticos (JEON; SUNG; CHUNG, 2016). Na hidrologia, diversos trabalhos avaliam a presença de pontos de mudança (DE LA CASA; NASELLO, 2010; JEON; SUNG; CHUNG, 2016; SAYEMUZZAMAN; JHA, 2014; VILLARINI et al., 2009, 2011a). Eles indicam que as mudanças abruptas são uma importante manifestação da variabilidade da série, podendo ser as principais responsáveis pelas mudanças encontradas nos parâmetros das distribuições de probabilidade (CHENG et al., 2014; XIE et al., 2019). De acordo com Luke et al. (2017) e Serinaldi e Kilsby (2015), as previsões feitas por modelos estacionários podem ser severamente afetadas pela presença de mudanças abruptas e/ou mudanças físicas conhecidas na série temporal. Sendo assim, sua verificação junto à análise de tendência da série é essencial para que sejam consideradas possíveis quebras de estacionariedade, bem como para auxiliar no entendimento dos efeitos das mudanças climáticas e atividades humanas no comportamento das variáveis climáticas (VILLARINI et al., 2011b; XIE et al., 2019).

As mudanças abruptas podem estar associados tanto a efeitos climáticos como antropogênicos (VILLARINI et al., 2011a). Para o caso dos eventos de precipitação, esses efeitos podem ser representados, por exemplo, pela presença de flutuações de longo prazo na série observada, alterações significativas no microclima ou, ainda, pela mudança no instrumento de medida. Geralmente, a série observada não possui um tamanho amostral grande o suficiente para identificar os diferentes padrões de oscilações de longo prazo e possíveis mudanças no regime climático. Dessa forma, ela acaba considerando apenas uma parcela dos dados, sem captar todas as mudanças de longo prazo no comportamento dos eventos extremos, podendo conter eventos de precipitação associados a diferentes regimes climáticos.

Para detectar a presença das mudanças abruptas nas séries de máxima precipitação (AMS), optou-se por utilizar, nesse estudo, o teste não paramétrico descrito por Pettitt (1979), que avalia pontos de mudança na média de uma série temporal quando sua exata posição na série é desconhecida. Esse método é baseado na estatística Mann-Whitney, que testa se duas amostras diferentes são derivadas da mesma população (GAO et al., 2011). Sua determinação é feita de acordo com as Equações 16 e 17.

$$U_{t,T} = \sum_{i=1}^t \sum_{j=t+1}^T \text{sgn}(x_i - x_j) \quad (16)$$

$$\text{sgn}(x_i - x_j) = \begin{cases} 1, & \text{se } x_i > x_j \\ 0, & \text{se } x_i = x_j \\ -1, & \text{se } x_i < x_j \end{cases} \quad (17)$$

A estatística $U_{t,T}$ é então calculada para cada valor de t , onde $1 \leq t \leq T$, sendo T o tamanho da série temporal, e x_i e x_j são os valores referentes às duas subamostras $x_i, \dots, x_t$ e $x_{t+1}, \dots, x_T$ da série. Basicamente, o teste de Pettitt avalia a hipótese nula de ausência de um ponto de mudança na série, através da estatística K_T (Equação 18), associada à uma probabilidade P (Equação 19), dadas por:

$$K_T = \max_{1 \leq t \leq T} |U_{t,T}| \quad (18)$$

$$P \cong 2 \exp \left[-\frac{6K_T^2}{T^3 + T^2} \right] \quad (19)$$

A avaliação do ponto de mudança foi aplicada para as estações estudadas com, no mínimo, 70 anos completos em sua série. As estações foram separadas em dois grupos, de acordo com a presença ou não de ponto de mudança significativo, utilizando o teste de Pettitt ao nível de significância de 5%. Assim como Villarini et al. (2011b), avaliou-se a presença de um único ponto de mudança na série temporal, evitando assim, dividir a série em muitas subséries que poderiam afetar a análise dos quantis estimados pelo pequeno tamanho amostral.

3.2.3.2 Comparação dos quantis de precipitação

Para cada uma das séries avaliadas segundo o teste de Pettitt, estimou-se os quantis de precipitação associados a diferentes períodos de retorno, a partir da distribuição GEV. Esse procedimento foi repetido para as 2 subséries, chamadas aqui de série antiga e série recente, definidas a partir da marcação do ponto de mudança. Para as séries sem ponto de mudança significativo ao nível de 5%, adotou-se duas subséries relativas à primeira (antiga) e à segunda (recente) metade da série temporal completa. Dessa maneira, para cada uma dessas estações,

associou-se 3 curvas de quantis de precipitação: a primeira delas associada às estimativas realizadas considerando a série completa, e as demais associadas às estimativas geradas considerando as subséries antiga e recente, respectivamente.

Diferentes fatores podem introduzir incertezas no processo de modelagem, como o tamanho da série de dados observados, o modelo de probabilidade utilizado e o método de estimativa de seus parâmetros (DEGAETANO; CASTELLANO, 2013). Sendo assim, para considerar as possíveis incertezas, foi determinado o intervalo de confiança para cada quantil de precipitação calculado, através da técnica de reamostragem com reposição chamada *bootstrap* a um nível de 95% ($CI_{95\%}$). Segundo DeGaetano e Castellano (2013), as técnicas de reamostragem conseguem simular diferentes realizações de eventos de precipitação, permitindo que diferentes séries, geradas a partir da associação aleatória dos dados observados, possam ser avaliadas. Dessa maneira, avalia-se a distribuição dos quantis estimados segundo essas diferentes séries, funcionando como uma eficiente ferramenta para a contabilização das incertezas (SERINALDI; KILSBY, 2015). Para avaliar o intervalo de confiança de um quantil de precipitação associado a um determinado período de retorno, realiza-se sua estimativa diversas vezes e determina-se o limite inferior e superior através dos percentis de 2,5% e 97,5%. Com isso, define-se $CI_{95\%}$, como sendo o intervalo cuja probabilidade de que o “verdadeiro” quantil de precipitação esteja contido nele é de 95% (SERINALDI; KILSBY, 2015).

3.2.3.3 Consequências da redução da série amostral

Um ponto negativo na utilização de somente dados representativos na modelagem dos eventos extremos é a redução da série amostral, que pode introduzir incertezas e diminuir a precisão das estimativas. Para identificar os possíveis impactos dessa prática, foram avaliadas as possíveis consequências da redução da série amostral para as estações avaliadas. Para isso, comparou-se os quantis de precipitação estimados considerando duas séries sintéticas reamostradas pelo método *bootstrap*: (1) série reduzida com metade do tamanho da série original e (2) série completa, considerando todos os dados disponíveis. Com essas séries sintéticas, estimou-se os quantis de precipitação associados aos períodos de retorno de 10, 25, 50, 100 e 200 anos. Esse procedimento foi repetido mil vezes para avaliar a distribuição e as incertezas dos quantis estimados via diagrama *box-plot*.

3.3 RESULTADOS E DISCUSSÃO

3.3.1 Análise de tendência temporal

Na Tabela 3, é apresentada a porcentagem média de estações que apresentaram tendência significativa, considerando conjuntamente as séries S_i e S_f de acordo com o teste MK. Em média, 62,4% das estações avaliadas apresentaram tendência significativa em suas séries de quantis de precipitação, sendo que S_f apresentou magnitudes ligeiramente maiores que S_i . Independentemente da série avaliada (S_i e S_f), o sinal da tendência encontrado foi, na maioria das vezes, o mesmo. Foi identificada uma maior presença de tendências positivas para os quantis de precipitação associados a baixos períodos de retorno (5 e 10 anos). Entretanto, para os outros períodos de retorno avaliados, encontramos uma situação diferente, com um número semelhante de estações apresentando tendências de redução e intensificação dos eventos, com uma ligeira predominância de tendência negativas (Figura 14).

Esse fato sugere que os eventos mais comuns (5 a 10 anos) estão, no geral, se intensificando no estado. Para o caso dos eventos menos frequentes (> 10 anos), os resultados foram diferentes quando consideradas as séries S_i e S_f . A primeira aponta para uma maior parcela de estações com tendências negativas, enquanto que a segunda indica um número semelhante de estações com tendências negativas e positivas. A S_i apresenta um menor grau de incertezas por incorporar, gradualmente, os dados observados, aumentando assim a série utilizada na modelagem. A S_f está associada a um maior grau de incertezas pelo fato de considerar séries com tamanho fixo para a modelagem.

Resultados semelhantes foram encontrados por Carvalho et al. (2014) e Zandonadi et al. (2016). O primeiro destes trabalhos identificou uma intensificação geral (tendência positiva) dos eventos extremos na região sudeste do Brasil. Entretanto, não diferenciou os eventos extremos de acordo com seus períodos de retorno. Já Zandonadi et al. (2016) apontaram para um forte aumento da precipitação em grande parte da bacia do rio Paraná (região sudeste/sul do país). Foram identificadas tendências positivas associadas a eventos de menores magnitudes para todas estações avaliadas na bacia. Para o caso de eventos mais raros, foram encontradas tendências positivas e negativas igualmente distribuídas.

Tabela 3 – Parcela das estações avaliadas (%) que apresentaram tendência significativa (nível de significância de 5%) considerando a média das tendências encontradas para as séries crescente e móvel.

Período de Retorno - Tr (anos)					
5	10	25	50	100	200
64.8%	64.2%	61.7%	61.4%	61.1%	61.1%

Fonte: Autor (2021).

As magnitudes das tendências encontradas para o estado de São Paulo aumentam com o aumento do período de retorno analisado. Essa intensificação pode ser atribuída às maiores incertezas presentes nas estimativas de eventos extremos com períodos de retorno maiores do que o tamanho amostral, localizando-se na faixa de extrapolação associada à menor precisão das estimativas (VOLPI, 2019). A presença de tendências significativas, tanto positivas como negativas, destaca a importância de considerar aspectos referentes ao comportamento temporal (intensificação/redução) das séries de precipitação no processo de caracterização. Essas mudanças possuem impacto direto na análise de frequência e, consequentemente, nas práticas de gestão e dimensionamento de sistemas hidráulicos, aumentando a probabilidade de ocorrência de super ou subestimativas dos quantis (PAPALEXIOU; KOUTSOYIANNIS, 2013; RAGNO et al., 2018).

Figura 14 – Tendências dos quantis de precipitação das séries S_i e S_f , considerando os períodos de retorno de 5, 10, 25, 50, 100 e 200 anos. A distribuição espacial das tendências encontradas foi espacialmente interpolada pelo método do inverso quadrado das distâncias (IDW).

Fonte: Autor (2021).

3.3.2 Análise da representatividade dos dados

Das 50 estações avaliadas com série amostral de no mínimo 70 anos, 4 (8%) apresentaram presença de ponto de mudança, de acordo com o teste de Pettitt ao nível de 5%. Para uma análise mais aprofundada, foram selecionadas quatro estações dentre aquelas sem mudanças abruptas significativas, para compará-las com as estações com ponto de mudança em sua série (Tabela 4). As séries foram selecionadas aleatoriamente dentre aquelas com predominância de tendências significativas e positivas para os diferentes períodos de retorno avaliados. Para fins de ilustração, os resultados das séries mais longas de cada subgrupo (com e sem mudança abrupta significativa) foram apresentados separadamente.

Tabela 4 - Dados das estações avaliadas em relação à presença de ponto de mudança. Linhas destacadas em cinza representam estações com ponto de mudança significativo.

ID	Estação	Período (Anos completos)	Latitude	Longitude	Tendência
e36	Araraquara	1937 - 2017 (74)	21° 47' S	48° 10' W	+ (S)
e609	São Paulo-1	1936 - 2017 (80)	23° 39' S	46° 38' W	+ (S)
e113	Campinas-1	1942 - 2017 (72)	22° 52' S	47° 04' W	+ (S)
e626	Sarutaia	1939 - 2017 (71)	23° 13' S	49° 27' W	+ (S)
e680	Torrinha	1937 - 2017 (75)	22° 23' S	48° 09' W	+ (S)
e155	Corumbataí	1940 - 2017 (75)	22° 12' S	47° 37' W	+ (S)
e115	Campinas-2	1890 - 2015 (126)	22° 53' S	47° 05' W	+ (S)
e607	São Paulo-2	1889-2003 (96)	23° 32' S	46° 38' W	+ (S)
+ (S) = Tendência positiva e significativa à nível de significância de 5%.					

Fonte: Autor (2021).

A partir das séries selecionadas, foram construídas as subséries antiga e recente divididas a partir do ponto de mudança (Figura 15). Para as séries sem ponto de mudança significativo, as séries antiga e recente referem-se à primeira e segunda metade, respectivamente, da série completa.

Diferenças significativas podem ser observadas nos valores médios das subséries e da série completa da estação com ponto de mudança significativo. Essas diferenças ficam ainda mais evidentes ao compararmos a distribuição das precipitações observadas via *box-plot*. Além disso, a presença de precipitações menos intensas na subsérie antiga influenciou também a distribuição dos dados da série completa, ainda que as observações recentes tenham apresentado um comportamento diferente e mais intenso. Já para as estações sem mudanças abruptas significativas, foram encontradas séries e subséries semelhantes, com distribuições e valores médios similares. O mesmo padrão foi observado para as demais estações avaliadas (Figura 16).

Segundo Villarini *et al.*, (2011a), diferentemente das variações lentas no comportamento do evento observado - indicadas pela análise de tendência - que possuem maior probabilidade de continuar no futuro, as mudanças abruptas de um regime para outro – indicadas pela análise de ponto de mudança - tendem a permanecer com o mesmo padrão até que uma nova mudança de regime seja observada. Isto é, tendências apresentam um

comportamento contínuo e mudanças abruptas, um comportamento periódico. Dessa maneira, para o dimensionamento de infraestruturas hidráulicas considerando séries com mudanças abruptas, pode-se considerar que, no futuro, os eventos observados podem se assemelhar tanto aos eventos mais antigos, como aos eventos mais recentes, a depender da existência de mudanças abruptas no futuro. Portanto, se faz necessário considerar, como alternativa no processo de caracterização, os diferentes regimes observados e suas consequências nos quantis estimados.

Figura 15 – Série temporal de precipitação observada (AMS) para as estações São Paulo-1 e Campinas-2, divididas no ponto de mudança e na metade, respectivamente. Distribuição dos dados da subsérie antiga (azul/verde), recente (amarelo/vermelho) e da série completa estão representados no diagrama box-plot (com/sem mudança abrupta).

Fonte: Autor (2021).

Figura 16 – Séries de precipitação observadas (AMS) das estações avaliadas. Esquerda: estações com mudanças abruptas. Direita: estações sem mudanças abruptas.

Fonte: Autor (2021).

3.3.2.1 Comparação dos quantis de precipitação

Para cada uma das estações selecionadas, comparou-se os quantis de precipitação, estimados pela distribuição GEV ajustada às subséries antiga, recente e à série completa. Na Figura 17, são apresentadas as estações selecionadas para detalhamento (São Paulo-1 e Campinas-2). A estação com mudança abrupta (São Paulo-1) apresentou diferenças significativas entre os quantis estimados pelas três séries temporais analisadas. Subestimativas de aproximadamente 15 mm e 60 mm foram encontradas quando comparados o quantil associado ao período de retorno de 100 anos da série recente com aqueles estimados pela série completa e pela série antiga, respectivamente.

Esses resultados sugerem como a existência de regimes diferentes nos dados observados pode afetar as estimativas feitas a partir da série completa, por agregarem informações de um comportamento não mais observado na condição atual, devido à presença

da mudança abrupta causadas por questões climáticas (como periodicidades e oscilações) ou antropogênicas. Essas mudanças representam uma redução de cerca de 50 (série completa) e 85 (série antiga) anos no intervalo médio de recorrência de um evento associado ao período de retorno de 100 anos, estimado através da série recente (Figura 17b). Padrões semelhantes foram observados nas demais estações com mudanças abruptas (Figura 18), com exceção da estação Sarutaia que apresentou quantis semelhantes estimados pela série completa e série recente. Comparando os quantis de 100 anos da série recente com os da série passada e completa, obtivemos uma diferença relativa média para todas as estações avaliadas de 32% e 13%, respectivamente.

Figura 17 - Estimativa dos quantis de precipitação associados a diferentes períodos de retorno para as estações São Paulo-1 (a-b) e Campinas-2 (c-d). Esquerda: comparação dos quantis estimados usando as subséries antiga (azul/verde), recente (amarelo/vermelho) e completa (preto) junto ao $CI_{95\%}$, calculado pelo método de reamostragem bootstrap. Direita: detalhes das diferenças nos períodos de retorno do quantis estimados.

Fonte: Autor (2021).

Essas diferenças, entretanto, não foram observadas para as séries sem uma mudança abrupta significativa. Para todas as estações avaliadas (Figura 18), foram encontradas

estimativas muito próximas feitas pelas séries antiga, recente e completa, com sobreposição das estimativas e seus respectivos intervalos de confiança. A diferença relativa média entre o quantil estimado pelas subséries recentes com aqueles estimados pelas subséries antigas foi de 2% para um evento associado ao período de retorno de 100 anos. A pior situação, obtida quando comparamos a estimativa das séries antiga e recente, levou a uma redução do período de retorno de 100 para 85 anos. Vale ressaltar que as estimativas feitas pelas subséries apresentaram maiores CI's, que indicam maiores incertezas em suas estimativas. Sendo assim, para séries sem mudanças abruptas, o uso da série completa é mais indicado, dado que séries reduzidas podem aumentar as incertezas nas estimativas

Todas as séries avaliadas apresentaram tendências positivas e significativas em seus quantis estimados (Tabela 4). Mesmo assim, as séries sem mudanças abruptas apresentaram estimativas muito próximas feitas pela série completa e pelas subséries. Esses resultados sugerem que a falta de representatividade dos dados, decorrente de mudanças abruptas, apresentam consequências significativamente maiores no processo de caracterização do que a presença de tendências. De fato, mesmo considerando séries sem tendências significativas, a existência de flutuações de longo prazo nos eventos de precipitação acaba afetando significativamente a estimativa de eventos futuros, pois considera eventos de precipitação de populações estatísticas diferentes (MARANI; ZANETTI, 2015).

Resultados semelhantes foram observados por Luke et al. (2017) para séries de vazão, indicando que, embora modelos estacionários apresentem melhores resultados do que os modelos não estacionários, eles apresentaram baixo desempenho para descrever as séries com mudanças físicas conhecidas, que podem ser indicadas pelas mudanças abruptas. Iliopoulou e Koutsoyiannis (2020) confirmaram, em um estudo para avaliar a eficácia das análises de tendência, que a consideração de eventos recentes, em detrimento da utilização da série inteira, é responsável pela melhor previsão das propriedades da precipitação futura. Além disso, verificaram que os modelos baseados na média dos valores observados superam, em geral, os modelos que consideram a presença de tendências lineares. Esse resultado sugere que, em muitos casos, a consideração de tendências na modelagem, como geralmente é feito nos modelos não-estacionários, pode apresentar reduzido desempenho.

Marani e Zanetti (2015), ao analisarem os padrões temporais de uma série de dados de chuva excepcionalmente grande (268 anos – Padova, Itália), indicaram que as séries normalmente disponíveis, com tamanho amostral reduzido, não conseguem representar adequadamente os eventos futuros, pois podem apresentar características e regimes diferentes

das condições futuras como consequência da presença de oscilações e periodicidades de longo prazo. Esses estudos corroboram com a ideia de que dados não homogêneos, advindos de populações estaticamente diferentes, podem afetar negativamente as estimativas feitas a partir da série completa.

Figura 18 – Diagramas de período de retorno para todas as estações avaliadas, comparando os quantis estimados pelas séries antiga, recente e completa. Esquerda e direita: estações com séries com e sem ponto de mudança significativo, respectivamente. Todos os gráficos apresentam o intervalo de confiança $CI_{95\%}$.

Fonte: Autor (2021).

3.3.2.2 Consequências da redução da série amostral

Avaliando os efeitos da redução da série amostral nas estimativas, foram identificadas diferenças mínimas entre os quantis médios e medianos estimados pela série reduzida e completa (Figura 19). Para ambas as condições (com ou sem ponto de mudança significativo), o impacto da redução esteve mais relacionado com a distribuição dos quantis estimados e, consequentemente, com as incertezas inerentes ao processo de caracterização. Em outras

palavras, pode-se dizer que a redução da série amostral afeta principalmente a precisão das estimativas, com menor impacto nos seus valores médios/medianos.

A diferença entre a amplitude dos $CI_{95\%}$ calculados para as séries reduzida e completa aumentou para os maiores períodos de retorno avaliados, conforme esperado, devido às maiores incertezas associadas às suas estimativas. Em seis das estações avaliadas (São Paulo-1, Campinas-2, Corumbatai, Torrinha, Sarutaia e São Paulo-2) a diferença entre os limites do $CI_{95\%}$ ficaram abaixo de 20 mm, com diferenças reduzidas para eventos mais comuns (< 25 anos). As duas estações restantes (Araraquara e Campinas-1) exibiram maiores amplitudes do $CI_{95\%}$ chegando a aproximadamente 40 mm para os eventos associados ao período de retorno de 200 anos. Entretanto, essas duas estações apresentaram também grande variabilidade entre os quantis estimados pelas séries antiga, recente e completa (Figura 18). Este fato sugere que os efeitos de dados não representativos, com propriedades bem diferentes das observadas atualmente, podem estar influenciando a variabilidade da simulação, visto que eles dados com regimes e propriedades bem diferentes podem ser selecionados para compor as séries reduzida e completa no processo de simulação. Assim, de maneira geral, pode-se concluir que a redução da série amostral afeta principalmente a variabilidade dos quantis estimados, aumentando as incertezas presentes no processo, conforme o esperado.

Figura 19 – Distribuição dos quantis de precipitação estimados pelas séries reduzida e completa, para os períodos de retorno de 10, 25, 50, 100 e 200 anos. Linhas tracejadas representam a diferença entre as amplitudes do $CI_{95\%}$ considerando as séries reduzidas e completas.

Fonte: Autor (2021).

3.4 CONCLUSÃO

Neste estudo, avaliou-se como mudanças temporais em séries observadas podem afetar o processo de estimativa de eventos extremos. Conforme visto no estudo de caso aplicado para o estado de São Paulo, as mudanças observadas nos eventos de precipitação afetam sua caracterização, impactando diretamente na análise de frequência. Assim, as infraestruturas hidráulicas, apoiadas no conceito de período de retorno na estacionariedade, acabam sendo muitas vezes sub ou superdimensionadas. O uso de modelos não estacionários para o cálculo das mudanças climáticas ainda é um assunto controverso, pois requer um entendimento completo das variáveis que influenciam o processo, complicando o procedimento de modelagem. Ademais, se modelos não estacionários são usados sem funções determinísticas que definem as relações entre as variáveis, maiores incertezas podem ser encontradas.

No entanto, variações temporais nas séries existem e precisam ser consideradas na caracterização de eventos extremos. Conforme visto no estudo de caso, a maior parte das estações do estado apresentam tendências lineares em suas séries de quantis estimados, sendo observada uma intensificação dos eventos com menores período de retorno e uma parcela similar de tendências negativas e positivas para os eventos com maior período de retorno. Ainda, eventos de precipitação extrema geralmente apresentam oscilações de longo prazo, com periodicidades muitas vezes maior do que o tamanho amostral da série observada. Nesses casos, há um impacto direto no processo de caracterização, pois são consideradas observações provenientes de diferentes regimes climáticos e com, consequentemente, diferentes comportamentos, para caracterizar os eventos futuros.

Uma abordagem possível, avaliada nesse trabalho, foi considerar as mudanças na construção das séries, avaliando a existência de diferentes comportamentos dentro da série temporal e suas influências nas estimativas realizadas. Como visto, caso existam mudanças abruptas nas propriedades temporais, sendo possível dividir a série original em subséries com características distintas, diferentes quantis de precipitação são encontrados. Sendo assim, essa abordagem pode ser útil para tomadores de decisões avaliarem diferentes cenários possíveis para caracterização dos eventos extremos, considerando diferentes populações de precipitação decorrentes de periodicidades e oscilações de longo prazo existentes nas séries observadas, bem como mudanças graduais. No entanto, é importante destacar que a redução da série amostral afeta negativamente as estimativas, principalmente em relação ao aumento das incertezas nos quantis estimados. Algumas situações requerem uma exploração dos dados disponíveis e de sua representatividade, a fim de fornecer mais informações relacionadas a diferentes abordagens de estimativa de eventos extremos.

Espera-se que a disponibilidade de dados seja suficiente no futuro, com o acréscimo de dados observados ao longo dos anos. Isso facilitará a caracterização dos diferentes comportamentos apresentados nas séries temporais, contribuindo para um melhor entendimento da estrutura de análise de frequência. Porém, a situação atual ainda fornece pouca base para essas influências, afetando nosso conhecimento sobre as propriedades e padrões temporais de chuva.

CAPÍTULO IV

**Avaliação da distribuição metaestatística de valores extremos para
caracterização da precipitação em um contexto de mudanças climáticas.**

4.1 INTRODUÇÃO

O estudo de eventos extremos de precipitação é fundamental para a gestão e o planejamento de recursos hídricos. A prática usual para a caracterização desses eventos baseia-se na análise estatística de frequência, que identifica uma distribuição de probabilidade para associar eventos de diferentes magnitudes a uma determinada frequência de ocorrência (LAIO; DI BALDASSARRE; MONTANARI, 2009; MARRA et al., 2019; OKOLI et al., 2018). A teoria de valores extremos (EVT) (COLES, 2001), largamente utilizada em modelagens hidrológicas, fornece uma base teórica e abrangente para a realização dessa análise de frequência, possibilitando a caracterização de diferentes extremos hidrológicos, como eventos de cheias e chuvas intensas (AMMAR et al., 2019; PAPALEXIOU; KOUTSOYIANNIS, 2013). Na EVT, as distribuições de probabilidade podem ser modeladas considerando dois principais tipos de séries temporais: as séries de máxima precipitação anual (AMS), também chamadas de séries de bloco de máxima, onde são utilizados os valores de máximo de um determinado bloco temporal; e as séries de duração parcial (POT), composta por todos eventos observados com magnitude acima de um limiar pré-determinado (AMMAR et al., 2019; GAUME, 2018; LANG; OUARDA; BOBÉE, 1999; MARRA et al., 2018; NGUYEN et al., 2017; NGUYEN; NGUYEN, 2019).

Em ambos os casos, a EVT assume que: (1) o número de eventos independentes e identicamente distribuídos (i.i.d) por bloco temporal (como por exemplo, igual a um ano) é grande o suficiente ($n \rightarrow \infty$) e (2) que as séries do tipo AMS e POT convergem para a distribuição generalizada de valores extremos (GEV) ou para a distribuição generalizada de Pareto (GPD), respectivamente (MARANI; IGNACCOLO, 2015; MINIUSSI; MARANI, 2020; SERINALDI; LOMBARDO; KILSBY, 2020; STRUPCZEWSKI et al., 2011; ZORZETTO; BOTTER; MARANI, 2016). Essas hipóteses, entretanto, são raramente atendidas, visto que o número de eventos de precipitação (n) em um ano é finito e, em geral, insuficiente para validar a suposição de larga disponibilidade de eventos e de convergência assintótica (KOUTSOYIANNIS, 2004; SCHELLANDER; LIEB; HELL, 2019; ZAGHLOUL et al., 2020). Além disso, ambas abordagens utilizam somente uma pequena parcela de dados da amostra completa para descrever o fenômeno de interesse, descartando um grande número de observações que poderiam aprimorar a precisão das estimativas (VOLPI et al., 2019; ZORZETTO; BOTTER; MARANI, 2016).

Uma abordagem alternativa, introduzida por Marani e Ignaccolo (2015), relaxa as considerações da teoria clássica de valores extremos, tornando possível a utilização de todos os

dados observados (MINIUSSI; MARANI; VILLARINI, 2020). Essa nova abordagem, chamada de distribuição metaestatística de valores extremos (MEV), assume que os eventos extremos são retirados de uma amostra finita e variável de eventos ordinários, considerados como todo evento de precipitação de um determinado bloco temporal - em geral, de um ano (MARRA; BORGA; MORIN, 2020). Dessa maneira, a MEV permite uma ampla utilização dos dados observados, sem assumir, no entanto, condições iniciais que raramente são atendidas (ZORZETTO; BOTTER; MARANI, 2016). Além disso, a MEV é modelada considerando as distribuições dos eventos ordinários em cada ano observado, levando em conta, dessa maneira, as diferentes propriedades apresentadas pela precipitação ao longo dos anos (MARANI; IGNACCOLO, 2015). Essa última propriedade possui extrema importância em um contexto de mudanças climáticas, onde são esperadas alterações no comportamento dos eventos extremos e variação da representatividade dos dados observados (FISCHER; KNUTTI, 2015).

Diferentes estudos avaliaram a eficiência da distribuição MEV para modelar eventos hidrológicos extremos, confirmando seu melhor desempenho em comparação com a EVT, especialmente para caracterizar eventos com período de retorno maior que o tamanho da amostra utilizada na modelagem (MARANI; IGNACCOLO, 2015; MARRA et al., 2018; MINIUSSI; MARANI; VILLARINI, 2020; SCHELLANDER; LIEB; HELL, 2019; ZORZETTO; BOTTER; MARANI, 2016). Entretanto, pouco se sabe ainda sobre o desempenho da MEV no atual contexto climático. Isto é, o desempenho superior da MEV em comparação com as metodologias usuais é também observado em um contexto de constantes alterações climáticas? A MEV é adequada para modelar séries temporais com alterações de suas propriedades ao longo dos anos?

Diante das questões acima levantadas, avaliou-se, nesse estudo, o desempenho da distribuição MEV para modelar extremos em meio a um contexto de mudanças climáticas. Especificamente, o desempenho da MEV foi comparado com o desempenho das metodologias clássicas (GEV-via AMS e GPD-via POT) para modelar dois tipos de séries com alterações climáticas: (1) séries com mudanças abruptas e (2) séries com alterações lineares de suas propriedades. Da mesma forma que Villarini et al. (2009), assumiu-se que a presença de mudanças abruptas e/ou tendências lineares rompem com a suposição de estacionariedade.

Para isso, foram geradas séries sintéticas de eventos ordinários de precipitação, simulando as duas condições de alterações climáticas avaliadas nesse estudo. Optou-se pela utilização da distribuição de Weibull para modelar os eventos ordinários, por ter sido indicada como adequada para modelar eventos de precipitação (PAPALEXIOU; AGHAKOUCHAK;

FOUFOULA-GEORGIOU, 2018; WILSON; TOUMI, 2005) e ter sido adotada em recentes estudos (MARRA et al., 2019; ZORZETTO; BOTTER; MARANI, 2016). Para conseguir representar eventos reais de precipitação, foram utilizados, no processo de geração das séries sintéticas, os parâmetros Weibull anuais de 699 estações de precipitação localizadas no estado de São Paulo, Brasil. Dessa forma, garante-se que os eventos sintéticos gerados se assemelhem a eventos de precipitação observados. Esse estudo teve como foco principal o avanço no entendimento da distribuição MEV, avaliando seu desempenho em diferentes cenários de alterações climáticas, a fim de garantir uma resposta sobre o uso dessa distribuição na caracterização de eventos extremos de precipitação no contexto atual.

4.2 ESTRUTURA TEÓRICA

Seja $X > 0$ uma variável aleatória que representa diferentes alturas de precipitação e $f(x)$ e $F(x)$ representam sua função densidade de probabilidade e função acumulada de probabilidade, respectivamente. Considera-se que em um bloco anual há n eventos chuvosos (isto é, onde $X > 0$), sendo n uma variável aleatória discreta. A teoria clássica de valores extremos baseia-se na análise dos eventos de máximo de um determinado bloco, expressos por Y , onde $Y_i = \max \{X_1, X_2, \dots, X_n\}$, com $i = 1, 2, \dots, N$ (número de anos completos com dados observados). Isto é, a variável aleatória Y representa os eventos de máxima precipitação diária anual. Assumindo que os n eventos ordinários que ocorrem em um ano são independentes, a distribuição de probabilidade dos eventos de máximo pode ser expressa por $H_n(y) = [F(y)]^n$.

Assume-se, na EVT, que o número de eventos ordinários em cada bloco (n) é grande o suficiente, com $n \rightarrow \infty$, de maneira que o limite $\lim_{n \rightarrow \infty} H_n(y)$ converge, embora não exaustivamente, para uma dentre três formas funcionais ($H(y)$), de acordo com o comportamento da cauda da distribuição original da variável X (NAGHETTINI; PINTO, 2007; SCHELLANDER; LIEB; HELL, 2019). A primeira forma funcional (EV1), também chamada de distribuição de Gumbel ou dupla exponencial, possui decaimento exponencial da cauda extrema de sua função densidade. A distribuição de Frechét (EV2), segunda forma funcional, possui decaimento na direção do extremo seguindo uma função polinomial e a distribuição de Weibull reversa (EV3), terceira forma funcional, é obtida quando X possui uma limitação na direção do extremo. Em termos práticos, a distribuição assume forma EV1 se a distribuição $F_x(x)$ é ilimitada superiormente e se aproxima de 1 tão rapidamente quanto a distribuição exponencial, para $n \rightarrow \infty$. A distribuição possui uma cauda superior polinomial, assumindo a forma EV2, se $F_x(x)$, além de ilimitada superiormente, aproxima-se de 1 mais lentamente que

a distribuição exponencial, para $n \rightarrow \infty$. Por fim, a distribuição é do tipo EV3 caso a variável X seja limitada superiormente (NAGHETTINI; PINTO, 2007).

4.2.1.1 Distribuição generalizada de valores extremos (GEV)

As três formas assintóticas de valores extremos (EV1, EV2 e EV3) podem ser consideradas como casos especiais da distribuição generalizada de valores extremos - GEV (Equação 20), que incorpora os três possíveis casos da função de valores extremos em uma única expressão:

$$H_{GEV}(y) = \exp \left\{ - \left(1 + k \cdot \frac{y - \beta}{\alpha} \right)_+^{-1/k} \right\} \quad (20)$$

onde $(\cdot)_+ = \max((\cdot), 0)$, β é o parâmetro de posição, α é o parâmetro de escala e k é o parâmetro de forma, que define qual das três formas a distribuição GEV irá assumir. Para $k = 1$, a distribuição GEV corresponde ao caso especial EV1 (Gumbel). Para $k > 0$, a GEV corresponde à distribuição EV2 e, para $k < 0$, à distribuição EV3 (MARANI; IGNACCOLO, 2015). A distribuição GEV é amplamente utilizada em análises hidrológicas, sendo recomendadas nas diretrizes nacionais de diferentes países, como Áustria, Espanha e Alemanha (ALAM et al., 2018; NGUYEN et al., 2017; SALINAS et al., 2014b, 2014a).

4.2.1.2 Distribuição generalizada de pareto (GPD)

Uma abordagem alternativa para a modelagem de valores extremos consiste na consideração de todos os eventos com valores acima de um determinado limite preestabelecido (POT). Considerando que o número de eventos de precipitação diária considerados na modelagem é grande o suficiente, sua distribuição pode ser aproximada pela distribuição generalizada de Pareto (GPD), expressa pela Equação 21, sendo φ o parâmetro de forma e $\sigma > 0$ o parâmetro de escala. Se $\varphi \leq 0$, temos que $0 < y < \infty$. Caso $\varphi > 0$, temos que $0 < y < \sigma/\varphi$ (SCHELLANDER; LIEB; HELL, 2019). A distribuição GPD é também amplamente utilizada em análises de frequência hidrológica (ÖZTEKIN, 2005). Vale ressaltar que ao assumir que os (i) valores y são determinados de acordo com a técnica POT, (ii) a distribuição segue a distribuição GPD e que (iii) o número n de eventos de precipitação segue a distribuição de Poisson, a POT leva a distribuição de máximo anual que possui a forma da distribuição GEV (ZORZETTO; BOTTER; MARANI, 2016).

$$H_{GPD}(y) = 1 - \left(1 + \varphi \cdot \frac{y}{\sigma}\right)^{-1/\varphi} \quad (21)$$

4.2.1.3 Distribuição metaestatística de valores extremos (MEV)

A distribuição metaestatística de valores extremos (MEV), proposta por Marani e Ignaccolo (2015), relaxa as suposições iniciais da teoria de valores extremos de convergência assintótica e número suficiente n de dados observados em cada bloco temporal (ZORZETTO; BOTTER; MARANI, 2016). Ao contrário da teoria tradicional de modelagem de valores extremos, que descarta uma grande quantidade de dados observados no processo de modelagem, a MEV possibilita a utilização de todos os dados disponíveis (MINIUSI; VILLARINI; MARANI, 2020). Nela, assume-se que o número de eventos por ano (n) e os parâmetros (θ) da distribuição dos eventos ordinários de um determinado ano (considerados aqui como qualquer evento com altura de precipitação maior que 0 mm) são variáveis aleatórias (MARRA et al., 2019). Dessa maneira, a MEV consegue considerar o processo aleatório de ocorrência de eventos de precipitação em cada bloco temporal avaliado, alterando, quando necessário, a distribuição de probabilidade dos eventos ordinários separadamente (SCHELLANDER; LIEB; HELL, 2019; ZORZETTO; BOTTER; MARANI, 2016). A formulação geral da distribuição acumulada de probabilidade da distribuição MEV, obtida através do teorema da probabilidade total, é dada por:

$$\zeta(x) = \sum_{n=1}^{\infty} \int_{\Omega_{\theta}} [F(x, \theta)]^n g(n, \theta) d\theta \quad (22)$$

sendo $g(n, \theta)$ a distribuição de probabilidade conjunta do número de eventos n e dos parâmetros θ (discreto em n e contínuo em θ) e Ω_{θ} é o conjunto de todos os possíveis valores dos parâmetros. Da mesma forma que Zorzetto, Botter e Marani (2016), a distribuição de Weibull (Equação 23) foi adotada para representar os eventos ordinários de precipitação (x) de cada ano j observado, sendo $j = 1, 2, \dots, T$, conforme indicada por Wilson e Toumi (2005) como adequada para modelar eventos ordinários de precipitação, por apresentar uma cauda mais pesada que consegue representar os eventos extremos (MARRA et al., 2018).

$$F(x) = 1 - \exp\left(-\frac{x}{C}\right)^w \quad (23)$$

sendo w e C os parâmetros de forma e escala da distribuição de Weibull, respectivamente. Pode-se obter uma expressão aproximada para $\zeta(x)$ substituindo a função densidade de probabilidade conjunta $g(n, \theta)$ por médias amostrais, ao invés de especificá-la. Com isso, a distribuição acumulada de probabilidade da MEV, considerando a distribuição Weibull para descrever os eventos ordinários, fica:

$$\zeta(x) = \sum_{n=1}^{\infty} \int_C \int_w \left[1 - \exp\left(-\frac{x}{C}\right)^w \right]^n g(n, C, w) dC dw \quad (24)$$

Pode-se obter, por fim, a distribuição de eventos de máxima precipitação a partir de uma aproximação da média amostral das distribuições de Weibull ($\zeta(x) \cong \zeta_m(x)$), calculadas ao longo dos T anos avaliados, de acordo com a seguinte expressão:

$$\zeta_m(x) = \frac{1}{T} \sum_{j=1}^T \left[1 - \exp\left(-\frac{x}{C_j}\right)^{w_j} \right]^{n_j} \quad (25)$$

sendo w_j , C_j e n_j os parâmetros de forma, escala e o número de eventos de precipitação ordinários de cada ano j de dados disponíveis, respectivamente, e T o número de anos avaliados. Na Figura 20 é apresentado o esquema gráfico para a construção da distribuição MEV.

Figura 20 – Esquema conceitual da construção da distribuição MEV aplicado à máxima precipitação diária. Parâmetros de Weibull (em azul e vermelho) e o número de dias chuvosos (verde) são determinados para cada ano, definindo uma distribuição acumulada anual de máxima precipitação diária. Distribuições acumuladas anuais (cinza) e suas projeções (preto) definem a variabilidade dos valores de C , w , e N . A distribuição MEV é obtida pela média das distribuições acumuladas de cada ano.

Fonte: adaptado de Zorzetto, Botter e Marani (2016).

4.3 MATERIAIS E MÉTODOS

O trabalho foi apresentado de acordo com as seguintes etapas: Na seção 4.3.1, foi apresentada a área de estudo, bem como a metodologia de determinação dos parâmetros das séries sintéticas. Na seção 4.3.2, descreveu-se o método para estimativa dos parâmetros das distribuições MEV, GEV e GPD. Na seção 4.3.3, apresentou-se a metodologia de geração das séries sintéticas sem mudanças abruptas e/ou tendências lineares em seus parâmetros. Na seção 4.3.4 e 4.3.5, foi descrita a metodologia de geração das séries sintéticas com mudanças abruptas e com variações lineares (tendências) de seus parâmetros, respectivamente.

4.3.1 Área de estudo

Assim como no capítulo II, utilizou-se os dados de precipitação disponibilizados pelo Departamento de Água e Energia Elétrica (DAEE), em seu banco de dados (disponível em <http://www.hidrologia.daee.sp.gov.br/>). O critério para a seleção de anos válidos a serem utilizados na modelagem foi o mesmo: foram selecionados apenas os anos que não possuísem nenhuma falha nos meses chuvosos (Outubro – Março) e, no máximo, 3 falhas nos meses menos chuvosos (Abril – Setembro). Nesse estudo, foram utilizadas todas as estações disponíveis no banco de dados do DAEE com mais de 30 anos, totalizando 699 estações avaliadas (Figura 5).

Para cada ano considerado como válido das séries avaliadas, modelou-se a distribuição de Weibull (Equação 23). Com isso, obteve-se a distribuição dos parâmetros w e C para a área de estudo, considerando todas as estações (Figura 21). Os parâmetros da distribuição de Weibull para cada bloco anual foram determinados por meio do método dos momentos ponderados (MMP), que possui um bom desempenho para a modelagem de pequenas amostras e é pouco sensível à presença de outliers (GREENWOOD; LANDWEHR; MATALAS, 1979). O MMP caracteriza-se também por atribuir um maior peso à cauda da distribuição de probabilidade (ZORZETTO; BOTTER; MARANI, 2016).

Esses intervalos de parâmetros foram utilizados como base para geração das séries sintéticas, considerando três possíveis casos: (i) séries sem a indução de pontos de mudança e tendência linear dos parâmetros, (ii) séries com ponto de mudança e (iii) séries com tendência linear nos parâmetros. Para avaliar a adequabilidade da distribuição Weibull para descrever os eventos ordinários de precipitação, aplicou-se os testes de aderência de Kolmogorov-Smirnov (KS), Cramer Von Mises (CVM) e Anderson-Darling (AD) para cada ano de cada estação avaliada. Todos eles são amplamente utilizados em análises hidrológicas com o intuito de verificar a adequabilidade de distribuições de probabilidade (BESKOW et al., 2015; RULFOVÁ et al., 2016). A partir disso, pode-se determinar o valor-p médio para cada estação. Valores-p abaixo de 0,05 representam a rejeição da hipótese nula do teste de aderência, de que a amostra é retirada da distribuição de referência (Weibull). Em geral, os três testes avaliados demonstraram uma boa adequabilidade da distribuição para descrever os dados da região de estudo (Figura 22), com poucas estações onde a hipótese nula foi rejeitada, isto é valor-p < 0,05.

Figura 21 - Distribuição dos parâmetros da distribuição de probabilidade de Weibull, considerando todos os anos válidos de todas as estações avaliadas.

Fonte: Autor (2021).

Figura 22 - Adequabilidade da distribuição de Weibull para descrever os eventos ordinários de precipitação, segundo os testes de aderência de Kolmogorov-Smirnov (KS), Anderson-Darling (AD) e Cramer Von Mises (CVM). A linha tracejada escura indica o limite onde rejeita-se a hipótese nula (valor-p = 0,05).

Fonte: Autor (2021).

4.3.2 Estimativa dos parâmetros das distribuições de probabilidade

A determinação dos parâmetros da distribuição GEV foi feita de acordo com o método dos momentos-L (MM-L), por ser menos sensível à presença de *outliers* e apresentar bons resultados quando modelados para amostras pequenas (HOSKING, 1990). A distribuição GPD, utilizada para as séries POT, teve seus parâmetros determinados de acordo com o método da máxima verossimilhança (MV). Para a construção das séries POT, estabeleceu-se, nesse estudo, o limite de tal forma que o número médio de eventos disponíveis por ano fosse igual a 4, semelhante ao critério utilizado por Zorzetto, Botter e Marani (2016). Buscou-se, com isso, garantir um limiar alto o suficiente para que as hipóteses da EVT fossem verificadas e também, baixo o suficiente para que fossem mantidas as vantagens da abordagem POT sobre a AMS, isto é, a utilização de um maior número de observações. Por fim, para a distribuição MEV, utilizou-se o método MMP para estimar os parâmetros da distribuição de Weibull relativa a cada bloco anual, conforme descrito na seção 4.3.1.

4.3.3 Séries sintéticas

A partir dos parâmetros da distribuição de Weibull (w e C) e do número de dias chuvosos (n), determinados para todas estações da área de estudo, construiu-se uma série sintética com 1000 anos de dados observados. Cada ano dessa série sintética teve seus eventos chuvosos gerados de acordo com a distribuição de Weibull, cujos parâmetros w , C e o número de eventos chuvosos foram selecionados aleatoriamente considerando os parâmetros determinados previamente a área de estudo (Figura 21).

Com a série sintética original (1000 anos de dados), foram geradas séries sintéticas reduzidas de 10, 20, 50, 100 e 200 anos de dados, selecionadas de acordo com a técnica de reamostragem com reposição chamada de *bootstrap*. Esse procedimento foi repetido 1000 vezes, com o intuito de avaliar o intervalo de confiança das estimativas feitas a partir das séries reduzidas de diferentes tamanhos. Cada uma dessas séries reduzidas foi modelada de acordo as distribuições MEV, GEV e GPD, com essa última representando a técnica de amostragem POT. Optou-se por essa abordagem para comparar os quantis de precipitação estimados por essas três técnicas considerando diferentes tamanhos de séries no processo de modelagem.

4.3.4 Séries sintéticas com mudanças abruptas

Procedimento semelhante foi realizado para a geração das séries sintéticas com mudanças abruptas. Entretanto, nesse caso, a série sintética original (com 1000 anos de dados) foi construída a partir da junção de duas subséries (inicial e final) de 500 anos cada, com o

intuito de simular a presença de uma mudança abrupta na metade da série original. Para isso, foram considerados intervalos diferentes de seleção dos parâmetros da distribuição de Weibull observados na área de estudo para cada uma das subséries, garantindo a mudança abrupta de comportamento. Devemos ressaltar que a utilização de amostras com diferentes parâmetros para a construção da série completa viola os pressupostos presentes na teoria de valores extremos e na formulação da MEV, ainda que a última consiga incluir múltiplas distribuições (variação interanual) em suas análises (MARRA et al., 2018). Entretanto, essa metodologia permite que simulemos séries com ponto de mudança, dividindo a série completa em duas subséries com comportamentos diferentes.

Para a subsérie inicial, relativa à primeira metade da série sintética completa, considerou-se o intervalo percentil P1-P2 para a seleção aleatória dos parâmetros da distribuição de Weibull utilizados para a geração dos eventos chuvosos. Para a subsérie final, relativa à segunda metade da série sintética completa, considerou-se o intervalo P3-P4 para escolha dos parâmetros (Figura 23).

Esse processo foi repetido seis vezes (Tabela 5), gerando diferentes séries sintéticas completas com diferentes variações entre os parâmetros de Weibull de suas subséries inicial e final (séries CP1, CP2, ..., CP6). Para avaliar o desempenho das abordagens MEV, GEV e POT bem como o efeito do tamanho amostral nas estimativas realizadas por cada um dos modelos, foram geradas 1000 séries sintéticas reduzidas (via *bootstrap*) de diferentes tamanhos (20, 50, 100 e 200 anos), a partir das séries sintéticas originais. A primeira metade da série sintética reduzida foi selecionada considerando a subsérie inicial e a segunda metade, considerando a subsérie final. A partir da distribuição das estimativas realizadas por cada uma das 1000 séries sintéticas reduzidas, pode-se determinar o intervalo de confiança ao nível de 95% (CI_{95%}).

Para exemplificar o processo de geração das séries sintéticas com mudanças abruptas, tomou-se como exemplo a série sintética CP1. A sua primeira metade, chamada de subsérie inicial (primeiros 500 anos), teve seus eventos gerados pela distribuição Weibull, considerando o intervalo percentil P1-P2 (percentil-25 ao percentil-40 dos parâmetros observados na área de estudo) para seleção aleatória dos parâmetros utilizados em cada um desses 500 anos. Para a segunda metade, chamada de subsérie final (últimos 500 anos), considerou-se o intervalo P3-P4 (percentil-75 ao percentil-90) como o intervalo de seleção dos parâmetros Weibull observados para geração de seus eventos de precipitação. A junção das subséries inicial e final deu origem a série CP1. O processo para geração das séries reduzidas

foi feito da mesma forma. Isto é, a primeira metade da série sintética foi gerada considerando o intervalo percentil P1-P2 e, a segunda, considerando o intervalo percentil P3-P4.

Tabela 5 - Intervalo de percentis definido para seleção dos parâmetros das subséries inicial e final, que compõem a série sintética completa. Intervalo P1-P2 foi utilizado para a construção da subsérie inicial e intervalo P3-P4 foi utilizado para a construção da subsérie final.

ID	P1 (%)	P2 (%)	P3 (%)	P4 (%)
CP1	25	40	75	90
CP2	35	55	45	65
CP3	40	50	50	60
CP4	20	60	40	80
CP5	10	60	40	90
CP6	30	70	50	90

Fonte: Autor (2021).

Com o intuito de confirmar a presença de mudanças abruptas nas séries sintéticas geradas (CP1, CP2, ..., CP6), foi aplicado o teste de Pettitt, detalhado no item (3.2.3.1). O teste de Pettitt é amplamente utilizado em análises de séries temporais hidrológicas para identificar quebras ou discontinuidades do comportamento na série (DE LA CASA; NASELLO, 2010; VILLARINI et al., 2011a; XIE et al., 2019). Os resultados indicaram que todas séries geradas apresentaram ponto de mudança significativo, considerando o nível de 5% de significância. Em todos os casos, o ponto de mudança ficou na região central (na posição 500 ± 10 da série sintética).

4.3.5 Séries sintéticas com tendência linear em seus parâmetros

Para a construção das séries com tendência linear, adotou-se uma janela móvel para determinar o intervalo de escolha aleatória do parâmetro de interesse. Para cada um dos 1000 anos da série sintética completa, a janela se moveu ao longo do intervalo definido pelo percentil inicial e final, atualizando o intervalo de escolha dos parâmetros. Diferentemente das séries com mudanças abruptas, onde os parâmetros w , C e n das subséries inicial e final possuíam o mesmo intervalo percentil de escolha, variou-se separadamente, nesse caso, os parâmetros w , C e n . Nesse caso, a variação da janela móvel foi aplicada a apenas um parâmetro da distribuição Weibull. Para os demais parâmetros, considerou-se o intervalo total de variação, representado pelo percentil inicial e final. Da mesma maneira do que na série com mudanças abruptas, esse procedimento foi repetido seis vezes (Tabela 6), gerando três séries sintéticas completas com tendência linear positiva no parâmetro C e três séries sintéticas completas com tendência linear positiva no parâmetro w .

Com o intuito de comparar o desempenho dos três modelos (MEV, GEV e POT), foram geradas 1000 séries sintéticas, de diferentes tamanhos amostrais (20, 50, 100 e 200 anos), de acordo com a técnica de reamostragem *bootstrap*. A partir da distribuição das estimativas realizadas para cada um dos modelos e tamanhos amostrais, determinou-se o intervalo de confiança $CI_{95\%}$. Na Figura 23, encontram-se os esquemas de construção das séries com mudanças abruptas (item 4.2.4) e séries com variações lineares dos seus parâmetros (item 4.2.5). As doze séries sintéticas geradas, com mudanças abruptas e tendências lineares em seus parâmetros, estão ilustradas no Apêndice C.

Para exemplificar o processo de geração das séries sintéticas com tendências lineares, tomou-se como exemplo a série sintética T1. O seu primeiro evento foi gerado considerando o intervalo marcado pelo percentil inicial somado à extensão da janela móvel para escolha aleatória dos parâmetros da distribuição Weibull. Isto é, o intervalo de escolha aleatória foi do percentil-20 ao percentil-45 dos parâmetros observados. Gradualmente, a janela móvel avançou, sendo que o último evento da série T1, foi gerado considerando o intervalo marcado pela extensão da janela móvel e o percentil final, isto é percentil-35 ao percentil-65 dos parâmetros observados na área de estudo. Vale ressaltar que para série T1, o único parâmetro selecionado através dessa metodologia foi o parâmetro C . Os demais parâmetros (w e n) foram escolhidos aleatoriamente considerando o intervalo entre o percentil inicial e final, isto é, percentil-20 ao percentil-60 dos parâmetros observados.

Tabela 6 – Tamanho da janela móvel, parâmetro variável (com tendência linear positiva) e intervalo inicial e final dos percentis utilizados para a construção das séries sintéticas completas com tendência linear positiva em seus parâmetros.

ID	Percentil inicial (%)	Percentil final (%)	Tamanho da janela móvel (%)	Parâmetro com tendência linear
T1	20	50	15	C
T2	20	70	25	C
T3	10	80	15	C
T4	20	50	15	w
T5	20	70	25	w
T6	10	80	15	w

Fonte: Autor (2021).

Para confirmar as consequências das variações lineares nos parâmetros da distribuição de Weibull, utilizada na geração das séries sintéticas, aplicou-se o teste de Mann-Kendall modificado por Hamed e Rao (1998), detalhado no item 3.2.2.1. Os resultados indicaram a presença de tendências significativas, considerando o nível de significância de 5%, em todas as séries sintéticas avaliadas (T1, T2, ..., T6). Nas três primeiras séries, foram encontradas tendências positivas, enquanto que, nas três últimas (T4, T5 e T6), a tendência encontrada foi negativa.

Figura 23 - Esquema da construção das (A) séries com mudanças abruptas e das (B) séries com tendências lineares em seus parâmetros. No primeiro caso, os blocos vermelho e verde representam o intervalo percentil de onde foram escolhidos os parâmetros de geração da primeira e segunda metade da série completa. No segundo caso, o bloco amarelo indica a janela móvel, que representa o intervalo percentil de onde foram escolhidos os parâmetros de geração dos valores de precipitação que compuseram a série completa.

Fonte: Autor (2021).

4.3.6 Validação da MEV para a área de estudo

Para avaliar a adequabilidade da distribuição MEV para descrever os eventos extremos da área de estudo, foi realizado o seguinte procedimento: primeiramente, selecionou-se na área de estudo as estações com tamanho amostral maior ou igual a 70 anos, totalizando 48 estações. Optou-se por trabalhar com séries longas, pois elas possibilitam a comparação dos dados observados com as extrapolações feitas considerando amostras reduzidas. Para cada uma das estações selecionadas, foram construídas 100 séries sintéticas reduzidas, a partir da reamostragem com reposição *bootstrap* considerando os diferentes tamanhos amostrais de 10, 20, 30, 40 e 50 anos. Essas séries reduzidas foram, então, modeladas de acordo com as distribuições GEV e MEV, utilizadas para estimar os valores extremos de precipitação relativos à série completa original. Optou-se por não utilizar a abordagem POT por ela ter apresentado desempenho muito semelhante a abordagem GEV (via AMS) nesse estudo. Condições semelhantes foram observadas em estudos anteriores (MARRA et al., 2018; ZORZETTO; BOTTER; MARANI, 2016)

Determinou-se, então, a RMSE entre os quantis observados na série original e os quantis estimados. As distribuições dos valores de RMSE obtidos para cada série reduzida e estação avaliada foram utilizadas para comparar o desempenho das duas distribuições avaliadas. Por terem sido utilizados diferentes tamanhos amostrais, foi possível considerar o desempenho dos modelos tanto para descrever os dados observados, como para estimar eventos presentes na faixa de extrapolação, onde o evento de interesse possui período de retorno maior que o tamanho amostral da série disponível.

4.4 RESULTADOS E DISCUSSÃO

4.4.1 Séries Sintéticas

Os desempenhos dos três modelos avaliados, para diferentes tamanhos amostrais usados no processo de modelagem, são indicados na Figura 24. A distribuição GEV representa a metodologia de modelagem para as séries de máxima precipitação anual (AMS). A distribuição GPD foi modelada considerando séries de duração parcial (POT) e a distribuição MEV foi modelada considerando todos os eventos ordinários presentes na série sintética. Nota-se que, para todos os tamanhos amostrais, a distribuição MEV superou as metodologias clássicas (AMS e POT). A superação fica ainda mais clara para pequenos tamanhos amostrais ($L = 10$ e 20 anos). Com o aumento no número de anos observados utilizados na modelagem, a diferença entre o desempenho dos modelos diminuiu. Ainda assim, a MEV obteve um melhor desempenho em comparação com as demais, pelo fato de apresentar menores incertezas em suas estimativas, com menor amplitude em seus $CI_{95\%}$. Os três modelos apresentaram desempenhos semelhantes para estimar quantis associados a períodos de retorno menores ou iguais ao tamanho amostral usado na modelagem. Entretanto, para quantis presentes na faixa de extrapolação, onde os modelos são utilizados para prever/extrapolar informações a partir dos dados observados (VOLPI, 2019), a distribuição MEV apresentou o melhor desempenho, com menores incertezas que as outras duas técnicas avaliadas.

As duas técnicas baseadas na EVT apresentaram desempenhos semelhantes. Entretanto, a distribuição GPD, modelada via POT, apresentou menor amplitude do $CI_{95\%}$, com o aumento do tamanho amostral utilizado na modelagem. Isso ocorre pelo fato da diferença entre o número de dados utilizados pelas duas modelagens aumentar juntamente com o aumento do número de anos observados utilizados. Sendo assim, a POT acaba apresentando um número significativamente maior de dados utilizado em sua modelagem.

Figura 24 – Quantis estimados pelas distribuições MEV, GEV (técnica AMS) e GPD (técnica POT) considerando diferentes tamanhos amostrais (L) utilizados no processo de modelagem (10 a 200 anos). Os pontos pretos indicam os quantis empíricos extraídos da série de 1000 anos, calculados pela fórmula de posição de plotagem de Weibull. O sombreado colorido representa o intervalo de confiança ao nível de significância de 95% ($CI_{95\%}$) e as linhas sólidas indicam a média dos quantis estimados por cada um dos modelos.

Fonte: Autor (2021).

4.4.2 Séries sintéticas com mudanças abruptas

Na avaliação dos desempenhos dos três modelos avaliados para modelar séries com mudanças abruptas, optou-se por indicar somente os resultados (Figura 25) para a primeira série avaliada (CP1). Os demais resultados foram apresentados no Apêndice D. Resultados semelhantes foram encontrados para todas as seis séries sintéticas avaliadas (Figura 27).

A distribuição MEV obteve um melhor desempenho para caracterizar quantis de precipitação quando modelada a partir de amostras reduzidas ($L = 20$ e 50 anos), apresentando incertezas extremamente reduzidas em comparação com as abordagens tradicionais. Entretanto, ainda que tenha apresentado menores incertezas em todos os tamanhos amostrais, a MEV apresentou piores estimativas médias, superestimando consistentemente eventos associados a maiores períodos de retorno. Além disso, as abordagens tradicionais tiveram suas incertezas significativamente reduzidas com o aumento do tamanho amostral. Com isso, elas acabaram superando a MEV quando modeladas considerando tamanhos amostrais ampliados ($L = 100$ e 200 anos), apresentando menores ocorrências de sub ou superestimativas, ainda que com uma amplitude maior dos intervalos de confiança.

As três abordagens obtiveram desempenho médio semelhante para caracterizar eventos com período de retorno menor ou igual ao tamanho amostral. Entretanto, a MEV apresentou incertezas significativamente menores. Em relação às metodologias comumente utilizadas, a metodologia POT apresentou resultados ligeiramente melhores que a AMS para o caso de tamanho amostral reduzido, devido às menores incertezas encontradas em suas estimativas.

Figura 25 - Quantis estimados pelas distribuições MEV, GEV (técnica AMS) e GPD (técnica POT) considerando diferentes tamanhos amostrais (L) utilizados no processo de modelagem (10 a 200 anos) para séries com mudanças abruptas. Os pontos pretos indicam os quantis empíricos extraídos da série de 1000 anos, calculados pela fórmula de posição de plotagem de Weibull. O sombreado colorido representa o intervalo de confiança ao nível de significância de 95% ($CI_{95\%}$) e as linhas sólidas indicam a média dos quantis estimados por cada um dos modelos.

Fonte: Autor (2021).

Para ilustrar as superestimativas da MEV, calculou-se a razão entre o evento observado e os eventos simulados pelos três modelos, considerando o período de retorno de 100 anos (Figura 26), para as seis séries avaliadas. Optou-se por utilizar os tamanhos amostrais de 20 e 100 anos, para avaliar o desempenho dos modelos tanto quando modelados considerando amostras pequenas como grandes. Para o tamanho amostral de 20 anos, a MEV apresentou amplitude de intervalo interquartil significativamente menor do que a AMS e POT, ainda que com um pior valor médio. Entretanto, para o tamanho amostral de 100 anos, as metodologias tradicionais reduziram suas incertezas. Nesses casos, elas superaram em desempenho a

distribuição MEV. Resultados similares foram encontrados por Marra et al., (2018) e Schellander, Lieb e Hell (2019), indicando que a MEV apresenta uma tendência de superestimar eventos extremamente raros.

Figura 26 - Razão entre eventos extremos de precipitação (período de retorno de 100 anos) estimados e observados para todas as séries sintéticas com mudanças abruptas (CP1 a CP6). MEV, GEV e POT foram modelados considerando tamanhos de amostra de 20 (topo) e 100 (base) anos.

Fonte: Autor (2021).

Uma das possíveis causas desse erro sistemático, previamente indicada por estudos recentes (MARANI; IGNACCOLO, 2015; MARRA et al., 2018), é a baixa quantidade de eventos ordinários presentes em um ano para estimar os parâmetros de Weibull. Ao usar um bloco com o conjunto de todos eventos ordinários observados para modelar a MEV, ao invés de blocos anuais, Marani e Ignaccolo (2015) não encontraram viés sistemático entre os eventos estimados e observados, mesmo quando avaliaram períodos de retorno extremamente altos. Este resultado sugere que, para estações com baixa quantidade de eventos ordinários por ano, uma alternativa possível é considerar blocos maiores (> 1 ano) para modelar a distribuição Weibull, diminuindo o viés sistemático nas estimativas.

Com o intuito de elucidar as diferenças entre os desempenhos dos três modelos avaliados, foi calculada a raiz do erro quadrado médio (RMSE, Equação 2) para as 1000 séries reduzidas, selecionadas via reamostragem *bootstrap*, de cada um dos 4 tamanhos amostrais avaliados ($L = 20, 50, 100$ e 200 anos). No caso, os dados observados são aqueles retirados da amostra completa e os dados simulados são aqueles simulados pelos três modelos utilizando as séries reduzidas. A distribuição dos RMSE calculados considerando as estimativas das séries reduzidas e os valores da série completa foram dispostos em um diagrama *violin-plot*, que é similar ao diagrama de caixa (*box-plot*), exceto por acrescentar a densidade de probabilidade dos dados no diagrama (Figura 27).

A distribuição das estatísticas RMSE confirmam os resultados encontrados na Figura 25. Isto é, para todos os tamanhos amostrais avaliados, a distribuição MEV apresentou uma menor variação dos desempenhos, representada por intervalos de confiança (Figura 25) e funções densidade de probabilidade (Figura 27) mais estreitos. Entretanto, com o aumento do tamanho amostral utilizado na modelagem, as metodologias tradicionais apresentaram uma melhora significativa em seus desempenhos, diminuindo as incertezas e aumentando a acurácia, representada pelos diagramas mais próximos do valor 0 que representa um desempenho ideal, de acordo com o RMSE. Nesses casos, elas superaram, ainda que ligeiramente, a distribuição MEV. Resultados semelhantes foram encontrados para todas os casos avaliados. Em todas as situações, a distribuição MEV apresentou melhor desempenho quando modelada com uma série curta de dados (até 50 anos).

Figura 27 – Distribuição das estatísticas RMSE (*violin plot*), obtida para os três modelos avaliados, considerando os quatro tamanhos amostrais. Caixas pretas centrais indicam a distância inter-quartil do diagrama box-plot. Pontos brancos representam o valor médio da estatística RMSE.

Fonte: Autor (2021).

Dessa maneira, pode-se entender que a distribuição MEV possui desempenho significativamente melhor para estimar quantis de precipitação utilizando uma série amostral curta, tanto para séries sem ou com mudanças abruptas, por apresentar um reduzido grau de incertezas. Para o primeiro caso, a distribuição MEV apresenta melhores desempenhos que a GEV e GPD também em relação à modelagem a partir de séries amostrais longas (100 a 200 anos). Porém, para o segundo caso (séries com mudanças abruptas), esse desempenho

(modelagem a partir de séries amostrais longas) não se manteve totalmente estável, sendo a MEV superada pelas abordagens tradicionais. Resultados semelhantes foram encontrados por Marra et al. (2018) e Zorzetto, Botter e Marani (2016), que confirmaram que a distribuição MEV superou as metodologias tradicionais nas estimativas de quantis presentes na faixa de extrapolação realizadas a partir de séries curtas. Além disso, assim como Marra et al. (2018), vimos que a metodologia POT não superou significativamente a metodologia AMS, visto que esperava-se que o acréscimo de dados observados fornecido pela metodologia POT aumentasse a precisão das estimativas.

4.4.3 Séries sintéticas com tendência linear em seus parâmetros

Assim como feito para o caso das séries com mudanças abruptas, foram apresentados (Figura 28) os resultados obtidos para a primeira série sintética avaliada (T1). Os resultados das demais séries avaliadas, que apresentaram comportamento semelhante ao da série T1, são apresentados no Apêndice E.

Novamente, a distribuição MEV superou às técnicas AMS e POT para caracterizar eventos extremos a partir de um reduzido tamanho amostral (20 e 50 anos), apresentando menores incertezas em suas estimativas. Entretanto, assim como para o caso das séries com mudanças abruptas, a MEV apresentou uma acurácia ligeiramente inferior. Dessa forma, pelo fato das abordagens tradicionais terem apresentado uma significativa diminuição no seu grau de incertezas com o aumento do tamanho amostral utilizado na modelagem, seus desempenhos acabaram se assemelhando ao desempenho da abordagem MEV para caracterizar eventos extremos a partir de amostras ampliadas ($L = 100$ e 200 anos).

Figura 28 – Quantis estimados pelas distribuições MEV, GEV (técnica AMS) e GPD (técnica POT) considerando diferentes tamanhos amostrais (L) utilizados no processo de modelagem (10 a 200 anos) para as séries com tendência linear nos parâmetros. Os pontos pretos indicam os quantis empíricos extraídos da série de 1000 anos, calculados pela fórmula de posição de plotagem de Weibull. O sombreado colorido representa o intervalo de confiança ao nível de significância de 95% ($CI_{95\%}$) e as linhas sólidas indicam a média dos quantis estimados por cada um dos modelos.

Fonte: Autor (2021).

A razão entre quantis estimados e observados (Figura 29) mostra que a MEV também superestimou eventos extremos com alto período de retorno para séries sintéticas com tendências lineares, para ambos os tamanhos de amostras avaliados (20 e 100 anos). Entretanto, as diferenças entre as três metodologias são significativamente menores do que as encontradas para as séries sintéticas com mudanças abruptas. Os resultados encontrados para tendências lineares no parâmetro de forma ou no parâmetro de escala foram semelhantes. Contudo, para o primeiro caso, as metodologias tradicionais apresentaram tendência de subestimativa, enquanto que, para o segundo caso, a tendência encontrada foi de superestimativa.

Figura 29 - Razão entre eventos extremos de precipitação (período de retorno de 100 anos) estimados e observados para todas as séries sintéticas com tendências lineares em seus parâmetros (T1 a T6). MEV, AMS e POT foram modelados considerando tamanhos de amostra de 20 (topo) e 100 (base) anos.

Fonte: Autor (2021).

A distribuição das estatísticas RMSE (Figura 30) confirmam a superação da distribuição MEV sobre as demais para o caso de modelagem com tamanho amostral reduzido ($L = 20$ ou 50 anos). Além disso, a MEV apresentou um menor grau de incertezas, representado pelo tamanho dos diagramas e densidade de probabilidade reduzidos. Com o aumento do número utilizado na modelagem, o desempenho dos três modelos se aproximou, com as distribuições MEV e GEV apresentando desempenho ligeiramente superior ao da distribuição GPD. Novamente, o acréscimo de dados proporcionado pelo método POT não resultou em uma melhora significativa no desempenho.

A distribuição GPD apresentou o pior desempenho entre os três modelos avaliados, principalmente para o caso de modelagem a partir de séries curtas. Essa diferença pode ter sido encontrada pelo fato do método de seleção do limite da técnica POT não ter sido o mais adequado. Assim, informações relevantes para a modelagem deixaram de ser consideradas, ou

foram consideradas em excesso, agregando dados não representativos às séries de duração parcial. No entanto, esse desempenho semelhante entre as metodologias tradicionais foi relatado em outros estudos mesmo quando uma seleção automática de limiar, como a abordagem proposta por Solari et al., (2017), foi utilizada (MARRA et al., 2018).

Figura 30 - Distribuição das estatísticas RMSE (*violin plot*) obtida para os três modelos avaliados, considerando os quatro tamanhos amostrais. Caixas pretas indicam a distância inter-quartil do diagrama box-plot. Pontos brancos representam o valor médio da estatística RMSE.

Fonte: Autor (2021).

Os resultados encontrados corroboram com aqueles encontrados por Marra et al. (2019), Schellander, Lieb e Hell (2019) e Zorzetto, Botter e Marani (2016). A distribuição MEV possui desempenho superior ao das metodologias usuais para caracterizar eventos extremos de precipitação a partir de amostras reduzidas, por apresentar menor sensibilidade à disponibilidade de dados. Entretanto, ela apresenta uma menor acurácia, superestimando, em algumas situações, os quantis de precipitação observados. Sendo assim, deve-se considerar o *trade-off* entre a redução da acurácia e o aumento da precisão. O mesmo comportamento se manteve também para as duas últimas situações, onde foram avaliadas séries sintéticas que representassem possíveis séries temporais em um contexto de mudanças climáticas. Isso pode ser explicado pelo fato da MEV considerar a variabilidade interanual dos eventos de precipitação em sua formulação (MARANI; IGNACCOLO, 2015), capturando de melhor forma as variações decorrentes das mudanças abruptas e das tendências lineares.

4.4.4 Caracterização da precipitação na área de estudo

4.4.4.1 Validação da MEV para a área de estudo

A distribuição MEV superou a distribuição GEV para caracterização dos eventos de extremos precipitação da área de estudo (Figura 31). Essa diferença é mais evidente para o caso das séries sintéticas com menores tamanhos amostrais (10, 20 e 30 anos). Esses resultados indicam que a distribuição MEV possui um melhor desempenho para estimar eventos presentes na faixa de extrapolação que a distribuição GEV na região de estudo. Com aumento do tamanho amostral, os desempenhos se aproximam, chegando ao último caso ($N = 50$ anos), em que a distribuição GEV apresenta desempenho médio melhor que a distribuição MEV. Entretanto, considerando as incertezas nas estimativas, via $CI_{95\%}$, pode-se afirmar que ambas as distribuições apresentaram desempenhos semelhantes, visto que a MEV, embora apresente desempenho médio ligeiramente pior, possui incertezas significativamente menores em suas estimativas. Em algumas estações, o desempenho obtido em ambas distribuições avaliadas foi insatisfatório, com erros de estimativas médios da ordem de ± 20 mm. Nas demais estações, o desempenho encontrado ficou dentro do esperado dada a consideração de eventos presentes na faixa de extrapolação, com erros médios abaixo de 10 mm.

Para elucidar as diferenças entre as distribuições MEV e GEV, comparamos, para todas as 699 estações avaliadas na área de estudo, os quantis de precipitação associados ao período de retorno de 100 anos, estimados pelas 2 distribuições avaliadas (Figura 32). Pelo fato

da maioria das estações possuem série temporal com tamanho menor ou igual a 100 anos, garante-se a comparação das estimativas associadas a eventos localizados na faixa de extrapolação.

Em grande parte do estado, foram encontradas diferenças significativas (> 40 mm) entre os dois modelos. Em 4 estações, essas diferenças foram superiores a 200 mm. Diferenças significativas também foram encontradas por Schellander, Lieb e Hell (2019), quando compararam os quantis de precipitação associados ao período de retorno de 50 anos utilizando dados de 1257 estações na Áustria. Embora seja difícil optar por um único modelo dentre as distribuições GEV e MEV, as diferenças significativas entre as duas metodologias, encontradas em algumas situações, ressaltam a necessidade de serem consideradas ambas metodologias no processo de caracterização de eventos extremos, acrescentando informações ao processo de seleção de uma distribuição de probabilidade mais adequada.

Figura 31 – Desempenho (via RMSE) das distribuições MEV e GEV para caracterizar os eventos extremos de precipitação da área de estudo, considerando 48 estações com tamanho amostral maior ou igual a 70 anos. Áreas sombreadas representam o intervalo de confiança do RMSE ao nível de significância de 95% (CI95%). A linha tracejada preta representa o valor do RMSE relativo ao desempenho ideal (RMSE = 0 mm).

Fonte: Autor (2021).

Figura 32 – Diferenças entre os quantis de precipitação associados ao período de retorno de 100 anos, estimados pelas distribuições MEV e GEV. (a) América do Sul, com o Brasil e o estado de São Paulo destacados em cinza claro e escuro, respectivamente. (b) Distribuição espacial das diferenças (MEV – GEV) entre os quantis de precipitação estimados pelos dois modelos. Dados pontuais de cada estação (pontos pretos) interpolados espacialmente pelo método IDW. Valores positivos (negativos) representam subestimativa (superestimativa) dos quantis estimados pela GEV em relação aos quantis estimados pela MEV.

Fonte: Autor (2021).

4.5 CONCLUSÃO

Este estudo teve como objetivo avaliar o desempenho da distribuição metaestatística de valores extremos (MEV) para descrever os eventos extremos de precipitação em um contexto de mudanças climáticas. Através de simulações realizadas com séries sintéticas, foi possível avaliar o comportamento da MEV, em comparação às metodologias baseadas na teoria dos valores extremos (GEV e POT), para descrever séries com diferentes perturbações temporais. Foram geradas séries sintéticas com mudanças abruptas e tendências lineares em seu comportamento temporal, representando possíveis consequências das mudanças climáticas nas séries temporais de eventos extremos.

Em geral, a distribuição MEV apresentou, em todas as condições avaliadas, um desempenho superior às abordagens tradicionais para caracterizar eventos extremos a partir de uma modelagem com amostras reduzidas. Com o aumento do tamanho amostral, o desempenho das três metodologias avaliadas se aproximou, com uma ligeira superação das abordagens usuais sobre a MEV, por ter apresentado uma menor acurácia em algumas das simulações. Entretanto, ainda assim, a MEV apresentou um reduzido grau de incertezas para todos os tamanhos amostrais, indicando que a metodologia é menos sensível à disponibilidade de dados.

Em relação ao experimento numérico, a MEV se mostrou capaz, assim como as abordagens tradicionais, para contabilizar as mudanças temporais nas séries sintéticas. Este resultado deriva da formulação da MEV, que contabiliza a variabilidade interanual dos eventos extremos de precipitação, considerando parâmetros variáveis ao longo dos anos para descrever os fenômenos. Portanto, a MEV é capaz de incorporar mudanças, como mudanças abruptas e tendências, no procedimento de modelagem, sendo, portanto, adequada para caracterizar eventos extremos em cenários de mudanças climáticas.

Além das simulações realizadas com séries sintéticas, avaliamos o desempenho das distribuições MEV e GEV para caracterizar eventos extremos de precipitação da região de São Paulo. Os resultados indicaram uma superação da distribuição MEV para caracterizar eventos extremos na área de estudo, especialmente quando modelada considerando tamanhos amostrais reduzidos. A distribuição MEV apresentou reduzidos valores médios de RMSE, além de uma reduzida variação da estatística, indicando um menor grau de incertezas. Embora as duas distribuições tenham apresentado desempenhos semelhantes para longas séries amostrais, foram encontradas diferenças significativas entre os quantis de precipitação associados ao período de retorno de 100 anos. Tomando como referência os quantis estimados pela

distribuição MEV, observarmos uma tendência de subestimativa dos eventos extremos estimados pela GEV em todo o estado. Em algumas estações, essa diferença superou 200 mm.

O desequilíbrio entre os quantis estimados pelo método tradicional e pela MEV, somado aos resultados obtidos nas simulações realizadas com as séries sintéticas, ressaltam a importância de uma atualização da metodologia de caracterização dos eventos extremos. A metodologia MEV, por sofrer menor influência de diferentes fontes de incerteza, como a reduzida disponibilidade de dados e possíveis erros de medida, pode ser considerada como uma alternativa no processo de caracterização de eventos extremos. Em muitas situações, modelos probabilísticos acabam sendo utilizados para estimar eventos presentes na faixa de extrapolação, devido à baixa disponibilidade de dados observados. Nesses casos, pode-se indicar a distribuição MEV como uma possível abordagem, por apresentar bom desempenho para modelar eventos presentes na faixa de extrapolação especialmente quando modelada a partir de um reduzido tamanho amostral.

Futuras pesquisas nessa direção são necessárias, avaliando diferentes aspectos que podem interferir no desempenho do modelo MEV para caracterizar eventos extremos de precipitação. O limitado número de eventos ordinários presente na amostra observada e a presença de *outliers* são exemplos de fatores que podem influenciar a adequabilidade da distribuição. O desempenho da MEV pode, ainda, ser comparado com o desempenho de outras distribuições de probabilidade, popularmente utilizadas nas análises hidrológicas, como as distribuições Log-Pearson 3 e Gumbel. Além disso, o desempenho da MEV para caracterizar eventos de precipitação considerando diferentes cenários futuros, simulados por modelos climáticos globais ou regionais, pode ser avaliado, fornecendo um bom panorama para o desempenho da MEV em diferentes aplicações.

CAPÍTULO V

Considerações finais

5.1 CONSIDERAÇÕES FINAIS

O presente estudo reúne diferentes abordagens para avaliar o processo de caracterização de eventos extremos de precipitação e suas consequências nas estimativas realizadas em um contexto de mudanças climáticas. A grande diversidade de técnicas e metodologias existentes para realizar esse processo dificultam a existência de um consenso em relação à melhor forma de realizá-lo. Diferentes métodos de construção de séries observadas, distribuições de probabilidade, métodos de estimativa de parâmetros e métricas para seleção de modelos adequados podem afetar significativamente as estimativas realizadas. Dessa forma, este estudo foi desenvolvido com o objetivo principal de fornecer uma base aos tomadores de decisões para um aprimoramento das diferentes etapas do processo caracterização de eventos extremos de precipitação.

No Capítulo II, diferentes métricas utilizadas para seleção de distribuições de probabilidade foram avaliadas, considerando a possibilidade de incorporar análises preditivas às análises descritivas já realizadas comumente nas análises hidrológicas. Com base em um estudo de caso aplicado no estado de São Paulo, avaliou-se, também, a validade da prática usualmente feita de considerar um único modelo de probabilidade para toda uma região.

Os resultados indicaram que a consideração de aspectos preditivos é vital para o aprimoramento da seleção de distribuições, visto que modelos, considerados adequados pelos testes descritivos, apresentaram desempenho significativamente diferente quando considerados também sua capacidade preditiva. Pela escassa disponibilidade de longas séries observadas no país, há a necessidade de serem considerados os aspectos preditivos, considerando que, em muitos casos, as distribuições de probabilidade são utilizadas para caracterizar eventos presentes na faixa de extrapolação.

A utilização de uma distribuição de probabilidade para toda uma região pode acarretar em estimativas substancialmente erradas, principalmente para eventos com alto período de retorno, caso o modelo não seja adequadamente avaliado. A abordagem de seleção regional, descrita no Capítulo II, apresentou-se como uma alternativa para essa prática, visto que ela considera o desempenho dos modelos a partir de uma visão abrangente, permitindo a visualização da adequabilidade das distribuições de probabilidade de acordo com uma escala regional. Como demonstrado, um modelo considerado como o mais adequado em um grande número de estações, como foi o caso das distribuições GUM e BUR, não necessariamente é o mais adequado para ser utilizado na região, por apresentar maior variação em seus desempenhos.

A análise de tendências dos quantis de precipitação estimados, realizada no Capítulo III, indicou que ambas as séries avaliadas (S_i e S_f) apresentaram tendências significativas. Em geral, considerando o estado de São Paulo, foi observada uma intensificação dos quantis de precipitação associados a eventos mais frequentes (com período de retorno de 5 e 10 anos) e uma parcela semelhante de estações com tendências positivas e negativas dos quantis associados a eventos mais raros (com período de retorno superior a 10 anos).

A análise da representatividade dos dados, somada à confirmação das alterações temporais dos eventos extremos na análise de tendência, indicou que dados antigos podem não ser mais representativos das condições recentemente observadas, podendo diminuir a precisão e a acurácia das estimativas de eventos futuros. Conforme mencionado no Capítulo III, eventos extremos de precipitação apresentam periodicidade e oscilações de longo prazo que muitas vezes não são captados pelas séries observadas. Essas possíveis mudanças abruptas de regime podem acabar tornando dados observados não mais representativos das condições atuais. Dessa forma, foi proposta nesse estudo uma abordagem alternativa para caracterizar eventos de precipitação, considerando tanto a série completa, como as subséries antigas e recente para uma visão mais ampla dos possíveis efeitos das alterações climáticas nos quantis estimados. Essa abordagem permite captar as diferentes possibilidades, representadas por dados observados de populações estatísticas diferentes, existentes no processo de caracterização.

Por fim, no Capítulo IV, avaliou-se o desempenho da distribuição MEV para caracterização de eventos extremos de precipitação considerando as mudanças climáticas a partir de uma análise aplicada a séries sintéticas de precipitação. A MEV, em comparação com as metodologias baseadas na teoria clássica de valores extremos, apresentou desempenho superior para caracterizar eventos extremos a partir de uma série amostral reduzida. À medida que a disponibilidade de dados aumenta, os desempenhos das distribuições GEV e GPD se aproximam, superando a distribuição MEV, que apresentou menor acurácia em algumas situações, superestimando eventos com alto período de retorno.

Pelo fato de usar todos os dados disponíveis, em contraste com a metodologia clássica que descarta uma grande parcela de observações, a distribuição MEV apresenta um menor grau de incertezas em suas estimativas. Além disso, por considerar a variação interanual dos eventos em sua modelagem, mostrou-se capaz para caracterizar séries características do contexto atual com mudanças abruptas e tendências lineares. Ademais, a MEV obteve desempenho superior às metodologias clássicas para caracterizar os eventos extremos de

precipitação da área de estudo, apresentando reduzidos valores na métrica de erro (RMSE), especialmente para a situação em que a disponibilidade de dados é baixa, onde há necessidade de caracterizar eventos presentes na faixa de extrapolação.

Dessa forma, de acordo com os resultados obtidos, pode-se indicar que: (i) a consideração da capacidade preditiva dos modelos é fundamental para o processo de caracterização, visto que em muitas situações os modelos são utilizados para caracterizar eventos presentes na faixa de extrapolação e que essa capacidade não é avaliada de acordo com os testes preditivos. (ii) A avaliação da representatividade dos dados, analisando estimativas feitas por séries antigas e recentes, pode aprimorar o processo de caracterização e ampliar a visão dos tomadores de decisão em relação as possíveis consequências das alterações existentes nas séries observadas, visto que a baixa disponibilidade de dados impossibilita, em muitas situações, que as oscilações e periodicidades de longo prazo sejam captadas. (iii) A distribuição MEV é capaz de caracterizar séries em um contexto de mudanças climáticas, com alterações temporais como as mudanças abruptas e tendências lineares, se apresentando como uma alternativa às metodologias tradicionais para situações com baixa disponibilidade de dados, por ter apresentando menor sensibilidade à modelagem baseada em um tamanho amostral reduzido.

O processo de caracterização de eventos extremos não é uma tarefa simples, sendo considerado, ainda, como um grande desafio nas práticas de Engenharia. As diferentes etapas desse processo podem interferir significativamente nas estimativas realizadas, com maior potencial de alteração no atual cenário climático. O presente estudo demonstrou que todas as etapas da caracterização afetam seu desempenho. Em síntese, foram mostradas alternativas aos procedimentos comumente realizados, buscando fornecer um avanço técnico-científico para as práticas de gerenciamento e dimensionamento de sistemas hidráulicos. As análises aqui propostas representam uma pequena contribuição para a Hidrologia, podendo ser modificadas e aperfeiçoadas para uso em outros problemas, como por exemplo, para a caracterização de eventos extremos de cheia.

O trabalho se limitou a analisar a caracterização feita com base em dados observados diários e modelos estacionários. Dessa maneira, a falta de (i) utilização de modelos não estacionários e (ii) dados provenientes de sensoriamento remoto aparecem como uma das principais limitações deste estudo. A indisponibilidade de longas séries de dados sub-diários impossibilitou a análise de chuvas intensas considerando diferentes durações (prática frequentemente exigida para o planejamento urbano). Além disso, o trabalho focou na análise univariada dos eventos extremos de precipitação, não levando em conta modelos que

consideram diferentes variáveis climáticas, como a análise multivariada realizada através das funções Copulas ou modelos não estacionários e suas diferentes covariáveis. Essas limitações, associadas a novas hipóteses e ideias, aparecem como uma oportunidade para a continuação da pesquisa, podendo aperfeiçoar os resultados aqui obtidos e avançar tecnicamente nos processos de planejamento e gestão urbana.

REFERÊNCIAS

ALAM, M. A. et al. Best-fit probability distributions and return periods for maximum monthly rainfall in Bangladesh. **Climate**, v. 6, n. 1, 2018.

AMMAR, M. E. et al. Future floods using hydroclimatic simulations and peaks over threshold: An alternative to nonstationary analysis inferred from trend tests. **Advances in Water Resources**, n. November, p. 103463, 2019.

ASADIEH, B.; KRAKAUER, N. Y. Global trends in extreme precipitation: Climate models versus observations. **Hydrology and Earth System Sciences**, v. 19, n. 2, p. 877–891, 2015.

BARATTI, E. et al. Estimating the flood frequency distribution at seasonal and annual time scales. **Hydrology and Earth System Sciences**, v. 16, n. 12, p. 4651–4660, 2012.

BARTIKO, D. et al. Spatial and seasonal patterns of flood change across Brazil. **Hydrological Sciences Journal**, v. 64, n. 9, p. 1071–1079, 2019.

BATISTA, M. L.; COELHO, G. Precipitação Máxima Diária Anual Na Região Sudeste Do Brasil : Distribuição De Probabilidade E Análise Espacial. p. 77, 2013.

BESKOW, S. et al. Multiparameter probability distributions for heavy rainfall modeling in extreme southern Brazil. **Journal of Hydrology: Regional Studies**, v. 4, n. PB, p. 123–133, 2015.

BLAIN, G. C. Cento e vinte anos de totais extremos de precipitação pluvial máxima diária em campinas, estado de são paulo: Análises estatísticas. **Bragantia**, v. 70, n. 3, p. 722–728, 2011.

BLÖSCHL, G. et al. Twenty-three unsolved problems in hydrology (UPH)—a community perspective. **Hydrological Sciences Journal**, v. 64, n. 10, p. 1141–1158, 2019.

BRUNSELL, N. A. A multiscale information theory approach to assess spatial-temporal variability of daily precipitation. **Journal of Hydrology**, v. 385, n. 1–4, p. 165–172, 2010.

BURNHAM, K. P.; ANDERSON, D. R. Multimodel inference: Understanding AIC and BIC in model selection. **Sociological Methods and Research**, v. 33, n. 2, p. 261–304,

2004.

CALDEIRA, T. L. et al. Probabilistic modelling of extreme rainfall events in the Rio Grande do Sul state. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 19, n. 3, p. 197–203, 2015.

CARVALHO, J. R. P. DE et al. Annual maximum daily rainfall trends in the midwest, southeast and southern Brazil in the last 71 years. **Weather and Climate Extremes**, v. 5, n. 1, p. 7–15, 2014.

CHAI, T.; DRAXLER, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. **Geoscientific Model Development**, v. 7, n. 3, p. 1247–1250, 2014.

CHENG, L. et al. Non-stationary extreme value analysis in a changing climate. **Climatic Change**, v. 127, n. 2, p. 353–369, 2014.

CHENG, L.; AGHAKOUCHAK, A. Nonstationary precipitation intensity-duration-frequency curves for infrastructure design in a changing climate. **Scientific Reports**, v. 4, p. 1–6, 2014.

COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v. 20, n. 1, p. 37–46, 1960.

COLES, S. **An Introduction to Statistical Modeling of Extreme Values**. London: Springer, 2001.

DE LA CASA, A.; NASELLO, O. Breakpoints in annual rainfall trends in Córdoba, Argentina. **Atmospheric Research**, v. 95, n. 4, p. 419–427, 2010.

DEGAETANO, A. T. Time-dependent changes in extreme-precipitation return-period amounts in the continental united states. **Journal of Applied Meteorology and Climatology**, v. 48, n. 10, p. 2086–2099, 2009.

DEGAETANO, A. T.; CASTELLANO, C. M. Recent and future changes in extreme rainfall in the Catskills region of New York. **Annals of the New York Academy of Sciences**, v. 1298, n. 1, p. 43–51, 2013.

EFRON, B.; TIBSHIRANI, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. **Statistical Science**, v. 1, n. 1, p. 54–77,

1986.

EWEMOJE, T. A.; EWEMOOJE, O. S. Best distribution and plotting positions of daily maximum flood estimation at Ona River in Ogun-Oshun river basin, Nigeria. **Agricultural Engineering International: CIGR Journal**, v. 13, n. 3, p. 1–11, 2011.

FISCHER, E. M.; KNUTTI, R. Anthropogenic contribution to global occurrence of heavy-precipitation and high-temperature extremes. **Nature Climate Change**, v. 5, n. 6, p. 560–564, 2015.

FISCHER, E. M.; KNUTTI, R. Observed heavy precipitation increase confirms theory and early models. **Nature Climate Change**, v. 6, n. 11, p. 986–991, 2016.

GAO, P. et al. Changes in streamflow and sediment discharge and the response to human activities in the middle reaches of the Yellow River. **Hydrology and Earth System Sciences**, v. 15, n. 1, p. 1–10, 2011.

GARDIMAN JUNIOR, B. S. et al. Análise de técnicas de interpolação para espacialização da precipitação pluvial na bacia do rio Itapemirim (ES) / Analysis of interpolation techniques for spatial rainfall distribution in river basin Itapemirim (ES). **Revista Ambiente**, v. 8, n. 1, p. 61–71, 2012.

GAUME, E. Flood frequency analysis: The Bayesian choice. **Wiley Interdisciplinary Reviews: Water**, v. 5, n. 4, p. e1290, 2018.

GUPTA, H. V. et al. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. **Journal of Hydrology**, v. 377, n. 1–2, p. 80–91, 2009.

HAKTANIR, T.; COBANER, M.; GORKEMLI, B. Assessment of right-tail prediction ability of some distributions by monte carlo analyses. **Journal of Hydrologic Engineering**, v. 18, n. 5, p. 499–517, 2013.

HAMED, K. A.; RAO, A. R. A modified Mann-Kendall trend test for autocorrelated data. **Journal of Hydrology**, v. 204, n. 1–4, p. 182–196, 1998.

HOSKING, J. R. M. L-Moments: Analysis and Estimation of Distributions Using Linear Combination of Order Statistics. **Journal of the Royal Statistical Society**, v. 52, n. 1, p. 105–124, 1990.

HOSKING, J. R. M.; WALLIS, J. R. **Regional frequency analysis**. Cambridge, United Kingdom: Cambridge University Press, 1997.

ILIOPOULOU, T.; KOUTSOYIANNIS, D. Projecting the future of rainfall extremes: Better classic than trendy. **Journal of Hydrology**, v. 588, 2020.

JACKSON, E. K. et al. Introductory overview: Error metrics for hydrologic modelling – A review of common practices and an open source library to facilitate use and adoption. **Environmental Modelling and Software**, v. 119, n. May, p. 32–48, 2019.

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, p. 158–171, 1955.

JEON, J. J.; SUNG, J. H.; CHUNG, E. S. Abrupt change point detection of annual maximum precipitation using fused lasso. **Journal of Hydrology**, v. 538, p. 831–841, 2016.

KAO, S. C.; GOVINDARAJU, R. S. A bivariate frequency analysis of extreme rainfall with implications for design. **Journal of Geophysical Research Atmospheres**, v. 112, n. D13119, 2007.

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.

KOUTSOYIANNIS, D. Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. **Hydrological Sciences Journal**, v. 49, n. 4, p. 575–590, 2004.

KOUTSOYIANNIS, D. **Probability and statistics for geophysical processes**. Athens, Greece: National Technical University of Athens, 2008.

KOUTSOYIANNIS, D.; MONTANARI, A. Negligent killing of scientific concepts: the stationarity case. **Hydrological Sciences Journal**, v. 60, n. 7–8, p. 1174–1183, 2015.

KRAUSE, P.; BOYLE, D. P.; BÄSE, F. Comparison of different efficiency criteria for hydrological model assessment. **Advances in Geosciences**, v. 5, p. 89–97, 2005.

KUMAR, S. et al. Evaluation of temperature and precipitation trends and long-term persistence in CMIP5 twentieth-century climate simulations. **Journal of Climate**, v. 26, n. 12, p. 4168–4185, 2013.

LAIO, F.; DI BALDASSARRE, G.; MONTANARI, A. Model selection techniques for the frequency analysis of hydrological extremes. **Water Resources Research**, v. 45, n. 7, p. 1–11, 2009.

LANG, M.; OUARDA, T. B. M. J.; BOBÉE, B. Towards operational guidelines for over-threshold modeling. **Journal of Hydrology**, v. 225, n. 3–4, p. 103–117, 1999.

LINS, H. F.; COHN, T. A. Stationarity: Wanted dead or alive? **Journal of the American Water Resources Association**, v. 47, n. 3, p. 475–480, 2011.

LIU, Y. R. et al. Development of a Bayesian-copula-based frequency analysis method for hydrological risk assessment – The Naryn River in Central Asia. **Journal of Hydrology**, v. 580, n. September 2019, 2020.

LUKE, A. et al. Predicting nonstationary flood frequencies: Evidence supports an updated stationarity thesis in the United States. **Water Resources Research**, v. 53, p. 5469–5494, 2017.

LY, S.; CHARLES, C.; DEGRÉ, A. Geostatistical interpolation of daily rainfall at catchment scale: The use of several variogram models in the Ourthe and Ambleve catchments, Belgium. **Hydrology and Earth System Sciences**, v. 15, n. 7, p. 2259–2274, 2011.

MADSEN, H.; RASMUSSEN, P. F.; ROSBJERG, D. Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events. 1: At-site modeling. **Water Resources Research**, v. 33, n. 4, p. 747–757, 1997.

MAKKONEN, L. Plotting positions in extreme value analysis. **Journal of Applied Meteorology and Climatology**, v. 45, n. 3, p. 334–340, 2006.

MARANI, M.; IGNACCOLO, M. A metastatistical approach to rainfall extremes. **Advances in Water Resources**, v. 79, p. 121–126, 2015.

MARANI, M.; ZANETTI, S. Long-term oscillations in rainfall extremes in a 268 year daily time series. **Water Resources Research**, v. 21, p. 639–647, 2015.

MARRA, F. et al. Metastatistical Extreme Value analysis of hourly rainfall from short records: Estimation of high quantiles and impact of measurement errors. **Advances in Water Resources**, v. 117, n. March, p. 27–39, 2018.

MARRA, F. et al. A simplified MEV formulation to model extremes emerging from

multiple nonstationary underlying processes. **Advances in Water Resources**, v. 127, n. April, p. 280–290, 2019.

MARRA, F.; BORGA, M.; MORIN, E. A Unified Framework for Extreme Subdaily Precipitation Frequency Analyses Based on Ordinary Events. **Geophysical Research Letters**, v. 47, n. 18, p. 1–8, 2020.

MARTINS, E. S.; STEDINGER, J. R. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. **Water Resources Research**, v. 36, n. 3, p. 737–744, 2000.

MERZ, R.; BLÖSCHL, G. Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. **Water Resources Research**, v. 44, n. 8, p. 1–17, 2008.

MILLY, P. C. D. et al. Stationarity is dead: Whither water management? **Science**, v. 319, n. 5863, p. 573–574, 2008.

MINIUSI, A.; MARANI, M. Estimation of Daily Rainfall Extremes Through the Metastatistical Extreme Value Distribution: Uncertainty Minimization and Implications for Trend Detection. **Water Resources Research**, v. 56, n. 7, p. 1–18, 2020.

MINIUSI, A.; MARANI, M.; VILLARINI, G. Metastatistical Extreme Value Distribution applied to floods across the continental United States. **Advances in Water Resources**, v. 136, n. December 2019, p. 103498, 2020.

MINIUSI, A.; VILLARINI, G.; MARANI, M. Analyses Through the Metastatistical Extreme Value Distribution Identify Contributions of Tropical Cyclones to Rainfall Extremes in the Eastern United States. **Geophysical Research Letters**, v. 47, n. 7, p. 1–9, 2020.

MONTANARI, A.; KOUTSOYIANNIS, D. Modeling and mitigating natural hazards: Stationarity is immortal! **Water Resources Research**, n. 50, p. 9748–9756, 2014.

MORBIDELLI, R. et al. Influence of temporal data aggregation on trend estimation for intense rainfall. **Advances in Water Resources**, v. 122, n. May, p. 304–316, 2018.

MUKHERJEE, S. et al. Increase in extreme precipitation events under anthropogenic warming in India. **Weather and Climate Extremes**, v. 20, n. July 2017, p. 45–53, 2018.

NAGHETTINI, M.; PINTO, E. J. A. **Hidrologia Estatística**. Belo Horizonte, Brazil: CPRM - Serviço Geológico do Brasil, 2007.

NASH, J. E.; SUTCLIFFE, J. V. River flow forecasting through conceptual models Part I - a discussion of principles. **Journal of Hydrology**, v. 10, p. 282–290, 1970.

NGUYEN, T. H. et al. A systematic approach to selecting the best probability models for annual maximum rainfalls – A case study using data in Ontario (Canada). **Journal of Hydrology**, v. 553, p. 49–58, 2017.

NGUYEN, T. H.; NGUYEN, V. T. VAN. Decision-Support Tool for Constructing Robust Rainfall IDF Relations in Consideration of Model Uncertainty. **Journal of Hydrologic Engineering**, v. 24, n. 7, p. 1–10, 2019.

NGUYEN, V. T. VAN; NGUYEN, T. H. Statistical Modeling of Extreme Rainfall Processes (SMExRain): A Decision Support Tool for Extreme Rainfall Frequency Analyses. **Procedia Engineering**, v. 154, p. 624–630, 2016.

OKOLI, K. et al. Model averaging versus model selection: estimating design floods with uncertain river flow data. **Hydrological Sciences Journal**, v. 63, n. 13–14, p. 1913–1926, 2018.

OKOLI, K. et al. Design flood estimation: Exploring the potentials and limitations of two alternative approaches. **Water (Switzerland)**, v. 11, n. 4, 2019a.

OKOLI, K. et al. A systematic comparison of statistical and hydrological methods for design flood estimation. **Hydrology Research**, p. 1–14, 2019b.

ÖZTEKIN, T. Comparison of parameter estimation methods for the three-parameter generalized Pareto distribution. **Turkish Journal of Agriculture and Forestry**, v. 29, n. 6, p. 419–428, 2005.

ÖZTEKIN, T. Wakeby distribution for representing annual extreme and partial duration rainfall series. **Meteorological Applications**, v. 14, p. 381–387, 2007.

PAPALEXIOU, S. M. Unified theory for stochastic modelling of hydroclimatic processes: Preserving marginal distributions, correlation structures, and intermittency. **Advances in Water Resources**, v. 115, p. 234–252, 2018.

PAPALEXIOU, S. M.; AGHAKOUCHAK, A.; FOUFOULA-GEORGIOU, E. A

Diagnostic Framework for Understanding Climatology of Tails of Hourly Precipitation Extremes in the United States. **Water Resources Research**, v. 54, n. 9, p. 6725–6738, 2018.

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.

PEDRON, I. T. et al. Trends and variability in extremes of precipitation in Curitiba – Southern Brazil. **International Journal of Climatology**, v. 37, n. 3, p. 1250–1264, 2017.

PETTITT, A. N. A Non-Parametric Approach to the Change-Point Problem. **Royal Statistical Society**, v. 28, n. 2, p. 126–135, 1979.

POOL, S.; VIS, M.; SEIBERT, J. Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. **Hydrological Sciences Journal**, v. 63, n. 13–14, p. 1941–1953, 2018.

RAGNO, E. et al. Quantifying Changes in Future Intensity-Duration-Frequency Curves Using Multimodel Ensemble Simulations. **Water Resources Research**, v. 54, n. 3, p. 1751–1764, 2018.

RAGNO, E. et al. A generalized framework for process-informed nonstationary extreme value analysis. **Advances in Water Resources**, v. 130, p. 270–282, 2019.

RULFOVÁ, Z. et al. A two-component generalized extreme value distribution for precipitation frequency analysis. **Journal of Hydrology**, v. 534, p. 659–668, 2016.

SADEGH, M.; RAGNO, E.; AGHAKOUCHAK, A. Multivariate Copula Analysis Toolbox (MvCAT): Describing dependence and underlying uncertainty using a Bayesian framework. **Water Resources Research**, v. 53, p. 5166–5183, 2017.

SALAS, J. D.; OBEYSEKERA, J.; VOGEL, R. M. Techniques for assessing water infrastructure for nonstationary extreme events: a review. **Hydrological Sciences Journal**, v. 63, n. 3, p. 325–352, 2018.

SALINAS, J. L. et al. Regional parent flood frequency distributions in Europe - Part 2: Climate and scale controls. **Hydrology and Earth System Sciences**, v. 18, n. 11, p. 4391–4401, 2014a.

SALINAS, J. L. et al. Regional parent flood frequency distributions in Europe - Part 1: Is the GEV model suitable as a pan-European parent? **Hydrology and Earth System Sciences**, v. 18, n. 11, p. 4381–4389, 2014b.

SANGAL, B. P.; BISWAS, A. K. The 3-Parameter Lognormal Distribution and Its Applications in Hydrology. **Water Resources Research**, v. 6, n. 2, p. 505–515, 1970.

SANTOS, E. B.; LUCIO, P. S.; SANTOS E SILVA, C. M. Estimating return periods for daily precipitation extreme events over the Brazilian Amazon. **Theoretical and Applied Climatology**, v. 126, n. 3–4, p. 585–595, 2016.

SAYEMUZZAMAN, M.; JHA, M. K. Seasonal and annual precipitation time series trend analysis in North Carolina, United States. **Atmospheric Research**, v. 137, p. 183–194, 2014.

SCHELLANDER, H.; LIEB, A.; HELL, T. Error Structure of Metastatistical and Generalized Extreme Value Distributions for Modeling Extreme Rainfall in Austria. **Earth and Space Science**, v. 6, n. 9, p. 1616–1632, 2019.

SERINALDI, F.; KILSBY, C. G. Stationarity is undead: Uncertainty dominates the distribution of extremes. **Advances in Water Resources**, v. 77, p. 17–36, 2015.

SERINALDI, F.; LOMBARDO, F.; KILSBY, C. G. All in order: Distribution of serially correlated order statistics with applications to hydrological extremes. **Advances in Water Resources**, v. 144, n. January, p. 103686, 2020.

SILVA, A. T. et al. Construction of confidence intervals for extreme rainfall quantiles. **WIT Transactions on Information and Communication Technologies**, v. 44, p. 293–304, 2012.

SILVA, D. D. DA et al. Chuvas intensas no Estado da Bahia. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 6, n. 2, p. 362–367, 2002.

SIMONOVIC, S. P. et al. A web-based tool for the development of Intensity Duration Frequency curves under changing climate. **Environmental Modelling and Software**, v. 81, p. 136–153, 2016.

SLATER, L. J.; VILLARINI, G. Recent trends in U.S. flood risk. **Geophysical Research Letters**, p. 1–9, 2016.

SOLARI, S. et al. Peaks Over Threshold (POT): a methodology for automatic threshold estimation using goodness of fit p-value. **Water Resources Research**, v. 53, n. 3, p. 2833–2849, 2017.

STRUPCZEWSKI, W. G. et al. On the tails of distributions of annual peak flow. **Hydrology Research**, v. 42, n. 2–3, p. 171–192, 2011.

VIGLIONE, A. et al. Flood frequency hydrology: 3. A Bayesian analysis. **Water Resources Research**, v. 49, n. 2, p. 675–692, 2013.

VILLARINI, G. et al. On the stationarity of annual flood peaks in the continental United States during the 20th century. **Water Resources Research**, v. 45, p. W08417, 2009.

VILLARINI, G. et al. Analyses of seasonal and annual maximum daily discharge records for central Europe. **Journal of Hydrology**, v. 399, n. 3–4, p. 299–312, 2011a.

VILLARINI, G. et al. Examining Flood Frequency Distributions in the Midwest U.S. **Journal of the American Water Resources Association**, v. 47, n. 3, p. 447–463, 2011b.

VIOLA, M. R. et al. Métodos de interpolação espacial para o mapeamento da precipitação pluvial. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 14, n. 9, p. 970–978, 2010.

VOLPI, E. On return period and probability of failure in hydrology. **WIREs Water**, v. 6:e1340, 2019.

VOLPI, E. et al. Save hydrological observations! Return period estimation without data decimation. **Journal of Hydrology**, v. 571, p. 782–792, 2019.

WILSON, P. S.; TOUMI, R. A fundamental probability distribution for heavy rainfall. **Geophysical Research Letters**, v. 32, n. 14, p. 1–4, 2005.

XIE, P. et al. Comparison of different methods for detecting change points in hydroclimatic time series. **Journal of Hydrology**, v. 577, n. May, p. 123973, 2019.

YE, L. et al. The probability distribution of daily precipitation at the point and catchment scales in the United States. **Hydrology and Earth System Sciences**, v. 22, n. 12, p. 6519–6531, 2018.

YUE, S. et al. The influence of autocorrelation on the ability to detect trend in hydrological series. **Hydrological Processes**, v. 16, n. 9, p. 1807–1829, 2002.

ZAGHLOUL, M. et al. Revisiting flood peak distributions: A pan-Canadian investigation. **Advances in Water Resources**, v. 145, n. August, p. 103720, 2020.

ZANDONADI, L. et al. Changes in precipitation extremes in Brazil (Paraná River Basin). **Theoretical and Applied Climatology**, v. 123, n. 3–4, p. 741–756, 2016.

ZORZETTO, E.; BOTTER, G.; MARANI, M. On the emergence of rainfall extremes from ordinary events. **Geophysical Research Letters**, v. 43, p. 8076–8082, 2016.

APÊNDICE A

Dados das estações avaliadas. A melhor distribuição refere-se àquela escolhida pela metodologia proposta. No Capítulo II, utilizamos todas as estações. No Capítulo III, utilizamos estações com no mínimo 50 e 70 anos válidos para as análises de tendência e ponto de mudança, respectivamente. No Capítulo IV, utilizamos todas estações para avaliar a distribuição dos parâmetros Weibull e estações com no mínimo 70 anos para caracterizar a área de estudo.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho da série	Latitude	Longitude	Melhor Distribuição
e1	Adamantina	C8-004	1956	2018	54	-21.688	-51.042	PE3
e2	Adamantina	C8-052	1971	2018	40	-21.449	-51.010	BUR
e3	Aguai	D3-022	1939	2000	49	-22.050	-46.967	GUM
e4	Aguas da Prata	C3-034	1942	2000	44	-21.950	-46.700	GLO
e5	Aguas de Lindoia	D3-024	1962	2000	30	-22.467	-46.633	GUM
e6	Aguas de Santa Barbara	D6-006	1955	2018	47	-22.883	-49.233	BUR
e7	Agudos	D5-041	1943	2018	32	-22.477	-48.994	GUM
e8	Alambari	E4-031	1936	2000	46	-23.550	-47.900	LNO
e9	Alfredo Marcodes	C8-014	1960	2018	65	-21.952	-51.412	BUR
e10	Altair	B6-025	1943	2017	58	-20.525	-49.059	GLO
e11	Altinópolis	C4-039	1942	2018	47	-21.017	-47.400	GUM
e12	Altinópolis	B4-005	1936	2004	63	-20.833	-47.300	PE3
e13	Alto Alegre	C7-011	1939	2018	67	-21.583	-50.167	LP3
e14	Americana	D4-004	1937	2018	64	-22.712	-47.318	GUM
e15	Americo de Campos	B6-033	1959	2018	50	-20.300	-49.767	LP3
e16	Amparo	D3-023	1961	2018	53	-22.715	-46.844	GUM
e17	Amparo	D3-031	1940	2018	61	-22.620	-46.843	GEV
e18	Analândia	D4-035	1937	2018	72	-22.130	-47.668	LP3
e19	Andradina	B8-004	1944	2018	67	-20.910	-51.376	GUM
e20	Angatuba	E5-017	1939	2018	69	-23.488	-48.417	GLO
e21	Angatuba	E5-007	1954	1997	42	-23.550	-48.300	LP3
e22	Angatuba	E5-019	1939	2003	61	-23.452	-48.578	GUM
e23	Angatuba	E5-030	1958	2000	37	-23.517	-48.550	BUR
e24	Anhembi	D5-037	1937	2018	60	-22.794	-48.128	GLO
e25	Anhembi	D5-044	1939	2018	72	-22.685	-48.114	BUR
e26	Anhembi	D5-086	1984	2018	32	-22.665	-48.254	LNO
e27	Aperecida	D2-007	1940	1997	48	-22.817	-45.233	GLO
e28	Aperecida	D2-065	1957	2011	44	-22.950	-45.250	BUR
e29	Apiai	F5-019	1960	2018	52	-24.505	-48.848	GUM
e30	Apiai	F5-005	1952	2000	44	-24.433	-48.850	PE3
e31	Aracatuba	C7-073	1970	2018	32	-21.333	-50.517	GEV
e32	Aracatuba	C7-009	1943	2018	58	-21.200	-50.450	GEV
e33	Aracatuba	C7-003	1957	2018	56	-21.061	-50.466	BUR
e34	Aracoia da Serra	E4-018	1954	2000	41	-23.483	-47.683	LNO
e35	Aramina	B4-032	1942	2000	52	-20.167	-47.833	LP3

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e36	Araraquara	C5-050	1937	2018	74	-21.783	-48.167	BUR
e37	Araraquara	C5-056	1941	2000	57	-21.917	-48.183	BUR
e38	Araraquara	C5-012	1937	2018	37	-21.667	-48.233	GUM
e39	Araraquara	C5-017	1931	2018	72	-21.888	-48.152	GUM
e40	Araras	D4-027	1936	2000	57	-22.300	-47.450	LP3
e41	Araras	D4-029	1936	2018	70	-22.255	-47.376	BUR
e42	Araras	D4-034	1958	2000	38	-22.267	-47.317	GUM
e43	Arealva	D6-087	1970	2018	39	-22.053	-49.046	GEV
e44	Artur Nogueira	D4-099	1970	2018	44	-22.573	-47.155	GUM
e45	Aruja	E3-050	1965	2018	48	-23.383	-46.350	GEV
e46	Aruja	E3-108	1961	2018	53	-23.433	-46.283	GUM
e47	Assis	D7-055	1970	2018	45	-22.672	-50.478	LP3
e48	Assis	D7-020	1966	2018	34	-22.639	-50.401	LNO
e49	Atibaia	E3-074	1960	2018	42	-23.150	-46.550	BUR
e50	Auriflama	B7-045	1970	2018	43	-20.691	-50.551	GUM
e51	Avai	D6-058	1944	2018	42	-22.167	-49.333	LNO
e52	Avare	E5-014	1939	2018	71	-23.099	-48.912	GLO
e53	Bady Bassit	B6-006	1942	2018	61	-20.917	-49.450	LNO
e54	Balsamo	B6-022	1951	2018	61	-20.733	-49.583	GUM
e55	Bananal	D1-001	1937	2018	65	-22.683	-44.317	BUR
e56	Barao de Antonina	E6-010	1937	2000	47	-23.633	-49.550	GEV
e57	Barbosa	C6-003	1931	2018	69	-21.268	-49.942	BUR
e58	Barra Bonita	D5-028	1959	1957	34	-22.517	-48.533	LP3
e59	Barra do Turvo	F5-033	1969	2018	41	-24.756	-48.505	LNO
e60	Barra do Turvo	F5-022	1970	2012	41	-25.061	-48.558	LP3
e61	Barretos	B5-002	1936	2018	62	-20.583	-48.552	LNO
e62	Barretos	B5-012	1946	2018	59	-20.628	-48.811	LNO
e63	Barrinha	C5-115	1954	2018	47	-21.174	-48.118	GUM
e64	Barueri	E3-031	1937	2017	62	-23.550	-46.867	LP3
e65	Bastos	C7-062	1946	2018	59	-21.925	-50.712	BUR
e66	Batatais	B4-040	1943	2018	61	-20.883	-47.617	BUR
e67	Bauru	D6-036	1939	1999	38	-22.317	-49.033	GUM
e68	Bauru	D6-057	1944	2001	41	-22.300	-49.150	GUM
e69	Bebedouro	B5-036	1943	2017	66	-21.003	-48.549	GLO
e70	Bertioga	E3-106	1948	2001	48	-23.850	-46.133	LP3
e71	Bertioga	E3-042	1938	2017	75	-23.750	-46.133	GLO
e72	Bertioga	E3-040	1940	2017	73	-23.775	-46.111	BUR
e73	Bilac	C7-085	1984	2018	31	-21.388	-50.452	LP3
e74	Birigui	C7-008	1957	2018	42	-21.335	-50.357	GLO
e75	Birigui	C7-028	1939	2018	41	-21.283	-50.333	GUM

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e76	Boa Esperanca do Sul	C5-117	1971	2018	40	-21.991	-48.390	LP3
e77	Boa Esperanca do Sul	C5-016	1931	2018	76	-21.874	-48.473	GUM
e78	Bofete	E5-016	1958	2018	47	-23.105	-48.262	GUM
e79	Bofete	E5-064	1970	2017	37	-23.065	-48.186	BUR
e80	Boituva	E4-046	1941	1998	40	-23.283	-47.667	GEV
e81	Boraceia	D5-035	1961	1997	34	-22.150	-48.750	GEV
e82	Borborema	C6-056	1943	2000	52	-21.700	-49.133	GEV
e83	Borborema	C5-082	1943	2018	68	-21.700	-49.033	GLO
e84	Botucatu	D5-075	1970	2018	36	-22.941	-48.298	GEV
e85	Botucatu	D5-029	1937	2018	65	-22.871	-48.652	GUM
e86	Botucatu	D5-019	1954	2018	61	-22.826	-48.438	GUM
e87	Braganca Paulista	D3-069	1971	2018	35	-22.962	-46.655	GEV
e88	Braganca Paulista	D3-063	1970	2018	37	-22.926	-46.534	BUR
e89	Brauna	C7-016	1939	2018	63	-21.566	-50.318	GUM
e90	Brejo Alegre	C7-086	1979	2018	32	-21.138	-50.193	GUM
e91	Brodosqui	C4-086	1969	2018	40	-21.083	-47.650	GLO
e92	Brotas	D4-040	1938	1998	39	-22.233	-47.983	LNO
e93	Buri	E5-051	1951	2001	44	-23.800	-48.583	GEV
e94	Buritizal	B4-002	1937	1998	60	-20.183	-47.717	LP3
e95	Buritizal	B4-065	1975	2018	35	-20.277	-47.676	LP3
e96	c5-Araraquara	C5-050	1938	2017	74	-21.783	-48.167	BUR
e97	Cabralia Paulista	D6-021	1938	2018	69	-22.450	-49.317	GEV
e98	Cabreuva	E4-030	1964	2003	32	-23.267	-47.117	GUM
e99	Cacapava	E2-048	1957	2018	52	-23.083	-45.667	GEV
e100	Cacapava	E2-002	1960	2001	36	-23.217	-45.733	GUM
e101	Cacapava	E2-113	1969	2012	36	-23.033	-45.783	GUM
e102	Cacapava	E2-034	1942	2018	70	-23.133	-45.717	LNO
e103	Cacapava	E2-001	1959	1999	34	-23.133	-45.750	GEV
e104	Cachoeira Paulista	D2-013	1955	2000	40	-22.667	-45.017	LP3
e105	Caconde	C3-014	1937	2018	63	-21.519	-46.647	GLO
e106	Cafelandia	C6-017	1937	2018	66	-21.813	-49.609	LP3
e107	Caiabu	D8-057	1984	2018	31	-22.017	-51.233	LP3
e108	Caiua	C8-010	1942	2018	54	-21.832	-51.985	LNO
e109	Cajati	F5-028	1950	2018	53	-24.717	-48.117	GLO
e110	Cajuru	C4-088	1951	2000	46	-21.317	-47.367	LNO

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e111	Campina do Monte							
e111	Alegre	E5-018	1939	1998	35	-23.600	-48.450	GLO
e112	Campinas	D4-047	1958	2018	48	-22.836	-47.071	PE3
e113	Campinas	D4-044	1941	2018	72	-22.871	-47.078	BUR
e114	Campinas	D4-046	1958	2018	57	-22.780	-47.038	GEV
e115	Campinas	IAC	1890	2015	124	-22.883	-47.083	LP3
e116	Campinas	D3-002	1931	2006	59	-22.931	-46.899	GLO
e117	Campos do Jordao	D2-005	1937	1997	45	-22.717	-45.533	BUR
e118	Campos do Jordao	D2-001	1936	2017	60	-22.717	-45.567	GUM
e119	Campos Novos Paulista	D6-106	1982	2018	33	-22.600	-50.000	GEV
e120	Cananeia	G5-001	1941	1998	44	-25.217	-48.033	GUM
e121	Cananeia	G4-003	1959	1997	32	-25.100	-47.917	GUM
e122	Cananeia	F4-029	1954	2018	51	-24.933	-47.950	PE3
e123	Cananeia	G4-002	1951	2000	37	-24.967	-47.950	GUM
e124	Canas	D2-037	1957	2000	35	-22.750	-45.050	GLO
e125	Candido Mota	D7-031	1938	2018	74	-22.883	-50.329	GEV
e126	Candido Rodrigues	C5-073	1941	2018	73	-21.326	-48.631	GLO
e127	Canitar	D6-083	1949	2018	53	-23.029	-49.785	LP3
e128	Capao Bonito	F5-041	1961	2001	35	-24.017	-48.267	BUR
e129	Capao Bonito	E5-047	1947	1999	45	-23.950	-48.217	GLO
e130	Capao Bonito	E5-071	1972	2011	33	-23.934	-48.361	LP3
e131	Capivari	D4-069	1946	2018	63	-23.004	-47.509	BUR
e132	Capivari	D4-092	1953	2004	43	-22.876	-47.456	GUM
e133	Caraguatatuba	E2-046	1943	2012	60	-23.633	-45.433	PE3
e134	c - Araraquara	C5-012	1937	2018	37	-21.667	-48.233	GUM
e135	Cardoso	B6-050	1970	2018	44	-20.151	-49.967	BUR
e136	Casa Branca	C4-072	1960	2018	43	-21.783	-47.058	BUR
e137	Casa Branca	C4-026	1936	2018	50	-21.900	-47.033	GUM
e138	Casa Branca	C4-097	1970	2018	42	-21.933	-47.133	BUR
e139	Cassia dos Coqueiros	C4-102	1972	2018	41	-21.283	-47.167	GLO
e140	Castilho	B8-002	1956	2018	51	-20.871	-51.496	GEV
e141	Catigua	C6-053	1941	2018	57	-21.056	-49.062	LNO
e142	Cedral	B6-037	1964	2018	49	-20.905	-49.267	GEV
e143	Cerqueira Cesar	E6-030	1951	2018	55	-23.040	-49.163	BUR
e144	Cesario Lange	E4-049	1939	2001	52	-23.217	-47.950	LP3
e145	Charqueada	D4-054	1939	2000	40	-22.517	-47.783	LP3

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e146	Chavantes	E6-003	1955	2018	59	-23.053	-49.769	GLO
e147	Colina	B5-028	1938	2018	75	-20.745	-48.555	LNO
e148	Colombia	B5-063	1974	2018	37	-20.260	-48.806	LP3
e149	Colombia	B5-061	1972	2018	31	-20.239	-48.653	GEV
e150	Conchal	D4-094	1967	2000	31	-22.333	-47.183	GUM
e151	Conchas	E5-001	1956	2018	51	-23.021	-48.011	BUR
e152	Coroados	C7-024	1944	2018	64	-21.350	-50.283	LNO
e153	Coroados	C7-023	1959	1999	40	-21.174	-50.236	GUM
e154	Coronel Macedo	E6-012	1939	1999	51	-23.633	-49.317	LNO
e155	Corumbatai	D4-043	1939	2018	75	-22.216	-47.620	GUM
e156	Cosmopolis	D4-052	1943	2018	71	-22.661	-47.218	GUM
e157	Cosmorama	B6-023	1941	2018	66	-20.479	-49.774	PE3
e158	Cotia	E3-034	1935	2018	66	-23.650	-46.950	BUR
e159	Cotia	E4-012	1939	2016	53	-23.683	-47.017	GEV
e160	Cotia	E3-027	1936	2018	68	-23.717	-46.967	PE3
e161	Cotia	E3-013	1936	2018	51	-23.567	-46.933	GUM
e162	Cravinhos	C4-007	1936	2018	62	-21.360	-47.711	BUR
e163	Cravinhos	C4-032	1954	2018	57	-21.333	-47.783	GUM
e164	Cristais Paulista	B4-055	1965	2008	36	-20.350	-47.350	BUR
e165	Cruzeiro	D1-006	1937	2000	49	-22.583	-44.983	LNO
e166	Cubatao	E3-101	1947	1992	32	-23.883	-46.417	BUR
e167	Cubatao	E3-038	1937	1996	46	-23.852	-46.376	GUM
e168	Cunha	E1-005	1969	2018	39	-23.183	-44.983	LP3
e169	Cunha	D1-005	1960	2000	33	-22.917	-44.817	GUM
e170	Cunha	E1-001	1958	2006	40	-23.083	-44.900	GLO
e171	Descalvado	C4-004	1936	2000	48	-21.933	-47.617	BUR
e172	Descalvado	C4-021	1939	2000	53	-21.950	-47.700	BUR
e173	Descalvado	C4-071	1955	2018	60	-21.977	-47.703	LP3
e174	Dois Corregos	D5-087	1936	1999	61	-22.367	-48.383	GEV
e175	Dolcinopolis	B7-041	1970	2018	43	-20.121	-50.512	GEV
e176	Dourado	D5-023	1954	2017	60	-22.129	-48.330	LNO
e177	Dourado	D5-053	1941	1999	54	-22.100	-48.300	GLO
e178	Dracena	C8-043	1950	2018	58	-21.467	-51.548	LP3
e179	Duartina	D6-020	1938	2000	57	-22.417	-49.417	GLO
e180	Echaporã	D7-046	1946	2018	58	-22.431	-50.199	BUR

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e181	Eldorado	F5-021	1961	2005	39	-24.700	-48.333	GEV
e182	Eldorado	F5-007	1937	2002	58	-24.517	-48.100	LP3
e183	Eldorado	F5-013	1939	2018	54	-24.596	-48.218	PE3
e184	Eldorado	F5-020	1944	1999	33	-24.600	-48.300	LP3
e185	Eldorado	F5-036	1973	2015	42	-24.480	-48.299	BUR
e186	Elias Fausto	E4-013	1939	2000	45	-23.033	-47.367	GUM
e187	Emilianopolis	C8-025	1953	2018	54	-21.833	-51.480	GEV
e188	Espirito Santo do Pinhal	D3-003	1931	2000	52	-22.283	-46.767	LP3
e189	Euclides da Cunha	D9-004	1966	2018	41	-22.550	-52.583	GEV
e190	Paulista	D9-022	1982	2018	35	-22.467	-52.533	GUM
e191	Fartura	E6-017	1939	2000	48	-23.383	-49.517	GEV
e192	Fernandopolis	B7-005	1959	2018	45	-20.293	-50.242	LP3
e193	Fernandopolis	B7-019	1963	2018	49	-20.230	-50.324	GEV
e194	Fernando Prestes	C5-041	1959	2000	39	-21.267	-48.683	BUR
e195	Fernao	D6-005	1936	2000	52	-22.367	-49.517	LP3
e196	Flora Rica	C8-059	1974	2018	40	-21.738	-51.441	LP3
e197	Florida Paulista	C8-002	1954	2018	48	-21.616	-51.168	BUR
e198	Florida Paulista	C8-030	1947	2018	64	-21.500	-51.150	GUM
e199	Franca	B4-001	1935	2017	63	-20.517	-47.400	GUM
e200	Franco da Rocha	E3-047	1937	2010	53	-23.333	-46.683	GUM
e201	Galia	D6-084	1951	2000	45	-22.417	-49.667	LP3
e202	Galia	D6-019	1938	2018	73	-22.306	-49.531	BUR
e203	Garca	D6-018	1938	2018	76	-22.200	-49.650	BUR
e204	Getulina	C7-009	1939	2018	67	-21.750	-50.117	GUM
e205	Guaicara	C6-008	1937	2018	65	-21.598	-49.805	GLO
e206	Guaimbe	C6-078	1944	2018	61	-21.788	-49.819	BUR
e207	Guaimbe	C6-071	1946	2018	51	-21.914	-49.898	LNO
e208	Guaira	B5-040	1968	2018	39	-20.350	-48.300	BUR
e209	Guaira	B5-015	1959	1999	33	-20.200	-48.500	LNO
e210	Guapiacu	B6-056	1976	2018	36	-20.800	-49.217	GEV
e211	Guapiara	F5-010	1939	1998	52	-24.183	-48.500	LP3
e212	Guaracai	C8-022	1943	2018	62	-21.027	-51.192	PE3
e213	Guara	B4-035	1943	2018	41	-20.417	-47.817	GEV
e214	Guarani D'Oeste	B7-014	1959	2018	55	-20.078	-50.342	GUM
e215	Guaranta	C6-051	1954	2018	50	-21.888	-49.528	BUR

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e216	Guaranta	C6-002	1958	2001	32	-21.917	-49.650	LP3
e217	Guararapes	C7-037	1950	2018	55	-21.350	-50.683	GUM
e218	Guararapes	C7-012	1962	2000	34	-21.250	-50.717	GLO
e219	Guararapes	C7-033	1944	2018	57	-21.267	-50.633	LNO
e220	Guararapes	B7-048	1970	2018	40	-21.071	-50.711	LP3
e221	Guararema	E3-094	1943	2010	48	-23.383	-46.067	GUM
e222	Guararema	E3-054	1957	2018	36	-23.317	-46.117	BUR
e223	Guaratingueta	D2-069	1957	1999	33	-22.717	-45.250	GEV
e224	Guaratingueta	D2-009	1956	2015	40	-22.900	-45.150	BUR
e225	Guaratingueta	D2-076	1957	2017	39	-22.750	-45.233	GUM
e226	Guaratingueta	D2-015	1957	2002	36	-22.733	-45.200	GLO
e227	Guaratingueta	D2-077	1957	2002	38	-22.717	-45.350	GEV
e228	Guaratingueta	D2-079	1958	2001	37	-22.783	-45.350	GUM
e229	Guaratingueta	D2-031	1957	1999	42	-22.833	-45.167	LP3
e230	Guaratingueta	D2-063	1957	1996	34	-22.700	-45.283	LNO
e231	Guarei	E5-027	1959	2018	50	-23.361	-48.176	GUM
e232	Guariba	C5-112	1970	2018	36	-21.353	-48.209	LNO
e233	Guaruja	E3-043	1937	2018	54	-23.950	-46.183	LP3
e234	Guaruja	E3-070	1937	2018	62	-24.000	-46.283	GUM
e235	Guaruja	E3-045	1942	2002	37	-23.933	-46.283	LP3
e236	Guarulhos	E3-002	1936	2018	71	-23.417	-46.400	LNO
e237	Guatapara	C5-020	1937	1999	56	-21.500	-48.033	GUM
e238	Holambra	D4-082	1950	2018	62	-22.633	-47.052	GUM
e239	Iacanga	C5-029	1937	2000	56	-21.900	-49.033	GUM
e240	Iacri	C7-046	1946	2000	47	-21.800	-50.617	PE3
e241	Iacri	C7-054	1946	2018	54	-21.870	-50.676	LNO
e242	Ibate	C5-035	1939	1999	54	-21.950	-48.000	GEV
e243	Ibira	C6-100	1972	2018	39	-21.108	-49.156	LP3
e244	Ibirarema	D7-012	1937	2000	44	-22.817	-50.067	GUM
e245	Ibitinga	C5-081	1943	2018	59	-21.785	-48.780	GLO
e246	Ibitinga	C5-101	1950	2000	41	-21.850	-48.800	GEV
e247	Ibitinga	C5-038	1939	2000	48	-21.783	-48.850	LNO
e248	Ibiuna	E4-116	1969	2018	43	-23.950	-47.217	GUM
e249	Ibiuna	E4-047	1939	2018	64	-23.650	-47.217	GLO
e250	Ibiuna	E4-131	1972	2018	36	-23.717	-47.117	LP3

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e251	Icem	B6-001	1936	2018	72	-20.341	-49.202	GUM
e252	Iepe	D8-006	1960	2018	47	-22.717	-51.133	GUM
e253	Iepe	D8-013	1944	2018	61	-22.667	-51.083	BUR
e254	Igarapava	B4-029	1942	2000	53	-20.000	-47.817	LP3
e255	Igarapava	B4-026	1942	2000	49	-20.033	-47.750	GEV
e256	Igarata	E3-242	1972	2018	43	-23.200	-46.150	LNO
e257	Iguape	F4-048	1968	2007	36	-24.483	-47.250	LNO
e258	Iguape	F4-028	1951	2018	62	-24.700	-47.567	LNO
e259	Iguape	F4-034	1969	2018	41	-24.533	-47.533	GEV
e260	Iguape	F4-030	1954	2002	39	-24.817	-47.733	GUM
e261	Ilhabela	E2-012	1944	2017	54	-23.783	-45.350	BUR
e262	Ilha Comprida	F4-014	1960	1996	31	-24.900	-47.800	LNO
e263	Indaiatuba	E4-124	1970	2018	45	-23.167	-47.133	GUM
e264	Indaiatuba	E4-015	1937	2018	60	-23.083	-47.217	GLO
e265	Indiana	D8-047	1950	2018	58	-22.155	-51.253	GLO
e266	Indiaporá	A7-002	1959	2001	39	-19.983	-50.250	LP3
e267	Ipaussu	E6-007	1941	2018	69	-23.029	-49.568	BUR
e268	Ipaussu	E6-021	1943	2000	49	-23.050	-49.617	LP3
e269	Ipero	E4-019	1957	2018	58	-23.333	-47.683	GUM
e270	Ipeuna	D4-074	1970	2018	40	-22.437	-47.723	GUM
e271	Iporanga	F5-031	1969	2018	44	-24.554	-48.429	GUM
e272	Iporanga	F5-032	1969	2018	35	-24.433	-48.583	LNO
e273	Ipuá	B5-052	1970	2018	45	-20.450	-48.017	LP3
e274	Irapuru	C8-042	1949	2018	62	-21.567	-51.350	BUR
e275	Itabera	F6-003	1939	2018	52	-24.050	-49.083	GLO
e276	Itabera	E6-020	1939	1999	47	-23.867	-49.133	GEV
e277	Itai	E6-022	1954	2000	43	-23.417	-49.100	GUM
e278	Itajobi	C6-040	1940	2018	67	-21.311	-49.057	GUM
e279	Itaju	C5-055	1940	1999	52	-21.983	-48.800	LP3
e280	Itanhaém	F3-005	1938	2018	48	-24.183	-46.795	LNO
e281	Itaoca	F5-011	1962	2000	35	-24.650	-48.833	GUM
e282	Itapeçerica da Serra	E3-014	1936	2018	48	-23.783	-46.917	BUR
e283	Itapetininga	E5-034	1947	2000	41	-23.767	-48.133	BUR
e284	Itapetininga	E5-015	1939	2018	77	-23.579	-48.043	GLO
e285	Itapetininga	E5-069	1945	2001	41	-23.467	-48.050	LP3

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e286	Itapetininga	E5-066	1970	2016	43	-23.783	-48.356	LNO
e287	Itapeva	E5-045	1946	2018	58	-23.963	-48.949	BUR
e288	Itapeva	F5-012	1939	1999	54	-24.017	-48.717	LP3
e289	Itapira	D3-041	1942	2000	49	-22.333	-46.717	LP3
e290	Itapira	D3-015	1940	2018	72	-22.442	-46.821	PE3
e291	Itapirapua Paulista	F6-007	1970	2016	33	-24.567	-49.167	LP3
e292	Itapolis	C5-093	1970	2018	41	-21.583	-48.817	GUM
e293	Itapolis	C5-119	1972	2018	43	-21.600	-48.967	GEV
e294	Itaporanga	E6-008	1959	2006	37	-23.709	-49.491	LP3
e295	Itapui	D5-039	1937	1999	58	-22.317	-48.717	GUM
e296	Itaquaquecetuba	E3-147	1951	1994	36	-23.483	-46.350	LNO
e297	Itaquaquecetuba	E3-091	1943	2018	61	-23.483	-46.367	GUM
e298	Itariri	F4-026	1953	1994	33	-24.300	-47.067	LP3
e299	Itariri	F4-006	1959	2018	52	-24.267	-47.167	PE3
e300	Itatiba	E3-105	1939	2018	42	-22.984	-46.434	GLO
e301	Itatinga	E5-073	1973	2017	38	-23.111	-48.597	GLO
e302	Itatinga	D5-040	1960	2017	48	-22.997	-48.685	GEV
e303	Itirapina	D4-036	1937	2018	70	-22.300	-47.744	GEV
e304	Itirapina	D4-014	1936	1999	42	-22.233	-47.800	GLO
e305	Itirapina	D4-033	1937	1997	51	-22.167	-47.900	LNO
e306	Itirapina	D4-037	1937	2017	70	-22.148	-47.795	GUM
e307	Itirapua	B4-053	1953	2018	50	-20.633	-47.217	GUM
e308	Itu	E4-023	1936	2000	50	-23.333	-47.333	GUM
e309	Itupeva	E4-062	1963	2002	31	-23.083	-47.050	LNO
e310	Ituverava	B4-048	1945	2000	35	-20.217	-47.900	PE3
e311	Ituverava	B4-034	1943	2018	55	-20.340	-47.773	LP3
e312	Jaborandi	B5-029	1941	2018	70	-20.678	-48.411	GEV
e313	Jaborandi	B5-037	1943	2018	59	-20.584	-48.370	BUR
e314	Jaboticabal	C5-096	1950	2000	42	-21.183	-48.183	BUR
e315	Jaboticabal	C5-028	1963	2018	49	-21.333	-48.316	GEV
e316	Jacarei	E2-031	1942	2017	50	-23.283	-45.950	BUR
e317	Jacupiranga	F4-022	1960	2004	36	-24.800	-47.967	LP3
e318	Jacupiranga	F4-017	1939	2018	48	-24.717	-48.017	GLO
e319	Jaguariuna	D3-042	1942	2018	61	-22.701	-46.982	BUR
e320	Jales	B7-008	1959	2018	46	-20.306	-50.546	GLO

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e321	Jambeiro	E2-025	1943	2016	36	-23.250	-45.683	GUM
e322	Jambeiro	E2-037	1943	1994	49	-23.283	-45.783	GEV
e323	Jardinópolis	C4-054	1943	2018	62	-21.017	-47.767	BUR
e324	Jarinu	E3-154	1953	2018	52	-23.000	-46.717	GLO
e325	Jau	D5-007	1936	1999	39	-22.302	-48.540	LNO
e326	Jau	D5-084	1976	2018	39	-22.225	-48.599	GEV
e327	Jeriquara	B4-023	1939	2001	41	-20.317	-47.583	GEV
e328	Joanópolis	D3-019	1940	1993	35	-22.967	-46.100	GEV
e329	Joanópolis	D3-054	1952	2018	55	-22.932	-46.268	GUM
e330	Jose Bonifacio	C6-088	1970	2018	44	-21.060	-49.684	LNO
e331	Julio Mesquita	C6-023	1941	2000	47	-21.917	-49.783	PE3
e332	Jumirim	E4-053	1941	1999	52	-23.150	-47.800	GEV
e333	Jundiaí	E3-053	1957	2018	41	-23.200	-46.983	GEV
e334	Juquia	F4-007	1956	2017	58	-24.167	-47.650	LP3
e335	Juquia	F4-002	1958	2018	53	-24.100	-47.683	GUM
e336	Juquia	F4-007	1958	2018	55	-24.167	-47.650	BUR
e337	Juquia	F4-037	1961	2018	47	-24.350	-47.717	GUM
e338	Juquitiba	E4-059	1947	2017	44	-23.933	-47.067	BUR
e339	Lagoinha	E2-007	1959	2006	42	-23.117	-45.200	BUR
e340	Lagoinha	E2-049	1948	1999	46	-23.083	-45.200	GLO
e341	Laranjal Paulista	E4-050	1943	2018	63	-23.033	-47.850	GLO
e342	Laranjal Paulista	D4-022	1956	1995	38	-22.917	-47.900	LNO
e343	Lavinia	C7-079	1955	2001	44	-21.000	-50.917	GLO
e344	Leme	D4-030	1936	2018	80	-22.161	-47.259	PE3
e345	Lencois Paulista	D5-081	1972	2017	40	-22.772	-48.877	GUM
e346	Lencois Paulista	D5-072	1951	1997	37	-22.600	-48.783	GEV
e347	Limeira	D4-064	1965	2018	38	-22.598	-47.376	GEV
e348	Lindóia	D3-014	1936	2017	48	-22.526	-46.651	PE3
e349	Lins	C6-015	1937	2005	50	-21.704	-49.689	LNO
e350	Lorena	D2-062	1957	1996	35	-22.767	-45.117	GUM
e351	Lorena	D2-035	1957	2018	54	-22.733	-45.083	LNO
e352	Lucélia	C8-018	1963	2018	49	-21.729	-51.011	LP3
e353	Luis Antonio	C4-056	1950	2018	64	-21.446	-47.896	BUR
e354	Luis Antonio	C4-096	1970	2018	34	-21.583	-47.700	GUM

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e355	Luisiania	C7-064	1946	2018	63	-21.697	-50.295	LP3
e356	Lutecia	D7-007	1961	2018	44	-22.338	-50.390	BUR
e357	Lutecia	D7-066	1972	2018	39	-22.461	-50.369	GUM
e358	Macedonia	B7-012	1959	2018	51	-20.195	-50.197	GUM
e359	Mairinque	E4-043	1942	2018	57	-23.433	-47.250	GUM
e360	Mairinque	E4-041	1940	2018	37	-23.550	-47.183	LNO
e361	Mairipora	E3-005	1937	1997	53	-23.317	-46.583	GLO
e362	Maraba Paulista	D8-035	1970	2018	43	-22.100	-51.967	GUM
e363	Maracai	D7-041	1944	2000	34	-22.667	-50.617	GEV
e364	Marilia	D6-098	1972	2018	39	-22.133	-49.917	BUR
e365	Marilia	D6-025	1939	2018	72	-22.213	-49.934	BUR
e366	Martinopolis	D8-052	1972	2018	41	-22.417	-51.167	GEV
e367	Martinopolis	C8-054	1971	2018	40	-22.001	-51.136	GEV
e368	Martinopolis	D8-041	1945	2018	68	-22.250	-51.167	BUR
e369	Matao	C5-074	1941	2018	72	-21.600	-48.350	GLO
e370	Maua	E4-148	1952	2004	38	-23.667	-46.483	LNO
e371	Maua	E3-237	1971	2007	31	-23.700	-46.483	PE3
e372	Mendonca	C6-036	1952	2018	59	-21.184	-49.583	GLO
e373	Mineiros do Tiete	D5-022	1937	1999	40	-22.417	-48.450	GLO
e374	Miracatu	F4-036	1961	2008	37	-24.267	-47.400	GEV
e375	Miracatu	F4-012	1965	2017	30	-24.149	-47.316	LP3
e376	Mirandopolis	C8-056	1958	2018	53	-21.014	-51.102	LP3
e377	Mirandopolis	C8-058	1972	2018	42	-21.134	-51.101	LP3
e378	Mirandopolis	B8-011	1958	2018	53	-20.934	-51.107	GUM
e379	Mirandopolis	B8-012	1958	1999	39	-20.917	-51.017	GEV
e380	Mirante do Paranapanema	D9-006	1970	2018	33	-22.300	-52.083	GEV
e381	Mirante do Paranapanema	D8-008	1955	2018	58	-22.300	-51.917	GEV
e382	Mirassol	B6-002	1937	2018	66	-20.796	-49.522	BUR
e383	Mirassolandia	B6-007	1965	2018	40	-20.618	-49.468	GEV
e384	Mococa	C3-040	1940	2000	57	-21.433	-46.850	PE3
e385	Mococa	C4-077	1961	1997	35	-21.633	-47.017	LP3
e386	Mococa	C4-040	1940	2000	57	-21.450	-47.000	GUM
e387	Mococa	C4-069	1955	2000	36	-21.383	-47.083	GLO
e388	Moji das Cruzes	E3-097	1961	2000	33	-23.517	-46.200	BUR
e389	Moji das Cruzes	E3-032	1937	1999	42	-23.583	-46.233	GUM
e390	Moji Mirim	D3-033	1941	2000	35	-22.450	-46.950	LP3

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e391	Moji Mirim	D3-009	1937	2017	59	-22.529	-46.953	PE3
e392	Moji Mirim	D3-008	1936	2018	43	-22.411	-46.944	BUR
e393	Mongagua	F3-002	1937	2018	49	-24.083	-46.617	GEV
e394	Monte Alegre do Sul	D3-027	1942	2018	73	-22.696	-46.672	LNO
e395	Monteiro Lobato	D2-020	1939	2018	69	-22.933	-45.833	LNO
e396	Monteiro Lobato	D2-026	1943	2000	52	-22.867	-45.783	GLO
e397	Monte Mor	D4-083	1951	2018	45	-22.944	-47.281	GUM
e398	Morro Agudo	B5-050	1967	2018	43	-20.618	-47.998	LNO
e399	Morro Agudo	B5-004	1940	2018	73	-20.726	-48.056	LNO
e400	Morungaba	D3-046	1947	2018	60	-22.873	-46.792	LP3
e401	Motuca	C5-009	1936	2000	60	-21.517	-48.150	LP3
e402	Monte Alto	C5-070	1941	2018	63	-21.259	-48.496	BUR
e403	Monte Aprazível	B6-039	1968	2018	43	-20.768	-49.701	BUR
e404	Monte Azul Paulista	B5-033	1943	2018	57	-20.900	-48.633	GUM
e405	Monte Castelo	C8-008	1958	2017	55	-21.300	-51.570	LNO
e406	Monte Verde Paulista	B5-035	1943	2018	69	-20.842	-48.793	GLO
e407	Murutinga do Sul	C8-051	1971	2018	33	-21.087	-51.331	GUM
e408	Narandiba	D8-040	1944	2018	63	-22.417	-51.517	BUR
e409	Natividade da Serra	E2-136	1972	2018	37	-23.571	-45.458	PE3
e410	Natividade da Serra	E2-024	1940	1999	53	-23.467	-45.350	GUM
e411	Natividade da Serra	E2-008	1940	2018	72	-23.383	-45.450	GLO
e412	Nazare Paulista	E3-099	1946	2018	68	-23.183	-46.400	LNO
e413	Nipoa	B6-034	1959	2018	58	-20.917	-49.783	GUM
e414	Nova Alianca	C6-031	1954	2000	38	-21.083	-49.550	BUR
e415	Nova Campina	F5-017	1939	2001	54	-24.133	-48.967	GUM
e416	Nova Europa	C5-048	1940	2018	70	-21.804	-48.603	PE3
e417	Novais	B5-003	1936	2001	57	-20.983	-48.917	LNO
e418	Nuporanga	B4-003	1937	2003	64	-20.650	-47.683	LNO
e419	Ocaucu	D6-094	1971	2018	45	-22.433	-49.917	GEV
e420	Oleo	D6-003	1944	2018	57	-22.950	-49.383	LNO
e421	Olimpia	B5-020	1938	2018	73	-20.752	-48.914	BUR
e422	Olimpia	B5-009	1943	2018	68	-20.609	-48.959	BUR
e423	Onda Verde	B6-003	1937	2018	74	-20.630	-49.308	GUM
e424	Onda Verde	B6-026	1943	2018	67	-20.617	-49.300	PE3
e425	Oriente	D7-032	1938	2000	52	-22.150	-50.090	LNO

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e426	Orindiuva	B6-029	1958	2018	43	-20.181	-49.368	BUR
e427	Orlandia	B4-015	1939	2018	72	-20.746	-47.890	GUM
e428	Osvaldo Cruz	C7-067	1953	2018	52	-21.798	-50.872	GUM
e429	Ourinhos	D6-011	1937	2018	70	-22.983	-49.833	GLO
e430	Ouro Verde	C8-055	1971	2018	33	-21.584	-51.708	LP3
e431	Palestina	B6-030	1958	2018	58	-20.391	-49.441	BUR
e432	Palmares Paulista	C5-084	1965	1999	33	-21.078	-48.800	GUM
e433	Palmeira do Oeste	B7-042	1970	2018	44	-20.411	-50.764	BUR
e434	Palmital	D7-053	1966	2000	30	-22.783	-50.217	GUM
e435	Palmital	D7-071	1974	2018	32	-22.816	-50.310	GEV
e436	Panorama	C8-019	1944	2018	64	-21.353	-51.858	GLO
e437	Paraguacu Paulista	D7-043	1953	2018	54	-22.421	-50.571	GLO
e438	Paraibuna	E2-130	1970	2010	36	-23.533	-45.683	BUR
e439	Paraibuna	E2-026	1943	2018	50	-23.367	-45.683	GEV
e440	Paraibuna	E2-039	1943	2003	47	-23.450	-45.583	GUM
e441	Paraibuna	E2-040	1943	1993	38	-23.500	-45.550	PE3
e442	Paraiso	C5-043	1962	2000	33	-21.017	-48.767	LNO
e443	Parapua	C7-001	1944	2018	66	-21.940	-50.940	BUR
e444	Parapua	C7-036	1945	1999	46	-21.783	-50.783	LP3
e445	Pardinho	E5-060	1970	2018	33	-23.085	-48.375	GEV
e446	Pariquera Acu	F4-018	1942	2018	75	-24.600	-47.883	PE3
e447	Pariquera Acu	F4-035	1961	2000	32	-24.617	-47.833	PE3
e448	Pariquera Acu	F4-016	1939	2018	58	-24.717	-47.883	GLO
e449	Parisi	B7-036	1969	2018	47	-20.305	-50.007	GEV
e450	Patrocinio Paulista	B4-061	1970	2018	43	-20.634	-47.284	GLO
e451	Paulistania	D6-091	1970	2018	46	-22.583	-49.400	GLO
e452	Paulo de Faria	B6-013	1958	2012	30	-20.033	-49.400	LNO
e453	Pederneiras	D5-018	1937	2018	69	-22.312	-48.891	LNO
e454	Pedra Bela	D3-035	1940	2018	47	-22.754	-46.435	LNO
e455	Pedregulho	B4-063	1974	2018	42	-20.283	-47.417	GEV
e456	Pedregulho	B4-037	1943	2000	47	-20.250	-47.483	GUM
e457	Pedreira	D3-052	1946	2017	60	-22.743	-46.938	LP3
e458	Pedro do Toledo	F4-019	1957	2018	50	-24.283	-47.233	GLO
e459	Penapolis	C7-010	1939	2018	67	-21.433	-50.067	GUM
e460	Pereira Barreto	B8-001	1937	2018	67	-20.717	-51.153	GUM

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e461	Pereiras	E4-010	1956	2017	50	-23.067	-47.967	LNO
e462	Peruibe	F4-027	1963	2018	35	-24.317	-47.000	LP3
e463	Piedade	E4-001	1954	2012	42	-23.733	-47.417	PE3
e464	Piedade	E4-130	1971	2017	41	-23.750	-47.483	LNO
e465	Pilar do Sul	E4-028	1937	2018	68	-23.833	-47.650	GEV
e466	Pindamonhangaba	D2-071	1957	1997	35	-22.967	-45.417	GUM
e467	Pindamonhangaba	D2-004	1936	2000	42	-22.833	-45.633	GLO
e468	Pindamonhangaba	D2-072	1957	2010	42	-22.817	-45.367	GEV
e469	Pindamonhangaba	D2-075	1958	1999	37	-22.767	-45.450	GUM
e470	Pindamonhangaba	D2-068	1957	2009	41	-22.767	-45.517	GLO
e471	Pindamonhangaba	D2-070	1957	2000	38	-22.917	-45.467	GUM
e472	Pindamonhangaba	D2-012	1958	1999	36	-22.933	-45.467	LP3
e473	Pindamonhangaba	D2-067	1957	1995	31	-22.850	-45.583	GUM
e474	Pindamonhangaba	D2-041	1945	1999	45	-22.900	-45.483	LNO
e475	Pindorama	C5-018	1936	2018	75	-21.219	-48.898	BUR
e476	Pinhalzinho	D3-036	1940	2018	54	-22.788	-46.576	GUM
e477	Piquerobi	C8-001	1936	2018	64	-21.886	-51.727	BUR
e478	Piracaia	E3-229	1970	2018	32	-23.067	-46.300	BUR
e479	Piracaia	E3-076	1942	1994	44	-23.050	-46.350	BUR
e480	Piracicaba	D4-061	1943	2000	52	-22.683	-47.767	GEV
e481	Piracicaba	D4-104	1973	2018	38	-22.718	-47.652	PE3
e482	Piracicaba	D4-056	1943	1994	42	-22.683	-47.700	GEV
e483	Piraju	E6-006	1960	1997	37	-23.217	-49.233	GUM
e484	Pirajui	D6-086	1942	2018	60	-22.050	-49.617	LP3
e485	Pirapozinho	D8-025	1947	2018	52	-22.267	-51.500	GLO
e486	Pirassununga	C4-085	1947	2000	44	-21.967	-47.467	GUM
e487	Pirassununga	C4-033	1939	2018	68	-22.032	-47.421	LNO
e488	Piratininga	D6-022	1939	1998	46	-22.417	-49.133	GEV
e489	Pitangueiras	B5-027	1940	2018	60	-20.998	-48.332	GEV
e490	Pitangueiras	C5-040	1941	2018	69	-21.050	-48.267	PE3
e491	Platina	D7-064	1971	2018	39	-22.608	-50.181	BUR
e492	Pompeia	D7-003	1936	2000	45	-22.106	-50.195	GUM
e493	Pontal	C5-024	1936	1999	57	-21.017	-48.033	LNO
e494	Pontalinda	B7-038	1969	2018	43	-20.440	-50.533	LP3
e495	Populina	A7-001	1959	2001	35	-19.933	-50.533	BUR

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e496	Porangaba	E5-063	1970	2018	41	-23.177	-48.124	LNO
e497	Porto Feliz	E4-118	1970	2018	46	-23.204	-47.522	BUR
e498	Potirendaba	C6-066	1950	2018	62	-21.044	-49.375	PE3
e499	Pradópolis	C5-104	1699	2018	39	-21.363	-48.064	LNO
e500	Presidente Alves	D6-059	1944	2000	47	-22.100	-49.433	GLO
e501	Presidente Prudente	D8-003	1936	2018	58	-22.100	-51.383	BUR
e502	Presidente Venceslau	C8-011	1955	2018	55	-21.879	-51.838	GEV
e503	Promissão	C6-059	1944	2018	45	-21.554	-49.867	LNO
e504	Quadra	E5-023	1939	2001	51	-23.301	-48.056	GLO
e505	Quata	D7-001	1936	2018	54	-22.258	-50.701	GUM
e506	Queiros	C7-074	1946	2018	61	-21.800	-50.233	LNO
e507	Queluz	D1-009	1956	1995	35	-22.533	-44.767	BUR
e508	Quintana	D7-033	1940	2000	55	-22.072	-50.304	LP3
e509	Rancharia	D7-073	1974	2018	34	-22.122	-50.828	LNO
e510	Rancharia	D7-068	1972	2018	45	-22.624	-50.905	GEV
e511	Rancharia	D7-036	1941	2018	61	-22.236	-50.889	GEV
e512	Rancharia	D7-065	1971	2018	36	-22.441	-50.991	LP3
e513	Redenção da Serra	E2-106	1961	2000	31	-23.200	-45.533	GEV
e514	Redenção da Serra	E2-028	1953	2018	60	-23.283	-45.533	GLO
e515	Reginópolis	C6-050	1943	2000	52	-21.888	-49.234	GUM
e516	Registro	F4-024	1953	2000	41	-24.500	-47.733	PE3
e517	Registro	F4-005	1937	2018	64	-24.500	-47.850	BUR
e518	Registro	F4-004	1959	1999	32	-24.400	-47.750	LP3
e519	Ribeira	F6-012	1975	2018	39	-24.617	-49.050	GUM
e520	Ribeira	F5-008	1954	1999	41	-24.567	-48.950	GEV
e521	Ribeira	F5-002	1936	1999	44	-24.657	-49.008	GUM
e522	Ribeirão Bonito	D5-003	1936	2018	65	-22.071	-48.177	LP3
e523	Ribeirão Branco	F5-039	1971	2018	30	-24.276	-48.909	LNO
e524	Ribeirão Branco	F5-027	1966	2018	41	-24.222	-48.769	LNO
e525	Ribeirão Corrente	B4-022	1940	2018	48	-20.467	-47.600	GUM
e526	Ribeirão do Sul	D6-032	1937	2018	69	-22.783	-49.933	LNO
e527	Ribeirão Grande	F5-025	1947	2004	36	-24.136	-48.363	LNO
e528	Ribeirão Preto	C4-075	1945	2018	61	-21.103	-47.758	GEV
e529	Ribeirão Preto	C4-057	1943	2018	54	-21.300	-47.933	GLO
e530	Ribeirão Preto	C4-034	1939	2018	61	-21.217	-47.867	BUR

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e531	Rifaina	B4-038	1952	2018	43	-20.079	-47.426	LP3
e532	Rinopolis	C7-045	1946	2013	45	-21.733	-50.717	GUM
e533	Rio Claro	D4-012	1936	2018	69	-22.406	-47.556	LNO
e534	Rio Claro	D4-016	1953	2018	63	-22.333	-47.492	LNO
e535	Rio Claro	D4-012	1936	2018	69	-22.406	-47.556	LNO
e536	Rio das Pedras	D4-068	1946	2017	62	-22.866	-47.608	BUR
e537	Riolandia	A6-001	1958	2000	48	-19.967	-49.683	LP3
e538	Riversul	E6-011	1939	1997	50	-23.833	-49.433	GLO
e539	Rosana	D9-001	1954	2000	35	-22.467	-52.883	GEV
e540	Roseira	D2-060	1957	1999	36	-22.933	-45.317	LNO
e541	Sales	C6-085	1966	2000	32	-21.350	-49.500	GUM
e542	Sales Oliveira	B4-012	1940	2018	65	-20.800	-47.767	GEV
e543	Salesopolis	E2-112	1967	2018	48	-23.567	-45.967	BUR
e544	Salesopolis	E2-054	1946	2018	64	-23.533	-45.850	BUR
e545	Salesopolis	E2-131	1970	2009	31	-23.567	-45.833	LP3
e546	Salmourao	C7-075	1955	2018	47	-21.621	-50.916	GUM
e547	Salto	E4-127	1971	2018	36	-23.197	-47.290	GUM
e548	Salto de Pirapora	E4-025	1937	2017	48	-23.650	-47.567	LNO
e549	Salto Grande	D6-089	1958	1997	37	-22.900	-50.000	GLO
e550	Sandovalina	D8-004	1957	2018	47	-22.517	-51.817	PE3
e551	Santa Branca	E2-029	1942	2018	56	-23.400	-45.883	BUR
e552	Santa Cruz do Rio Pardo	D6-102	1974	2018	38	-22.583	-49.550	LP3
e553	Santa Cruz do Rio Pardo	D6-035	1955	2018	52	-22.900	-49.617	GUM
e554	Santa Cruz do Rio Pardo	D6-028	1937	2000	48	-22.950	-49.533	PE3
e555	Santa Gertrudes	D4-059	1941	2018	60	-22.478	-47.514	GLO
e556	Santa Isabel	E3-049	1937	2018	58	-23.333	-46.233	GLO
e557	Santa Lucia	C5-042	1960	2018	52	-21.659	-48.041	BUR
e558	Santa Lucia	C5-011	1936	2018	68	-21.684	-48.082	GEV
e559	Santa Maria da Serra	D5-062	1945	2000	41	-22.567	-48.150	PE3
e560	Santana da Parnaíba	E3-020	1956	2018	33	-23.450	-46.917	LP3
e561	Santa Rosa de Viterbo	C4-103	1972	2018	38	-21.550	-47.383	LNO
e562	Santa Rosa de Viterbo	C4-043	1942	2018	71	-21.433	-47.333	GEV
e563	Santo Anastacio	C8-009	1942	2018	64	-21.969	-51.643	BUR
e564	Santo Andre	E3-037	1937	1996	50	-23.783	-46.300	LP3
e565	Santo Antonio do Pinhal	D2-082	1943	1998	51	-22.817	-45.733	GUM

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e566	Santo Antonio do Pinhal	D2-029	1943	2018	67	-22.767	-45.700	BUR
e567	Santos	E3-041	1939	2017	71	-23.883	-46.221	LNO
e568	Sao Bernardo do Campo	E3-109	1944	1998	41	-23.850	-46.500	GEV
e569	Sao Bernardo do Campo	E3-142	1950	2009	41	-23.750	-46.533	BUR
e570	Sao Caetano do Sul	E3-022	1937	2004	60	-23.633	-46.583	LNO
e571	Sao Caetano do Sul	E3-085	1943	2018	62	-23.617	-46.550	PE3
e572	Sao Carlos	D5-076	1960	2017	49	-22.068	-48.046	BUR
e573	Sao Carlos	D4-106	1978	2017	30	-22.094	-47.975	BUR
e574	Sao Carlos	C4-092	1960	2017	43	-21.886	-47.786	BUR
e575	Sao Carlos	C4-019	1959	2018	50	-21.748	-47.768	GLO
e576	Sao Carlos	D4-015	1936	1998	55	-22.017	-47.900	PE3
e577	Sao Joao da Boa Vista	C3-031	1943	2018	62	-21.953	-46.808	GLO
e578	Sao Joaquim da Barra	B4-018	1937	2018	70	-20.519	-47.975	BUR
e579	Sao Jose da Bela Vista	B4-021	1954	2018	53	-20.600	-47.633	GLO
e580	Sao Jose do Barreiro	D1-003	1943	2002	46	-22.650	-44.583	GUM
e581	Sao Jose do Rio Pardo	C3-029	1940	2000	41	-21.533	-46.800	BUR
e582	Sao Jose do Rio Pardo	C3-035	1944	2018	59	-21.600	-46.900	GEV
e583	Sao Jose dos Campos	E2-036	1943	2000	50	-23.050	-45.900	GEV
e584	Sao Jose dos Campos	E2-102	1960	2005	38	-23.217	-45.783	GEV
e585	Sao Jose dos Campos	E2-057	1960	2000	39	-23.267	-45.833	LP3
e586	Sao Jose dos Campos	E2-099	1960	2018	51	-23.183	-45.800	LP3
e587	Sao Jose dos Campos	D2-021	1939	2001	51	-22.917	-45.967	GUM
e588	Sao Jose dos Campos	E2-032	1942	2003	41	-23.183	-45.883	LP3
e589	Sao Jose do Rio Preto	B6-004	1942	2004	48	-20.800	-49.383	LNO
e590	Sao Lourenco da Serra	E3-067	1958	2018	50	-23.783	-46.933	BUR
e591	Sao Luis do Paraitinga	E2-135	1972	2014	38	-23.367	-45.200	GEV
e592	Sao Luis do Paraitinga	E2-055	1947	2000	47	-23.250	-45.200	GEV
e593	Sao Manuel	D5-047	1940	2017	72	-22.737	-48.568	GUM
e594	Sao Miguel Arcanjo	E4-029	1936	2018	62	-23.867	-48.000	LP3
e595	Sao Miguel Arcanjo	E4-134	1972	2018	42	-23.733	-47.933	GLO
e596	Sao Miguel Arcanjo	F4-031	1960	2018	44	-24.050	-48.000	LNO
e597	Sao Miguel Arcanjo	F4-001	1936	2014	59	-24.000	-47.950	GUM
e598	Sao Paulo	E3-003	1937	2018	47	-23.517	-46.683	GUM
e599	Sao Paulo	E3-082	1940	1998	53	-23.417	-46.567	GLO
e600	Sao Paulo	E3-052	1940	2016	43	-23.633	-46.650	PE3

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e601	Sao Paulo	E3-254	1979	2018	32	-23.883	-46.750	GUM
e602	Sao Paulo	E3-081	1940	1997	49	-23.400	-46.583	GEV
e603	Sao Paulo	E3-059	1939	2003	39	-23.917	-46.650	LP3
e604	Sao Paulo	E3-071	1899	2003	55	-23.467	-46.633	GEV
e605	Sao Paulo	E3-090	1943	2003	53	-23.583	-46.650	GUM
e606	Sao Paulo	E3-010	1954	2000	37	-23.517	-46.733	LP3
e607	Sao Paulo	E3-036	1888	2004	96	-23.533	-46.633	LNO
e608	Sao Paulo	E3-033	1957	2008	36	-23.467	-46.717	LNO
e609	Sao Paulo	E3-035	1936	2018	80	-23.650	-46.633	GUM
e610	Sao Paulo	E3-243	1972	2018	43	-23.833	-46.733	GEV
e611	Sao Paulo	E3-030	1936	1997	33	-23.400	-46.750	GUM
e612	Sao Paulo	E3-096	1943	1993	37	-23.517	-46.650	GUM
e613	Sao Paulo	E3-007	1936	2004	36	-23.500	-46.617	GUM
e614	Sao Paulo	E3-006	1936	2018	49	-23.650	-46.700	GUM
e615	Sao Pedro	D4-060	1942	2000	46	-22.533	-47.917	LNO
e616	Sao Pedro do Turvo	D6-100	1972	2018	41	-22.583	-49.817	LP3
e617	Sao Pedro do Turvo	D6-095	1971	2018	39	-22.750	-49.733	LNO
e618	Sao Sebastiao Sao Sebastiao da	E2-045	1943	2018	53	-23.767	-45.417	BUR
e619	Grama	C3-011	1943	2018	55	-21.700	-46.817	GLO
e620	Sao Simao	C4-087	1950	2018	49	-21.433	-47.583	GUM
e621	Sao Simao	C4-052	1942	2000	56	-21.367	-47.433	GEV
e622	Sao Simao	C4-091	1965	2000	30	-21.500	-47.600	LP3
e623	Sao Vicente	E3-228	1939	2018	41	-23.954	-46.450	LP3
e624	Sao Vicente	E3-056	1938	2018	45	-23.967	-46.367	LNO
e625	Sarapui	E4-032	1936	1999	52	-23.650	-47.833	BUR
e626	Sarutaia	E6-002	1938	2018	71	-23.229	-49.464	LNO
e627	Sebastianopolis do Sul	B6-048	1969	2018	44	-20.664	-49.926	LNO
e628	Serrana	C4-083	1947	2018	50	-21.217	-47.600	LNO
e629	Serra Negra	D3-012	1936	2018	76	-22.601	-46.705	GLO
e630	Sertaozinho	C5-114	1954	2018	44	-21.100	-48.067	GUM
e631	Sete Barras	F4-025	1961	2018	55	-24.283	-47.950	GUM
e632	Sete Barras	F4-015	1939	2001	39	-24.383	-47.933	GLO
e633	Severinia	B5-005	1956	2000	43	-20.817	-48.783	GUM
e634	Silveiras	D1-020	1969	2018	34	-22.667	-44.850	LP3
e635	Socorro	D3-030	1942	2000	56	-22.598	-46.516	GLO

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e636	Sorocaba	E4-056	1961	1998	32	-23.500	-47.433	BUR
e637	Santa Cruz das Palmeiras	C4-029	1936	2000	62	-21.833	-47.250	GLO
e638	Santa Adelia	C5-027	1937	2018	74	-21.274	-48.766	GLO
e639	Santa Albertina	B7-013	1959	2018	51	-20.036	-50.731	LP3
e640	Santo Antonio da Alegria	C4-001	1959	2018	51	-21.100	-47.150	LP3
e641	Santo Antonio do Aracangua	B7-039	1970	2018	42	-20.866	-50.469	LNO
e642	Santo Antonio do Jardim	D3-071	1972	2018	44	-22.155	-46.714	GUM
e643	Santa Rita do Passa Quatro	C4-041	1943	2000	48	-21.717	-47.500	GUM
e644	Santa Barbara do Oeste	D4-087	1953	2000	43	-22.833	-47.433	BUR
e645	Santa Barbara do Oeste	D4-088	1953	2018	47	-22.758	-47.425	LNO
e646	Santa Cruz da Conceicao	D4-032	1936	2000	56	-22.117	-47.417	GLO
e647	Santa Fe do Sul	B7-024	1951	2000	43	-20.217	-50.917	BUR
e648	Suzanopolis	B8-016	1965	2001	34	-20.417	-51.100	PE3
e649	Tabapua	B6-009	1957	2018	55	-20.988	-49.010	GUM
e650	Tabatinga	C5-107	1970	2018	34	-21.733	-48.534	BUR
e651	Taciba	D8-028	1946	2018	56	-22.383	-51.283	PE3
e652	Taguai	E6-016	1939	2016	59	-23.449	-49.411	GEV
e653	Tamandua	B5-034	1943	2000	51	-20.717	-48.933	GUM
e654	Tanabi	B6-010	1958	2017	56	-20.478	-49.577	GUM
e655	Tanabi	B6-008	1956	2000	40	-20.617	-49.650	GEV
e656	Tapirai	F4-021	1954	2018	48	-24.050	-47.583	LP3
e657	Tapiratiba	C3-024	1940	2000	42	-21.467	-46.817	GLO
e658	Tapiratiba	C3-024	1940	2000	42	-21.467	-46.817	LP3
e659	Tapiratiba	C3-030	1947	2017	62	-21.467	-46.750	GLO
e660	Taquaritinga	C5-105	1970	2018	38	-21.467	-48.617	LP3
e661	Taquarituba	E6-013	1939	2018	65	-23.539	-49.246	GLO
e662	Taquarivai	E5-046	1947	1996	41	-23.917	-48.700	GLO
e663	Tatui	E4-020	1954	2011	45	-23.317	-47.783	BUR
e664	Tatui	E4-026	1957	2018	51	-23.400	-47.767	GUM
e665	Tatui	E5-062	1970	2018	42	-23.384	-48.028	GLO
e666	Tatui	E4-036	1939	2017	58	-23.367	-47.850	LP3
e667	Taubate	E2-092	1954	2015	50	-23.200	-45.467	GUM
e668	Taubate	E2-091	1954	1997	41	-23.150	-45.383	LNO
e669	Taubate	E2-095	1957	2018	45	-23.067	-45.500	GUM
e670	Taubate	E2-097	1957	1996	32	-23.133	-45.417	LNO

Fonte: elaborado pelo autor.

ID	Cidade	Prefixo DAEE	Início	Fim	Tamanho	Latitude	Longitude	Melhor Distribuição
e671	Taubate	E2-022	1963	2012	43	-23.033	-45.567	GEV
e672	Tejupa	E6-015	1939	2000	49	-23.333	-49.367	GUM
e673	Teodoro Sampaio	D9-014	1969	2018	46	-22.567	-52.417	PE3
e674	Teodoro Sampaio	D9-020	1976	2018	35	-22.617	-52.167	GUM
e675	Teodoro Sampaio	D9-003	1966	2018	38	-22.517	-52.183	LP3
e676	Terra Roxa	B5-024	1940	2018	70	-20.793	-48.357	LNO
e677	Tiete	E4-120	1970	2018	41	-23.100	-47.717	GLO
e678	Tiete	E4-037	1939	2001	54	-23.000	-47.717	BUR
e679	Timburi	E6-019	1939	1999	36	-23.200	-49.617	GLO
e680	Torrinha	D5-006	1936	2018	75	-22.396	-48.164	GUM
e681	Tres Lagoas	B8-011	1971	2016	38	-20.779	-51.657	BUR
e682	Tupa	C7-004	1960	2018	51	-21.962	-50.464	GEV
e683	Tupa	C7-043	1942	2000	55	-21.933	-50.533	BUR
e684	Tupa	C7-066	1948	2018	49	-21.901	-50.598	GLO
e685	Ubatuba	E2-009	1957	2010	41	-23.383	-45.117	GEV
e686	Ubatuba	E1-004	1945	1998	40	-23.383	-44.833	GUM
e687	Ubatuba	E2-052	1945	2018	47	-23.433	-45.067	GUM
e688	Ubirajajara	D6-040	1965	2018	38	-22.533	-49.650	GUM
e689	Uchoa	B6-037	1967	2000	30	-20.967	-49.183	GUM
e690	Urania	B7-006	1958	2018	53	-20.270	-50.652	LP3
e691	Urupes	C6-041	1940	2018	67	-21.206	-49.292	GLO
e692	Valentim Gentil	B7-011	1959	2018	52	-20.441	-50.083	GUM
e693	Valparaíso	C7-034	1944	2018	66	-21.250	-50.867	LNO
e694	Vargem	D3-018	1937	2018	68	-22.876	-46.411	PE3
e695	Vargem Grande do Sul	C3-009	1936	2018	52	-21.833	-46.900	LNO
e696	Vinhedo	E3-017	1936	1995	48	-23.033	-46.967	BUR
e697	Vista Alegre do Alto	C5-092	1969	2018	39	-21.157	-48.629	GEV
e698	Votuporanga	B6-032	1963	2017	41	-20.524	-49.955	LP3
e699	Votuporanga	B6-036	1965	2018	45	-20.440	-49.980	GLO

Fonte: elaborado pelo autor.

APÊNDICE B

Frequência acumulada dos desempenhos das distribuições de probabilidade candidatas nos diferentes testes descritivos e preditivos avaliados. Melhores modelos são aqueles que obtiveram as melhores posições no ranqueamento com maior frequência, isto é, cujas curvas de frequência acumulada sobem mais intensamente.

Figura B1 – Frequência acumulada dos desempenhos das distribuições de probabilidade nos diferentes testes descritivos avaliados.

Fonte: Autor (2021).

Figura B2 - Frequência acumulada dos desempenhos das distribuições de probabilidade nos diferentes testes preditivos avaliados.

Fonte: Autor (2021).

APÊNDICE C

Séries sintéticas completas (1000 anos) com (i) mudanças abruptas ou (ii) tendência lineares em seus parâmetros.

Figura C1 – Séries sintéticas completas com mudanças abruptas.

Figura C2 – Séries sintéticas completas com tendências lineares em seus parâmetros.

APÊNDICE D

Desempenho dos modelos MEV, GEV e GPD (modelado via POT) nas estimativas dos quantis de precipitação para as seis séries sintéticas com mudanças abruptas (Capítulo IV). Nos gráficos com os quantis de precipitação estimados, os sombreados representam os intervalos de confiança, ao nível de significância de 95%. As linhas sólidas indicam a média das estimativas realizadas pelas 1000 séries sintéticas reduzidas geradas. Os pontos pretos representam os dados observados, com seus períodos de retorno estimados empiricamente pela fórmula de posição de plotagem de Weibull.

Figura D1 - Quantis de precipitação estimados considerando a série CP1.

Fonte: Autor (2021).

Figura D2 - Quantis de precipitação estimados considerando a série CP2.

Fonte: Autor (2021).

Figura D3 - Quantis de precipitação estimados considerando a série CP3.

Fonte: Autor (2021).

Figura D4 - Quantis de precipitação estimados considerando a série CP4.

Fonte: Autor (2021).

Figura D5 - Quantis de precipitação estimados considerando a série CP5.

Fonte: Autor (2021).

Figura D6 - Quantis de precipitação estimados considerando a série CP6.

Fonte: Autor (2021).

APÊNDICE E

Desempenho dos modelos MEV, GEV e GPD (modelado via POT) nas estimativas dos quantis de precipitação para as seis séries sintéticas com tendências lineares em seus parâmetros (Capítulo IV). Nos gráficos com os quantis de precipitação estimados, os sombreados representam os intervalos de confiança, ao nível de significância de 95%. As linhas sólidas indicam a média das estimativas realizadas pelas 1000 séries sintéticas reduzidas geradas. Os pontos pretos representam os dados observados, com seus períodos de retorno estimados empiricamente pela fórmula de posição de plotagem de Weibull.

Figura E1 - Quantis de precipitação estimados considerando a série T1.

Fonte: Autor (2021).

Figura E2 - Quantis de precipitação estimados considerando a série T2.

Fonte: Autor (2021).

Figura E3 - Quantis de precipitação estimados considerando a série T3.

Finte: Autor (2021).

Figura E4 - Quantis de precipitação estimados considerando a série T4.

Fonte: Autor (2021).

Figura E5 - Quantis de precipitação estimados considerando a série T5.

Fonte: Autor (2021).

Figura E6 - Quantis de precipitação estimados considerando a série T6.

Fonte: Autor (2021).