

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA CIVIL E AMBIENTAL**

**DETECÇÃO DE TENDÊNCIAS EM SÉRIES DE EXTREMOS
HIDROLÓGICOS CONSIDERANDO EFEITOS DE
AUTOCORRELAÇÃO TEMPORAL E MULTIPLICIDADE
DE TESTES**

RENATO SOUZA AMORIM

ORIENTADOR: DIRCEU SILVEIRA REIS JUNIOR

**DISSERTAÇÃO DE MESTRADO EM TECNOLOGIA
AMBIENTAL E RECURSOS HÍDRICOS**

BRASÍLIA/DF: ABRIL/2018

**UNIVERSIDADE DE BRASÍLIA
FACULDADE DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA CIVIL**

**DETECÇÃO DE TENDÊNCIAS EM SÉRIES DE EXTREMOS
HIDROLÓGICOS CONSIDERANDO EFEITOS DE
AUTOCORRELAÇÃO TEMPORAL E MULTIPLICIDADE
DE TESTES**

RENATO SOUZA AMORIM

**DISSERTAÇÃO DE Mestrado submetida ao
Departamento de Engenharia Civil e Ambiental da
Faculdade de Tecnologia da Universidade de Brasília
como parte dos requisitos necessários para a
obtenção do grau de Mestre em Tecnologia
Ambiental e Recursos Hídricos.**

APROVADA POR:

**Prof. Dirceu Silveira Reis Júnior, PhD (ENC-UnB)
(Orientador)**

**Prof. Carlos Henrique Ribeiro Lima, PhD (UnB)
(Examinador Interno)**

**Prof. Wilson dos Santos Fernandes, PhD (UFMG)
(Examinador Externo)**

BRASÍLIA/DF, 27 DE ABRIL DE 2018.

FICHA CATALOGRÁFICA

AMORIM, RENATO SOUZA

Detecção de Tendências em Séries de Extremos Hidrológicos Considerando Efeitos de Autocorrelação e Multiplicidade de Testes [Distrito Federal] 2018.

xvi, 165 p., 210 x 297 mm (ENC/FT/UnB, Mestre, Tecnologia Ambiental e Recursos Hídricos, 2018).

Dissertação de Mestrado – Universidade de Brasília. Faculdade de Tecnologia.

Departamento de Engenharia Civil e Ambiental.

1. Mudanças Climáticas

3. Autocorrelação Amostral

I. ENC/FT/UnB

2. Detecção de Tendências Monotônicas

4. Multiplicidade de testes de Hipóteses

II. Título (série)

REFERÊNCIA BIBLIOGRÁFICA

AMORIM, R. S. (2018). Detecção de Tendências em Séries de Extremos Hidrológicos Considerando Efeitos de Autocorrelação Temporal e Multiplicidade de Testes. Dissertação de Mestrado em Tecnologia Ambiental e Recursos Hídricos, Publicação PTARH.DM-XXX/18, Departamento de Engenharia Civil e Ambiental, Universidade de Brasília, Brasília, DF, 167 p.

CESSÃO DE DIREITOS

AUTOR: Renato Souza Amorim.

TÍTULO: Detecção de Tendências em Séries de Extremos Hidrológicos Considerando Efeitos de Autocorrelação Temporal e Multiplicidade de Testes.

GRAU: Mestre

ANO: 2018

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte dessa dissertação de mestrado pode ser reproduzida sem autorização por escrito do autor.

Renato Souza Amorim
renatosamorim@gmail.com

AGRADECIMENTOS

Agradeço, primeiramente, ao meu orientador Dirceu Reis, por todos os ensinamentos e conselhos que resultaram em melhorias neste trabalho e na minha formação acadêmica. Agradeço a ele o desafio e a confiança depositada.

Agradeço também ao engenheiro Saulo Aires de Souza pelo apoio com as rotinas computacionais, parceria durante o andamento do projeto CAPES: “*Impacto das Mudanças Climáticas em Extremos Hidrológicos (SECAS e CHEIAS)*”, discussões e avaliações que ajudaram a enriquecer este trabalho.

À equipe do DNIT, em especial às colegas e amigas Ana Luísa Nunes e Vanessa da Matta, que, seja revisando o texto ou assumindo minhas atividades diárias do DNIT por um tempo, permitiram o término dessa etapa.

Aos engenheiros Calvin Creech e Matt Dirksen, do Corpo de Engenheiros dos Estados Unidos (USACE), pela amizade e companheirismo.

À minha família, e principalmente meus pais pelo amor e apoio incondicionais, mesmo estando distantes.

RESUMO

Os efeitos da alteração na frequência e intensidade de eventos hidrológicos extremos (cheias e secas) têm motivado a realização de estudos de detecção de tendências para subsidiar o processo de gestão dos recursos hídricos. Diversos fatores influenciam a capacidade de detecção de tendências verdadeiras, tais como a presença de autocorrelação temporal, a questão da multiplicidade de testes em análises regionais, e o fato de que o poder do teste empregado na análise e o grau de tendenciosidade na estimativa da magnitude da tendência dependem de características da série histórica (tamanho, graus de variabilidade e assimetria, distribuição geradora, e magnitude real de tendência). A falta de conhecimento sobre como tais fatores afetam a análise de detecção resulta em equívocos na identificação de estacionariedade. Este trabalho é dividido em três partes. Na primeira parte, procura-se avaliar, com base em estudos de Simulação Monte Carlo, como o tamanho da série histórica (n), magnitude da tendência (b), o coeficiente de variação (C_v) e o coeficiente de assimetria da série histórica, este último função do parâmetro de forma (k) da distribuição Generalizada de Valores Extremos (GEV), influenciam, numa análise local, o poder do teste Mann-Kendall (MK), a probabilidade de se obter uma estimativa da magnitude da tendência com sinal oposto ao real (erro tipo S) empregando o estimador sugerido por Sen (1968) e o grau de tendenciosidade dessa estimativa (erro tipo M). A segunda parte, também baseada em Simulação Monte Carlo, foca na questão da análise regional de tendência, procurando avaliar o desempenho de metodologias baseadas no conceito do False Discovery Rate (FDR), sugeridas na literatura (Benjamini e Hockberg, 1995, 200) para amenizar o problema da multiplicidade de testes de hipóteses, que tendem a resultar em número exagerado de resultados falso positivos. A terceira parte do trabalho ilustra o uso de tais metodologias num estudo de detecção de tendências monotônicas em extremos de variáveis hidrometeorológicas na bacia do Rio São Francisco. Os resultados da primeira parte permitiram mapear o poder do teste MK para diferentes combinações de n , C_v , k e b para séries de extremos do tipo GEV, possibilitando um melhor entendimento da probabilidade de se identificar uma estação com tendência quando de fato a tendência está presente. Por exemplo, uma série com 30 anos, $C_v = 1$, $k = -0,3$ e tendência relativa anual de 2%, a probabilidade do teste MK identificar essa tendência é de apenas 0,45. Se essa mesma série tivesse um comprimento de 50 anos, a probabilidade de detecção aumentaria para 0,90. Esses mesmos resultados sugerem que a probabilidade de se estimar uma tendência com o sinal

oposto ao real, quando a estação é declarada significativa, depende do poder do teste e pode assumir valores superiores a 5%, podendo chegar a até mesmo 40% para poder do teste inferior a 10%. Verificou-se também que existe uma tendência de superestimação da magnitude das tendências, e que o valor estimado, em média, pode ser 1,5 vezes maior que o valor real quando o poder do teste é igual a 0,40, podendo ainda ser até 5 vezes maior do que o valor real, em média, quando o poder do teste é igual a 0,05. Em relação à análise regional, os resultados mostram que o teste MK aplicado de forma independente em cada estação resulta em um número elevado de falsos positivos, muito além do que seria razoável. Por exemplo, numa série com $n = 40$ anos, relação $b/Cv = 0,02$, e $k = -0,3$, a proporção de falsos negativos chega a 39%, enquanto que os métodos baseados no FDR resultam em proporções menores do que 4%. Essa diferença possui um custo, já que o MK empregado de forma independente consegue identificar corretamente 92% das estações com tendência, enquanto os métodos baseados no FDR atingem um pouco mais de 50%. Na análise da bacia do Rio São Francisco, verificou-se um elevado número de estações com tendências decrescentes nas séries de vazões mínimas, com uma proporção de aproximadamente 50%. Na maioria dessas estações, a magnitude da tendência ultrapassa a taxa de 10% por década. Esses resultados foram obtidos com o emprego da técnica de Trend-Free Pre-Whitening (TFPW) para lidar com a autocorrelação temporal e com o uso dos métodos baseados no conceito de FDR para manter a proporção de falsos positivos menor do que 5%

ABSTRACT

The effects of a change in the frequency and intensity of hydrologic events (floods and droughts) have motivated trend detection studies to support water resources management planning. Several factors influence the ability to detect true trends, such as the presence of temporal autocorrelation, the question of multiplicity of tests in regional analyses, and the fact that the utility of the test used in the analysis and the degree of bias in the estimation of the magnitude of the trend depend on characteristics of the historical series (size, degrees of variability and asymmetry, chosen distribution, and actual magnitude of the trend). The lack of knowledge about how these factors affect the detection analysis leads to errors in the identification of stationarity. This work is divided into three parts. In the first part, an evaluation of trends is completed based on Monte Carlo simulation studies, which vary the size of the historical series (n); the magnitude of the trend (b); the coefficient of variance (Cv) and the coefficient of asymmetry of the historical series; the form of the function (k) of the distribution of Generalized Extreme Values (GEV); local influence of the power of the Mann-Kendall (MK) test; the probability of obtaining an estimate of the magnitude of the trend with an opposite sign of the actual (error type S) employing the estimator suggested by Sen (1968); and the degree of bias of this estimate (type M error). The second part of this work is also based on Monte Carlo simulations, which focus on the issue of regional trend analysis, seeking to analyze the performance of methodologies based on the concept of False Discovery Rate (FDR), suggested in the literature (Benjamini and Hockberg, 1995, 200) to mitigate the problem of multiplicity of hypothesis tests, which tend to result in higher numbers of false positives. The third part of the work illustrates the use of these methodologies in a case study to detect monotonic tendencies in extreme hydro-meteorological variables in the São Francisco River basin. The results of the first part allowed mapping of the power of the MK test to different combinations of n , Cv , k and b for extreme series of GEV, enabling a better understanding of the likelihood of identifying a season with a tendency when a true trend it is present. For example, a series with 30 years, $Cv = 1$, $k = -0.3$ and an annual relative trend of 2%, the likelihood of the MK test identifying this trend is only 0.45. If this same series had a length of 50 years, the likelihood of detection would increase to 0.90. These same results suggest that the likelihood of estimating a trend with the opposite sign of the true trend (when the station is determined to be significant) depends on the power of the test and can assume values greater than 5%, and can reach 40%

for test powers less than 10%. It was also verified that there is a tendency of overestimation of the magnitude of the trends, and that the estimate value (on average) can be 1.5 times greater than the actual value when the test power is equal to 0.40, and can still be up to 5 times greater than the actual value (on average) when the power of the test is equal to 0.05. In relation to the regional analysis, the results show that the MK test applied independently at each station results in a high number of false positives, far beyond what would be reasonable. For example, in a series with $n = 40$ years, ratio of $b/Cv = 0.02$, and $k = -0.3$, the proportion of false negatives reaches 39%, while FDR-based methods result in proportions less than 4%. This difference has a cost, since MK employed independently manages to correctly identify 92% of the seasonal trends, while the FDR-based methods reach a little over 50%. In the analysis of the São Francisco River basin, a large number of stations with declining trends in the minimum flow series were verified, with a ratio of approximately 50%. In most of these stations, the magnitude of the trend exceeds the rate of 10% per decade. These results were obtained with the use of the Trend-Free Pre-Whitening (TFPR) technique to deal with temporal autocorrelation and using methods based on the concept of FDR to maintain the proportion of false positives to less than 5%.

SUMÁRIO

1. INTRODUÇÃO	1
2. OBJETIVOS	6
2.1. OBJETIVO GERAL	6
2.2. OBJETIVOS ESPECÍFICOS	6
3. REVISÃO BIBLIOGRÁFICA	7
3.1. EVENTOS HIDROLÓGICOS EXTREMOS	7
3.2. INDICES DE EXTREMOS HIDROMETEOROLÓGICOS	10
3.3. DETECÇÃO DE TENDÊNCIAS MONOTÔNICAS	11
3.3.1. Referencial Teórico	12
3.3.2. Teste de Hipótese de Estacionariedade	15
3.3.3. Estimativa da Magnitude de Tendências: B_SEN	19
3.3.4. Perspectiva Histórica: Avaliação das Técnicas de Detecção	20
3.4. O PROBLEMA DA AUTOCORRELAÇÃO	23
3.4.1. O Impacto da Correlação nos Resultados de Detecção	23
3.4.2. Pre-Whitening	24
3.4.3. Trend-Free-Pre-Whitening	25
3.4.4. Perspectiva Histórica: A questão da Autocorrelação	26
3.5. A QUESTÃO DA MULTIPLICIDADE	32
3.5.1. Impacto da Multiplicidade nos Resultados de Detecção	32
3.5.2. Evolução Histórica	34
3.5.3. O FDR Aplicado à Hidrometeorologia	38
3.5.4. O Controle pelo FDR	47
4. METODOLOGIA	53
4.1. ESTRATÉGIA DE AÇÃO	53

4.2. CASO SINTÉTICO	56
4.2.1. A Distribuição Generalizada de Valores Extremos - GEV	58
4.2.2. O Processo de Geração e Estimativa do Poder do Teste.....	60
4.2.3. Estimativa do B_SEN, Erros Tipo S e M	63
4.2.4. Geração de Campo Sintético e Aplicação de Técnicas Baseadas no FDR	64
4.3. CASO REAL	68
4.3.1. Seleção da Área de Estudo.....	70
4.3.2. Base de Dados	71
4.3.3. Seleção e Estimativa de Índices Hidrológicos Extremos.....	77
5. RESULTADOS E DISCUSSÃO.....	81
5.1. AVALIAÇÃO DE DESEMPENHO DO TESTE DE MANN-KENDALL	
APLICADO A SÉRIES DE EXTREMOS	81
5.2. AVALIAÇÃO DE DESEMPENHO DO ESTIMADOR B_SEN APLICADO	
A SÉRIES DE EXTREMOS.....	90
5.3. AVALIAÇÃO DE DESEMPENHO DAS TÉCNICAS QUE UTILIZAM O	
FDR COMO CONTROLE DE ERROS DO TIPO I – BH95 E BH_ADAP	99
5.4. AVALIAÇÃO DE METODOLOGIAS DE DETECÇÃO DE TENDÊNCIAS	
EM SÉRIES AUTOCORRELACIONADAS.....	107
5.5. EFEITOS DA MULTIPLICIDADE SOBRE A PROPORÇÃO	
DECLARADA DE ESTAÇÕES NÃO-ESTACIONÁRIAS	126
6. CONCLUSÕES E RECOMENDAÇÕES.....	136
REFERÊNCIAS BIBLIOGRÁFICAS	142
APÊNDICE I.....	149
APÊNDICE II	155

LISTA DE TABELAS

Tabela 3.1 - Possíveis resultados ao se realizar múltiplos testes de hipóteses	14
--	----

LISTA DE FIGURAS

Figura 3.1 - Representação esquemática de como as mudanças climáticas (a) crescimento da média e (b) aumento da variância afeta a distribuição de probabilidades da temperatura diária (Zwiers <i>et. al.</i> , 2015, adaptado).....	9
Figura 3.2 - Fluxograma da estratégia de ação para estudos de mudança climática. A cronologia de execução das atividades proposta é da esquerda para a direita (Hay <i>et. al.</i> , 2015, adaptado)	9
Figura 3.3 - Ilustração dos erros tipo I e II em um teste de hipóteses unilateral (Naghetini e Pinto, 2007)	14
Figura 4.1: Fluxograma da Estratégia de Ação desenvolvida para a realização dos objetivos geral e específicos.....	54
Figura 4.2 - Área Padrão considerada na geração do campo sintético: os pontos vermelhos representam as estações pluviométricas selecionadas para a bacia do São Francisco e o retângulo azul (canto inferior esquerdo) é a delimitação de zona em que foram inseridas tendências nas séries estacionárias geradas via Simulação Monte Carlo.	67
Figura 4.3 - Representação esquemática de uma iteração do procedimento aplicação de metodologias de controle pelo FDR – Caso Sintético.....	68
Figura 4.4 – Série temporal de Q7min (1937 – 2013) da estação 40150000 para o período pré (linha decrescente tracejada) e pós construção (linha crescente sólida) da PCH de Cajuru, em 1959 (linha tracejada vermelha).	73
Figura 4.5 – (a) Localização das estações fluviométricas selecionadas pelos critérios (i) e (ii) e barramentos existentes; (b) estações que foram eliminadas pelo uso da premissa (iii)....	74
Figura 4.6 – (a) Localização das estações pluviométricas e (b) fluviométricas selecionadas para integrarem a base amostral	75
Figura 4.7 – Número (Total Remanescente) de Estações de Pluviométricas (preto) e Fluviométricas (cinza) de Faixas de Tamanho Amostral. As linhas tracejada e sólida	

representam, respectivamente, o total acumulado de Estações disponíveis para o estudo à medida que o número de anos aumenta.....	76
Figura 5.1 – Taxas de rejeição de séries temporais com tendências de magnitude $b = 0.00$ 2.00 10.00, tamanhos amostrais $n = [20\ 40\ 60\ 80\ 100]$, $\alpha = 0.05$, $\kappa = -0.30$ e (a) $Cv = 0.20$, (b) $Cv = 0.60$ e (c) $Cv = 1.00$.	84
Figura 5.2 – Taxas de rejeição de séries temporais com tendências de magnitude $b = 0.00$ 2.00 10.00, tamanho amostral $n = 50$, $\alpha = 0.05$, $Cv = 0.20$ 0.20 1.00 e (a) $\kappa = -0.30$, (b) $\kappa = 0.0$ e (c) $\kappa = +0.30$	86
Figura 5.3 – Taxas de rejeição de séries temporais com tendências de magnitude $b = 0.00$ 2.00 10.00%, tamanhos amostrais $\kappa = [-0.30\ -0.10\ 0.00\ 0.10\ 0.30]$, $\alpha = 0.05$, $n = 50$ e (a) $Cv = 0.20$, (b) $Cv = 0.60$ e (c) $Cv = 1.00$.	87
Figura 5.4 – Curva Magnitude da Tendência – Taxa de rejeição de séries do tipo GEV com parâmetros de forma iguais a $[-0.3\ -0.1\ 0\ .0.1\ 0.3]$, $\alpha = 0.05$, $Cv = 0.4$ e $n = 50$	88
Figura 5.5 – Curvas de igual poder do teste de MK para diferentes relações sinal-ruído (bCv), tamanhos amostrais (n) e parâmetros de forma da GEV ($\kappa = -0.30, 0.0, +0.30$)	89
Figura 5.6 – Probabilidade do Erro Tipo S (sinal) dado que a estação foi declarada não-estacionária ($\alpha = 0.05$) para $\kappa = -0.30$ e diferentes valores de $n = [20\ 40\ 60\ 80\ 100]$ e $Cv = 0.2$ (a), 0.6 (b) e 1.0 (c).....	92
Figura 5.7 - Probabilidade do Erro Tipo S (sinal) dado que a estação foi declarada não-estacionária ($\alpha = 0.05$) para $n=50$ e diferentes valores de $Cv = [0.2\ 0.4\ 0.6\ 0.8\ 1.0]$ e $\kappa = -0.3$ (a), 0.0 (b) e 0.3.0 (c).....	93
Figura 5.8 - Probabilidade do Erro Tipo S (sinal) dado que a estação foi declarada não-estacionária ($\alpha = 0.05$) para $n=50$ e diferentes valores de $\kappa = [-0.3\ -0.1\ 0\ 0.1\ 0.3]$ e $Cv = 0.2$ (a), 0.6 (b) e 1.0 (c).....	94
Figura 5.9 – Curva que expressa a relação entre o poder do teste e a probabilidade do Erro Tipo S significativo (%).....	95

Figura 5.10 - Curva que expressa a relação entre o poder do teste e o valor médio estimado por B_SEN dividido por b ($BSEN/b$).....	96
Figura 5.11 -Largura relativa do IC do Erro tipo M, dado que a estação foi declarada não-estacionária ($\alpha = 0.50$) para $\kappa = -0.3$, $b = 0.006$ e diferentes valores de $Cv = [0.2 (0.2) 1.0]$ e $n = [20 40 60 80 100]$	97
Figura 5.12 - Largura relativa do IC do Erro tipo M dado que a estação foi declarada não-estacionária ($\alpha = 0.50$) para $Cv = 0.6$, $b = 0.006$ e diferentes valores de $\kappa = [-0.3 -0.1 0 0.1 0.3]$ e $n = [20 40 60 80 100]$	98
Figura 5.13 – Valores médios (%) de NTP , FDR e $FNDR$ ao se aplicar as metodologias tradicionais, BH95 e BH_Adap à campos sintéticos cuja as séries possuem magnitude da tendência ($b = 0.002$), coeficiente de variação ($Cv = 0.2$), parâmetro de forma ($\kappa = -0.30$) e tamanhos amostrais ($n = 40, 60$ e 80).....	100
Figura 5.14 - Valores médios (%) de NTP , FDR e $FNDR$ ao se aplicar as metodologias tradicionais, BH95 e BH_Adap à campos sintéticos cuja as séries possuem magnitude da tendência ($b = 0.002$), coeficiente de variação ($Cv = 0.4, 0.6, 0.8$), parâmetro de forma ($\kappa = -0.30$) e tamanhos amostrais ($n = 60$)	104
Figura 5.15 - Valores médios (%) de NTP , FDR e $FNDR$ ao se aplicar as metodologias tradicionais, BH95 e BH_Adap à campos sintéticos cuja as séries possuem magnitude da tendência ($b = 0.006$), coeficiente de variação ($Cv = 0.4$), parâmetro de forma ($\kappa = -0.30, 0, +0.30$) e tamanhos amostrais ($n = 60$)	105
Figura 5.16 - Valores médios (%) de NTP , FDR e $FNDR$ ao se aplicar as metodologias tradicionais, BH95 e BH_Adap à campos sintéticos cuja as séries possuem magnitude da tendência ($b = 0.004, 0.006, 0.008$), coeficiente de variação ($Cv = 0.4$), parâmetro de forma ($\kappa = -0.30$) e tamanhos amostrais ($n = 60$).....	106
Figura 5.17 – Taxas de Rejeição das Abordagens (a: ignorar autocorrelação), (b: aplicar Pre_Whitening), (c: aplicar Trend-Free-Pre-Whitening) para os Indices de Precipitação	107
Figura 5.18 – Taxas de Rejeição das Abordagens (a: ignorar autocorrelação), (b: aplicar Pre-Whitening), (c: aplicar Trend-Free-Pre-Whitening) para os Indices de Vazão.....	108

Figura 5.19 – Percentuais de estações pluviométricas de significância estatística declarada pelo MK-TFPW subdivididos em faixas de magnitude de tendência expressas pelo B_SEN_relativo (a): 0 – 5%, (b) 5 – 10% e (c) >10%.....	113
Figura 5.20 - Percentuais de estações pluviométricas de significância estatística declarada pelo MK-TFPW subdivididos em faixas de magnitude de tendência expressas pelo B_SEN_Relativo (a): 0 – 5%, (b) 5 – 10% e (c) >10%.....	114
Figura 5.21 – Estações significativas para R1 após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativos por (a).....	117
Figura 5.22 - Estações significativas para SDII após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativos por (a).....	119
Figura 5.23-- Estações significativas para R10 após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativos por (a).....	121
Figura 5.24 – Séries temporais de R50 para as estações 1144014 (laranja) e 1144005 (azul)	122
Figura 5.25 – Séries temporais de R50 para as estações 1344002 (laranja) e 1343008 (azul)	123
Figura 5.26- Estações significativas para Q7min após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativos por (a).....	125
Figura 5.27: Proporção de estações declaradas como não-estacionárias ($\alpha=5\%$) , utilizando-se procedimentos locais de análise (não considera a multiplicidade) e o BH95. Para situações em que a correlação temporal é significativa foi aplicado o TFP.....	128

Figura 5.28: Proporção de estações declaradas como não-estacionárias ($\alpha=5\%$), utilizando-se procedimentos locais de análise (não considera a multiplicidade) e o BH95. Para situações em que a correlação temporal é significativa foi aplicado o TFP..... 129

Figura 5.29 – Variação do percentual de rejeição dos índices extremos de pluviometria obtidos para o BH95 e BH_Adap em relação ao MK+TFPW (referência) por magnitudes de B_SEN_relativos 130

Figura 5.30 - Variação do percentual de rejeição dos índices extremos de pluviometria obtidos para o BH95 e BH_Adap em relação ao MK+TFPW (referência) por magnitudes de B_SEN_relativos 131

1. INTRODUÇÃO

Eventos hidrológicos extremos, tais como cheias e secas prolongadas, afetam diretamente todas as facetas da sociedade, incluindo a economia, meio-ambiente e a cultura. Dada a abrangência dos seus efeitos, legisladores, gestores e tomadores de decisão estão cada vez mais interessados em conhecer a natureza e o comportamento desses eventos para diferentes escalas de tempo e espaço.

A interrupção de estradas, os alagamentos de centros urbanos e a escassez de água para consumo humano são apenas alguns dos impactos que podem ter a intensidade, frequência e duração alteradas, por exemplo, devido a alterações no padrão de precipitação e temperatura observados em uma dada região. Diante disso, surgem discussões sobre a necessidade de incorporação de tais alterações no comportamento de extremos de variáveis hidroclimáticas no dimensionamento de novas infraestruturas, nos processos de avaliação e gestão de riscos (Birkmann e Welle, 2015) e na elaboração de estudos de viabilidade de empreendimentos.

Os sistemas hidrometeorológicos são afetados não apenas pelas mudanças climáticas, como também pela ação antrópica, por meio, principalmente, da modificação na cobertura vegetal e do uso e ocupação do solo. Juntos, esses fatores interferem nos aspectos espaciais e temporais terrestres do ciclo hidrológico, incluindo a frequência e a intensidade dos eventos extremos (Kundzewicz *et al.*, 2007). Enquanto o uso do solo está diretamente relacionado a impactos sobre a rugosidade da superfície, propriedades do solo, albedo e evapotranspiração (Brown *et al.*, 2005), a mudança climática também atua sobre o ciclo hidrológico, modificando seus componentes básicos, tais como precipitação, evaporação, umidade do solo, disponibilidade hídrica subterrânea e magnitude e tempo de escoamento superficial (Dooge, 1992).

Face ao cenário atual de mudanças climáticas apontado por trabalhos ligados à detecção de tendências em variáveis hidrometeorológicas (Lettenmaier *et al.*, 1994; Krakauer e Fung, 2008; Vogel *et al.*, 2011; Alkama *et al.*, 2013; Hossain, 2014; Madsen *et al.*, 2014; Bayazit, 2015) e com vistas ao desenvolvimento de uma gestão e planejamento sustentável de

recursos hídricos, é essencial entender como e por que as características dos eventos hidrológicos extremos podem se alterar no espaço e tempo. Atentas a essa necessidade, a Organização Meteorológica Mundial (WMO) e a Organização das Nações Unidas para a Educação, Ciência e Cultura (UNESCO) vêm chamando a atenção para a necessidade de realização de pesquisas de âmbito nacional e internacional, em seus países signatários, para fomentar as análises de séries temporais de dados hidrológicos com respeito à variabilidade e mudanças no clima (WMO, 2009).

Entendido como o primeiro passo de uma série de outros que permitirão um maior conhecimento dos fatores intervenientes que atuam sobre os eventos climáticos extremos e sua dinâmica, o processo de detecção consiste em analisar uma série de dados observados e identificar quais tipos de mudanças, caso existam, foram constatadas em diferentes aspectos da mesma, tais como alterações nas séries de máximos ou mínimos de chuva e vazão, aumento ou redução do número de dias com precipitações intensas, entre outras.

A detecção de mudanças em series hidrológicas pode ter caráter local ou regional e baseia-se em testes de hipóteses, que empregam conjecturas a respeito do comportamento populacional de uma variável hidrológica extrema, por exemplo. A realização de testes permite identificar se a série observada apresenta ou não mudanças estatisticamente significantes, embora a decisão de apontá-la como estacionária ou não sempre estará sujeita a erros. Esclarece-se que uma série é dita estacionária quando ao se excluir as flutuações aleatórias, as observações amostrais são invariantes, com relação à cronologia de suas ocorrências (Naghetini e Pinto, 2007).

Em estudos de caráter local, em que o objetivo é entender o que acontece em uma dada estação de medição, sem se preocupar com o que acontece nas proximidades, os erros supracitados dependem do teste empregado, do nível de significância escolhido para a realização do teste, assim como de características da série em avaliação, tais como: tamanho amostral, magnitude da mudança, quando presente, assimetria, distribuição de probabilidade geradora da amostra e seu grau de autocorrelação.

Apesar do esforço da comunidade científica em entender como esses fatores afetam os possíveis erros envolvidos nos principais testes empregados em estudos de detecção, ainda existem muitas lacunas no conhecimento. Por exemplo, a relação do poder estatístico de diferentes testes, assim como dos erros Tipo I e Tipo II associados aos fatores anteriormente mencionados, é relativamente bem entendida no caso de variáveis normais, embora a questão da autocorrelação ainda não esteja bem resolvida, como será discutido mais à frente. Mas o mesmo não pode ser dito em relação a outras distribuições, como a Generalizada de Valores Extremos (GEV), muito empregada em séries de extremos, necessitando haver um esforço maior para entender como os testes se comportam em amostras oriundas dessa distribuição.

Mais recentemente, em estudo ligados à área da medicina, iniciou-se uma discussão acerca de outros tipos de erros que podem prejudicar as análises dos resultados obtidos em estudos de detecção a nível local. Trata-se dos erros Tipo S e Tipo M (Gelman e Carlin, 2014). O primeiro está relacionado com a estimativa equivocada (contrária) do sinal (S) de mudança, ou seja, quando da realização de um teste de hipótese, existe a probabilidade de se obter uma estimativa de tendência crescente (decrecente) quando na verdade a mesma é decrescente (crescente). A probabilidade do erro Tipo S está associada a fatores similares àqueles mencionados anteriormente, mas esses aspectos são pouco discutidos e entendidos, embora sejam importantes. O erro Tipo M refere-se a uma tendência em se superestimar a magnitude das tendências, especialmente quando o poder do teste é baixo. Esses dois erros não são discutidos na literatura de detecção em variáveis hidrometeorológicas, apesar da importância de seu entendimento no processo de análise de resultados dos testes. Afinal, tão importante quanto a detecção é a estimativa, de forma precisa, da magnitude da mudança, que irá sinalizar aos gestores sua gravidade e a necessidade ou não de intervenção imediata.

Por último, mas não menos importante, pode-se mencionar a questão da presença de autocorrelação na série histórica da variável de interesse. Sua presença aumenta a chance de se obter um resultado significativo do ponto de vista estatístico, mesmo quando a tendência não está presente nos dados (falso positivo). No caso de extremos hidrometeorológicos, essa questão é bastante relevante quando se deseja detectar tendências em vazões mínimas, por exemplo. As soluções usualmente empregadas são a (i) modificação da série original por meio da remoção da estrutura de correlação ou (ii) a obtenção de uma distribuição nula

compatível com a presença de autocorrelação. Apesar de poder distorcer de uma forma significativa os resultados obtidos em estudos de detecção a nível local, essa questão é muitas vezes negligenciada, especialmente em estudos no Brasil (Detzel *et al.*, 2011; Scofield *et al.*, 2014 e Rosin *et al.*, 2015)

Embora a análise de detecção de tendências em variáveis hidrometeorológicas a nível local seja bastante comum, muitos estudos realizados no mundo possuem um caráter regional ou até mesmo nacional, geralmente com o objetivo de identificar regiões onde existe um número elevado de estações com tendência. A análise regional de detecção de tendências padece de aspectos ligados à aplicação de múltiplos testes de hipóteses, simultaneamente, para estações de uma dada região, em que a finitude e a existência de interdependência entre os pontos de monitoramento comprometem a capacidade de detecção de tendência verdadeiras. Esse problema, conhecido como multiplicidade de testes, passou a ser efetivamente considerado a partir dos estudos de Livezey e Chen (1983). Diante da necessidade de se obterem resultados de detecção de tendência regionais mais verossímeis para subsidiar o processo decisório e de gestão de recursos hídricos, a evolução das técnicas que lidam com o problema da multiplicidade, ao longo do tempo, levou ao surgimento de técnicas de controle global de erros em testes de hipóteses, tais como o *Family-Wise Error* (FWER) e o *False Discovery Error* (FDR) (Benjamini e Hockberg, 1995) e suas variações (Benjamini e Hockberg, 1997; Benjamini e Hockberg, 2000 e Benjamini e Yekutieli, 2001).

O emprego das técnicas de controle de erros tipo I pelo *FDR* voltado para a hidrologia é relativamente recente (Ventura *et. al.*, 2004), não podendo o mesmo ser dito de outras áreas da ciência, como a epidemiologia e a genética (Storey e Tibshirani, 2003 e Efron, 2008). Com a vantagem de controlar a proporção de falsos positivos em relação ao número de rejeições, independente da quantidade de estações apontadas como não-estacionárias, as referidas abordagens evitam a elevada taxa de rejeições equivocadas, que são usualmente encontradas quando a questão da multiplicidade não é considerada.

Assim, diante das lacunas de conhecimento identificadas e das dificuldades relatadas à capacidade de detecção de tendências verdadeira e estimativa da mudança, sinal e magnitude, o presente trabalho visa contribuir na avaliação do desempenho de métodos

estatísticos quando aplicados a séries de extremos hidrometeorológicos, levando-se em consideração os efeitos da presença de autocorrelação serial e da multiplicidade de testes. Para tal, as metodologias testadas serão aplicadas a séries sintéticas, obtidas via simulação Monte Carlo, e reais, de índices de extremos hidrometeorológicos extremos estimados para estações pluviométricas e fluviométricas localizadas na bacia do São Francisco. Esclarece-se que a escolha da bacia do São Francisco como área de estudo está atrelada a sua relevância estratégica para a sociedade brasileira e a sua inserção no projeto de pesquisa CAPES em andamento: “*Impacto das Mudanças Climáticas em Extremos Hidrológicos (SECAS e CHEIAS)*”.

Assim, este trabalho é dividido em 5 partes principais: **Objetivos**, em que são delineados os objetivos principais e específicos do estudo com base no histórico apresentado; **Revisão Bibliográfica**, em que são apresentados os estudos já desenvolvidos sobre o tema com suas vantagens e limitações, bem como a base teórica utilizada neste trabalho; **Metodologia**, em que são indicados os passos que foram seguidos para alcançar os objetivos propostos, **Resultados e Discussão**, em que se apresentam os principais resultados e, por fim as **Conclusões**.

2. OBJETIVOS

2.1. OBJETIVO GERAL

O presente estudo se propõe a avaliar métodos estatísticos de detecção de tendências monotônicas em séries de extremos hidrometeorológicos, que visam controlar a rejeição equivocada da hipótese nula, *erro tipo I*, quando múltiplos testes de hipóteses são simultaneamente aplicados, levando-se em consideração a presença de correlação temporal e espacial, a fim de subsidiar o planejamento e a gestão de recursos hídricos.

2.2. OBJETIVOS ESPECÍFICOS

Buscando-se alcançar o objetivo geral, foram definidos os seguintes objetivos específicos:

- 1) Avaliar, por meio de estudo de Simulação Monte Carlo, como o tamanho da série histórica (n), magnitude da tendência (b), coeficiente de variação (C_v) e parâmetro de forma (κ) da distribuição generalizada de valores extremos (GEV) afetam o Poder do Teste de Mann-Kendall e a capacidade do mesmo em manter o percentual de erros do tipo I dentro do valor nominal de significância, quando inexistente tendência.
- 2) Avaliar, por meio de estudo de Simulação Monte Carlo, como o tamanho da série histórica (n), magnitude da tendência (b), coeficiente de variação (C_v) e parâmetro de forma (κ) da distribuição generalizada de valores extremos (GEV) afetam os erros do tipo S e M associados ao estimador de magnitude da mudança proposto por Sen (1968) – B_SEN.
- 3) Identificar como métodos recentemente desenvolvidos e aplicados nas áreas de genética e epidemiologia, baseados no conceito de controle de falsos positivos (*False Discovery Rate* – FDR), se comportam em estudos regionais de detecção de tendência monotônicas em séries de extremos hidrometeorológicos e posterior aplicação de tais métodos na bacia do Rio São Francisco.

3. REVISÃO BIBLIOGRÁFICA

A presente seção apresenta o que já foi realizado por outros pesquisadores e quais foram suas conclusões, no que diz respeito a estudos que se relacionam aos objetivos geral e específicos propostos no capítulo 2. Em suma, esse capítulo irá apresentar conceitos, delimitar o problema, o estado atual dos conhecimentos sobre a temática de estudo, as lacunas existentes e a contribuição que se espera com a realização do presente trabalho em relação ao campo de estudo.

Assim, este capítulo foi subdividido em cinco seções, a saber: (3.1) Eventos Hidrológicos Extremos, (3.2) Índices Extremos Hidrometeorológicos, (3.3) detecção de tendências monotônicas, (3.4) Problema da Autocorrelação e (3.5) Questão da Multiplicidade.

3.1. EVENTOS HIDROLÓGICOS EXTREMOS

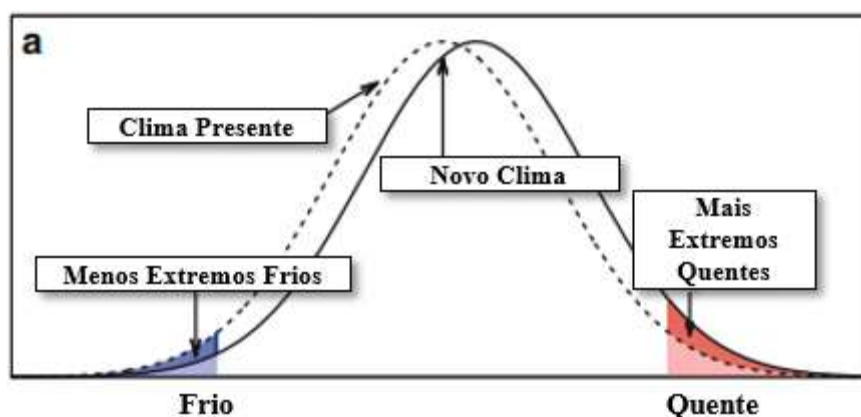
Por definição, extremos climáticos são eventos raros; entretanto, as mudanças no clima têm resultado em alterações na sua frequência e intensidade (Seneviratne *et al.*, 2012). Seja pela ação de forças externas, como o aumento da emissão de gases do efeito estufa, ou resultado da variabilidade natural do clima, o cenário de mudança visto para algumas variáveis hidrológicas, tais como a vazão e precipitação, varia de local para local. Essa constatação tem fomentado debates acerca da consideração de não-estacionariedade em problemas que vão desde a gestão de recursos hídricos até a elaboração de projetos de engenharia.

Ao empregar métodos baseados na teoria dos valores extremos para o dimensionamento de projetos de estruturas hidráulicas, como diques e reservatórios, é usual assumir que o comportamento da série observada até o momento de projeto se perpetuará ao longo da vida útil do empreendimento, e, conseqüentemente, que os riscos hidrológicos associados permanecerão os mesmos. Esse pode não ser o caso de algumas localidades, que experimentam mudanças significativas em variáveis hidrológicas, por exemplo, na vazão, em período pós-construção de uma dada infraestrutura. Assim, negligenciar as alterações detectadas pode não ser a melhor alternativa, pois essa omissão pode resultar no colapso das estruturas hidráulicas e/ou em perdas de vidas humanas, razão pela qual discute-se a necessidade de se considerar as eventuais alterações observadas nas variáveis hidrológicas

durante a elaboração de projetos de engenharia e/ou readequação de estruturas hidráulicas para um cenário de mudança.

O progresso no conhecimento das implicações das mudanças nos extremos hidrológicos depende fortemente da capacidade de detectar e reduzir as incertezas no entendimento das alterações constatadas no período instrumental (Zwiers *et al.*, 2011). Para melhor compreender esse fenômeno é necessário promover a melhoria na qualidade e disponibilidade das séries de dados observados, o desenvolvimento de melhores modelos físicos, avanços no poder das técnicas estatísticas, refinamento na compreensão dos processos físicos que produzem os extremos e melhoria na capacidade de atribuir causas às mudanças detectadas.

As mudanças nas séries hidrológicas podem se manifestar de maneira abrupta (saltos), gradual (monotônicas) ou através de formas mais complexas. Podem ser observadas nos valores médios, na variabilidade (variância, extremos, persistência) ou em flutuações na distribuição anual (mudanças de sazonalidade e regimes fluviais), conforme ilustrado na Figura 3.1. Mudanças abruptas normalmente ocorrem como resposta a uma alteração repentina na bacia, tal como construção de reservatórios, desvios em cursos d'água, etc. Elas também podem surgir a partir do uso de novos processos de medição das variáveis aleatórias nas estações de monitoramento e alterações na curva-chave. Já as mudanças hidrológicas graduais, objeto de atenção deste trabalho, tipicamente têm origem em causas que paulatinamente se modificam, como é o caso de urbanizações, alterações do uso e ocupação do solo e variabilidade climática natural, apenas para citar os principais (WMO, 2009).



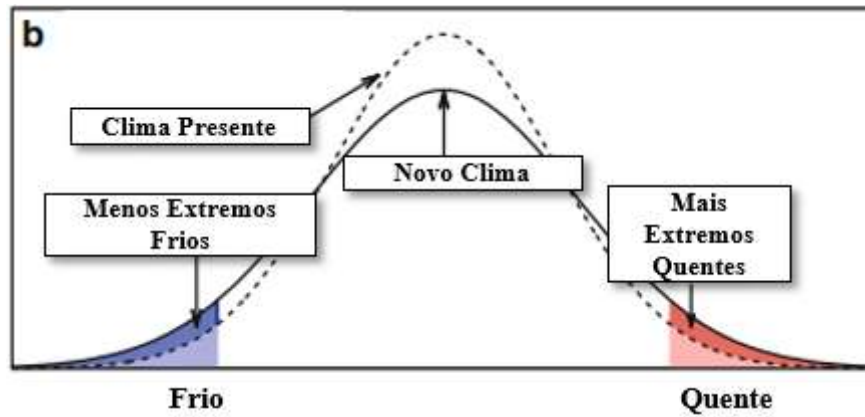


Figura 3.1 - Representação esquemática de como as mudanças climáticas (a) crescimento da média e (b) aumento da variância a afeta da distribuição de probabilidades da temperatura diária (Zwiers *et. al.*, 2015, adaptado)

O fluxograma da Figura 3.2 é baseado na subdivisão temática proposta por Hay *et al.* (2015) e ilustra a estratégia de ação para compreender os desdobramentos e consequências das mudanças climáticas, após a detecção das mesmas.

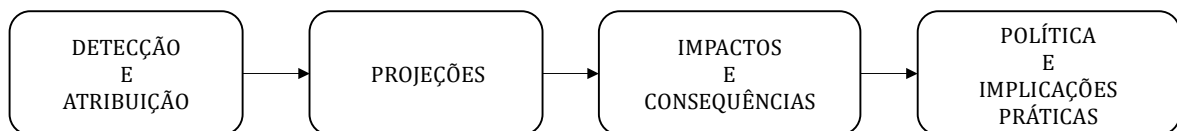


Figura 3.2 - Fluxograma da estratégia de ação para estudos de mudança climática. A cronologia de execução das atividades proposta é da esquerda para a direita (Hay *et. al.*, 2015, adaptado)

Para a realização da totalidade das tarefas envolvidas nessas 4 etapas principais da estratégia de ação, é necessário o envolvimento de uma equipe multidisciplinar, capaz de avaliar os efeitos da mudança sob diferentes aspectos, tal como o *Expert Team on Climate Change Detection and Indices* (ETCCDI), parte integrante do CLIVAR/WMO (do inglês, *Climate Variability & Predictability Programme*). Assim, dada a necessidade de domínio de diversos campos da ciência para a integralização do processo em pauta, optou-se, neste trabalho, em restringir o foco de atuação em uma porção específica do primeiro passo metodológico, a *detecção*.

3.2. INDICES DE EXTREMOS HIDROMETEOROLÓGICOS

As análises globais de mudanças climáticas de longo-termo em extremos foram negligenciadas por décadas, graças à inexistência de base diária de dados abrangente e suficientemente longa para subsidiar sua detecção. Ao invés disso, o foco das avaliações estava voltado para a busca de alterações nos valores médios de variáveis hidrometeorológicas, devido à capacidade da rede de estações de monitoramento em prover eficiente cobertura do globo terrestre e com a periodicidade mensal para dados de temperatura e precipitação (Peterson, 2005).

Assim, diante desse cenário, a Organização Meteorológica Mundial (WMO), em um esforço conjunto com o CCI/CLIVAR/JCOMM (do inglês, *Commission for Climatology/ Climate Variability & Predictability Programme/ Joint Technical Commission for Oceanography and Marine Meteorology*), criou, na primeira metade da década de 2000, um grupo de trabalho denominado ETCCDI (do inglês, “*Expert Team on Climate Change Detection and Indices*”). Esse grupo tinha o objetivo de elaborar uma metodologia de análise de mudanças climáticas que fosse capaz de conciliar as especificidades existentes em cada um dos continentes, ou até mesmo países, sem que o aspecto global fosse negligenciado (Zhang *et al.*, 2011).

O ETCCDI elaborou 27 índices climáticos, dentre os quais 11 estão relacionados com a precipitação e 16 referem-se à temperatura do ar. Relata-se que tais índices foram desenvolvidos para caracterizar diferentes partes da distribuição de probabilidade de uma dada variável, sendo que os de interesse para este trabalho são aqueles que contêm informações à respeito das caudas distributivas, ou seja, relacionados a extremos, principalmente de precipitação: a razão entre o total anual precipitado e o número de dias chuvosos (*SDII*); o número de dias com a precipitação superior a um dado valor “x”, em que x é igual a 10, 20 e 50 mm (*R_x*); número de dias consecutivos sem chuva ($< 1\text{ mm}$) (*CDD*); número de dias consecutivos chuvosos (*CWD*); número de dias chuvosos em um ano (*RI*); máximo anual acumulado de precipitação, dado um período $t=1$ e 5 dias (*R_xt*); o número de dias com precipitação superior ou igual a um dado percentil $p=95$ e 99% (*R_p*) e total anual de precipitação (*PRCPTot*).

De acordo com Zwiers *et al.* (2011), os índices climáticos formulados pelo ETCCDI são de simples uso e cálculo, capazes de apresentar informações relacionadas à mudança de maneira compacta e aceitável para diversos tipos de usuários. Quando aplicáveis, esses índices serão utilizados no presente trabalho para expressar os aspectos climáticos observados na pluviometria da área de estudo. Partindo do mesmo princípio e procurando as mesmas vantagens relatadas por esses autores, foram definidos índices específicos relacionados à vazão que tentam abranger a máxima, média e mínimas diárias anuais, por exemplo, Q_{\max} , $Q_{\text{méd}}$ e $Q_{7\text{min}}$, que são, respectivamente, a descarga máxima anual, descarga média e descarga média mínima de sete (7) dias consecutivos.

Salienta-se que maiores detalhes a respeito dos índices integrantes do ETCCDI selecionados para a pluviometria e aqueles adotados para a vazão serão apresentados em seção posterior.

3.3. DETECÇÃO DE TENDÊNCIAS MONOTÔNICAS

Detectar uma mudança climática equivale a identificar, a partir do emprego de metodologias de base estatística, alterações amostrais significativas em diferentes escalas de tempo e espaço. A tarefa de detecção é conceitualmente simples, mas de difícil aplicação, tanto pelas limitações de disponibilidade de base amostral quanto pelas dificuldades intrínsecas à própria atividade. Diversos pesquisadores no mundo têm se dedicado a esse campo de estudo, fato esse comprovado pelo grande número de trabalhos de avaliação de estacionariedade de variáveis hidrológicas publicados em periódicos internacionais (Lettenmaier *et al.*, 1994; Krakauer e Fung, 2008; Vogel *et al.*, 2011; Alkama *et al.*, 2013; Hossain, 2014; Madsen *et al.*, 2014; Bayazit, 2015) e no Brasil (Muller *et al.*, 1998; Detzel *et al.*, 2011; Aires de Souza *et al.*, 2013; Alves *et al.*, 2013; Araújo e Aires de Souza, 2014; Scofield *et al.*, 2014; Rosin *et al.*, 2015).

Dentre os vários métodos que podem ser empregados na detecção de tendências, distinguem-se 2 grupos como aqueles que são de maior utilização: (i) os baseados exclusivamente em testes de hipóteses, paramétricos e não-paramétricos, e (ii) os que procuram ajustar um conjunto de observações de uma variável aleatória a uma distribuição de probabilidades

específica, cujos parâmetros distributivos podem variar no tempo. O presente trabalho enfoca as metodologias de detecção que se enquadram no primeiro grupo.

As técnicas que compõem esse primeiro grupo prescindem do estabelecimento prévio de níveis de confiança e estão sujeitas a fatores que afetam a capacidade de detecção, tais como o tamanho amostral, magnitude da mudança, assimetria, variância e autocorrelação serial. As mesmas podem ser utilizadas para o cumprimento de recomendação do Projeto A2 do Programa Mundial Climático - Água (WCP), cujo objetivo é o estímulo da detecção de tendências em dados hidrológicos em escala mundial.

Assim, o presente tópico foi subdividido de modo que o referencial teórico, os testes de detecção empregados, os trabalhos de avaliação de desempenho dos testes de estacionariedade já realizados e a influência dos fatores que interveem na capacidade de percepção do sinal de mudança verdadeiro pudessem ser evidenciados. Ressalta-se que o efeito da correlação temporal será discutido em seção específica.

3.3.1. Referencial Teórico

Estudos de detecção de tendência monotônica são normalmente realizados com base em testes de hipóteses paramétricos ou não-paramétricos. Testes paramétricos assumem que a amostra, por exemplo, a série temporal de uma variável hidrometeorológica, é proveniente de uma distribuição específica, enquanto os não-paramétricos não fazem qualquer especificação prévia sobre a distribuição da qual são originados os dados amostrais.

Em geral, em testes dessa natureza, a hipótese nula (H_0) assume que a série em questão é estacionária, ou seja, não possui tendência, enquanto a hipótese alternativa (H_a) prevê que a mudança está de fato presente na série. A decisão sobre a presença ou não de estacionariedade na amostra é baseada na distribuição da hipótese nula da estatística do teste (T) específica para cada tipo de teste.

A distribuição nula é o nome que se dá à distribuição da estatística do teste T , quando a hipótese nula é verdadeira. A hipótese nula é geralmente rejeitada quando o valor amostral

de T se localiza na cauda da distribuição, indicando uma baixa probabilidade de se observar tal amostra, caso essa conjectura fosse realmente verdadeira. Embora não muito provável, é possível observar uma série estacionária, cujo valor de T esteja na cauda da distribuição nula ($p < 0.05$). Vem daí a definição do erro tipo I em testes de hipótese e sua relação com a definição da região de rejeição do teste, R .

Existe uma relação direta entre a região de rejeição R da estatística teste e a probabilidade de se cometer um erro tipo I, que ocorre quando se rejeita erroneamente a hipótese nula. A probabilidade de se cometer esse engano define o nível de significância do teste, α , que pode ser compreendido como a materialização do risco que se está disposto a correr em permitir que determinada hipótese seja rejeitada equivocadamente, conforme a Equação 3.1.

$$P(\text{Erro do Tipo I}) = P(T \in R | H_0) = \alpha \quad (3.1)$$

Outro erro passível de ser cometido quando se realiza um teste de hipóteses é o de tipo II, definido como a probabilidade de não rejeição da hipótese nula, dado que ela é de fato falsa (Eq. 3.2).

$$P(\text{Erro do Tipo II}) = P(T \notin R | H_a) = \beta \quad (3.2)$$

Na ausência de erros, ou seja, se as hipóteses nulas verdadeira e a falsa são, respectivamente, não rejeitada e rejeitada, a probabilidade de tomada dessa decisão é complementar àquelas atribuídas aos erros do tipo I e II, ou seja, $1-\alpha$ e $1-\beta$. Em especial, a probabilidade complementar a β , poder do teste, é outro parâmetro de interesse por representar a propensão de se rejeitar uma hipótese nula falsa, ou seja, reflete o “poder” de detecção real de evidências que levem a rejeição do teste realizado (Eq. 3.3).

$$\text{Poder do Teste} = P(T \in R | H_a) = 1 - \beta \quad (3.3)$$

Fortemente relacionados, os erros tipo I e tipo II apresentam comportamentos antagônicos, ou seja, o aumento da probabilidade de ocorrência de um implica na imediata redução da outra. Essa relação é claramente ilustrada na Figura 3.3, em que uma diminuição no valor de α (deslocamento do $T_{crítico}$ para a direita) implicaria em aumento no poder do teste $1-\beta$, e vice-versa.

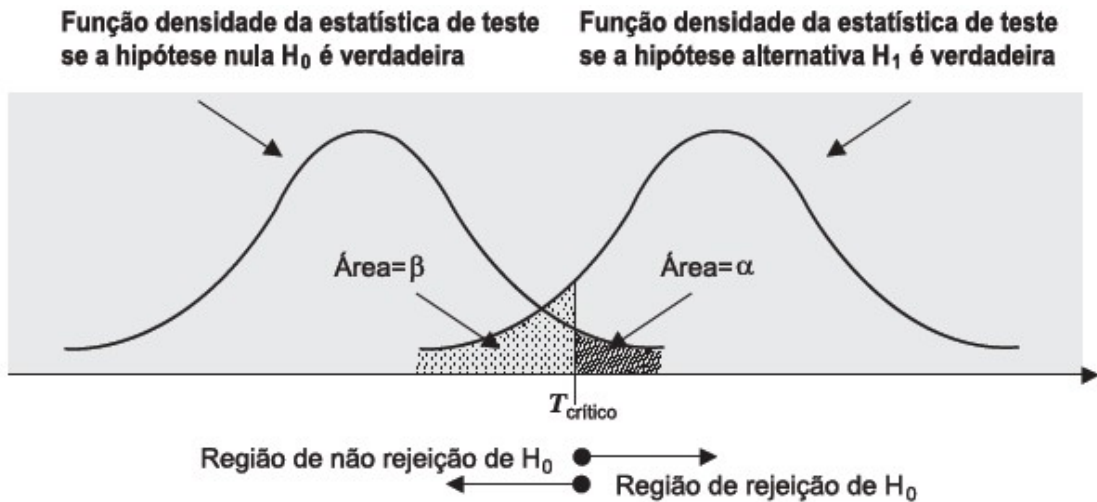


Figura 3.3 - Ilustração dos erros tipo I e II em um teste de hipóteses unilateral (Naghetini e Pinto, 2007)

A avaliação de estacionariedade pode ser de caráter regional, caso em que, invariavelmente é necessário realizar testes de hipóteses simultâneos em várias estações de uma determinada região. A Tabela 3.1 resume os resultados possíveis para as m avaliações de significância realizadas, bem como a quantidade de cada uma das opções factíveis de decisão: (i) nH_0 , número de hipóteses nulas verdadeiras; (ii) n_{FP} , número de hipóteses nulas rejeitadas erroneamente (erro tipo I, falso positivo); (iii) nH_a , número de hipóteses alternativas verdadeiras e, (iv) número de hipóteses nulas falsas não-rejeitadas (erro tipo II, falso negativa).

Tabela 3.1 - Possíveis resultados ao se realizar múltiplos testes de hipóteses

	Decisão		Soma
	Não-Rejeitar	Rejeitar	
Hipótese Nula Verdadeira (H_0)	nH_0	n_{FP}	m_0
Hipótese Alternativa (H_a)	n_{FN}	nH_a	$m - m_0$
Soma	$m - n_{rej}$	n_{rej}	m

3.3.2. Teste de Hipótese de Estacionariedade

Tratando-se de tendências monotônicas, doravante referidas apenas como tendências, diversas metodologias que fazem uso de testes de hipóteses podem ser utilizadas para a sua detecção na base amostral. No presente trabalho, optou-se por enfatizar três técnicas para melhor detalhamento de seus processos executivos e, por fim, selecionar aquela que será utilizada nas análises que serão procedidas para o atingimento do objetivo geral proposto: *t-test*, *Rho de Spearman* e *Mann-Kendall*.

O primeiro, *t-test*, assim como a grande parte dos testes paramétricos, é baseado na suposição de normalidade populacional e na avaliação da magnitude do coeficiente angular, b , obtido por meio de análise de regressão linear (OLS) com emprego da base amostral. Na prática, o procedimento proposto pelo teste consiste em determinar se o valor de b é estatisticamente diferente de zero ou não, dado um nível de significância α e uma hipótese nula, $H_0: b = 0$.

Por sua vez, o teste não-paramétrico de Rho de Spearman baseia-se na ideia de que uma tendência eventualmente existente na série de dados hidrológicos ao longo de um intervalo de tempo t pode ser detectada pela correlação entre a série e o índice de tempo. Formalmente, definido por Lehmann (1975), a estatística do teste de Spearman pode ser expressa conforme a Equação 3.4:

$$r_s = 1 - 6 \sum_{i=1}^n (m_t - T_t)^2 / (N^3 - N) \quad (3.4)$$

onde m_t é a ordem de classificação dos elementos de uma série de dados hidrológicos X_t , rearranjados em ordem crescente, e T_t são os índices tempo iguais a 1,2, ..., n.

Naghetini e Pinto (2007) apontam que é possível demonstrar que a distribuição de r_s pode ser aproximada por uma Normal de média igual a $E[r_s] = 0$ e variância dada por $VAR[r_s] = 1/(n - 1)$ para valores de $n > 10$ e sob a hipótese nula de que não há correlação entre m_t e T_t .

Assim, a estatística de teste de Rho de Spearman pode ser formulada como:

$$T = \frac{r_s}{\sqrt{\text{Var}[r_s]}} \quad (3.5)$$

O teste de Rho de Spearman e o *t-test* são bilaterais, sendo que a rejeição da hipótese nula se dará se $|T| > Z_{1-\alpha/2}$ para um nível de significância α .

O teste de Mann-Kendall assume como hipótese nula que a amostra $X = \{X_1, X_2, \dots, X_n\}$ é formada por variáveis independentes, provenientes de uma mesma distribuição estatística e sem tendência. Já a hipótese alternativa aponta para a existência de tendência monotônica nas séries temporais analisadas. Mann (1945) e Kendall (1975) desenvolveram a estatística de teste *S* baseando-se no fato de que se a hipótese nula é verdadeira, de modo que a ordem das variáveis que foi observada na amostra seria apenas uma possibilidade das $n!$ equiprováveis que poderiam ter acontecido. Então, sugeriu-se o uso da estatística *S*, conforme apresentada abaixo:

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{sign}(x_j - x_i) \quad (3.6)$$

onde:

$$\text{sign}(x_j - x_i) = \begin{cases} 1, & \text{se } x_j > x_i \\ 0, & \text{se } x_j = x_i \\ -1, & \text{se } x_j < x_i \end{cases} \quad (3.7)$$

Nota-se que a estatística de teste *S* não depende da magnitude dos valores observados para X , mas sim de sua classificação, o que significa que nenhuma suposição é feita a respeito da natureza da distribuição de probabilidade que gerou a amostra. Isso explica o porquê do teste MK não ser muito sensível a valores amostrais discrepantes e ser capaz de manter o poder de detecção, mesmo quando a distribuição populacional é bastante assimétrica.

Quando a hipótese nula é verdadeira, Kendall (1975) mostrou que, para valores de $n > 10$, a distribuição de S pode ser aproximada por uma distribuição Normal, $S \sim N(0, \sigma_0^2)$, com média nula e variância dada por:

$$\sigma_0^2 = \frac{n(n-1)(2n+5) - \sum_{j=1}^m t_j j(j-1)(2j+5)}{18} \quad (3.8)$$

onde t_j é o t -ésimo empate na série e m é o número de grupos de empates. Para melhor ilustrar a formação dos chamados grupos de empate e a determinação de t , apresenta-se o seguinte exemplo: suponha-se que a série de dados amostrais seja composta pelos seguintes valores discretos da variável aleatória $X = \{1, 1, 2, 2, 3, 5, 3, 3\}$. Nesse caso, têm-se dois (2) grupos de empates (m): Grupo 1: agrupamentos em que ocorreram apenas uma a igualdade, subconjuntos $\{1,1\}$ e $\{2,2\}$; Grupo 2: existência de dois empates, $\{3,3,3\}$. Para a situação em questão, os valores de t são 2 e 1 para, respectivamente, Grupo 1 e Grupo 2.

Como a distribuição de S é conhecida, quando a hipótese nula é verdadeira, a região de rejeição R pode ser facilmente definida. Primeiro, dado que S é normalmente distribuído, então $S/\sigma \sim N(0,1)$. Kendall (1975) sugere a adição ou subtração de uma unidade para a correção de continuidade, conforme mostrado abaixo

$$Z = \begin{cases} \frac{S-1}{\sigma}, & \text{se } S > 0 \\ 0 & , \text{se } S = 0 \\ \frac{S+1}{\sigma}, & \text{se } S < 0 \end{cases} \quad (3.9)$$

onde Z é a variável aleatória da Normal padrão.

Assim, para um teste bicaudal com o nível de significância igual a $\alpha\%$, se $Z < Z_{\alpha/2}$ ou $Z > Z_{1-\alpha/2}$, a hipótese nula é rejeitada e assume-se que a tendência é estatisticamente significativa para a estação de monitoramento avaliada. Diz ainda que o p-valor pode ser estimado por:

$$p_{valor} = 2\{\min[\Phi(z), 1 - \Phi(z)]\} \quad (3.10)$$

Dentre os três testes apresentados, o *t-test* é aquele que apresenta melhores resultados quando a amostra é de fato normalmente distribuída, embora as diferenças ao compararmos com o Rho de Spearman e Mann-Kendall não sejam muito grandes (Yue e Pilon, 2004). Entretanto, o *t-test* apresenta algumas desvantagens: (i) é muito sensível a presença de disparidades amostrais, *outliers*, por considerar, conforme descrito, os valores observados da variável aleatória diretamente na determinação da estatística de teste, β ; (ii) perde seu poder à medida que a violação da hipótese de normalidade aumenta, fato esse bem possível de se ocorrer quando se lida com séries de extremos hidrometeorológicos.

Adicionalmente, Yue e Pilon (2004) mostraram que os dois testes não-paramétricos baseados em classificações eram menos suscetíveis a presença de *outliers*, o que era esperado, já que ambos não são baseados diretamente nos valores amostrais. Além disso, os autores mostraram que o Rho de Spearman e o Mann Kendall superam o desempenho do *t-test*, quando as séries analisadas não apresentam a distribuição populacional Normal, ou seja, são do tipo Gumbel, LP3, EV2 e Weibull, apenas para citar algumas.

Ainda quanto aos testes não-paramétricos, estudos de simulação Monte Carlo conduzidos por Yue *et al.* (2002a) mostraram também que Rho de Spearman e Mann-Kendall possuem desempenhos similares, mas que, por alguma razão, o teste de Mann-Kendall tornou-se o mais popular para estudos de detecção de tendências monotônicas em dados hidrometeorológicos.

Diante do exposto, será utilizado o teste de Mann-Kendall como técnica para as avaliações de estacionariedade realizadas neste trabalho, excetuando-se em situações que a questão da multiplicidade for levada em consideração.

3.3.3. Estimativa da Magnitude de Tendências: B_SEN

Como alternativa ao coeficiente de regressão β obtido pelo método dos mínimos quadrados (OLS), Sen (1968) sugere o emprego de estimador não-paramétrico baseado no Coeficiente de Correlação Tau de Kendall para a determinação da magnitude de tendências, a fim de evitar a influência de pontos amostrais discrepantes na estimativa realizada de b .

O estimador proposto por Sen (1968), $Sen(b)$, prevê que seja calculada para cada um dos $n(n-1)/2$ pares de observação amostrais possíveis (X_j, X_i) , a chamada *inclinação do par ordenado*, d_{ij} , dada por:

$$d_{ij} = (X_j - X_i)/(j - i) \quad (3.11)$$

Onde i e j são os índices de tempo das posições amostrais ocupadas, respectivamente, por X_i e X_j , sendo que $1 \leq i \leq j \leq n$.

Para as situações em que as séries temporais apresentam falhas, ou seja, na ausência de observação da variável aleatória para certo intervalo de tempo, deve-se proceder na definição dos índices i e j , conforme exemplo descrito a seguir.

Considere o seguinte conjunto de pares ordenados, anos e vazão máxima anual, para uma estação fictícia: $\{(1957, 350,4); (1959, 502,4). (1960, 380,5)\}$. Assim, no caso apresentado, o primeiro valor de d_{ij} será obtido a partir da resolução da seguinte expressão: $(502,4 - 350,4)/(3 - 1)$. É importante notar que, dada a ausência de informações disponíveis para o ano de 1958, o índice i do valor X_i de 1959, que ocupa a segunda (2ª) posição da série de dados, é igual a 3. Isso nos leva a concluir que o valor do índice de tempo imediatamente posterior a falha, j , pode ser obtido pela soma do i anterior à lacuna amostral e o número de falhas detectado.

Formalmente, define-se o estimador de Sen (b) como sendo a mediana de todos d_{ij} calculados, a saber:

$$b = med\{d_{ij}\} \quad (3.12)$$

3.3.4. Perspectiva Histórica: Avaliação das Técnicas de Detecção

Conhecer os aspectos intervenientes na capacidade de se detectar ou não tendências é fundamental para proceder a análise dos resultados obtidos pela aplicação de testes estatísticos de avaliação de estacionariedade. Podendo afetar a taxa de rejeição da hipótese nula, fatores como o tamanho serial, magnitude da não-estacionariedade, nível de significância, coeficiente de variação e até o tipo da distribuição que melhor se ajusta à base amostral, foram exaustivamente estudados e tiveram seus efeitos descritos por vários estudos já realizados. Posto isso, o objetivo dessa seção é o de apresentar, de maneira sintética, a descrição dos procedimentos utilizados pelos diferentes autores ao se avaliar o impacto de uma determinada variável no processo de detecção, bem como sinalizar as lacunas de conhecimento por eles deixadas e seus resultados.

Partindo-se da idéia supracitada, Yue *et al.* (2002a), a partir de simulações Monte Carlo (MC), investigaram o poder de testes não-paramétricos usualmente empregados em estudos hidrológicos, no caso, Mann-Kendall e Rho de Spearman, para amostras independentes, ou seja, sem a presença de correlação temporal. A partir da geração de 2,000 séries sintéticas para cada uma das seguintes possíveis configurações: distribuição normal, log-normal, Pearson III, EV1, EV2, EV3; variância ($\sigma = (0.1i)^2$, $i = 1 (1) 10$); tamanho amostral ($n = 10 (1) 100$), magnitude de tendência ($T_t = bt$, $b = -0.01 (0.002) 0.01$), $t = 0, 1, 2, \dots, n$) e nível de significância ($\alpha = 0.002, 0.005, 0.01, 0.025 (0.025) 0.20$), Yue *et al.* (2002a) demonstraram que o poder do teste para séries normais cresce à medida que o valor absoluto da tendência se eleva e o tamanho da série aumenta. Para a variação amostral, nota-se um efeito contrário: um crescimento na variância implica em redução na capacidade de identificação de tendências estatisticamente significantes do teste. Ademais, Yue *et al.* (2002a) apontaram que, para séries não-estacionárias, o tipo de distribuição e a assimetria da série temporal também influenciam no poder dos testes avaliados (EV3 e Log-Normal são, respectivamente, as distribuições de maior e menor poder, quando confrontadas sob as mesmas condições). Ressalta-se que a constatação de dependência distributiva é

intuitivamente controversa, já que as metodologias de avaliação empregadas em sua essência são ditas não-paramétricas, ou seja, independem da natureza da qual os dados provem.

Análogo ao proposto por Yue *et al.* (2002a), Yue *et al.* (2004) propõem a realização de avaliação do poder dos testes estatísticos: *t*-test, MK e testes baseados em reamostragem com reposição (*BS-slope*, *BS-MK*), via experimentos MC, para séries independentes normalmente distribuídas ou não (Pearson III, Gumbel, EV2 e Weibull), quando existe a presença de tendências monotônicas lineares ou não-lineares. Para o caso de amostras normais e linearmente não-estacionárias, adotaram-se os seguintes valores para as séries simuladas: média ($\mu = 1.0$), coeficiente de variação ($C_V = 0.5$), magnitude da tendência ($T_t = bt = 0.00 (0.004)0.02$), nível de significância ($\alpha = 0.05$) e tamanho da amostra ($n = 50$). Especificamente para as situações de não-normalidade distributiva definiram-se para a EV2: $\kappa = -0.3$; Weibull: $\omega = 0.6$; Pearson III (P3): $\gamma = 1.5$, sendo as demais grandezas adotadas idênticas àquelas anteriormente apresentadas. Os resultados das simulações realizadas indicaram que (a) o *t*-test e o *BS-slope* possuem o mesmo poder, assim como MK e BS-MK, (b) para uma base de dados normalmente distribuída, o poder daqueles métodos baseados na avaliação de significância estatística do coeficiente angular da reta ajustada entre os pontos amostrais, *t*-test e o *BS-slope*, foi ligeiramente maior do que as demais técnicas avaliadas e (c) considerando-se a não-normalidade das séries temporais, os testes de MK e BS-MK foram os que demonstraram possuir maior poder de detecção. Adicionalmente, o estudo realizado por Yue *et al.* (2004) apontou também que a capacidade dos testes em detectar tendências verdadeiras é ligeiramente sensível ao formato da não estacionariedade, ou seja, se ela é linear ou não.

Além dos aspectos intervenientes no ato de detectar elencados, é importante se ter em mente que, mesmo quando a componente de mudança está presente, não significa que a detectaremos para um dado nível de significância arbitrado. Pittock (1980) e Chiew e McMahon (1993) examinaram a capacidade de se detectar alterações em séries de dados hidrológicos. Eles encontram que é improvável que uma tendência seja identificada se esta não se prolongar continuamente por um determinado intervalo de tempo.

Assim, partindo-se da dificuldade real em se detectar tendências estatisticamente significantes, principalmente quando o sinal de mudança é fraco, mas a variabilidade

amostral é consideravelmente grande, Radziejewski e Kundzewicz (2004) propuseram a geração de séries sintéticas compostas pela associação de duas parcelas: (i) base – proveniente de um processo aleatório normalmente distribuído ($N(0,1)$) e independente e (ii) mudança – oriundo de um processo determinístico e de dois tipos: gradual (linear) e abrupto (salto). Realizada de maneira controlada, a adição entre as parcelas supracitada objetivou determinar duas porções distintas na série temporal artificial, cujo limite entre ambas, o ponto de mudanças, marca o início da parcela serial que foi afetada pela mudança (“período de mudança”) ou o término do intervalo em que a interferência da tendência inexistente (“período de referência”). Variando entre 10 e 70 anos (o período de referência é fixo e igual a 30) e igual à 0.05 vezes o valor do desvio padrão (SD) da parcela aleatória da série por ano, no caso monotônico, e 1 SD para situações de salto, foram definidos, respectivamente, o tamanho do período de mudança (“*run up time*”) e seu gradiente de alteração. Aplicando-se os testes de Mann-Kendall, Rho de Spearman e regressão linear para às séries monotônicas, e CUSUM e desvios acumulados (Buishand, 1982) àquelas que possuem salto, Radziejewski e Kundzewicz (2004) concluíram que a incapacidade de um teste estatístico em detectar a presença de mudança não implica necessariamente em sua inexistência, visto que eles não são capazes de detectar tendências de baixa magnitude, ou que se prolongam por um curto espaço de tempo após o início da mudança.

Como se observa pela perspectiva histórica apresentada, os fatores intervenientes na tarefa de detecção, conforme já mencionado, foram quase em sua totalidade mapeados, seus efeitos estudados e abordados pelos diversos estudos já procedidos. Entretanto, considerando-se o foco em extremos hidrológicos dispensado neste trabalho e a falta de uma abordagem específica que contemple a influência do tamanho, coeficiente de variação e assimetria amostral associadas a distribuições extremas não-estacionárias no poder de detecção do teste de Mann-Kendall, optou-se pela realização de tal avaliação. Em linhas gerais, o que se propõe é a geração, via simulação Monte Carlo (MC), de conjuntos amostrais de séries temporais distribuídos conforme uma GEV, parâmetros amostrais n , C_v e γ e magnitude da tendência (b). De posse das amostras sintéticas, aplica-se o teste de MK e determina-se o seu poder pela razão entre o número de hipóteses rejeitadas e a totalidade de séries geradas, no caso 10,000 para cada uma das combinações possíveis entre n , C_v , γ e b . Salienta-se que

maiores detalhes a respeito da análise MC mencionada serão apresentados em seção posterior desta dissertação.

3.4. O PROBLEMA DA AUTOCORRELAÇÃO

3.4.1. O Impacto da Correlação nos Resultados de Detecção

A presença de autocorrelação amostral viola a suposição de independência da grande maioria dos testes de hipóteses para detecção de tendências, comprometendo, portanto, a acurácia e concisão dos resultados obtidos a partir de sua aplicação direta. Como exemplo desse problema, cita-se o teste de Mann-Kendall (MK), o qual foi concebido para séries de dados independentes e cuja capacidade de controle sobre o Erro Tipo I é perdida, quando aplicado às séries constituídas de observações correlacionadas. Em outras palavras, diz-se que a probabilidade de detecção equivocada de tendências significantes pelo uso do teste de MK torna-se maior do que o nível de significância previamente estabelecido (Von Storch, 1995). É importante ressaltar que em casos de múltiplos testes, a não consideração da autocorrelação pode ocasionar, por exemplo, a superestimação do número de estações de monitoramento consideradas com tendências estatisticamente significantes de uma região qualquer.

Dada a complexidade de eliminação do efeito exercido pela correlação serial sobre a detecção de tendências, a literatura aponta que existem duas estratégias para lidar com o problema relatado. A primeira delas consiste na aplicação de métodos que modificam a série originalmente observada, removendo a autocorrelação e, posteriormente, para o conjunto de dados modificado, realiza-se o teste de hipótese para a avaliação da mudança. O Pre-Whitening (PW), desenvolvido por Kulkarni e Von Storch (1995), e o Trend-Free-Pre-Whitening, introduzido por Yue *et al.* (2002b), são exemplos da metodologia supracitada.

Já a segunda maneira de lidar com a autocorrelação não prevê alteração da série original, mas prescinde de ajustes a serem realizados na distribuição de probabilidades da estatística de teste quando a hipótese nula é verdadeira. Essa abordagem pode ocorrer de diferentes formas, por exemplo, através da correção da variância da estatística de teste de MK como

realizado por Hamed e Rao (1998), ou empregando-se uma nova distribuição nula para a estatística de teste, conforme derivado analiticamente por Hamed (2009) ou obtido pelo processo de reamostragem com reposição agrupada, *bootstrap* em blocos (Onoz e Bayazit, 2012).

Dada a vastidão de técnicas sugeridas para a mitigação do problema ocasionado pela presença da autocorrelação amostral, na sequência deste capítulo é apresentada uma revisão sintética de estudos de desempenho de algumas das metodologias empregadas para lidar com a correlação temporal em testes de detecção, como o Pre-Whitening (PW) e Trend-Free-Pre-Whitening (TFPW). Ademais, será discutido o referencial teórico de cada um dos dois métodos supracitados, bem como seus passos metodológicos de aplicação.

3.4.2. Pre-Whitening

O procedimento de PW é muito simples de ser aplicado e prescinde, para a obtenção de bons resultados, de que o conjunto de dados da amostra avaliada, X_t , possa ser expresso adequadamente por um modelo autoregressivo lag 1, AR(1), cuja definição formal é dada por

$$X_t = \theta X_{t-1} + \epsilon_t \quad (3.13)$$

onde θ é o coeficiente de AR(1) e ϵ_t é o erro normalmente distribuído do modelo, $N(0, \sigma_\epsilon^2)$.

De maneira sistemática, a metodologia proposta por Kulkarni e Von Storch (1995) pode ser realizada através dos seguintes passos:

- 1) Estimativa da autocorrelação amostral (\hat{r}_1) por $r_1 = \frac{\sum_{i=1}^{n-1} (X_i - \bar{X})(X_{i+1} - \bar{X})}{\sum_{i=1}^n (X_i - \bar{X})^2}$
- 2) Se \hat{r}_1 é estatisticamente diferente de zero, o teste de MK é aplicado para a série modificada $X_t^{mod} = X_t - \hat{r}_1 X_{t-1}$ para $t = 2, \dots, n$. Caso contrário, o teste de Mann Kendall é aplicado à série original e a avaliação de significância é realizada conforme demonstrado na seção 3.3.2.

Quanto à aplicabilidade do PW para a eliminação da influência da correlação serial, conforme descrito acima, Yue e Wang (2002) demonstraram, através de estudos de Simulação Monte Carlo, que:

- Quando a tendência existe em uma série temporal, o impacto da correlação, positiva ou negativa, é dependente do tamanho amostral, da magnitude da autocorrelação e tendência;
- Para séries de tamanho reduzido, ou seja, $n \leq 50$, a presença de correlação positiva faz com que a possibilidade de rejeição da hipótese nula seja maior, e um efeito oposto é observado para aquela em que se detecta a presença de correlação negativa;
- Para os casos em que o tamanho da amostra e a magnitude da tendência são significativos, a correlação não é um fator preponderante de influência no teste de MK. Assim, é preferível utilizá-lo na avaliação de X_t ao invés do X_t^{mod} . O intuito disso é evitar as possíveis distorções advindas aplicação do PW ao teste empregado na detecção das tendências;
- A remoção de uma correlação positiva, através do uso do PW, impacta na eliminação conjunta de parte da tendência, ocasionando uma redução em sua magnitude. Caso contrário, o valor obtido para o modulo da tendência será inflado.
- Se a série temporal não puder ser satisfatoriamente modelada através de um AR (1) ou mesmo inexista tendência, o procedimento de PW pode não ser suficiente para reduzir os efeitos da correlação temporal sobre o teste de MK.

3.4.3. Trend-Free-Pre-Whitening

Dadas as limitações constatadas no procedimento de *Pre-Whitening*, Yue *et al.* (2002) sugerem o uso do Trend-Free-Pre-Whitening, após a realização de uma série de estudos de simulação Monte Carlo. As conclusões resultantes do trabalho citado mostraram que remover a autocorrelação, conforme definido pelo procedimento do PW, implica na redução da magnitude da tendência presente na série modificada, ocasionado, por consequência, uma substancial diminuição no poder de detecção do teste.

Com o intuito de mitigar o problema relatado, Yue *et al.* (2002) sugerem a remoção da tendência estimada da série originalmente observada antes da aplicação dos procedimentos previstos pelo PW. De maneira geral, o TFPW prescinde da realização dos seguintes passos metodológicos:

- 1) Estimativa da magnitude da tendência da série amostral, \hat{b} , a partir da aplicação do método não-paramétrico proposto por Sen (1968);
- 2) Remoção da tendência da série original (X_t) e, conseqüente, obtenção de amostra modificada, a partir da aplicação da expressão: $X_t^{mod} = X_t - \hat{b}t$, em que t é o intervalo de tempo;
- 3) Estimativa da autocorrelação amostral (r_1) da série modificada, X_t^{mod} ;
- 4) Se r_1 não é estatisticamente diferente de zero, então o teste de MK é aplicado à série original. Caso contrário, o procedimento de PW é aplicado a $X_t^{mod} \rightarrow Y_t^{mod} = X_t^{mod} - r_1 X_{t-1}^{mod}$;
- 5) A tendência é novamente devolvida à série $Y_t = Y_t^{mod} + \hat{b}t$;
- 6) Aplicação do teste de MK à Y_t .

Para TFPW, após a remoção da correlação, a avaliação da significância da tendência eventualmente existente na amostra se processará conforme descrito na seção 3.3.2.

3.4.4. Perspectiva Histórica: A questão da Autocorrelação

Com o objetivo de eliminar a influência da correlação serial para o teste de MK, Kulkarni e Von Storch (1995) e Von Storch (1995) propuseram modificar a série original, definida como autoregressiva lag-1 (AR (1)), a partir da remoção da autocorrelação amostral estimada antes da aplicação do teste de detecção não-paramétrico supracitado. Denominando-o como Pre-Whitening (PW), Von Storch (1995) demonstrou ainda, por meio de experimentos Monte Carlo de geração de séries AR (1) para diferentes valores de correlação temporal (ρ), a forte relação existente entre os resultados de detecção e a magnitude de ρ e a capacidade do PW de reduzir os efeitos da autocorrelação para o teste de MK, dada a inexistência de tendências na série analisada. A influência da autocorrelação é manifesta pela incapacidade de manutenção das taxas de rejeição próxima aquelas definidas à priori, principalmente na

presença de tendências estatisticamente significantes, conforme demonstrado por Von Storch (1995) e constatado em estudos posteriores, que também procuraram estudar em como lidar com os efeitos da correlação serial na atividade de detecção (Douglas *et al.*, 2000, Yue *et al.*, 2002b, Bayazit e Önoz, 2007, Khaliq *et al.*, 2009, Hamed, 2009 e Onoz e Bayazit, 2012). Apenas para ilustrar o problema relatado, cita-se novamente os experimentos de Von Storch (1995), em que a taxa de rejeição falsa esperada era de 5% ($\alpha = 5\%$) e, dada a presença de correlação serial igual a 0.3, obteve-se 15%, chegando-se a valores superiores a 0.30 de rejeição quando $\rho = 0,60$.

Investigando a presença de tendências em séries temporais de vazão nos Estados Unidos, cheias e estiagem, bem como o efeito da correlação espacial em seus resultados de detecção regional, Significância de Campo, Douglas *et al.* (2000) demonstraram que a aplicação do PW pode ocasionar redução na taxa de detecção de tendências estatisticamente significantes para o teste de MK. Salienta-se que esse efeito foi constatado nas séries temporais de vazão do período de estiagem, que possuíam significativa persistência temporal, sendo que algumas regiões que haviam sido declaradas como não-estacionárias antes da aplicação do PW passaram a ser consideradas como estatisticamente desprovidas de tendência.

Visando investigar a interação observada por Douglas *et al.* (2000) entre tendências monotônicas, lineares, e o processo autoregressivo lag-1, quando ambos existem em uma série temporal, Yue *et al.* (2002b) realizaram estudos, via simulação Monte Carlo, de avaliação dos efeitos: (i) do processo AR (1), positivo ou negativo, sobre o erro tipo I, (ii) da tendência sobre a autocorrelação e, por fim, (iii) da correlação serial e não-estacionariedade combinados. Para as duas primeiras análises, (i) e (ii), os experimentos procedidos demonstraram, respectivamente, que a existência de autocorrelação afeta a variância da estatística de teste de Mann-Kendall estimada (S) e a presença de tendência estatisticamente significativa modifica a magnitude da correlação serial calculada. Já para o caso (iii), Yue *et al.* (2002b) propuseram que amostras do tipo AR (1) fossem geradas com diferentes graus de correlação temporal ($\rho_1 = 0.2$ (0.2) 0.8), magnitude de tendência ($b = 0.002$ (0.002) 0.008) e com tamanho amostral ($n = 100$), a partir da expressão:

$$X_t = \mu_X + \rho_1(X_{t-1} - \mu_X) + \varepsilon_t + bt \quad (3.13)$$

onde μ_x é a média de X_t e igual a 1.0, ε_t é o ruído branco com média $\mu_\varepsilon = 0$ e a variância $\sigma_\varepsilon^2 = \sigma_x^2(1 - p_1^2)$, na qual $\sigma_x^2 = 0.5$ é a variância de X_t .

Totalizando 1,000 séries geradas e estimando-se os coeficientes de correlação amostral lag-1 dos ranks (r_1) e a magnitude da tendência (b), conforme proposto por Sen (1968), para cada uma das séries sintéticas, Yue *et al.* (2002b) mostraram que, para um processo AR (1), cujo coeficiente de autocorrelação é positivo, após a aplicação do PW, parte da tendência presente na amostra é removida juntamente com a tentativa de se eliminar a correlação temporal. Isso possibilita, no caso de estudos de detecção, a aceitação de hipóteses nulas de estacionariedade, quando de fato, elas são falsas. Face a essa constatação, Yue *et al.* (2002b) propõem, como alternativa, a remoção da tendência da série temporal antes da aplicação do PW, objetivando evitar os efeitos da não-estacionariedade sobre a autocorrelação serial e na estimativa do coeficiente AR (1). Conhecida como TFPW, a abordagem alternativa ao PW relatada foi aplicada por esses autores para a avaliação de tendências em séries correlacionadas de vazões médias e mínima anuais de algumas bacias intocadas de Ontário, Canadá. Os resultados indicaram que, com os procedimentos existentes e anteriores ao proposto, os pesquisadores podem ter incorretamente identificado tendências estatisticamente significantes.

A limitação relatada por Yue *et al.* (2002b) para o PW levou Bayazit e Önoz (2007) a questionarem a validade ou não de se aplicar o referido procedimento ao se realizar análises de detecção de tendências (“To prewhiten or not to prewhiten in trend analysis?”). Assim, objetivando-se determinar quando o PW pode ser utilizado sem resultar em perda real de poder de detecção, Bayazit e Önoz (2007) aplicaram o teste de MK a cada conjunto de 20,000 séries temporais normalmente distribuídas, geradas por meio de um modelo AR(1) ($r = 0, 0.2, 0.4, 0.6, 0.8$), tamanho amostral ($n = 25, 50, 75, 100$), média igual 1, coeficiente de variação ($C_v = 0.1, 0.3, 0.5, 0.7, 0.9$) e tendências lineares ($b = 0.002, 0.004, 0.006, 0.008, 0.01, 0.015, 0.020, 0.025$). Como resultado da simulação Monte Carlo realizada, os autores, em linhas gerais, sinalizam para a impossibilidade de emprego indiscriminado do PW, devendo ser evitado quando o teste apresenta alto poder, isto é, quando o coeficiente de variação é muito baixo, a magnitude da tendência é alta e o tamanho da amostra é grande. Excetuando-se a situação anteriormente mencionada, Bayazit

e Önoz (2007) apontam que o Pre-Whitening é capaz de prevenir a falsa detecção de uma tendência inexistente e identificá-la, caso exista, sem ocasionar significativa redução no poder do teste.

No estudo realizado por Khaliq *et al.* (2009), as metodologias do PW, TFPW, Correção de Variância (VC) e Bootstrap em blocos (BBS) foram aplicadas às séries de vazão média diária anual de 301 estações de monitoramento de rios canadenses. Os resultados obtidos indicaram que o PW é conservador na identificação de estações com tendências significativas e concluiu-se que o emprego do BBS e VC é preferível em relação aos demais procedimentos testados. Baseando-se não só na impossibilidade de uso irrestrito do PW, já mencionada por Bayazit e Önoz (2007), ou na incapacidade do TFPW em preservar o nível de significância nominal, a indicação das abordagens do BBS e VC pelos autores foi baseada no desempenho superior demonstrado por ambas quando as amostras avaliadas possuíam dependência temporal de ordem superior à unidade, ou seja, não podiam ser expressas como um AR (1). Ademais, especificamente com relação ao BBS, Khaliq *et al.* (2009) apontam para a existência de diferenças negligenciáveis entre o tamanho do bloco considerado ótimo pela publicação em questão e aquele determinado pela relação $\kappa + 1$, sendo κ o número da correlação serial significativa presente na amostra.

Além da discussão sobre a viabilidade de aplicação ou não do Pre-Whitening relatada em Bayazit e Önoz (2007), Hamed (2009) debate a factibilidade da estimativa simultânea dos coeficientes angular da tendência linear (b) e de autocorrelação (r), seguida de uma correção de viés em r , como forma de aumento da efetividade do PW. Baseando-se no fato de que a existência de tendência amostral resultará em uma superestimação de r e conseqüente perda de poder do teste, o autor é favorável ao emprego do procedimento supracitado e recomenda ainda que o nível de significância do teste de MK seja elevado, caso a capacidade de detecção de tendências reais pelo Pre-Whitening for a maior das preocupações durante determinada análise procedida. Aspectos adicionais, como a importância da correção de viés proposta para amostras de tamanho reduzido e a escolha cuidadosa do modelo de autocorrelação para simular uma dada base de dados, também foram destacados pelo resultado do experimento Monte Carlo procedido por Hamed (2009).

Alternativamente aos procedimentos que pressupõem a modificação de séries temporais correlacionadas, no caso o PW e o TFPW, Onoz e Bayazit (2012) propuseram avaliar o desempenho do *Bootstrap* em blocos, associado ao teste de MK (BBS_MK), como ferramenta de prevenção ao acometimento de detecções falsas. A partir da geração de 10,000 amostras do tipo AR(1), média igual a unidade, coeficientes de variação ($C_v = 0.2$ (0.3) 0.8), tamanhos seriais ($n = 25, 50, 100$) e coeficientes de autocorrelação ($r = 0.1$ (0.1) 0.6), os autores, como um primeiro passo para a obtenção da distribuição da estatística do teste de MK (S) livre de tendência, determinaram para cada uma das amostras o valor de S . Na sequência, via *bootstrap* em blocos, Onoz e Bayazit (2012) propuseram a geração de mais 2,000 amostras, cujos valores de S foram estimados para então obter a distribuição empírica da estatística de teste de MK quando a hipótese nula é verdadeira, dado um valor de C_v , n e r .

Conforme mencionado, com o objetivo de investigar o poder do BBS_MK, Onoz e Bayazit (2012) adicionaram às séries sintéticas, até então estacionárias, tendências lineares positivas de magnitude $b = 0.005$ ou $b = 0.010$. Adotando-se um nível de significância igual a $\alpha = 0.05$ (5%), procedeu-se a realização de teste de verificação de estacionariedade unicaudal, em que a rejeição da hipótese nula se deu quando a estatística de teste de MK da série original fosse maior do que os 100 maiores valores de S das 2,000 séries geradas por meio de processos de reamostragem com reposição, *Bootstrap em blocos*. Em outras palavras, a rejeição de H_0 ocorreu quando o valor de S da série original estava compreendido entre os limites da zona de rejeição R .

De acordo com os resultados das simulações procedidas e de testes realizados utilizando séries de vazão não estacionárias de rios turcos, Onoz e Bayazit (2012) indicam que o BBS_MK apresenta poder similar ao constatado para o PW e TFPW. Adicionalmente, foi determinado o tamanho ótimo do bloco que minimiza a ocorrência de erros do Tipo I para esse teste. Os autores também relatam que a capacidade do BBS_MK em manter a taxa de rejeição da hipótese nula próxima ao nível de significância arbitrado está relacionada à consideração do tamanho amostral e coeficiente de autocorrelação lag-1 na escolha do comprimento do bloco.

Não obstante, no Brasil, estudos de comparação de desempenho de metodologias de remoção da influência da correlação serial no teste de Mann-Kendall também foram desenvolvidos. Dentre estes, destaca-se o realizado por Blain (2014), em que se procurou avaliar a capacidade do Pre-Whitening (PW), Trend-Free Pre-Whitening (TFPW) e o Trend-Free Pre-Whitening Modificado (MTFPW) quando os mesmos são aplicados a séries autocorrelacionadas na busca por tendências monotônicas. Por meio de Simulação Monte Carlo (MC) e consequente aplicação das metodologias supracitadas a amostras sintéticas geradas para a estimativa e posterior avaliação dos erros tipo I e II, o autor aponta para a superioridade de poder e maior ocorrência de falsas rejeições de hipótese nula do TFPW. Ademais, em sintonia com os resultados das demais publicações aqui relatadas, o PW claramente limitou a influência da correlação serial sobre o teste de MK, entretanto, com prejuízos para a capacidade de detecção. Sobre o MTFPW, Blain (2014) relata um melhor equilíbrio entre as probabilidades dos erros tipo I e II, fato esse compartilhado pelos resultados obtidos pela avaliação de estacionariedade procedida para os valores mensais de precipitação, mínimos e máximos de temperatura do ar de estações climáticas da cidade de Ribeirão Preto, no estado de São Paulo. Adicionalmente, esclarece-se que para a simulação MC realizada pelo autor foram geradas 10.000 amostras para cada uma das possíveis combinações entre os parâmetros variáveis, no caso n e r , por meio da seguinte expressão:

$$x_r = E(X) + r(x_{t-1} - E(X)) + \xi_t + T \quad (3.13)$$

onde $E(X)$, r e T são, respectivamente, a média (1.0), o coeficiente de autocorrelação lag-1 e a tendência; ξ_t é o ruído branco de média zero e variância igual a $\sqrt{[Var(X)(1 - r^2)]}$; $n = \{30,60,90\}$; $r = \{0.0,0.2, 0.4, 0.6\}$ e coeficiente de variação igual a 0.5

É nesse contexto de aparente inexistência de estudos que contemplam o problema da autocorrelação associada a eventos hidrológicos extremos que o presente trabalho está inserido. Assim, procurando contribuir para o preenchimento dessa lacuna de conhecimento, será avaliado o desempenho do PW e TFPW quando aplicados a amostras de variáveis hidrológicas extremas reais e, possivelmente, correlacionadas no tempo. Para tal, os testes supracitados serão utilizados na avaliação de estacionariedade de séries diárias anuais de

índices extremos de precipitação e vazão de estações localizadas na bacia do São Francisco, e terão os seus resultados comparativamente analisados.

3.5. A QUESTÃO DA MULTIPLICIDADE

3.5.1. Impacto da Multiplicidade nos Resultados de Detecção

Estudos de detecção de tendência quase sempre consistem na aplicação simultânea de testes de hipóteses a várias estações de monitoramento de uma dada variável hidrológica de uma dada região. O nível de significância, α_L , associado ao teste de hipótese local, é sempre especificado previamente, e reflete o risco que se está disposto a tolerar em relação a cometer um Erro do Tipo I, ou seja, declarar uma determinada estação como não-estacionária, quando, de fato, não existe nenhuma tendência associada a ela. Assim, pode-se dizer que se a hipótese nula é verdadeira, a probabilidade de se cometer o Erro Tipo I é igual a α_L .

A definição da região de rejeição R de um teste de hipótese local para um dado α_L é direta, mas, quando o objetivo é o controle do Erro Tipo I regional, o cenário torna-se mais complicado. Nem mesmo está claro qual o erro que deve ser controlado. Como opção de mecanismos estatísticos de controle, a literatura usualmente recomenda o *Per-Comparison Type I Error* (PCER), o *Family-Wise Type I Error* (FWER) ou o *False Discovery Rate* (FDR).

O procedimento baseado no PCER simplesmente ignora os problemas originados pela multiplicidade, já que a decisão de aceitação ou rejeição da hipótese nula é feita individualmente, ou seja, utilizando-se apenas a informação contida na estação que se encontra em análise. Por esse procedimento não considerar a perspectiva global das estações presentes em uma região, invariavelmente ocorrem violações do valor nominal, α_R , de significância regional. Por exemplo, suponha que um estudo de detecção de tendências utilizando o PCER e um nível de significância local, α_L , igual a 0,05, é realizado em uma região que contenha 50 estações. Se todas as hipóteses nulas assumidas são verdadeiras e as estatísticas dos testes são independentes, a probabilidade de se cometer pelo menos uma

falsa rejeição $P_r(nFD)$ segue uma distribuição binomial (Livezey e Chen, 1983) e é dada por:

$$P_r(nFD \geq 1) = 1 - (1 - \alpha_L)^m = 1 - (1 - 0.05)^{50} = 0.92 \quad (3.14)$$

Assim, nota-se que o procedimento baseado no PCER controla a falsa detecção em um nível local, compatível com o valor de α_L arbitrado, não sendo, portanto, uma abordagem adequada para lidar com problemas que envolvem a realização de múltiplos testes. Isso fez com que os testes de hipótese baseados no PCER sejam denominados como testes não corrigidos.

Uma concepção de controle de erros empregada com bastante frequência em questões que envolvem a aplicação de múltiplos testes é a Family-Wise Error Rate (FWER). A FWER é definida como a probabilidade de se ter ao menos uma falsa rejeição, quando realizados, simultaneamente, testes de hipóteses a m estações em uma região. O emprego do controle FWER é uma estratégia bastante atraente, mas, por outro lado, pode ser extremamente restritiva, já que, se a evidência amostral não é muito forte, a rejeição da hipótese nula não ocorre, o que resulta em baixo poder de detecção. Esse problema se torna ainda maior quando o número de estações a serem avaliadas cresce. O baixo poder torna-se um grande inconveniente, dado que a idéia central de se realizarem esses testes é identificar as estações que, de fato, possuem tendências. Salienta-se que o teste de Bonferroni é um exemplo de método que utiliza esse mecanismo específico de controle, o FWER.

Alternativamente, Benjamini e Hockberg (1995) sugeriram uma abordagem diferente para lidar com os problemas advindos da multiplicidade. Ao invés do FWER, eles propuseram que o controle fosse exercido através da *False Discovery Rate* (FDR), definida como a razão entre a quantidade de hipóteses rejeitadas erroneamente (falsos positivos) e a totalidade daquelas rejeitadas.

3.5.2. Evolução Histórica

A desigualdade de Bonferroni é usualmente empregada para a fixação de um limite superior global (conjunto) quando são realizados múltiplos testes de significância, e resulta em rejeição de uma hipótese nula específica (H_i) para cada $p_i \leq \alpha/n$ ($i = 1, \dots, n$) (Miller, 1981). Embora vários métodos multivariados tenham sido desenvolvidos para as situações de inferência estatística múltipla, o procedimento de Bonferroni (BF) é ainda bastante valioso por ser de simples utilização, não requerer a suposição de nenhum modelo distributivo e permitir verificar, individualmente, os resultados locais de cada um dos testes realizados, rejeição ou não. No entanto, o BF é conservador e carece de maior poder de detecção quando testes correlacionados são avaliados. Dada as limitações relatadas, adaptações à metodologia clássica foram, ao longo do tempo, desenvolvidas por Holms (1979), Simes (1986) e Hockberg (1988), com o intuito de manter as principais qualidades existentes em BF e lidar com as limitações inerentes ao procedimento.

Holms (1979), utilizando como base teórica a inequação de Boole aplicada à teoria de múltiplas inferências estatísticas, desenvolveu um procedimento de rejeição sequencial, no qual as hipóteses nulas são rejeitadas, uma a uma, até que não se possa mais realizar nenhuma rejeição. A partir do ordenamento crescente dos p -valores obtidos nos testes individuais, rejeitam-se aqueles para o qual $p(i) \leq \alpha / (n - i + 1)$. Permitindo-se a incorporação de pesos (c), o procedimento de Holms (PH) hierarquiza as hipóteses nulas de acordo com o seu grau de relevância, a partir da introdução de uma estatística S , $S = p(i)/c(i)$. De forma análoga à descrita para a metodologia não-ponderada, a rejeição da hipótese nula se dará quando $S(i) \leq \alpha / \sum c(i)$. Esse arranjo proposto, o PH, resultou, em quase todas as situações, em ganhos na probabilidade de rejeição de uma hipótese falsa, ou seja, aumento do poder do teste.

Simes (1986) propõem testar a hipótese nula global $H_0 = \bigcap_{i=1}^n H_i$ a um nível de significância α pré-definido, em termos dos p -valores ordenados das estatísticas do teste individuais, sugerindo a rejeição de H_0 , se $p(i)$ é igual ou inferior a $i \alpha/n$, para ao menos um i . Claramente mais poderoso do que o procedimento clássico (BF), dado que sua região de rejeição contém a de Bonferroni, o Método de Simes, via simulação, provou ser vantajoso

quando as estatísticas de teste são correlacionadas. Em relação à *FWER*, o autor mostra que o procedimento modificado mantém o controle a um nível α pré estabelecido apenas no caso de independência. Contudo, Simes (1986) conjecturou que o controle relatado permanece para situações em que as estatísticas são dependentes com algumas distribuições específicas: o valor absoluto da normal multivariada e um tipo de distribuição gama multivariada. Se a hipótese global H_0 é rejeitada, as afirmações referentes às hipóteses individuais testadas devem ser realizadas a partir do emprego da metodologia sugerida por Holm (1979).

Hockberg (1988), procurando estabelecer uma solução para o problema de realizar afirmações individuais a respeito das hipóteses testadas em Simes (1986), empregou o princípio estabelecido por Marcus, Peritz e Gabriel (1976) para estender a validade do procedimento relatado para o nível local. Assim, para a avaliação de significância de um dado subconjunto H' obtido do universo de hipóteses testadas H , de modo que $H' \subseteq H$ e $n' < n$, o procedimento de Simes estendido prevê que a hipótese global testada, H_0' , será avaliada em todos os subconjuntos formadores de H' (H''), e resultará em rejeição se tiverem o seu conjunto de p -valores ordenados (p_{i_j}) limitados ou iguais a $j \alpha/n''$. O autor mostra ainda que o emprego do procedimento estendido de Simes é mais poderoso do que o proposto por Holm (1979) quando da realização de análises localizadas da inferência estatística.

Com o passar do tempo, algumas dificuldades não puderam ser sanadas com o advento das metodologias propostas por Bonferroni e suas variações, tais como: (i) os procedimentos clássicos que controlam o *FWER*, de maneira forte, tendem a possuir substancialmente menor poder de detecção do que o *PCER* para um mesmo nível e, (ii) frequentemente o controle exercido pelo *FWER* não é necessário. Essas dificuldades levaram os procedimentos clássicos de múltipla comparação (MCP's) ao desuso e propiciaram o surgimento de uma nova abordagem para o problema da multiplicidade de testes. Corroborando ainda mais para a diminuição de uso relatada, Shaffer (1995), ao publicar os resultados obtidos em Seeger (1968), mostrou que a utilização do *FWER* não garante o controle de erros a uma taxa previamente arbitrada (α) quando algumas hipóteses nulas são verdadeiras e outras falsas ($m_0 < m$).

Muitos problemas de multiplicidade recaem sobre a necessidade de quantificação (número) de rejeições errôneas, e não apenas em determinar se qualquer erro foi cometido. Frente a esse fato e à constatação da relação inversa existente entre a gravidade de incorrer em um erro tipo I e o número de hipóteses rejeitadas, surgiu a necessidade de desenvolver uma nova taxa de controle de erros. Assim, a proporção esperada de erros dentre as hipóteses rejeitadas, denominada *false discovery rate* (FDR), foi proposta como mecanismo de mediação de falsos positivos. O termo FDR passou a ser utilizado somente após Soriç (1989), e foi popularizado por Benjamini e Hockberg (1995).

A publicação de Benjamini e Hockberg (1995) teve sua origem ligada a dois artigos, cuja preocupação central residia em realizar m testes de hipóteses, simultaneamente, desconhecendo-se a quantidade daquelas que eram efetivamente verdadeiras, m_0 . O primeiro deles, de autoria de Schweder e Spjøtvoll (1982), sugere que a estimativa do valor de m_0 seja realizada a partir da determinação do coeficiente angular de uma reta, visualmente ajustada, ao conjunto de pares ordenados formados pelos p_valores $(1-p)$ e o número de p_valores maiores do que p (N_p), sendo p o valor da função de probabilidade acumulada F_t para a estatística Z_t . Quando plotados, os p_valores associados a hipóteses nulas verdadeiras, assumidos como uniformemente distribuídos (U [0,1]) e regidos pela relação $E(N_p) = m_0(1 - p)$, encontram-se visivelmente alinhados, o que torna possível distingui-los daqueles rejeitados, $m - m_0$.

O segundo trabalho, realizado por Soriç (1989), opôs-se fortemente ao emprego de testes de hipóteses locais quando muitos deles eram simultaneamente realizados. A métrica de valor esperado de falsas descobertas, a saber, a quantidade de erros do tipo I cometidos (n_{FP}) dividida pelo número total de rejeições (n_{rejec}), $E(n_{FP})/E(n_{rejec})$, foi utilizada pelo autor para chamar atenção para a existência de falsos positivos dentre aquelas hipóteses declaradas como rejeitadas.

Mais tarde, supondo independência amostral, o conceito do *False Discovery Rate* (FDR) também foi empregado por Benjamini e Hockberg (1995) como métrica de controle de erros do tipo I. Criando-se uma metodologia *step-up*, denominada doravante BH95, os autores sugerem o ordenamento crescente dos p_valores ($p_{(i)}$) e rejeitam a hipótese nula de

estacionariedade (H_i), se a razão entre o número de ordem (i) e o total de estações avaliadas (m), multiplicado pelo nível de significância regional (q), for superior ao $p_{(i)}$ estimado.

Mesmo após a publicação de Benjamini e Hockberg (1995), a aceitação da idéia do FDR permaneceu lenta. Somente anos mais tarde, graças ao desenvolvimento da genética e, conseqüente, do crescimento do número de hipóteses a serem testadas, foi dado o impulso necessário para a difusão do uso de ferramentas que procuram conciliar o controle múltiplo e o poder de detecção. Nesse cenário, o FDR passou a largamente utilizado, inclusive a partir do ano 2000, fomentando o aparecimento de grupos de estatísticos interessados.

Com o intuito de aprimorar a técnica seminal do controle pelo FDR apresentada por Benjamini e Hockberg (1995), esforços estão sendo empreendidos na tentativa de: (i) desenvolver novos métodos de estimação do número de hipóteses nulas verdadeiras (m_0), (ii) lidar com a dependência presente nas estatísticas de teste (Yekutieli e Benjamini, 1999) e (iii) obter uma nova teoria e metodologias para o controle do *FDR*, com base na abordagem bayesiana.

Visando-se prover uma nova estimativa para o valor de m_0 , Benjamini e Hockberg (2001) sugerem a execução de procedimento de controle adaptativo simples, em que o número desconhecido das hipóteses nulas verdadeiras é estimado com base na técnica da menor declividade (LSL, do inglês *Lowest Slope*), em associação com a regra de parada: $S_i < S_{i-1}$. O valor estimado de m_0 será aquele menor e inteiro que atenda $\hat{m}_0 = \min[(1/S_j + 1), m]$. A partir daí esse valor é utilizado para a realização do procedimento proposto por Benjamini e Hockberg (1995), o que, portanto, restringe o seu uso às situações de independência.

Com o objetivo similar ao de Benjamini e Hockberg (2001), Storey (2002) sugeriu o uso de $\#(p - \text{valore} > c)/(1 - c)$. Ainda nesse sentido, ajustar diferentes modelos de mistura (do inglês, *mixture models*) à distribuição das estatísticas de teste, aos p-valores, ou seus z-valores transformados, e estimar a proporção da distribuição que pode ser atribuída aos valores sobre a hipótese nula verdadeira, são algumas das direções tomadas por Genovese e Wasserman (2002) e Efron (2008).

É importante se ter em mente que a evolução na estimativa do valor de m_0 não assegura o controle do *FDR* em nível desejado. Assim, ajustes no estimador também se fazem necessários, tais como em: Storey *et al.* (2004) e Benjamini *et al.* (2006).

A abordagem de Benjamini e Yekutieli (2001) foi fundamental para provar que o emprego do simples procedimento de Benjamini e Hockberg (1995) é de uso seguro para grande parte das situações que surgem na prática (dependência positiva). Esses mesmos autores propuseram uma alteração sutil no BH95, de modo que o procedimento passe exercer controle efetivo de erros para todas as formas de estrutura de dependência da estatística de teste. Outra forma de lidar com a dependência é através do emprego de técnicas de *bootstrap*, conforme realizado por Yekutieli e Benjamini (1999) e Storey e Tibshirani (2003).

Atualmente, muitos estudos e pesquisas são dedicados à incorporação de ideias bayesianas ao *FDR*, buscando-se o desenvolvimento de nova teoria e metodologias específicas. A abordagem empírica bayesiana da *false discovery rate*, extensivamente estudada por Efron (2008), se mostra promissora e constitui um campo de estudo para futuros trabalhos ligados ao controle de erros em problemas de multiplicidade de testes.

Conectado à tendência histórica de preponderância de emprego do *FDR*, verificada principalmente a partir dos anos 2000, o presente trabalho irá focar exclusivamente nessa taxa de controle de erros, buscando mostrar de maneira mais detalhada aspectos de sua definição formal, elucidar passos executivos e peculiaridades das técnicas mais frequentemente empregadas (BH95 e BH_Adap) e, por fim, avaliar o desempenho das mesmas.

3.5.3. O FDR Aplicado à Hidrometeorologia

Largamente utilizada na genética e epidemiologia, a abordagem de controle de erros do tipo I por meio do emprego do conceito do *FDR* é relativamente nova na hidrologia, sendo que apenas nos últimos anos começaram a surgir publicações que abordam esse aspecto, como os trabalhos desenvolvidos por Alpert *et al.* (2002), Ventura *et al.* (2004), Renard *et al.* (2008), Fatichi (2009), Khaliq *et al.* (2009), Miao *et al.* (2012), Sun *et al.* (2014), Ganguly

e Ganguli (2015), Gudmundsson e Seneviratne (2015), Cammalleri (2016), Merz *et al.* (2016), Humphrey *et al.* (2016) e Mallya *et al.* (2016).

Alpert *et al.* (2002) investigaram a veracidade e a extensão do comportamento aparentemente paradoxal entre o crescimento das precipitações diárias extremas e o decréscimo no total anual precipitado no Mediterrâneo reportado por estudos predecessores. Os autores dividiram as precipitações diárias observadas em 265 estações (1951-1995) localizadas em diferentes zonas do Mediterrâneo (Espanha (182), Itália (42), Chipre (3) e Israel (38) em seis grupos, que variam de fraca: 0-4 mm/d a torrencial: > 128 mm/d. Essa medida objetivava avaliar as contribuições das diferentes categorias para o total precipitado e como essas contribuições (percentuais) alteram-se com o tempo. Diante disso, Alpert *et al.*, (2002) sugeriram o emprego do BH95, como umas das metodologias, para detectar a presença de tendências nas referidas séries de percentuais.

Tentando avaliar o controle da proporção da rejeição falsa de hipóteses quando realizados múltiplos testes aplicados a dados climatológicos, Ventura *et al.* (2004) apresentaram, aplicaram e, por fim, confrontaram a abordagem do “*false discovery rate*” com as metodologias de análise de tendência temporais clássicas, avaliação da significância local ($\alpha = 5\%$) ou *PCER* e *Bonferroni-Corrigido* ($\alpha = n^{-1} \times 5\%$), por meio do uso de exemplo real: detecção de tendência na variância da temperatura média diária filtrada por banda de passagem em 500 hPa estudada por Paciorek *et al.* (2002).

Além disso, visando investigar a melhor maneira de se aplicar a abordagem do *FDR* à dados espacialmente correlacionados, comuns na climatologia, os autores realizaram o estudo comparativo de desempenho de três (3) metodologias distintas. Os procedimentos propostos por Benjamini e Hockberg (1995), Yekutieli e Benjamini (1999) e Benjamini e Yekutieli (2001) e nomeando-os, respectivamente, por *FDR-Indep*, *FDR-Corr* e *FDR-General*, foram aplicados por Ventura *et al.* (2004) a diferentes conjuntos amostrais sintéticos, que procuravam abranger as diversas magnitudes da correlação espacial que podem ser observadas em dados climatológicos: **Tipo A**, sem dependência espacial; **Tipo B**, dependência espacial consistente com aquela observada por Paciorek *et al.* (2002) para

vários índices de atividade de tempestades de inverno durante 1949 – 99 e **Tipo C**, forte dependência espacial.

Baseando-se nas séries temporais utilizadas no estudo realizado por Paciorek *et al.* (2002), Ventura *et al.* (2004) optaram pela remoção da porção não-estacionária de todas as $n = 3.024$ amostras disponíveis como primeiro passo para a geração dos conjuntos A, B e C. Relata-se que as tendências removidas tiveram as suas magnitudes estimadas com base em ajustes de regressão linear da variância da temperatura em função do tempo. Como resultado do procedimento descrito, foram obtidas séries temporais espacialmente correlacionadas e sem nenhum sinal ($Y_i(t), i = 1, \dots, n$, em que i e t são, respectivamente, a localização de uma dada estação considerada e o tempo).

As séries sintéticas ($Y_i(t)^*$), de base anual e tamanho amostral igual a 51 anos, visto que o período de registro disponível de dados adotados como referência abrangia os anos de 1949-1999 (Paciorek *et al.*, 2002), foram obtidas por procedimentos de reamostragem com reposição, *bootstrap*. Portanto, escolheu-se, ao acaso, um ano dentre todos os possíveis $t = \{1949, 1950, \dots, 1999\}$, sendo o primeiro elemento da série temporal sintética aquele valor da variável hidrológica associado ao ano sorteado (t^*). Como exemplo, suponha que os pares ordenados de precipitação máxima anual (mm) e tempo (ano) de um dado ponto de monitoramento são dados por: [1,000, 1950], [990,1951] e [500, 1952]. Ao se escolher aleatoriamente $t^*=1952$, o valor de $Y_i(1) = 500$. Ressalta-se que esse procedimento deve ser repetido até que o tamanho amostral sintético seja igual ao observado, em que um mesmo ano pode ser sorteado mais de uma vez. Atenta-se que o emprego de um mesmo valor de t^* para todas as estações preserva a estrutura da correlação espacial originalmente existente na amostra observada, sendo o contrário também verdadeiro. Posto isso, Ventura *et al.* (2004) utilizou-se dessa constatação para a geração das séries sintéticas do tipo B e A. Já para aquelas do tipo C, partiu-se das amostras obtidas para B, em que a inserção da forte correlação se deu a partir da substituição do $Y_i(t_j^*)$, para cada local i e ano j , pela média dos $Y_k(t_j^*)$ de todas as estações “ k ” distantes em 1913 km (0.3 radianos) do ponto i .

Criadas as séries temporais estacionárias e espacialmente independentes ou correlacionadas, Ventura *et al.* (2004) propuseram a adição de um tendência linear diferente de zero (b_i) às

estações que tiveram as suas séries A, B e C sinteticamente geradas, de modo que a proporção de hipóteses nulas positivas variasse entre 0 e 99% (0, 1, 10, 25, 50, 75, 90 e 99%) e fosse expressa por:

$$X_i(t_j) = Y_i(t_j^*) + b_i t_j \quad (3.15)$$

Ressalta-se que tanto para os dados correlacionados quanto para os independentes utilizou-se o mesmo sinal (b_i), diferindo-se apenas na estrutura do erro de $Y_i(t)^*$. Nesse caso, os autores optaram pela seleção de valores de b_i que resultassem na manutenção da razão sinal – ruído constante para todas as estações com a magnitude da tendência diferindo de zero, isto é, b_i foi escolhido de modo que a proporção da variabilidade serial (R^2) da regressão de $X_i(t)$, em t , explicada pela não-estacionariedade, fosse a mesma em cada localidade. Devido à diferença da magnitude dos ruídos aleatórios estação a estação, Ventura *et al.* (2004) optaram pelo emprego de um valor fixo $b_i = b$ para cada uma dos pontos de monitoramento. Finalizado o processo de geração de séries, estacionárias ou não, Ventura *et al.* (2004) aplicaram as metodologias do *FDR-Indep*, *FDR-Corr* e *FDR-General* aos diferentes tipos de dados sintéticos (A, B e C). Lembra-se que a proporção de hipóteses nulas positivas (0 - 99%) e a proporção da variabilidade da série associada à tendência ($R^2 = 10\%$, 15% e 20%) foram adotados como variáveis. Assim, portanto, a comparação das metodologias supracitadas foi realizada com base nos valores médios dos resultados obtidos para 10.000 conjuntos de dados simulados.

A respeito da simulação procedida, os autores concluíram que a metodologia do *FDR-Indep*, a mais simples das avaliadas, foi capaz de controlar a taxa de rejeição falsa de hipóteses nulas, mesmo tendo as suas premissas de aplicação violadas pela existência de correlação espacial entre as séries testadas, a saber: (i) os p_valores são mutuamente independentes quando a hipótese nula é verdadeira e (ii) a totalidade dos p_valores são independentes entre si. Quanto aos demais procedimentos analisados, *FDR-Corr* e *FDR-General*, os resultados obtidos por Ventura *et al.* (2004) mostram que o desempenho destes é ligeiramente superior em relação ao do *FDR-Indep*. No entanto, a complexidade e o elevado custo computacional envolvido na realização de ambas levaram à ponderação e à seleção do *FDR-Indep* como a de maior vantajosidade pelos autores. Ainda assim, com o intuito de melhoria no poder de

detecção, sugeriu-se uma simples modificação de procedimentos do *FDR-Indep* para tornar seu controle mais restritivo (ver na seção 3.5.4.1).

Embasado pelas análises dos resultados obtidos na simulação procedida, Ventura *et al.* (2004), conforme mencionado, propuseram a alteração da taxa de rejeição regional de q para $(1 - a)^{-1}q'$ do *FDR-Indep*, visto que a aplicação do procedimento supracitado resulta em limitação superior do *FDR* igual a $(1 - a)q$ (Genovese e Wasserman, 2004). Sendo a definido como a proporção desconhecida de hipóteses verdadeiras (n_{H_A}/n), a sugestão de alteração no valor da métrica de controle recai na necessidade de se assegurar que o *FDR* esteja o mais próximo possível do experimentalmente verificado, evitando-se o acometimento de equívocos de rejeição ou aceitação das hipóteses nulas.

O valor da estimativa de a (\hat{a}), que é desconhecido e necessário para a aplicação do *FDR-Indep* modificado, foi definido pelos autores com base nos estudos realizados por Storey (2002) e Genovese e Wasserman (2004) da seguinte maneira:

$$\hat{a} = I^{-1} \sum_{i=1}^I \max \left[0, \frac{\hat{F}_p(x_i) - x_i}{1 - x_i} \right] \quad (3.16)$$

$$x_i = x_0 + (1 - x_0) \frac{(i - 1)}{I}$$

onde x_i é subdividido em I espaços regulares entre x_0 e 1 (usualmente $x_0 = 0.8$ e $I = 20$), e a distribuição de probabilidade cumulativa dos p _valores utilizados na realização do teste é:

$$\hat{F}_p = n^{-1} \sum_{i=1}^n I_{[0,x]}(p_i) \quad (3.17)$$

sendo que $I_s(p_i) = 1$, se $p_i \in S$, e $I_s(p_i) = 0$, caso contrário.

Os resultados obtidos por Ventura *et al.* (2004), principalmente em seu estudo de caso real, que evidenciaram a discrepância existente entre as metodologias avaliadas (foram obtidas 941, 19 e 338 estações com tendências crescentes, respectivamente, para *PCER*, *Bonferroni-Corrigido* e *FDR-Indep*), demonstraram que procedimentos com base no controle exercido pelo *FDR* são aplicáveis à hidrologia e reforçaram a importância da consideração da multiplicidade à comunidade científica.

Renard *et al.* (2008) descrevem quatro métodos regionais para avaliar a significância de campo, sendo um deles o *false discovery rate*, e três técnicas de análise de consistência regional na detecção de tendência em extremos hidrológicos. Os autores ainda comparam, via simulação Monte Carlo, o desempenho dos seguintes procedimentos de avaliação de significância de campo: *número efetivo de estações*, *técnicas de bootstrap*, *cópula gaussiana* e o *false discovery rate* e os aplica a séries de dados diários de vazão de 195 estações fluviométricas na França. Quanto ao *FDR*, ponto de interesse deste trabalho, os autores apontam que quando a mudança regional afeta apenas um número limitado de estações, o *false discovery rate* é o mais poderoso dentre os procedimentos testados. Entretanto, à medida que o número de não estacionariedades cresce, os métodos baseados no bootstrap e cópula tem o seu desempenho equiparado ao *FDR*, tornando-se em alguns casos até mais poderosos para a detecção de mudanças regionais generalizadas. Renard *et al.* (2008) mostram ainda que a aplicação do *FDR* resulta em taxas de rejeição aceitáveis, embora note-se uma tendência em se tornar mais conservador para séries de dados dependentes. Por exemplo, quando a nível de significância nominal do teste ($\alpha' = 10\%$) e a correlação espacial (ρ) varia de 0 a 0.9, observa-se que as taxas de rejeição obtidas são, respectivamente, 9.4% e 5.6%.

Fatichi (2009) e Cammalleri (2016) reconhecem a necessidade de lidar com os efeitos da multiplicidade, portanto, ambos preveem o emprego do *FDR* como etapa do processo de avaliação de estacionariedade em seus estudos. Ao avaliar a presença de mudanças no regime de precipitação da região da Toscana (Itália), Fatichi (2009) propõem o uso de 40 índices de precipitação e posterior aplicação dos testes de MK, Cox-Stuart, Fisher e Daniels, todos que negligenciam o problema da multiplicidade (abordagem clássica). Na sequência, os autores recomendam os usos das técnicas do *false discovery rate* proposto por Benjamini

e Hockberg (1995) e por Ventura et. al. (2004), FDR-mod. Os resultados obtidos por Fatichi (2009), em geral, não mostram a existência de não estacionariedade para os padrões de pluviometria da região da Toscana. Entretanto, os autores chamam a atenção para alguns índices que tiveram estações apontadas como detentoras de tendências fracas ao se negligenciar a questão da multiplicidade e que assim não permaneceram ao se utilizar o BH95 e o FDR-mod. Como já mencionado, Camalleri et al (2016) também utilizam o BH95 ($q = 0.05$) para analisar a presença de tendências significantes no déficit de água presente no solo de toda a Europa, a partir de dados de grade meteorológica que possibilita a produção de mapas de 5km de resolução da umidade do solo na região da raiz (do inglês, *root zone soil moisture*).

Khaliq et. al. (2009) reforçam que as técnicas paramétricas e não-paramétricas empregadas na identificação de mudanças temporais em séries de variáveis hidrológicas são afetadas, quando aplicadas na escala local, pela presença de correlação temporal (autocorrelação). Já para uma dada região, a interpretação de resultados de múltiplos testes é sensível à existência da correlação espacial presente na rede de monitoramento empregada em uma dada análise. A baixa atenção dispensada a tais aspectos na grande maioria dos estudos de tendência realizados motivou os autores a apresentar uma revisão de algumas das metodologias de detecção, que podem ser utilizadas na presença/ausências de autocorrelação e correlação espacial, por exemplo, o FDR. Ao aplicar essas técnicas a séries de vazão média de alguns rios do Canadá, Khaliq et. al. (2009) concluíram que negligenciar a correlação espacial e temporal pode induzir a conclusões equivocadas, devendo-se, portanto, evitar estudos de tendências que não levam em consideração esses aspectos.

Miao *et al.* (2012), procurando identificar mudanças nas características da vazão do rio Amarelo (China), aplicaram os testes de Mann-Kendall (MK) e Pettitt a séries de 23 estações fluviométricas de tamanho serial igual a aproximadamente 50 anos. Os resultados obtidos pelo uso do MK ($\alpha = 0.05$) apontam para a existência de tendências monotônicas decrescentes entre os anos de 1958-2008 para 21 dos 23 pontos de monitoramento testados. Os autores apontam ainda para a influência da autocorrelação e da multiplicidade de testes sobre a atividade de detecção, o que os levou a aplicar o TFPW às séries antes do uso MK e o BH95 ($q = 0.05$) ao conjunto de estações avaliadas. Esclarece-se que especificamente para

este estudo, Miao *et al.* (2012) obtiveram os mesmos resultados ao se utilizar o MK e o BH95.

Sun *et al.* (2014) propuseram aplicar uma abordagem de controle de erros do tipo I pelo *FDR* na avaliação de tendências múltiplas no espaço a uma variável climática, no caso o nível médio diário de 8 horas de ozônio troposférico de 631 estações no Oeste dos Estados Unidos (1997-2005). O estudo, que buscou desenvolver uma estrutura teórica e computacional unificada para o controle de falsos positivos quando da realização de testes múltiplos, reforçou a adequabilidade de metodologias que empregam o *FDR* nas análises de estacionariedade de variáveis hidrológicas. Definiu-se também uma nova métrica de controle pelo *false discovery rate* aplicada a *clusters* (*false cluster rate*).

Podendo ser implementado por meio de algoritmos computacionais bayesianos, a estrutura proposta por Sun *et al.* (2014) teve êxito em exercer um controle mais acurado do erro e um desempenho superior quando confrontado com os métodos convencionais.

Já Ganguli e Ganguly (2015), objetivando entender e caracterizar secas meteorológicas constatadas em séries de dados observados em estações pluviométricas nos Estados Unidos, aplicaram a metodologia proposta por Benjamini e Hockberg (1995) aos registros de precipitação médios mensais (1218) e horários (1023) de séries que integram, respectivamente, a base do USHCN (United States Climatology Network, 1926-2013) e do NCDC (National Climate Data Center, 1950-2009) para a avaliar a significância de campo de variáveis que procuram representar diferentes aspectos da seca, tais como: severidade, duração e frequência.

Trabalho similar ao realizado por Ganguli e Ganguly (2015), só que para o continente europeu, foi desenvolvido por Gudmundsson e Seneviratne (2015) ao avaliar a presença ou não de mudança na frequência de secas em diferentes escalas de tempo (1 a 36 meses) por meio do SPI (do inglês, Standardized Precipitation Index). Os autores avaliaram a significância local após o emprego combinado do estimador de tendência Theil-Sen e o teste de Mann-Kendall ($\alpha=0.05$) e a significância de campo pelas técnicas que empregam o *FDR* como mecanismo de controle de erros do tipo I.

Merz *et. al.* (2016), diante da repetida ocorrência de inundações em um curto período de tempo na Alemanha tais como: rio Reno em 1993 e 1995, Elba e Danúbio em 2002 e 2003, propõem-se a avaliar a hipótese de que as cheias na Europa Central possuem períodos em que mesmas são mais frequentes (rico) e outros nem tanto (pobre). Essa conjectura foi testada a partir da realização de testes de significância em clusters temporais de ocorrências de cheias (pico acima de um determinado limiar) para 68 pontos de monitoramento localizados por toda a Alemanha (1932 a 2005). Os autores utilizaram-se de métodos baseado no índice de dispersão e na estimativa da taxa de ocorrência de kernel com teste significância para determinar se existem períodos ricos e pobres em cheias. Ainda nesse estudo, Merz *et. al.* (2016) utilizaram a metodologia baseada no *false discovery rate* definida por Benjamini e Hockberg (1995) para avaliar a significância de campo, que é declarada quando para pelo menos uma das estações testadas a hipótese nula é rejeitada.

As técnicas de detecção que utilizam o FDR foram ainda empregadas em Humphrey *et al.* (2016). Em linhas gerais, os autores propuseram-se a avaliar a presença de tendências, ciclo sazonais e anomalias sub-sazonais no armazenamento global de água. Após decompor em componentes (tendência linear, interanual, sazonal e sub-sazonal) os dados de grade provenientes do GRACE (do inglês, *Global Water Storage Variability*) da altura equivalente de água mensal, Humphrey *et. al.* (2016) aplicaram o teste de Mann-Kendall modificado à parcela denominada tendência linear, no caso da avaliação de significância local (uma célula da grade), e o BH95 ($q = 0.01$) devido à alta correlação espacial existente nos dados do GRACE, procurando-se evitar o crescimento do número de hipóteses nulas rejeitadas erroneamente (erro tipo I). Adicionalmente, o BH95 também foi aplicado aos dados sazonais.

Mallya *et. al.* (2016), ao investigarem as características da seca para a região indiana de monções, utilizaram-se de dados de grade diários de precipitação do IMD (do inglês, *India Meteorological Department*) e mensais da Universidade do Delaware (EUA) para determinar o SPI, SPEI (*Standardized precipitation-evapotranspiration index*), GMM-DI (*Gaussian mixture model-based drought index*) e o HMM-DI (*hidden Markov model-based drought index*) para o período que se estende de 1901-2004. Os autores propuseram que as tendências relacionadas a seca e sua variabilidade fossem analisadas para três períodos

distintos: 1901-1935, 1936-1971 e 1972-2004 por meio do emprego do teste de Mann-Kendall modificado, que leva em consideração a presença de autocorrelação amostral, e de técnicas que utilizam o FDR como mecanismo de controle de erros tipo I, a fim de lidar com eventual existência de correlação espacial na amostra. Os resultados obtidos por Mallya *et al.* (2016) apontam que a severidade e a frequência das secas durante as décadas mais recentes (1972-2004) têm se agravado, independentemente da base de dados e da metodologia de avaliação de estacionariedade empregada

O presente trabalho está inserido nesse contexto de crescente avanço da aplicação das metodologias de controle de erros tipo I pelo *FDR* à hidrologia e da preocupação com a questão da multiplicidade. Esse estudo levou em consideração as conclusões de seus predecessores e visa contribuir para o preenchimento das lacunas deixadas pelas pesquisas aqui descritas. Foram avaliadas as diferentes abordagens que empregam o *FDR* quando associado a variáveis de extremos hidrológicos, caso sintético e real, na tentativa de se definir padrões regionais de não-estacionariedade.

Nas seções subsequentes, o controle exercido pelo FDR será abordado em mais detalhes. Será também feita uma descrição pormenorizada das técnicas que o utilizam e que serão empregadas neste trabalho para a consecução dos objetivos já elencados, a saber, o *FDR-Indep* (BH95) e o BH_Adap.

3.5.4. O Controle pelo FDR

Uma abordagem alternativa ao controle de erros pelo *FWER* surgiu da necessidade, em vários problemas de multiplicidade, de se controlar o número de rejeições equivocadas realizadas e da idéia de que a gravidade de um falso positivo está inversamente relacionada à quantidade total de hipóteses rejeitadas. Buscando conciliar os anseios relatados em única métrica de controle, propôs-se que a proporção de erros do tipo I cometidos dentre a totalidade das hipóteses rejeitadas, $n_{FP}/n_{rejeitado}$, fosse utilizada.

Expressa por uma variável aleatória Q , a razão supracitada assumirá o valor nulo quando inexisterem rejeições, ou seja, $m = m_0$. Dado que valores do n_{FP} e n_{H_a} são desconhecidos,

mesmo após a realização de testes e análise dos dados, define-se o FDR como sendo o valor esperado de Q (Q_e).

$$Q_e = E(Q) = E\left[\frac{n_{FP}}{n_{FP} + n_{H_a}}\right] \quad (3.18)$$

Oriundas de sua definição formal, duas propriedades emergem como relevantes para a taxa de controle expressa na Equação (3.18)

- a) Se todas as hipóteses nulas são verdadeiras, o FDR é equivalente a $FWER$. Nesse caso, $n_{H_a} = 0$ e $n_{FP} = n_{rejeitado}$, então se $n_{FP} = 0$, diz-se que $FDR = 0$. Agora, se $n_{FP} > 0$, então $FDR = 1$, conduzindo a $P(n_{FP} \geq 1) = E\left(\frac{n_{FP}}{n_{rejeitado}}\right) = FDR$. Assim, portanto, o controle exercido pelo FDR assemelha-se àquele exercido pelo $FWER$

- b) Quando o $n_{FP} + n_{H_0} < m$, o valor de FDR é menor ou igual ao $FWER$. Nessa situação, se $n_{FP} > 0$, então $n_{FP}/n_{rejeitado} \leq 1$, conduzindo a $\chi_{(n_{FP} \geq 1)} \geq FDR$. Matematicamente, aplicando-se o valor esperado a ambos os lados da expressão anteriormente apresentada, obtém-se $P(n_{FP} \geq 1) \geq FDR$. Dada a demonstração realizada, é possível concluir que o emprego de qualquer procedimento que utilize o $FWER$ como controle está também submetido, indiretamente, às restrições impostas pelo FDR . Ademais, se esse for o único mecanismo de controle em uso, espera-se um ganho no poder de detecção à medida que o número de hipóteses ditas como não-verdadeiras aumenta, n_{H_a} .

Exercer controle sobre a variável Q é o objetivo desejado quando se realiza múltiplos testes de hipótese simultaneamente. Entretanto, exercer tal influência é difícil para algumas situações. O caso, em que $m = m_0$ e, portanto, $n_{FP}/n_{reje} = 1$, exemplifica as limitações de se controlar o valor de Q ao se utilizar a expressão (3.18) sem adaptações.

Situações como a demonstrada levaram à proposição de novas formulações para o FDR e, conseqüentemente, resultaram em novas abordagens de controle dessa taxa de erro, tal como $FDR_+ = E(n_{FP}/n_{reje})|n_{reje} > 0$, defendido em Storey (2002).

Assim, o presente trabalho, seguindo a ordem cronológica do desenvolvimento das técnicas relatadas, irá se concentrar em apresentar os diferentes procedimentos executivos desenvolvidos para lidar com o problema da multiplicidade de testes a partir do controle do *FDR*.

3.5.4.1. BH95

Com o intuito de evitar problemas no controle de Q , Benjamini e Hockberg (1995), ao optarem pelo uso do *False Discovery Rate* como taxa de controle de erros de sua metodologia, definiram o *FDR* da seguinte forma:

$$FDR = E[Q] = E[V/R | R > 0] P_r(R > 0) \quad (3.19)$$

onde V é o número de rejeições equivocadas da hipótese nula e R é o total de rejeições.

Por se tratar de um procedimento sequencial, as decisões de rejeição ou não da hipótese nula são realizadas uma a uma, estação por estação, começando a partir daquela que possui o maior p-valor e, conseqüentemente, menores evidências de que H_0 será rejeitada. Assim, considerando m hipóteses (nulas) testadas, $H_{(1)}, H_{(2)}, \dots, H_m$, e os respectivos p-valores associados ordenados, $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)}$, ocorrerá rejeição de todas as hipóteses nulas se $P_m \leq \alpha_R$. Caso contrário, estima-se um limite superior de rejeição (k), tal que seu valor é igual ao maior i para o qual a relação abaixo seja satisfeita

$$P_i \leq \frac{i}{m} \alpha_R \quad (3.20)$$

Esse procedimento, nomeado BH95 no presente trabalho, é capaz de manter o controle sobre FDR dentro de um nível previamente arbitrado (α_R), desde que as estatísticas de teste sejam independentes. O BH95 apresenta maior poder de detecção se comparado ao obtido na aplicação do procedimento de Hockberg (1988). Entretanto, à medida que o número de testes a serem avaliados aumenta, verifica-se um decréscimo em sua capacidade de detecção de hipóteses nulas verdadeiras (Benjamini e Hockberg, 1995).

É importante ressaltar que dentro da mesma abordagem sugerida, outros procedimentos podem ser desenvolvidos, a exemplo do método adaptativo que incorpora as ideias para a estimativa de m_0 de Hockberg e Benjamini (1990), que será abordado em seção subsequente.

3.5.4.2. BH_Adaptativo

Comprovadamente eficaz para estatísticas de teste independentes, o BH95 se torna bastante conservador quando as hipóteses a serem avaliadas são, de fato, falsas. Dada a limitação relatada, Benjamini e Hockberg (2000) propuseram a execução de um procedimento de controle adaptativo simples, em que o número desconhecido de hipóteses nulas (m_0) é estimado conforme proposto por Hockberg e Benjamini (1990). A partir daí esse valor é utilizado para a realização do BH95, o que, portanto, restringe o seu uso às situações de independência.

Motivado pela abordagem gráfica sugerida por Schweder e Spjøtvoll (1982), Hockberg e Benjamini (1990) se basearam na seguinte idéia para a definição da metodologia de estimativa de m_0 : Se todas as hipóteses nulas são verdadeiras, ou seja, $m = m_0$, e as estatísticas de testes são independentes, o conjunto de p-valores pode ser considerado como a realização de uma amostra ordenada sob uma distribuição uniforme, $U [0,1]$. Assim, o valor esperado para o i -ésimo p-valor é $E(P_{(i)}) = i/(m + 1)$ e o gráfico de $p(i)$ por i deverá exibir uma relação linear, cuja declividade da reta é igual a $S = 1/(m + 1)$ e passa pela origem e o ponto $(m + 1,1)$.

Relata-se que, em situações que $m_0 < m$, os p-valores correspondentes às falsas hipóteses nulas tendem a ser menores do que aqueles quando verdadeiros, concentrando-se, portanto, à esquerda do gráfico esboçado. À direita, a relação entre os pares ordenados permanece aproximadamente linear, com inclinação $b = 1/(m_0 + 1)$. Assim, utilizando-se um conjunto formado pelos maiores p-valores, ajusta-se uma linha reta através do ponto $(m + 1,1)$, com a declividade B , para obter o valor estimado de m_0 por $\hat{m}_0 = 1/\hat{B}$.

Sabendo como proceder para a realização da estimativa do valor de m_0 , uma pergunta parece sem resposta: quantos pontos, considerados como os de maior p_i , devem ser utilizados para o ajuste linear proposto? Como resposta, Hockberg e Benjamini (1990) propuseram que tal ajuste fosse realizado de maneira iterativa e progressiva, de modo que se busque em cada uma das tentativas, reduzir o número de p-valores empregado no processo. A atividade termina quando se nota o primeiro decréscimo no valor de B estimado, não sendo, assim, necessário realizar mais nenhuma iteração, ou seja, excluir mais um valor p_i . Após a etapa de exclusões, será definido como o número de pontos ideal, aquele que resultou em um menor valor de declividade da reta ajustada antes do decréscimo relatado.

Alternativamente, Benjamini e Hockberg (2006) propõem o emprego da técnica da menor declividade (S_i) (LSL, do inglês *Lowest Slope*), em associação com a regra de parada anteriormente descrita, para a detecção do ponto de mudança, ou seja, de término do comportamento linear entre $p_{(i)}$ e i . A base teórica dessa técnica é bastante similar à utilizada na determinação de m_0 em Hockberg e Benjamini (1990). Os autores sugerem o cálculo de S_i , começando-se em $i = 1$, em direção à valores cada vez maiores de i até que a condição de parada seja atingida, $S_i < S_{i-1}$ (Benjamini e Hockberg, 2006). O valor estimado de m_0 será aquele menor e inteiro que satisfaz as seguintes condições:

$$\hat{m}_0 = \min[(1/S_j + 1), m] \quad (3.21)$$

De maneira sistemática, o procedimento do BH_Adap pode ser resumido nos seguintes passos:

- 1) Ordenar, de modo crescente, os p-valores ($p_{(i)}$);
- 2) Comparar cada $p_{(i)}$ a $i \alpha_R/m$. Se, para todas as avaliações realizadas $p_{(i)} > i \alpha_R/m$, nenhuma hipótese será rejeitada e análise realizada se encerra;
- 3) Calcular as declividades $S_i = (1 - p_{(i)})/(m + 1 - i)$
- 4) Começando com $i = 1$, prosseguir no passo 3 até que $S_i \geq S_{i-1}$. Quando, pela primeira vez, $S_i < S_{i-1}$, parar e estimar $\hat{m}_0 = \min[1/S_j + 1, m]$;

- 5) Agora, iniciando-se do maior p-valor (p_m), comparar cada $p_{(i)}$ a $i \alpha_R / \hat{m}_0$ até atingir o primeiro p-valor que satisfaça $p_{(k)} \leq k \alpha_R / \hat{m}_0$. Rejeitam-se todas as k hipóteses cujos p-valores são menores do que $p_{(k)}$.

Quanto à capacidade de controle de erros pelo *FDR*, a abordagem adaptativa sugerida por Benjamini e Hockberg (2006) padece das mesmas limitações impostas ao BH95, afinal, o procedimento de rejeição utilizado é análogo ao adotada pelo seu antecessor, diferindo apenas na técnica empregada para a estimativa do número de hipóteses nulas verdadeiras (m_0).

Confirmando-se a supremacia do poder de detecção do teste dos métodos que utilizam o controle de erros pelo *FDR* sobre o *FWER*, Benjamini e Hockberg (2006) apontam, via simulação, que, à medida que a proporção de hipóteses nulas, de fato verdadeiras, decresce, aumenta-se a capacidade de detecção do BH-Adaptativo (BH_Adap). Ademais, comprovou-se que o poder do teste associado ao BH_Adap é maior do que aquele constatado para o BH95.

4. METODOLOGIA

Neste capítulo será apresentada a metodologia utilizada para alcançar os objetivos geral e específicos propostos para o presente trabalho. Primeiramente, na seção 4.1, denominada *Estratégia de Ação*, é apresentado sucintamente um panorama geral de todas as atividades propostas, elencando quais delas estão associadas a cada um dos três objetivos específicos definidos e, por meio de numeração, estabelece-se a cronologia de execução de todas as tarefas a serem realizadas até que o objetivo geral seja atendido.

Na sequência, os passos relatados na *Estratégia de Ação* são melhor detalhados e subdivididos em dois grupos: *Caso Sintético* e *Caso Real*. O tópico dedicado ao Caso Sintético está subdividido em quatro partes: (i) *A Distribuição Generalizada de Valores Extremos – GEV*, (ii) *O Processo de Geração e Estimativa do Poder do Teste*, (iii) *Estimativa do B_{SEN} , Erros Tipo S e M*, (iv) *Geração do Campo Sintético e Aplicação de Técnicas Baseadas no FDR*. Já o referente ao Caso Real se divide em três subseções: (i) *Seleção da Área de Estudo*, (ii) *Base de Dados* e (iii) *Seleção e Estimativa de Índices Hidrológicos Extremos*.

4.1. ESTRATÉGIA DE AÇÃO

A estratégia de ação foi concebida considerando-se nove passos metodológicos básicos e duas vertentes independentes de atividades: (i) *Caso Sintético* e a (ii) *Real*, que foram assim nomeadas por utilizarem como insumo, respectivamente, dados resultantes das simulações Monte Carlo (artificial) e séries de dados observados de realizações das variáveis hidrológicas de interesse (real). Adicionalmente, aponta-se que a similaridade entre as atividades a serem realizadas para (i) e (ii) ocorrem somente a partir do passo metodológico de número oito (8) e as tarefas associadas a cada um dos objetivos específicos foram agrupadas por meio de chaves na Figura 4.1.

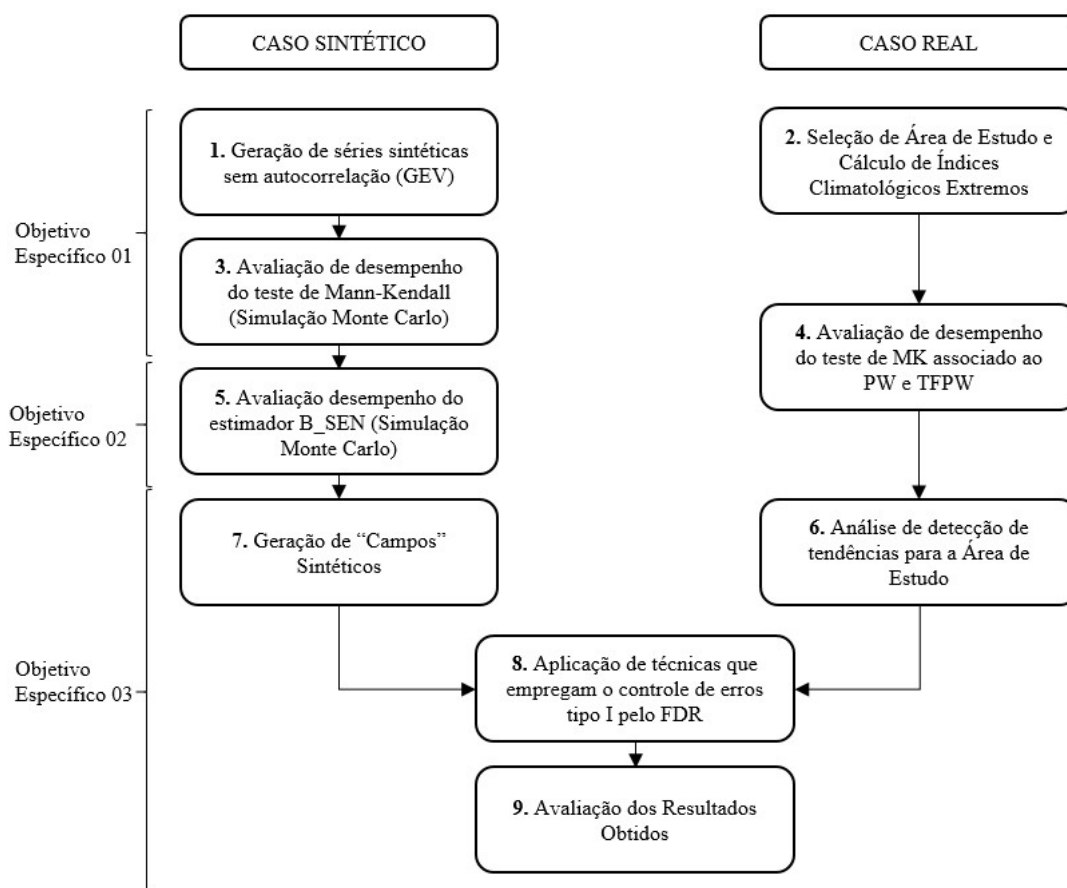


Figura 4.1: Fluxograma da Estratégia de Ação desenvolvida para a realização dos objetivo geral e específicos

Primeiramente, via simulação Monte Carlo, foram geradas séries sintéticas distribuídas segundo uma GEV (do inglês, *Generalized Extreme Values*), cujo parâmetro de posição é assumido como variável no tempo, enquanto os demais, escala e forma, não. No intuito de captar a característica extrema das séries temporais hidrometeorológicas reais dos máximos e mínimos, gerou-se um total de 10,000 séries para cada uma das possíveis combinações dos parâmetros n , C_v , κ e b (Passo 01). Esses parâmetros foram definidos dentro das faixas de variação dos valores usualmente constatados na hidrologia.

No Passo 03, foram realizados testes de estacionariedade para todas as séries temporais geradas no Passo 01. Buscou-se identificar padrões comportamentais que permitissem determinar de que maneira fatores possivelmente intervenientes na tarefa de detecção – tamanho serial, magnitude da tendência, coeficiente de variação e assimetria associados a

uma distribuição do tipo GEV – se relacionam com o Poder do Teste de Mann-Kendall (MK) e a sua capacidade em manter o percentual de erros do tipo I dentro do valor nominal de significância. Assim, as tarefas referentes aos Passos 01 e 03 integralizam o atendimento ao objetivo específico 01.

A avaliação de desempenho do estimador de tendência proposto por Sen (1968), doravante denominado B_SEN, foi feita utilizando as séries geradas no Passo 01. O B_SEN foi estimado para amostras de diferentes tamanhos (n), coeficientes de variação (C_v), parâmetros de forma (κ) e magnitude da tendência (b), fazendo uso dos erros tipo S e M, definidos por Gelman e Carlin (2014) como métricas de avaliação. Conhecendo a priori o valor de b , foi possível avaliar a proporção de eventuais inconsistências de sinais (erro S) e desvios de magnitude (erro M) previstos por B_SEN. Essas atividades constituem o Passo 05 e possibilitaram alcançar o objetivo específico 02.

Em paralelo, o *Pre-Whitening* (PW) e o *Trend-Free-Pre-Whitening* (TFPW), ambas metodologias associadas ao teste de Mann-Kendall, foram aplicadas a séries de índices climáticos extremos anuais de vazão e precipitação, calculadas com base nos registros históricos de estações localizadas na bacia hidrográfica do rio São Francisco (Passos 02 e 04). De posse dos resultados de significância estatística obtidos a partir desse procedimento, foi avaliada a existência de padrões regionais, núcleos ou regiões propensas a decréscimos ou acréscimos em algum dos índices climáticos calculados, procurando, se possível, elaborar hipóteses sobre como as não-estacionariedade detectadas se correlacionam a alterações do espaço e na dinâmica da precipitação e vazão (Passo 06).

O campo sintético, constituído pelo agrupamento de 179 estações fictícias (número equivalente ao total de postos pluviométricos utilizados neste trabalho e localizadas na bacia do São Francisco), as quais estão associadas a séries do tipo GEV ($\xi \propto \kappa$), foi gerado 2,000 vezes para cada um dos arranjos possíveis entre os valores de n , C_v , b e κ . Esclarece-se que as séries formadoras de um campo qualquer são independentes e foram geradas com base no mesmo conjunto de parâmetros assumidos como variáveis durante o processo de Simulação MC (n , C_v , b e κ). Em outras palavras, as 179 amostras que integram um campo sintético possuem o mesmo tamanho amostral, e o coeficiente de variação, magnitude de tendência e

parâmetro forma são assumidos como idênticos durante a etapa de geração serial. Salienta-se que, assim como as séries dos índices climatológicos determinados para as estações da bacia do rio São Francisco, os campos consistem em insumo para etapas subsequentes da estratégia de ação deste trabalho (Passo 07).

A aplicação das metodologias que realizam o controle de erros tipo I pelo *FDR*, *BH95* e *BH_Adap*, foi realizada para diversas configurações de campo sintético (n , C_v , b e κ), sendo o seu desempenho avaliado por meio de métricas definidas em conformidade com Ventura *et. al.* (2004), a saber: N_{TP} , *FDR* e *FNDR*. Por se tratar de técnicas ainda pouco empregadas no campo da hidrologia, adotou-se como referência comparativa de desempenho o procedimento dito tradicional (clássico), cujo controle está baseado na taxa de falsos positivos (*FPR*).

O passo metodológico 8, por ser um elo de ligação entre ambas vertentes propostas nesse trabalho, foi aplicado tanto aos campos sintéticos quanto às séries dos índices climatológicos extremos de pluviometria e fluviometria obtidos no Passo 02.

Como o objetivo específico 02 abrange os passos 02, 04, 06 e 07, referentes aos casos real e sintético, além do passo 08, comum a ambas as vertentes, a avaliação dos resultados foi realizada depois da finalização de todas as fases pretéritas e registrada em seção específica: Resultados e Discussões. Os achados mais significativos foram expressos no capítulo de Conclusões (Passo 09).

Salienta-se que as seções subsequentes referentes à *Metodologia*, assim como proposto na estratégia de ação, estão subdivididas em *Caso Sintético* e *Real*. Abordagem semelhante à relatada também foi adotada para o item de *Resultados e Discussões*.

4.2. CASO SINTÉTICO

A presente seção está dividida em 3 subseções, assim nomeadas: (i) *A Distribuição Generalizada de Valores Extremos – GEV*, (ii) *O Processo de Geração: Estimativa do Poder do Teste* e (iii) *Aplicação das Metodologias de Controle pelo FDR*. Nesta, procura-se

elucidar e pormenorizar aspectos metodológicos considerados relevantes e particulares de cada uma das atividades apontadas pela Figura 4.1, considerando-se apenas aquelas relacionadas ao *Caso Sintético*, ou seja, que possuem como insumo básico para a sua realização dados sintéticos de características extremas. Esclarece-se que as atividades de caráter avaliativo, como as de número 3, 5 e 9, serão contempladas apenas na seção de *Resultados*, por se tratarem exclusivamente de análise de resultados obtidos após o término de tarefas previstas no fluxograma supracitado.

A subseção 4.2.1 versa sobre as principais características da distribuição Generalizada de Valores Extremos e aponta para algumas peculiaridades essenciais para o sucesso da geração de séries sintéticas do tipo GEV, tal como o uso da média igual a unidade ($\mu = 1.0$). Essa subseção se refere ao passo metodológico 01 previsto no fluxograma da Figura 4.1, que somente se integraliza após a realização do previsto em 4.2.2. Como o próprio título antecipa, o item 4.2.2. contém a metodologia utilizada para a geração de séries não correlacionadas espacialmente, sem e com tendência linear, apresentando os valores de tamanho amostral (n), coeficiente de variação (C_v), magnitude de tendência (b) e parâmetro de forma da GEV (κ) utilizados na simulação Monte Carlo (MC) proposta. Relata-se ainda que esta unidade apresenta a descrição da métrica de estimativa da taxa de rejeição empregada, após a aplicação do teste de Mann-Kendall a cada grupo de 10,000 séries sintéticas geradas.

Já a subseção 4.2.3 apresenta todos os aspectos metodológicos envolvidos no desenvolvimento do denominado “campo sintético” (Passo 07), bem como a relação das técnicas que utilizam o “*false discovery rate - FDR*” empregadas às regiões fictícias criadas, e as métricas de avaliação de desempenho adotadas em 08.

Acrescenta-se que, devido à grande quantidade de dados envolvida nos passos metodológicos previstos para o *Caso Sintético*, principalmente relacionados às Simulações Monte Carlo realizadas, foi necessário o desenvolvimento de rotinas computacionais específicas, por meio do MatLab®, capazes não só de gerar as séries sintéticas, mas também de aplicar os testes Mann-Kendall, BH95 e BH_Adap, por exemplo, à base amostral.

4.2.1. A Distribuição Generalizada de Valores Extremos - GEV

A distribuição Generalizada de Valores Extremos engloba três distribuições frequentemente utilizadas na descrição probabilística de variáveis meteorológicas e hidrológicas máximas e mínimas, denominadas EV1 (Gumbel), EV2 (Fréchet) e EV3 (Weibull), cuja função de distribuição acumulada é dada por:

$$F(x) = \begin{cases} \exp\left\{-\left[1 - \kappa \frac{(x - \xi)}{\alpha}\right]^{1/\kappa}\right\}, & \kappa \neq 0 \\ \exp\left\{-\exp\left[-\frac{(x - \xi)}{\alpha}\right]\right\}, & \kappa = 0 \end{cases} \quad (4.1)$$

onde ξ , α e κ são os parâmetros de posição, escala e forma, respectivamente. Se $\kappa < 0$ (EV2), a distribuição possui limitação inferior ($x \geq \xi + \alpha/\kappa$), mas, se $\kappa > 0$ (EV3), nota-se a existência de restrição superior $x \leq \xi + \alpha/\kappa$. Para $\kappa = 0$, a GEV converge, no limite, para uma distribuição de Gumbel ilimitada, ou seja, $-\infty \leq x \leq \infty$.

Ao contrário de outras distribuições empregadas na avaliação de extremos, tal como a LP3, a GEV possui expressão analítica para o cálculo de quantis:

$$x_p = \begin{cases} \xi + \frac{\alpha}{\kappa} \{1 - [-\ln p]^\kappa\}, & \kappa \neq 0 \\ \xi - \alpha \ln[-\ln p], & \kappa = 0 \end{cases} \quad (4.2)$$

onde p é a probabilidade acumulada e os parâmetros da GEV, ξ , α e κ , podem ser empregados no cômputo dos valores da média (μ), desvio padrão (σ) e coeficiente de assimetria (γ) populacionais, a saber:

$$\begin{aligned} \mu &= \xi + \frac{\alpha}{\kappa} [1 - \Gamma(1 + \kappa)] \\ \sigma &= \frac{\alpha}{|\kappa|} [\Gamma(1 + 2\kappa + \Gamma^2(1 + \kappa))]^{1/2} \\ \gamma &= \text{sinal}(\kappa) \frac{-\Gamma(1 + 3\kappa) + 3\Gamma(1 + \kappa)\Gamma(1 + 2\kappa) - 2\Gamma^3(1 + \kappa)}{[\Gamma(1 + 2\kappa) - \Gamma^2(1 + \kappa)]^{3/2}} \end{aligned} \quad (4.3)$$

Diferentemente das equações apresentadas para o cálculo de μ e σ , que são dependentes de mais de um parâmetro da distribuição GEV, inclusive sendo restritas à valores de $\kappa \geq -1/3$, a expressão utilizada para a estimativa do coeficiente de assimetria possui relação exclusiva com o parâmetro de forma κ . Essa dependência única favorece o processo de geração de amostras do tipo GEV a partir de um determinado valor fixado de assimetria populacional.

Contudo, o coeficiente de variação ($Cv = \sigma/\mu$), que é uma estatística populacional que também deve ser controlada durante a obtenção aleatória de séries temporais, depende dos três parâmetros distributivos da GEV, ξ , α e κ . Isso pode ser mostrado a partir da manipulação das equações acima apresentadas, sendo inclusive possível reescrevê-las de modo a correlacioná-lo a dois fatores, o parâmetros de forma (k) e uma variável adimensional ($\phi = \xi/\alpha$):

$$Cv = \frac{\frac{1}{|\kappa|} [\Gamma(1 + 2\kappa + \Gamma^2(1 + \kappa))]^{1/2}}{\xi/\alpha + \frac{1}{\kappa} [1 - \Gamma(1 + \kappa)]} \quad (4.4)$$

Então, para gerar amostras aleatórias baseadas em uma distribuição GEV de coeficientes de assimetria (γ_0) e variação (Cv_0), primeiramente determina-se k_0 , com base na Eq. 4.3, e $\phi_0 = \xi/\alpha$ utilizando-se a Eq. 4.4. Com os valores de k_0 e ϕ_0 definidos, pode-se, por exemplo, escolher $\alpha_0 = 1$ e $\xi_0 = \phi_0$. Salienta-se que essa última parte é arbitrária, pois qualquer escolha de ξ e α que resultem em $\xi/\alpha = \phi_0$ proverá um conjunto de amostras distribuídas segundo uma GEV oriundas de população de $Cv = Cv_0$. Entretanto, para a realização de Simulação Monte Carlo, escolhe-se valores de ξ_0 e α_0 que resultem em $\mu = 1$. Informa-se que o uso do μ unitário está relacionada à geração de séries temporais não-estacionárias, mais especificamente à magnitude das tendências adotadas neste trabalho.

Logo, por conveniência e sem perder a generalidade, ξ_0 e α_0 devem ser calculados de modo a ter $\mu = 1$:

$$\alpha_0 = \frac{\kappa_0}{\kappa_0 \phi_0 + [1 - \Gamma(\Gamma 1 + \kappa_0)]} \quad (4.5)$$

$$\xi_0 = \frac{\kappa_0 \phi_0}{\kappa_0 \phi_0 + [1 - \Gamma(\Gamma 1 + \kappa_0)]} \quad (4.6)$$

4.2.2. O Processo de Geração e Estimativa do Poder do Teste

O objetivo dessa seção é relatar os procedimentos utilizados para a geração de diferentes amostras aleatórias do tipo GEV e apresentar a métrica empregada na avaliação da capacidade de controle de erros tipo I e poder do teste de MK, quando o mesmo é aplicado na detecção de tendências monotônicas (lineares) em séries extremas, geradas para diferentes combinações de n , C_v , b e κ , fatores possivelmente intervenientes na acurácia da análise proposta.

Assim, neste trabalho, optou-se pela realização de experimentos Monte Carlo (MC), sendo as séries geradas capazes de retratar o comportamento e a variabilidade natural observada em variáveis hidrológicas extremas. Para tal, a seleção dos valores assumidos para parâmetros ditos variáveis durante o processo MC, n , C_v , b e κ , foi pautada naqueles usualmente utilizados/vistos na hidrologia. Diante disso, escolheu-se os tamanhos amostrais, coeficientes de variação, parâmetros de forma e magnitudes de tendência iguais a, respectivamente, ($n = 20$ (20) 100), ($C_v = 0.2$ (0.2) 1.0), ($\kappa = -0.3, -0.1, 0, 0.1, 0.3$) e ($b = -0.01$ (0.002) 0.01). A seleção do parâmetro de forma da GEV baseou-se no estudo desenvolvido por Martins e Stedinger (2000), em que se define o intervalo compreendido entre -0.30 e +0.30 como razoável para a definição dos valores de κ em aplicações hidrológicas.

Em posse dos valores previamente arbitrados (n , C_v , κ e b), gerar amostras estacionárias independentes e distribuídas conforme uma GEV é relativamente simples, dada a existência

de expressão analítica para a determinação dos quantis. Isso permite que seja empregado o método da Transformação Inversa, que possui dois passos metodológicos básicos:

1. Gerar p utilizando-se uma distribuição uniforme entre 0 e 1 ($p \sim U(0,1)$);
2. Calcular x_p com base na Equação 4.2

Mais especificamente, para gerar uma série do tipo GEV com n observações de uma determinada população com coeficientes de assimetria (γ_0) e variação (Cv_0), são necessários seguir as seguintes etapas metodológicas:

1. Calcular k_0 utilizando-se da última equação em 4.3 para $\gamma = \gamma_0$;
2. Determinar ϕ_0 utilizando-se da Equação 4.4 com $Cv = Cv_0$ e $\kappa = \kappa_0$;
3. Calcular os valores de α_0 e ξ_0 – por conveniência, o cálculo foi realizado utilizando as Equações 4.5 e 4.6;
4. Gerar 10,000 amostras do tipo GEV seguindo os dois passos executivos básicos do método da Transformação Inversa; para tal, emprega-se ξ_0 , α_0 e κ_0 durante a estimativa de x_p , por meio da Equação 4.2.

No caso da geração de amostras do tipo GEV com a presença de tendências lineares, um termo não-estacionário é adicionado à série independente obtida por meio dos procedimentos anteriormente descritos. Supondo-se X_t como sendo o conjunto das observações independentes e distribuídas segundo uma GEV de parâmetros populacionais ξ_0 , α_0 e κ_0 e média igual a unidade ($\mu = 1$), a série monotônica (Y_t) pode ser expressa por:

$$Y_t = X_t + b \times t \quad (4.7)$$

onde b é a magnitude da tendência ou a taxa da mudança em relação à média por unidade de tempo. Por exemplo, se $b = 0.01$ e a série é de base anual, então tem-se 1% e 10%, respectivamente, para o ano e a década de taxa de variação. Esclarece-se que caso fossem utilizados valores de μ diferentes da unidade, por exemplo $\mu = 10$, o termo não-estacionário de (4.7), $b \times t$, representaria um percentual de mudança médio pequeno, o que poderia levar a não-percepção de tendência quando de fato a mesma está presente na série.

Salienta-se que essa configuração permite a extrapolação dos resultados a serem obtidos a partir da realização da simulação Monte Carlo proposta, visto que, para um dado valor de b , que representa a taxa de mudança relativa da média por unidade de tempo, e $\theta_0 = (\xi_0, \alpha_0, \kappa_0)$, associado aos valores populacionais γ_0 e Cv_0 da série, conclusões sobre o desempenho dos métodos de detecção podem ser atribuídas a quaisquer outras séries que possuam as mesmas características γ , Cv e percentual de mudança em relação à média.

Quanto à avaliação de desempenho do teste de Mann-Kendall, esclarece-se que esta será pautada na estimativa de seu poder. Contando com 10,000 séries sintéticas de extremos de tamanho n (GEV) para cada um dos possíveis arranjos entre os valores de Cv , κ e b , a taxa de rejeição da hipótese nula pode ser estimada a partir da seguinte equação:

$$R_{rej} = N_{rej}/N \quad (4.8)$$

onde N é o número total de experimentos simulados (10,000) e N_{rej} é a quantidade de hipóteses, cujo resultado localiza-se na região crítica R , determinada para um nível de significância igual a 5% ($\alpha_L = 0.05$).

A escolha para o número de total de séries simuladas foi realizada com base em estudos Monte Carlo similares ao proposto neste trabalho e mostrado em 3.3.4. Ainda assim foi avaliada a relação custo computacional e desvio entre os valores de R_{rej} obtidos para 2,000, 5,000, 10,000 e 20,000 séries geradas para um dado conjunto de n , Cv , κ e b . Como resultado, observou-se que o tempo de processamento até 10,000 séries pouco variava, assim como os resultados da taxa de rejeição da hipótese nula. Não se pode dizer o mesmo para 20,000 amostras, em que apesar da proximidade de resultados, o dispêndio de tempo envolvido no processo de geração foi aproximadamente o dobro do obtido para as demais situações testadas.

Todas as rotinas computacionais utilizadas nesta etapa de geração de séries via Simulação Monte Carlo e estimativa do poder do teste foram desenvolvidas com auxílio do *software* MatLab®.

4.2.3. Estimativa do B_SEN, Erros Tipo S e M

Ao recomendarem que a avaliação de estacionariedade seja complementada pela análise dos erros do *Tipo S* (sinal) e *Tipo M* (Magnitude), ao invés do uso exclusivo da abordagem convencional de poder (erros tipo I e tipo II), Gelman e Carlin (2014) apontam para a possibilidade de cometer equívocos de sinal e magnitude da mudança durante a estimativa, principalmente para situações em que as amostras estudadas apresentam elevado ruído e tamanho reduzido.

Visto que não foram encontrados trabalhos no campo da hidrologia que relatam o uso e o impacto de tais erros, *S* e *M*, durante a tarefa de detecção, o presente trabalho procurará avaliar os aspectos intervenientes à capacidade de estimativa de sentido, positivo e negativo, bem como do valor da tendência calculada pelo estimador proposto por Sen (1968), B_{Sen} . Para tanto, foram utilizadas séries distribuídas conforme GEV, geradas por meio de procedimento metodológico descrito em 4.2.2, e o cálculo de B_{SEN} foi feito com base nas equações da seção 3.3.3. Esclarece-se que os parâmetros de forma (κ), coeficientes de variação (C_v), média, tamanhos amostrais (n) e magnitudes de tendência foram análogos àqueles relatados naquela mesma seção.

Assim, no intuito de evidenciar a relação de interdependência entre os fatores assumidos como intervenientes à acurácia de B_{SEN} : n , C_v , κ e b , foram utilizadas as métricas relatadas por Gelman e Carlin (2014), definidas por tais autores da seguinte forma:

- (a) *Erro Tipo S* - representa a probabilidade de a estimativa do sinal de mudança apresentar sinal incorreto, dado que o mesmo é estatisticamente significativo diferente de zero, conforme a equação abaixo:

$$Erro\ Tipo\ S\ | \ Significativo = N_{Erro\ tipo\ S} / N_{Significativas} \quad (4.9)$$

onde $N_{Erro\ tipo\ S}$ é o número de vezes em que o sinal da tendência estimado por B_{SEN} é contrário ao assumido durante o processos de geração serial (b) e $N_{Significativas}$ é o

número de séries declaradas como não estacionárias após a aplicação do teste de Mann-Kendall para um dado conjunto de parâmetros n , b , C_v e κ , sendo $b \neq 0$;

(b) *Erro Tipo M* - relacionado à magnitude da tendência estimada; representa a expectativa do valor absoluto da estimativa dividido pelo tamanho efetivo, se estatisticamente significante diferente de zero, conforme a equação a seguir:

$$L_M = \frac{IC_{97,5\%} - IC_{2,5\%}}{b} \quad (4.10)$$

onde L_M é a métrica indireta de avaliação do erro tipo M, denominada Largura Relativa do Erro Tipo M, $IC_{97,5\%}$ e $IC_{2,5\%}$ são os valores de valor da magnitude de mudança estimados por B_SEN associados aos percentis de, respectivamente, 97.5 e 2.5% para um dado n , C_v , κ e b , sendo também $b \neq 0$.

Enfatiza-se que a probabilidade do Erro tipo S e M foram determinada considerando apenas a parte das séries que foram apontadas como não estacionárias pelo teste de Mann-Kendall, ou seja, utilizou-se apenas aquelas, dentre as 10,000 amostras simuladas via MC, que foram declaradas como estatisticamente significantes ($\alpha = 0.05$). A ideia por trás dessa análise de significância a priori é reproduzir o sequenciamento de procedimentos usualmente adotados para a tarefa de detecção: primeiro, aplicação de teste de estacionariedade, seguido de avaliação da magnitude e sinal da mudança, e restringir o total de séries a aquele em que uma definição equivocada no sentido e/ou grau da tendência seja responsável por erros de tomada de decisão frente a um clima em mudança.

4.2.4. Geração de Campo Sintético e Aplicação de Técnicas Baseadas no FDR

Neste trabalho, campo sintético é o nome dado ao conjunto formado por m séries tipo $GEV(\xi, \alpha, \kappa)$ de tamanho n e obtidas, via Simulação Monte Carlo, para um certo coeficiente de variação (C_v), média unitária ($\mu=1.0$), parâmetro de forma (κ) e magnitude de tendência (b). Salienta-se que o procedimento utilizado para a obtenção do campo é análogo ao descrito em 4.2.2 e que, para cada uma das possíveis combinações de $n = [40 \ 60 \ 80 \ 100]$, $C_v =$

0.2 (0.2)1.0, $\kappa = [-0.3, -0.1, 0, 0.1, 0.3]$ e $b = [-0.008 (0.002) 0.008]$, foram gerados 2,000 conjuntos de m (179) séries.

Os campos sintéticos são divididos em dois grupos: (i) correlação espacial e tendência nula ($\rho = 0$ e $b = 0$) e (ii) correlação espacial nula e tendência não-nula ($\rho = 0$ e $b \neq 0$),. Ambos tiveram sua não estacionariedade avaliada por meio da aplicação do teste de Mann-Kendall, BH95 e BH_Adap, sendo que a taxa de rejeição da hipótese nula desses três procedimentos foi dada por:

$$N_{TN} = \overline{N_{rej}}/N \quad (4.11)$$

onde $\overline{N_{rej}}$ é o número médio de rejeições obtido pela aplicação do teste de Mann-Kendall aos 2,000 campos sintéticos gerados para cada n , Cv , κ e b e N é a quantidade m de séries formadoras do campo.

Além do percentual de rejeição, em conformidade com o proposto por Ventura *et al.* (2004), adotaram-se também como métricas de avaliações o FNDR e o FDR. O FNDR pode ser conceitualmente definido como o valor esperado da proporção de falsos negativos dentre o total de hipóteses aceitas, ao passo que o FDR é o valor esperado do percentual de falsos positivos em relação à quantidade de rejeições. Essas métricas podem ser matematicamente expressas pelas equações abaixo.

$$FNDR = E[FNDP] = n_{FN}/n_{aceitas} \leq \alpha \quad (4.12)$$

onde n_{FN} é o número de vezes em que a hipótese alternativa é verdadeira, mas a série não foi rejeitada (erro tipo II) e $n_{aceitas}$ é o total de amostras que não foram rejeitadas.

$$FDR = E[FDP] = n_{FP}/n_{rejeitadas} \leq q \quad (4.13)$$

onde n_{FP} é o número de vezes em que a hipótese nula é verdadeira, mas a série foi rejeitada (erro tipo I) e $n_{rejeitadas}$ é o total de amostras que foram rejeitadas.

Esclarece-se que os valores de FNDR e FDR foram obtidos por meio da média dos 2,000 valores calculados das razões apresentadas em 4.12 e 4.13, ou seja, para cada um dos campos de mesmo n , Cv , κ e b gerados, foram determinados $n_{FN}/n_{aceitas}$ e $n_{FP}/n_{rejeitadas}$. Após a estimativa dos 2,000 valores de $FNDR$ e FDR , a média foi calculada, obtendo-se os valores das métricas procurados.

No caso daqueles campos cuja magnitude de $b \neq 0$, definiu-se uma zona não-estacionária de abrangência conhecida na área padrão, a qual se estende por 100 km em ambos eixos coordenados (x e y) e medidos a partir da origem (0,0). Entende-se como área-padrão aquela representada pelo conjunto de estações espacialmente distribuídas, segundo coordenadas geográficas dos pontos de monitoramento de pluviometria utilizados neste trabalho para a bacia do rio São Francisco, em que a origem do eixo cartesiano XY se situa na estação de maior latitude e longitude dentre as 179 consideradas. Diz-se “área padrão” porque é essa a configuração espacial que foi utilizada para a geração de todos os campos sintéticos (Figura 4.2). Em outras palavras, as estações que se localizam no interior do retângulo fictício (100x100 km, linhas azuis) têm suas séries expressas por $X_t + b \times t$, em que as magnitudes de tendências adicionadas às amostras estacionárias (X_t) são iguais a $b = [-0.008 -0.006 -0.004 -0.002 0.002 0.004 0.006 0.008]$.

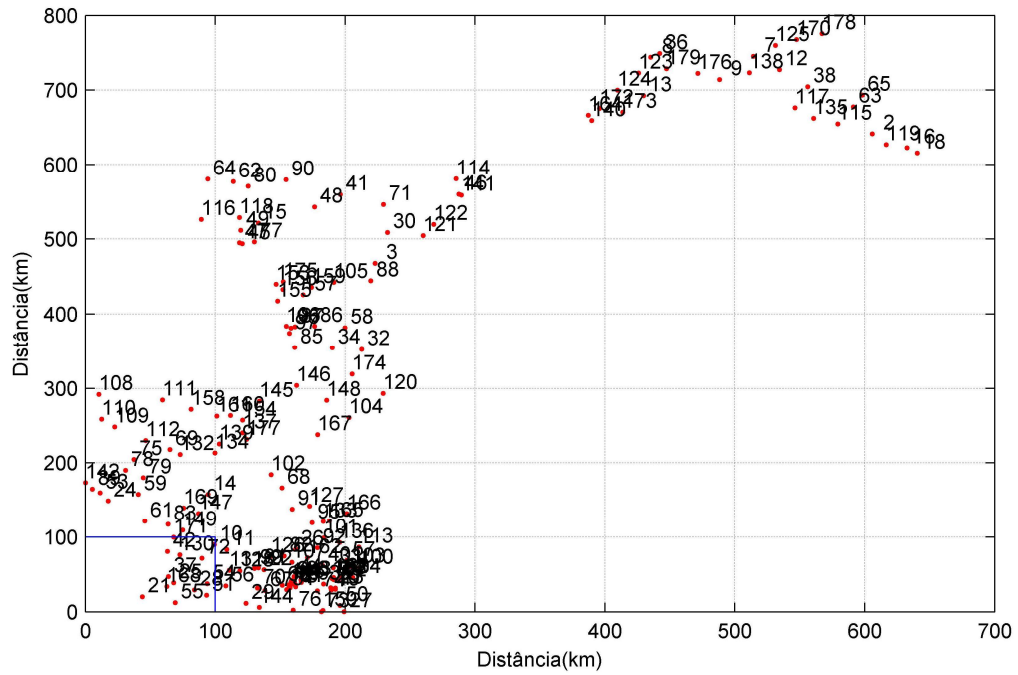


Figura 4.2 - Área Padrão considerada na geração do campo sintético: os pontos vermelhos representam as estações pluviométricas selecionadas para a bacia do São Francisco e o retângulo azul (canto inferior esquerdo) é a delimitação de zona em que foram inseridas tendência nas séries estacionárias geradas via Simulação Monte Carlo.

De maneira prática, as tarefas necessárias à geração dos campos sintéticos e posterior aplicação das metodologias de controle de erros tipo I pelo *FDR* foram sintetizadas em quatro passos e cronologicamente desenvolvidas em conformidade com o fluxograma apresentado na Figura 4.3.

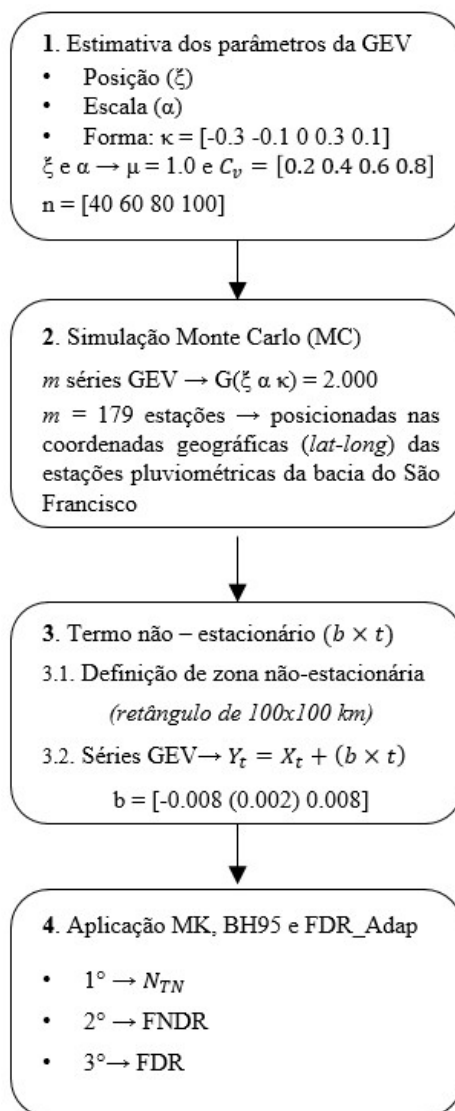


Figura 4.3 - Representação esquemática de uma iteração do procedimento aplicação de metodologias de controle pelo FDR – Caso Sintético

4.3. CASO REAL

Para o Caso Real definido neste trabalho, foram utilizados dados provenientes de observações reais de variáveis hidrológicas extremas de precipitação e vazão. Com base nesses dados, as atividades realizadas vão desde a escolha de uma região de estudo até aplicação de métodos de detecção regional de tendência que consideram a questão da multiplicidade com base no *FDR*.

A presente seção foi estruturada em 3 (três) subitens, a saber: (i) *Seleção da Área de Estudo*, (ii) *Base de Dados* e (iii) *Estimativa de Índices Hidrológicos Extremos*, e foi organizada de modo a facilitar a compreensão do leitor em relação aos passos metodológicos adotados. Os subitens foram ordenados em conformidade com a cronologia prevista das atividades, ou seja, seguindo o encadeamento temporal e lógico de execução das tarefas propostas e apresentadas no fluxograma da Figura 4.1.

Dito isso, a primeira subseção, *4.3.1. Seleção da Área de Estudo*, apresenta uma descrição sucinta de aspectos da pluviometria, vazão, relevo e economia da bacia do rio São Francisco, bem como as peculiaridades que motivaram sua escolha para este trabalho – seja pela relevância estratégica e econômica, seja pelo fato de estar inserida dentro de projeto de pesquisa *CAPES*. Outra razão de a bacia do rio São Francisco se distinguir das demais é a presença de rede de estações de monitoramento de chuva e vazão cobrindo quase a totalidade de sua extensão territorial, o que possibilita captar a variabilidade regional das variáveis hidrológicas extremas a serem analisadas.

Já na seção *4.3.2- Base de Dados* são esclarecidos aspectos relativos aos critérios adotados para a triagem das estações de interesse dentro do universo daquelas existentes na Bacia do São Francisco, ao tipo de estação (pluviométrica ou fluviométrica), ao número absoluto e localização geográfica dos pontos de monitoramentos considerados, dentre outros.

Na sequência, a subseção *4.3.3* apresenta os Índices Hidrológicos Extremos previamente selecionados, seus conceitos e expressões de cálculo. Os referidos índices podem se referir à vazão ou à precipitação e são, em sua grande maioria, provenientes do ETCCDI. Sua estimativa é feita a partir de dados de base diária coletadas por agências governamentais e que foram escolhidas em conformidade com o proposto na seção antecedente.

Esclarece-se que, por se tratar de apenas uma aplicação direta de metodologias exaustivamente discutidas na literatura e já bastante difundidas no meio acadêmico, o Passo 04, previsto no fluxograma da seção 4.1, não carece de seção específica na subseção 4 deste trabalho. Informa-se apenas que serão aplicadas à totalidade das séries temporais dos índices hidrológicos extremos o teste de Mann-Kendall associado às técnicas que visam mitigar o

efeito da autocorrelação sobre a taxa de rejeição, a saber: Pre-Whitening (PW) e o Trend-Free-Pre-Whitening (TFPW). Arbitrando-se o nível de significância local igual a 5% ($\alpha_L = 0.05$), os resultados obtidos, bem como a sua avaliação, serão considerados no capítulo *Resultados e Discussões*.

Situação semelhante ocorre em relação à aplicação de técnicas de controle do erro tipo I que empregam o *FDR* (Passo 06). Visto que, nesse passo, as metodologias já descritas no Caso Sintético foram aplicadas aos dados reais considerados, torna-se desnecessária a repetição dos procedimentos, sua base teórica e passo-a-passo executivo. Assim, tanto para o passo 06 quanto para o passo 08, serão apresentadas seções específicas apenas no capítulo de *Resultados e Discussões*.

4.3.1. Seleção da Área de Estudo

A seleção da bacia do Rio São Francisco como estudo de caso deste trabalho se deve à sua relevância estratégica para a sociedade brasileira e ao projeto de pesquisa CAPES em andamento, “*Impacto das Mudanças Climáticas em Extremos Hidrológicos (SECAS e CHEIAS)*”, que também enfoca essa região.

O rio São Francisco constitui uma importante rota de transporte de mercadorias do interior para o litoral. As 9 usinas hidrelétricas nele instaladas são responsáveis por 15% do total da capacidade hidrelétrica instalada no Brasil e possuem potencial de geração de 10.000 MW, o que supre as demandas doméstica, industrial, agrícola da região. O grande número de barragens de usos múltiplos situadas ao longo da calha principal do rio São Francisco também possibilita o controle de cheias. Cabe destacar que a bacia hidrográfica do rio São Francisco se tornou uma bacia doadora, na tentativa de reduzir o risco de escassez de água enfrentada pelos estados vizinhos da região semiárida brasileira. Esses estados, localizados ao norte da bacia de estudo, contam com aproximadamente 10 milhões de residente sem uma fonte água confiável, e foram beneficiados por um dos maiores projetos de infraestrutura da história do país, a transposição do rio São Francisco, que iniciou suas operações no ano de 2017.

Localizado em sua maior parte no Nordeste do Brasil, o rio São Francisco possui 2,860 km de extensão, flui de sul a norte e drena uma área de 630,000 km². Trata-se, portanto, da maior bacia brasileira que se situa exclusivamente em território nacional. Ao longo de sua trajetória, o rio atravessa regiões com diferentes regimes climáticos. Na sua cabeceira, localizada no estado de Minas Gerais (Região Alta), a precipitação média anual é superior a 2,000 mm, ao passo que, na zona semiárida dos estados da Bahia e Pernambuco, em que se situam, respectivamente, as porções Média e Médio-baixa da bacia, esse valor dificilmente excede os 350 mm anuais. Já nas imediações de seu delta, a precipitação anual média aumenta, atingindo valores de 1,300 mm.

Quanto à vazão, nota-se que seu valor médio de longo-termo em sua foz é de 2,810 m³/s, sendo que 70% desse total provém de suas regiões úmida Alta e de parte da zona Média da bacia, ambas no estado de Minas Gerais. O restante da porção Média, no estado da Bahia, contribui com 22%, sendo que os 8% restantes da vazão total anual originam-se da porção Média-Baixa e Baixa da área de estudo.

A população que vive na área abrangida pela bacia do rio São Francisco concentra-se na região Alta, na parte sul, e é de aproximadamente 13 milhões de pessoas, distribuídas em 521 municípios, dentre os quais Pirapora (MG), Juazeiro (BA) e Petrolina (PE) são os principais.

Assim, dada a relevância da região de estudo, detectar qualquer possível mudança no padrão de variáveis hidrometeorológicas é o primeiro de uma sucessão de passos mais abrangentes que poderão orientar o planejamento e a gestão de recursos hídricos na tentativa de tornar a sociedade mais resiliente no futuro.

4.3.2. Base de Dados

Os dados utilizados neste estudo abrangem as séries temporais de precipitação e vazão provenientes de estações operadas por 3 diferentes entidades brasileiras: Agência Nacional de Águas (ANA), Instituto Nacional de Meteorologia (INMET) e Operador Nacional do Sistema Elétrico (ONS).

A seleção dos pontos de monitoramento foi baseada nos seguintes critérios: (i) a estação deveria possuir, minimamente, 30 anos de dados, não necessariamente sequenciais, sem falhas, e (ii) pelo menos 6 anos da amostra deveriam estar compreendidos no período de 2000-2016, a fim de assegurar que o passado recente esteja representado na série a ser avaliada. Adicionalmente, apenas para as estações fluviométricas, um terceiro critério de triagem foi adotado, a saber, (iii) a presença ou não de barramento de curso d'água nas imediações de ponto de monitoramento. Visto que os resultados de significância estatística e/ou magnitude e sentido de mudança nessas estações podem estar sob a influência de regime artificialmente gerado pelas regras operacionais do barramento, as mesmas foram excluídas da base de dados. Nesses casos, a ocorrência de remanso, que altera a relação cota-vazão de estação localizada em região afetada pelas flutuações de nível de reservatório, pode resultar em equívocos na vazão indiretamente estimada (curva-chave) e, portanto, induzir à detecção de uma tendência que, na realidade, inexistente ou é estatisticamente não significativa. Esse não é o único fenômeno que pode alterar a percepção de existência, sinal e/ou sentido da tendência; a regularização das vazões advinda da existência de UHE, por exemplo, bastante comum na calha principal do rio São Francisco, resulta, intuitivamente, em períodos secos menos extremos, o que é captado equivocadamente pelos testes estatísticos procedidos como uma não-estacionariedade crescente.

Na Figura 4.4, a linha azul representa a série de $Q7_{min}$ (1937-2013) associada à estação 40150000. É possível observar, para o período pré-operação da PCH de Cajuru (-20.238°, -44.75), 1937 a 1959, o decréscimo da magnitude da variável em análise (linha tracejada), o que ilustra o efeito do barramento sobre as vazões mínimas. Diferente do observado para os anos de inexistência do referido barramento, à esquerda de linha tracejada vermelha, nota-se um rápido aumento de $Q7_{min}$ entre 1953 e 1956, que perdura, em menor grau, até o restante dos registros históricos disponíveis (linha sólida). Credita-se a diferença de comportamento à construção e posterior operação do barramento de Cajuru, que regulariza a vazão defluente da PCH e, conseqüentemente modifica a magnitude das vazões estimadas em 40150000, que se localiza a jusante. Assim, diante das dificuldades em dissociar os efeitos ocasionados pela existência do barramento de outros fatores potencialmente intervenientes na flutuação de $Q7_{min}$, as estações fluviométricas em situação análoga não foram consideradas nas análises realizadas neste trabalho.

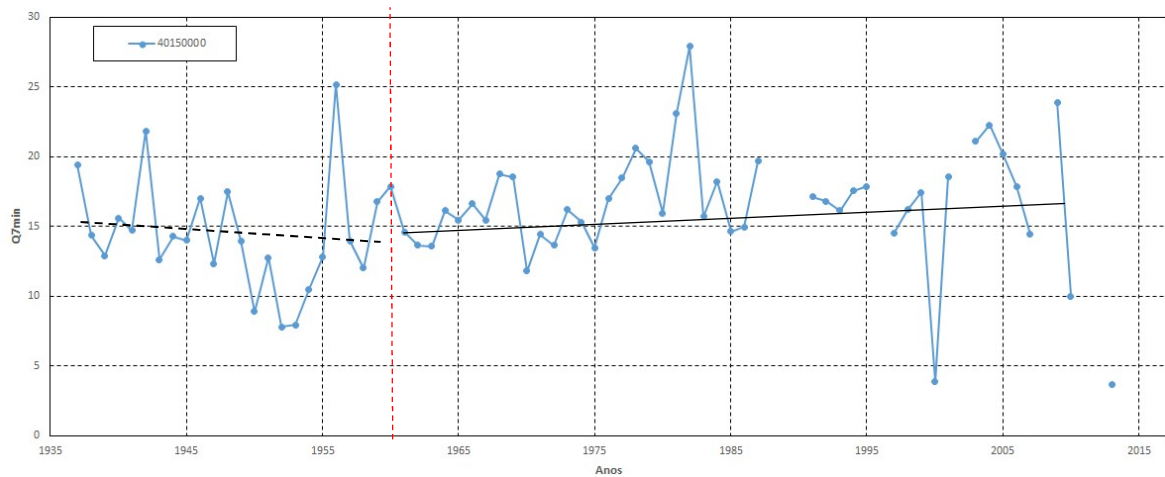


Figura 4.4 – Série temporal de Q7min (1937 – 2013) da estação 40150000 para o período pré (linha decrescente tracejada) e pós construção (linha crescente sólida) da PCH de Cajuru, em 1959 (linha tracejada vermelha).

O processo de definição da existência ou não da influência de uma dada barragem sobre as medições realizadas em uma estação fluviométrica constou, primeiramente, da inspeção visual da Figura 4.5, a fim de selecionar os pontos de monitoramento chamados de potencialmente mais afetados (PMA), ou seja, aqueles que se situam no mesmo curso d'água e são os mais próximos geograficamente de um dado barramento. Assim, para cada barragem, foi selecionado um PMA. Em paralelo, a data de início de operação de cada uma das 27 UHE's e PCH's existentes na bacia do rio São Francisco foi investigada em diversos sítios eletrônicos. Ao plotar os dados diários de vazão das PMA's ao longo do tempo, foi possível constatar visualmente a existência de alteração no padrão de descarga dos períodos pré e pós construção do barramento. Esse processo é o mesmo que foi realizado para Q7min da estação 40150000, conforme a Figura 4.4, e permite apenas duas respostas: o barramento afeta ou não afeta o comportamento da série observada. Caso a resposta seja positiva, o PMA é eliminado da base amostral e a segunda estação mais próxima da UHE e situada no mesmo curso d'água é avaliada. O procedimento continua até que se verifique um ponto de monitoramento em que a influência do barramento é indistinguível das variações naturais ou de tendência existente na amostra.

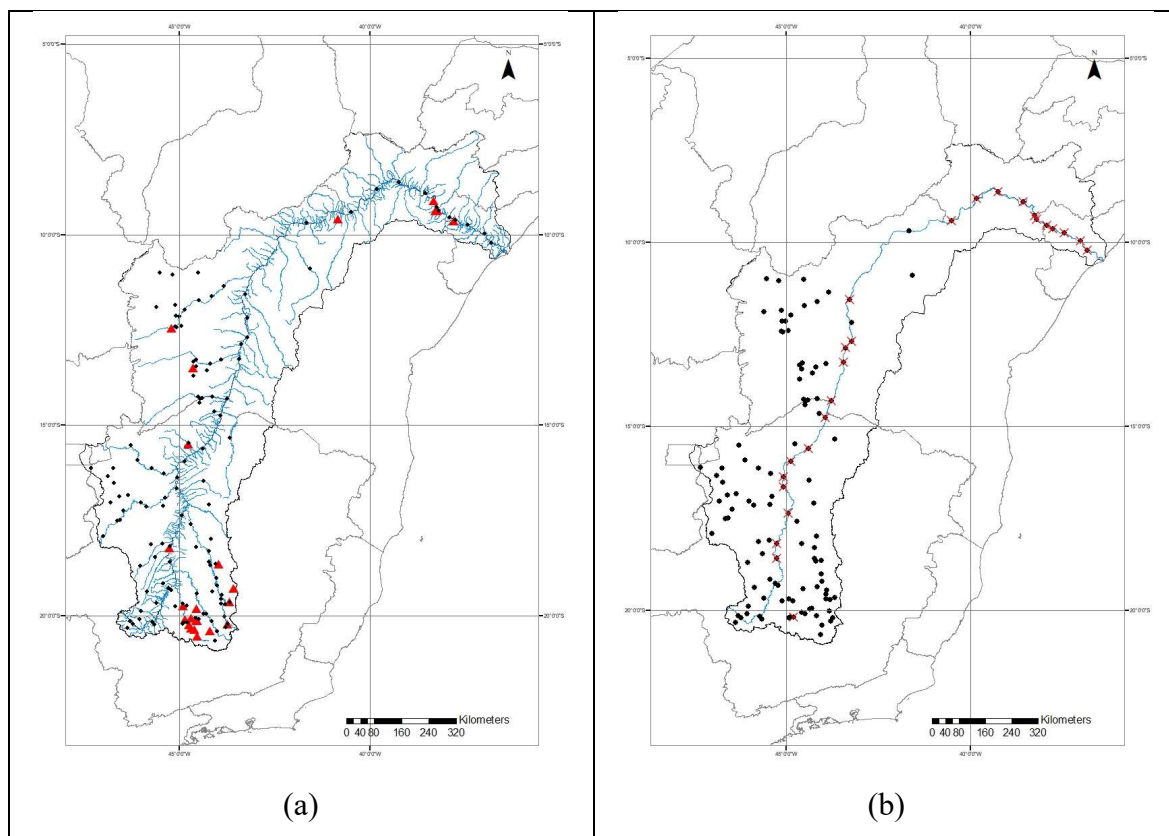


Figura 4.5 – (a) Localização das estações fluviométricas selecionadas pelos critérios (i) e (ii) e barramentos existentes; (b) estações que foram eliminadas pelo uso da premissa (iii).

Assim, baseando-se nos critérios apresentados, a base de dados construída contém 179 e 100 estações, respectivamente, de chuva e vazão, abrangendo, minimamente, o período de 1986-2006, embora as observações em algumas séries tenham sido iniciadas na década de 30. A Figura 4.6 ilustra a localização dos pontos de monitoramento de vazão e precipitação selecionados e contém um indicativo a respeito do tamanho amostral. Adicionalmente, a existência de 27 barragens, majoritariamente distribuídas no estado de Minas Gerais e ao longo da calha principal do rio São Francisco, levou à exclusão de 25 estações de fluviometria triadas com base apenas nas premissas (i) e (ii). Com base nesses critérios, apenas quatro dos postos que se localizam no curso d'água principal da bacia de estudo foram considerados na análise, a saber: 40100000, 40050000, 40037000 e 40032000.

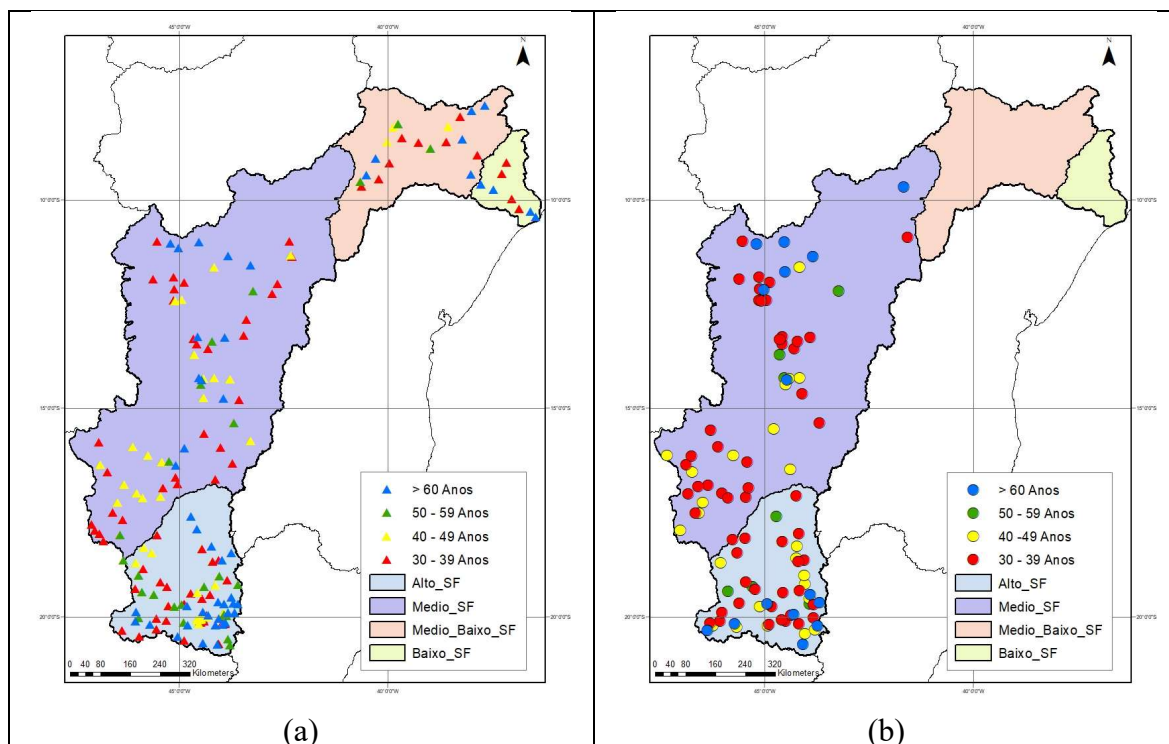


Figura 4.6 – (a) Localização das estações pluviométricas e (b) fluviométricas selecionadas para integrarem a base amostral

Conforme se nota pela Figura 4.6, a base amostral selecionada é heterogênea em relação ao tamanho da série observada, sendo que a distribuição das estações pluviométricas é espacialmente mais homogênea do que a dos postos fluviométricos, visto que inexistem pontos de monitoramento de vazão no Médio Baixo e Baixo São Francisco. Apesar dessa diferença, a região que concentra o maior número de estações, independentemente de seu tamanho amostral e tipo, é a correspondente ao Alto São Francisco.

A prevalência de postos cujo tamanho amostral se situa na faixa de maiores ou iguais a trinta (30) e menores do que quarenta (40) anos, tanto para precipitação quanto para a vazão, fica evidente na Figura 4.7 ($P_{[30,40[} = 73$ (~41%) e $Q_{[30,40[} = 54$ (54%)). Pela análise dessa figura, nota-se ainda a existência de descompasso entre a proporção de séries cujo comprimento varia entre 50 e 60 anos ($P_{[50,60[} = 24$ (~13%) e $Q_{[50,60[} = 7$ (7%)) e maiores que 60 anos ($P_{[60,....[} = 48$ (27%) e $Q_{[60,....[} = 7$ (7%)), que não é compartilhado por aquelas de pluviometria e fluviometria que integram o intervalo $[40,50[$ ($P_{[40,50[} = 34$ (19%) e $Q_{[40,50[} = 24$ (24%)).

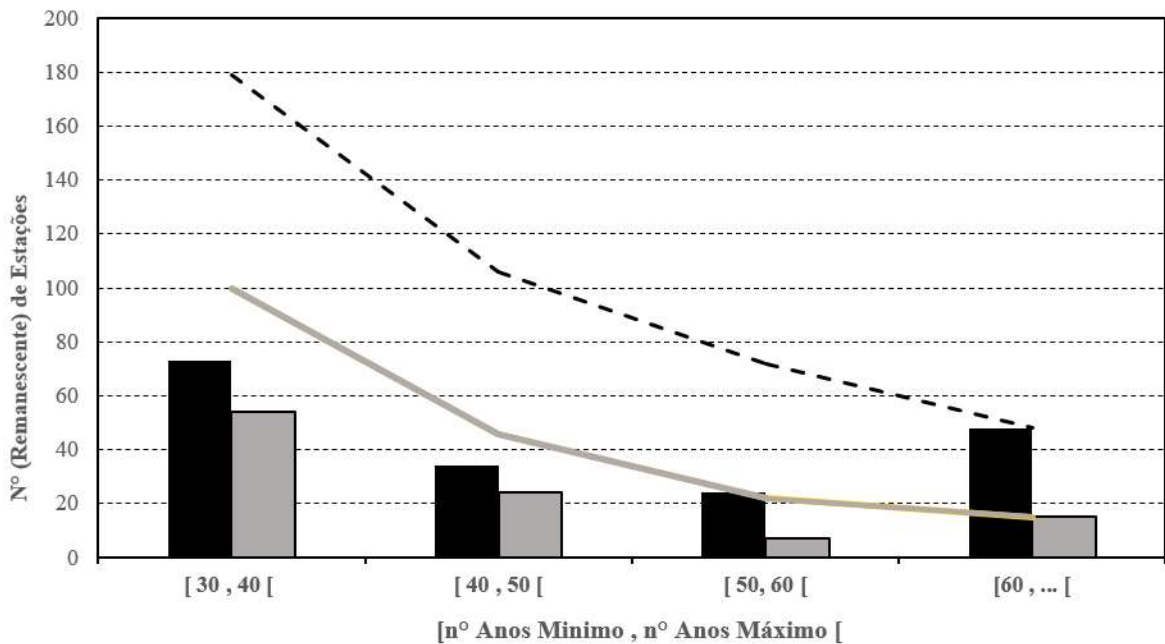


Figura 4.7 – Número (Total Remanescente) de Estações de Pluviométricas (preto) e Fluviométricas (cinza) de Faixas de Tamanho Amostral. As linhas tracejada e sólida representam, respectivamente, o total acumulado de Estações disponíveis para o estudo à medida que o número de anos aumenta.

Observa-se, pelo traçado das linhas tracejada e sólida da Figura 4.7, que o número total de estações remanescentes decresce rapidamente à medida em que se aumenta o valor do tamanho amostral mínimo exigido para que uma estação seja considerada apta. Isso justifica a adoção de estações de menor comprimento serial, no caso, aquelas compreendidas entre 30 e 50 anos, para a composição da base de dados. Diz-se isso já que, ao se avaliar a presença de tendências na Bacia do São Francisco, é desejável que o conjunto amostral escolhido satisfaça, de maneira equilibrada, dois aspectos básicos a priori: (i) representatividade da área de estudo, ou seja, que o número de estações pluviométricas e fluviométricas analisado seja capaz de representar, com o máximo de confiança possível, a dinâmica das chuvas e vazões observadas na bacia de estudo e suas nuances e (ii) que as amostras utilizadas tenham tamanho suficiente para que o poder de detecção dos testes utilizados nas diversas abordagens previstas sejam minimamente afetados, evitando-se as falsas detecções de tendência, problemática relatada em seções anteriores.

Maiores detalhes referentes às estações pluviométricas e fluviométricas utilizadas neste trabalho podem ser vistos no Apêndice I – Estações Pluviométricas e Fluviométricas, que apresenta individualmente o comprimento amostral, bem como a presença ou não de falhas anuais nas séries disponíveis para análise

4.3.3. Seleção e Estimativa de Índices Hidrológicos Extremos

A seleção dos índices extremos adotados neste trabalho foi norteadada pela tentativa de abranger/caracterizar, de maneira concisa, diferentes aspectos da precipitação e vazão presentes nas séries históricas das estações da bacia do rio São Francisco, bem como de possibilitar o alinhamento do estudo às iniciativas globais de estudos de mudança do clima, como o ETCCDI.

Para tanto, dentre os 27 índices extremos listados no sítio eletrônico da entidade supracitada, buscou-se selecionar aqueles que são aplicáveis à realidade brasileira – excluindo, por exemplo, os índices relacionados à ocorrência de neve –, e que fossem capazes de condensar características das variáveis de interesse deste trabalho, tais como frequência, duração e magnitude. Assim, foram selecionados dez índices extremos referentes à precipitação: PRCPTot, R1, R10, R20, R50, RX1d, RX5d, SDII, CDD e CWD.

Focado basicamente na precipitação e temperatura, o ETCCDI negligencia a vazão. Diante disso, foi necessário criar índices específicos que permitissem expressar os comportamentos extremos da variável hidrológica em questão, máximos e mínimos, dentro de um intervalo de tempo desejado. Como resultado, definiram-se seis índices extremos relacionado à vazão, a saber: Qméd, (QX1d) Qmáx, Q7min, Q30 min, QX5d e QX30d.

Assim, os dezesseis índices extremos de base anual foram estimados para cada um dos anos hidrológicos de uma dada estação, independente do tipo, por meio das seguintes expressões:

a) **PRCPTOT**, *total anual precipitado em dias úmidos*:

Seja RR_{ij} o total precipitado em um dia i em um período j . Se l representa o número de dias em j , então:

$$PRCPTOT = \sum_{i=1}^l RR_{ij} \quad (4.14)$$

b) R1, *Contagem anual de dias em que a precipitação é superior ou igual a 1 mm*

Seja RR_{ij} o total precipitado em um dia i em um período j . Conta-se o número de dias em que:

$$RR_{ij} \geq 1mm \quad (4.15)$$

c) R10, *Contagem anual de dias em que a precipitação é superior ou igual a 10 mm*

Seja RR_{ij} o total precipitado em um dia i em um período j . Conta-se o número de dias em que:

$$RR_{ij} \geq 10mm \quad (4.16)$$

d) R20, *Contagem anual de dias em que a precipitação é superior ou igual a 20 mm*

Seja RR_{ij} o total precipitado em um dia i em um período j . Conta-se o número de dias em que:

$$RR_{ij} \geq 20mm \quad (4.17)$$

e) R50, *Contagem anual de dias em que a precipitação é superior ou igual a 50 mm*

Seja RR_{ij} o total precipitado em um dia i em um período j . Conta-se o número de dias em que:

$$RR_{ij} \geq 50mm \quad (4.18)$$

f) Rx1d, *Máximo mensal precipitado em 1 (um) dia:*

Seja RR_{ij} o total precipitado em um dia i em um período j . O valor máximo diário para o período j é:

$$Rx1d = \max(RR_{ij}) \quad (4.19)$$

g) Rx5d, *Máximo mensal precipitado em 5 dias consecutivos:*

Seja RR_k o total precipitado em um intervalo de 5 dias terminado em k . Então o valor máximo de 5 (cinco) dias para um período j é:

$$Rx5d = \max(RR_k) \quad (4.20)$$

h) SDII, *índice de intensidade de precipitação simples*:

Seja RR_{wj} a precipitação diária em dias úmidos ($RR \geq 1mm$) em um período j . Se W representa o número de dias úmidos em j , então:

$$SDII_j = \frac{\sum_{w=1}^W RR_{wj}}{W} \quad (4.21)$$

i) CDD, *Máxima duração do período de seca, ou seja, número máximo de dias consecutivos com total precipitado inferior a 1 mm ($RR < 1 mm$)*

Seja RR_{ij} o total precipitado em um dia i em um período j . Conta-se o maior número de dias consecutivos em que:

$$RR_{ij} < 1 mm \quad (4.22)$$

j) CWD, *Máxima duração do período úmido, ou seja, número máximo de dias consecutivos com total precipitado superior ou igual a 1 mm ($RR \geq 1 mm$)*

Seja RR_{ij} o total precipitado em um dia i em um período j . Conta-se o maior número de dias consecutivos em que:

$$RR_{ij} \geq 1 mm \quad (4.23)$$

k) Qméd, *Vazão média anual*

Seja Q_{ij} a vazão média em um dia i em um período j . O valor médio diário para o período j é:

$$Q_{méd} = \text{média}(Q_{ij}) \quad (4.24)$$

l) Qx1day (Qmáx), *Vazão máxima anual*

Seja Q_{ij} a vazão média em um dia i em um período j . O valor máximo diário para o período j é:

$$Q_{x1day} = \text{maximo}(Q_{ij}) \quad (4.25)$$

m) Q7min, *Vazão mínima de média diária de 7 dias consecutivos*

Seja Q_{kj} a vazão média de um intervalo de 7 dias terminado em k . Então o valor mínimo de 7 (sete) dias para um período j é:

$$Q_{min} = \min(Q_{7j}) \quad (4.26)$$

n) Q30min, Vazão mínima de média diária de 30 dias consecutivos

Seja Q_{kj} a vazão média de um intervalo de 30 dias terminado em k . Então o valor mínimo de 30 (trinta) dias para um período j é:

$$Q_{min} = \min(Q_{30j}) \quad (4.27)$$

o) Qx5day, Vazão máxima para um período de 5 dias consecutivos

Seja Q_k a vazão máxima em um dia i para um intervalo de 5 consecutivos terminado em k . Então o valor máximo diário para o período j é:

$$Q_{x5d} = \max(Q_{kj}) \quad (4.28)$$

p) Qx30day, Vazão máxima para um período de 30 dias consecutivos

Seja Q_k a vazão máxima em um dia i para um intervalo de 30 consecutivos terminado em k . Então o valor máximo diário para o período j é:

$$Q_{x30d} = \max(Q_{kj}) \quad (4.29)$$

Esclarece-se que a definição do ano hidrológico (j) foi realizada estação a estação, sendo seu início definido como o mês seguinte àquele que, historicamente, apresenta a maior ocorrência de mínimos de vazão ou precipitação para, respectivamente, pontos de monitoramento de fluviometria e pluviometria.

5. RESULTADOS E DISCUSSÃO

O presente capítulo foi subdividido em cinco subseções, sendo que as três primeiras apresentam os resultados e discussões referentes ao Caso Sintético – avaliação do desempenho dos testes de (5.1) MK, (5.2) B_SEN e (5.3) BH95 e BH_Adap, quando aplicados à séries extremas (GEV e diferentes valores de n , C_v , κ e b) – e as duas outras estão relacionadas ao Caso Real. Neste, procurou-se identificar quais os efeitos (5.4) da autocorrelação e (5.5) da multiplicidade sobre os testes de estacionariedade aplicados, estabelecendo padrões de mudança, caso existentes, para a bacia do São Francisco, com base nos resultados provenientes da associação do MK com o PW e TFPW, bem como BH95 e BH_Adap para 10 índices extremos de vazão e 6, de precipitação.

5.1. AVALIAÇÃO DE DESEMPENHO DO TESTE DE MANN-KENDALL APLICADO A SÉRIES DE EXTREMOS

Em conformidade com as etapas metodológicas descritas na seção 4.2.2, foi realizada simulação Monte Carlo (MC) para avaliar o poder do teste de Mann-Kendall. Apenas para recapitular, o experimento supracitado resultou na geração de 10.000 séries temporais independentes, distribuídas segundo uma GEV, para tamanhos amostrais $n = 20$ (10) 100, com média $\mu = 1$, coeficientes de variação $Cv = 0.1$ (0.1) 1.0 e $\kappa = [-0.3, -0.1, 0, 0.1, 0.3]$. Ademais, a cada uma das possíveis combinações entre as variáveis anteriormente elencadas para a geração de amostras, foram acrescentadas tendências lineares de magnitude pré-definidas ($T_t = b \times t$, $b = -0.01$ (0.002) 0.01, $t = 0, 1, 2, \dots, n$). Relata-se que dentre os valores previstos para b inclui-se o zero, que resultará em séries estacionárias, e cuja proporção de rejeição da hipótese nula esperada limita-se ao nível de significância do teste assumido (5%).

Procedida a simulação supracitada, os resultados foram organizados de modo que os fatores intervenientes na capacidade de detecção do teste de MK analisados – tamanho, coeficiente de variação amostral, parâmetro de forma da GEV e magnitude da tendência – fossem evidenciados. Para tanto, foram elaborados gráficos de barra agrupados e invariavelmente da forma: (*magnitude da tendência x taxa de rejeição*), mostrando o comportamento

global do percentual de rejeição do teste de MK à medida em que ocorrem acréscimos ou decréscimos da variável de interesse ao se variar b e Cv ou κ .

Esclarece-se que, no caso dos gráficos do tipo barra, a magnitude da tendência b será expressa por meio de seu valor de mudança decadal relativa. Por exemplo, para um $b = 0.01$ tem-se 10% de mudança na magnitude ao longo da década, ao passo que $b = 0.006$ corresponde a 6% e assim por diante.

A Figura 5.1 mostra a relação entre poder, magnitude da tendência e tamanho amostral para um dado nível de significância ($\alpha = 0.05$), parâmetro de forma ($\kappa = -0.3$) e coeficientes de variação ($Cv = 0.2, 0.6$ e 1.0). Nota-se, de maneira geral, que o poder do teste é uma função crescente tanto para o valor absoluto da magnitude da tendência quanto para o tamanho serial. Em outras palavras, à medida que o tamanho da série aumenta, o poder do teste cresce, o que indica um aumento da capacidade de detecção de tendência pelo teste de MK. Esse padrão fica ainda mais evidente ao se manterem fixos o coeficiente de variação e parâmetro de forma, sendo possível observar que a menor e a maior taxa de rejeição da hipótese nula estão associadas ao mínimo e máximo tamanho amostral simulados, ou seja, quando $n = 20$ e $n = 100$. Como exemplo, cita-se a diferença de aproximadamente 42% entre as taxas de rejeição determinadas para $n = 20$ e $n = 100$, a maior calculada dentre as possíveis para os diferentes comprimentos de séries geradas, quando $b = -0.01$ e $Cv = 0.2$. Esclarece-se que essa constatação se estende para todos os valores de magnitude de tendência testadas, independentemente do valor de Cv considerado.

Dada a existência de eixo de simetria em $x = 0$ ($b = 0$), optou-se por mostrar na Figura 5.1 apenas os resultados de rejeição da hipótese nula para as magnitudes de tendência definidas como positivas, 2.00, 4.00, 6.00, 8.00 e 10.00% (magnitude de mudança decadal relativa), e nulas (0.00). Assim, percebe-se que as taxas de rejeição são mínimas para as amostras estacionárias, quando considerado um coeficiente de variação específico e toda a faixa de b utilizada para o processo de simulação Monte Carlo de geração serial. Com valores próximos ao nível de significância do teste, a proporção de séries declaradas como não estacionárias para $b = 0$ é compatível com o esperado, e sofre menos impacto das variações do tamanho e coeficiente de variação amostral do que aquelas sabidamente dotadas de tendência. Diz-se

isso porque, para $n = 20$ e $n = 100$, o percentual de rejeição determinado pouco se altera, quando $b = 0$, mesmo com a elevação de C_v . Isso não ocorre para amostras dotadas de tendência, já que, além do expressivo aumento da rejeição ao se elevar o tamanho da amostra, observa-se redução com acréscimos no coeficiente de variação (Por exemplo: $C_v = 0.6 \rightarrow n=20 (0.048)$ e $n=100 (0.47)$ e $C_v = 1.0 \rightarrow n=20 (0.046)$ e $n=100 (0.21)$, ambas para $b=2.00\%$ e $\kappa = -0.3$).

A Figura 5.1 sinaliza ainda que ocorre redução do total de séries rejeitadas à medida que o valor do coeficiente de variação se eleva C_v (0.2, 0.6 e 1.0), provavelmente graças ao aumento do ruído amostral. O crescimento da dificuldade em se detectar tendências devido à variação positiva de C_v fica evidente pelo confronto do percentual de rejeição obtido para séries cujo tamanho e magnitude de mudança são iguais a, respectivamente, 100 e 2.00%, por exemplo. A redução das taxas de rejeição à metade a cada 0.40 que se eleva na magnitude do coeficiente de variação ($C_v(0.20) = 0.99$, $C_v(0.60) = 0.45$ $C_v(1.0) = 0.21$) para esse caso indica a possibilidade de extensão das evidências relatadas para valores de C_v , b e n diferentes dos aqui apresentados.

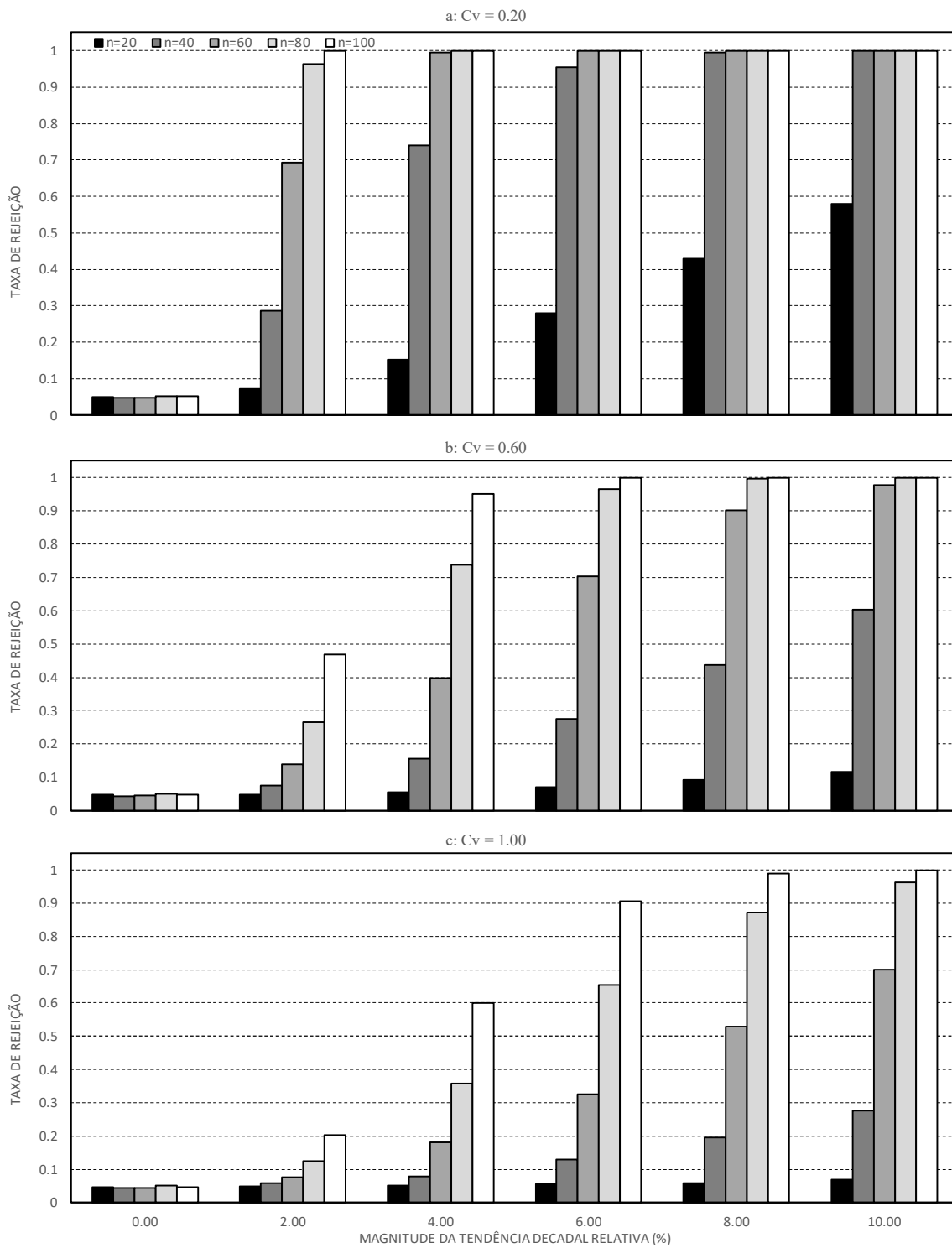


Figura 5.1 – Taxas de rejeição de séries temporais com tendências de magnitude $b = 0.00$ (2.00) 10.00, tamanhos amostrais $n = [20\ 40\ 60\ 80\ 100]$, $\alpha = 0.05$, $\kappa = -0.30$ e (a) $C_v = 0.20$, (b) $C_v = 0.60$ e (c) $C_v = 1.00$.

Visando confirmar a constatação vista na Figura 5.1 para o Cv e expressar a relação global entre taxa de rejeição, magnitude da tendência e coeficiente de variação, o gráfico apresentado na Figura 5.2 foi elaborado para amostras de tamanho igual a 50, nível de significância de 0.05, parâmetro de forma de ($\kappa = -0.3, 0.0$ e $+0.30$) e diferentes Cv 's (0.2 (0.2) 1.0) Esse gráfico permite observar que, para um valor de b fixo, o poder do teste é uma função decrescente quando avaliada sob a perspectiva do coeficiente de variação de uma determinada série temporal. Isso ratifica que o poder do teste reduz com o crescimento da variabilidade dentro de uma dada amostra, representado pelo aumento de Cv , tornando-se mais difícil detectar a existência de tendências estatisticamente significantes.

Já a Figura 5.3 mostra a dependência existente entre taxa de rejeição, magnitude da tendência e parâmetro de forma (κ) para um nível de significância ($\alpha = 0.05$), coeficiente de variação ($Cv = 0.4$) e tamanho ($n = 50$). Observa-se claramente que, para um valor fixo de κ , a capacidade de detecção é uma função crescente em relação ao valor absoluto da magnitude da tendência. Em outras palavras, quanto maiores foram os módulos de b , maior será a facilidade em se detectar. Isso pode visto ao verificar a evolução das taxas de rejeição, por exemplo, das séries com $k = -0.30$ à medida que b assume magnitudes de tendência ascendentes e $Cv = 0.20$, a saber: 0.00 (0.05), 2.00 (0.485), 4.00 (0.945), 6.00 (0.998), 8.00 (1.00) e 10.00% (1.00).

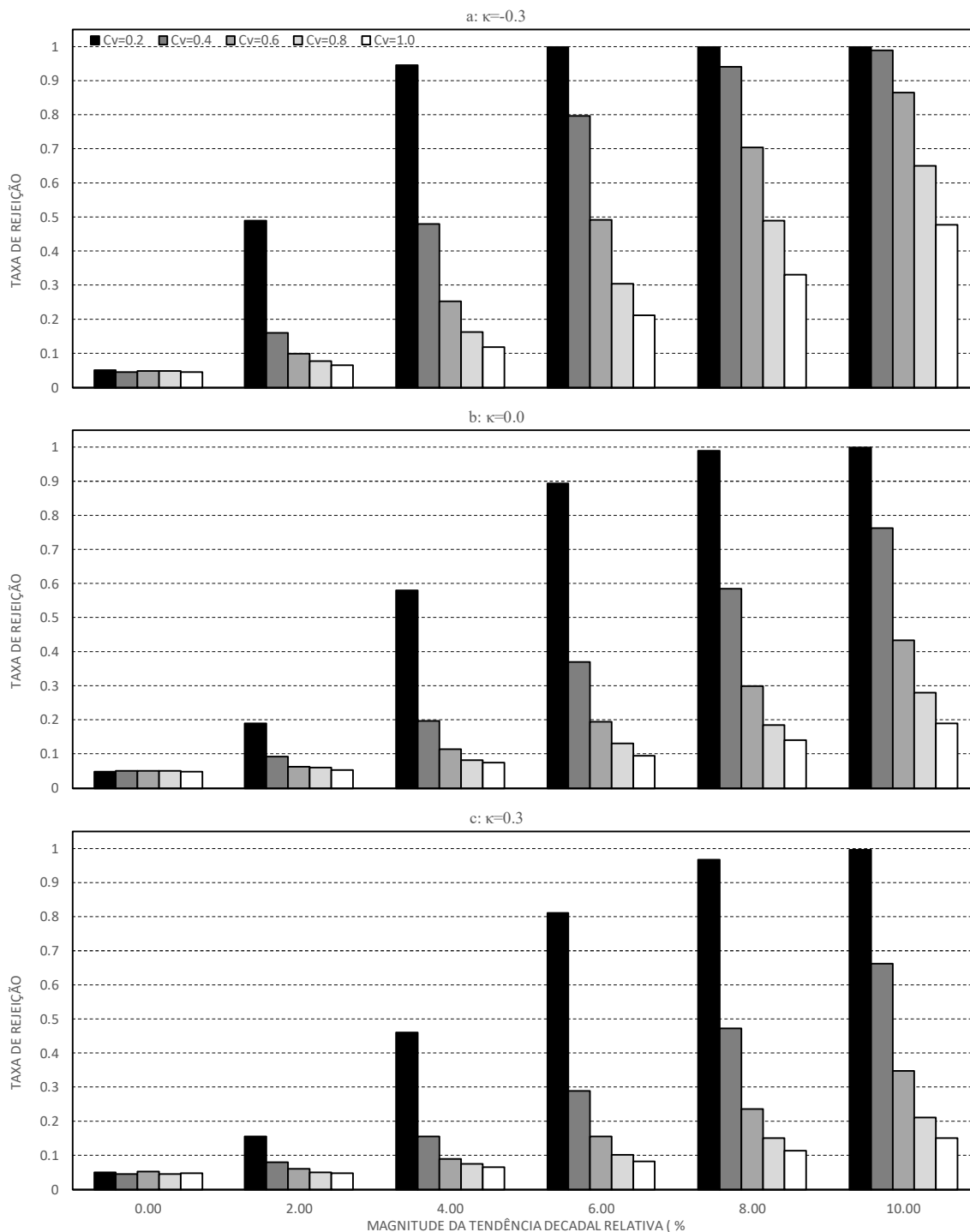


Figura 5.2 – Taxas de rejeição de séries temporais com tendências de magnitude $b = 0.00$ (2.00) 10.00, tamanho amostral $n = 50$, $\alpha = 0.05$, $C_v = 0.20$ (0.20) 1.00 e (a) $\kappa = -0.30$, (b) $\kappa = 0.0$ e (c) $\kappa = +0.30$.

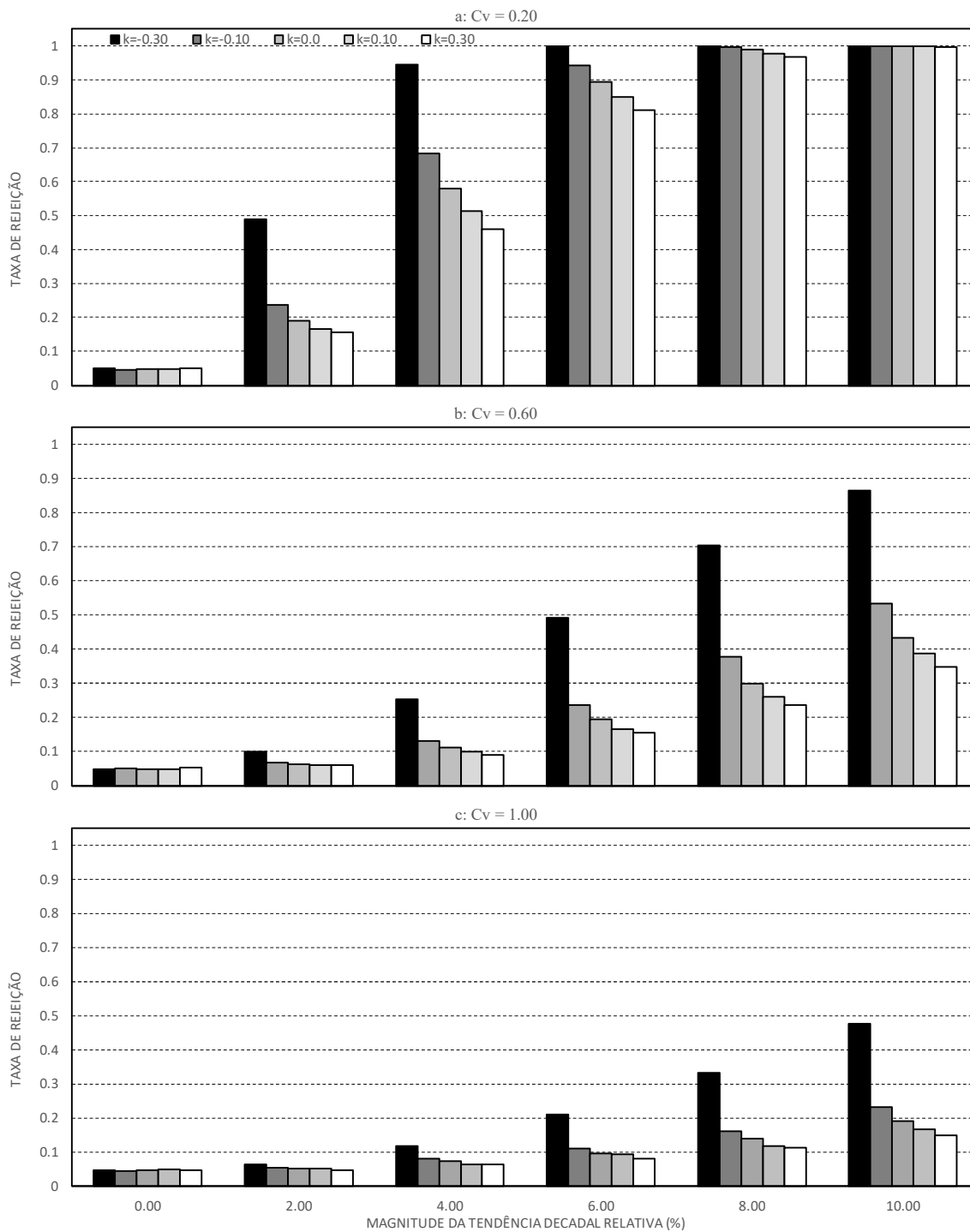


Figura 5.3 – Taxas de rejeição de séries temporais com tendências de magnitude $b = 0.00$ (2.00) 10.00%, tamanhos amostrais $\kappa = [-0.30 - 0.10 0.00 0.10 0.30]$, $\alpha = 0.05$, $n = 50$ e (a) $C_v = 0.20$, (b) $C_v = 0.60$ e (c) $C_v = 1.00$.

Diferente dos valores de κ utilizados por Yue *et al.* (2002a), $\kappa = [-1.0 \ 0.0 \ 1.0]$, a Figura 5.4 retrata, em um mesmo par de eixos coordenados, a relação explicitada pela Figura 5.4, de redução da taxa de rejeição da hipótese nula à medida que o parâmetro de forma da distribuição do tipo GEV torna-se negativo e crescente em módulo. Procurando utilizar magnitudes de k compatíveis com a realidade hidrológica, o presente trabalho obteve resultados similares aos descritos por aquele autor, apesar de aparente violação de restrições impostas ao momento de terceira ordem r da distribuição *GEV*, já que o mesmo se encontra limitado a $k < -1/3$. A Figura 5.4 evidencia que as taxas de rejeição decaem quando o parâmetro de forma tem a sua magnitude aumentada na direção positiva, sendo os máximos e mínimos, respectivamente, constatados para $\kappa = -0.3$ e $\kappa = 0.30$.

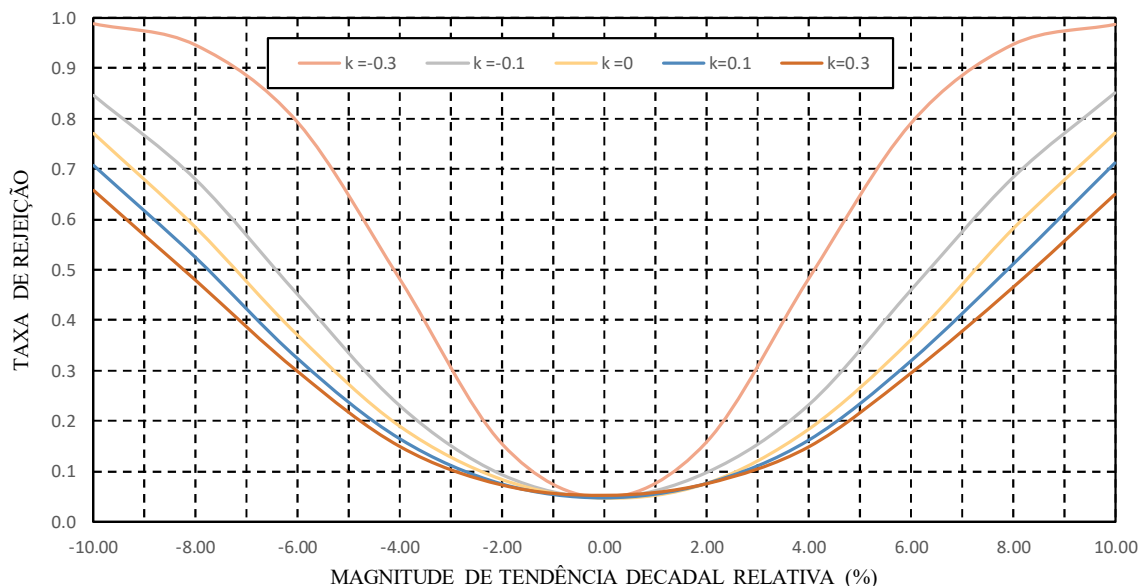


Figura 5.4 – Curva Magnitude da Tendência – Taxa de rejeição de séries do tipo GEV com parâmetros de forma iguais a $[-0.3 \ -0.1 \ 0 \ 0.1 \ 0.3]$, $\alpha = 0.05$, $C_v = 0.4$ e $n = 50$).

A Figura 5.5 procura sintetizar as constatações até aqui relatadas. Plotando-se linhas de iguais valores de poder do teste em eixos coordenados x e y, respectivamente, tamanho amostral, em anos, e a razão entre o sinal-ruído (b/C_v), para valores de κ iguais a -0.30, 0.0 e +0.30, nota-se que o crescimento dessa razão resulta em ganho de poder quando considera-se um tamanho amostral e κ fixos. Isso é consistente com a variação esperada na capacidade de detecção ao se elevar o C_v . Ao se manter fixo o valor de b , a relação sinal-ruído se eleva ao se diminuir os valores do coeficiente de variação, que são inferiores ou iguais a unidade na simulação MC procedida. O mesmo comportamento visto para C_v se observa para κ , ao

se manterem fixos os demais fatores influentes na detecção. Isso pode ser observado na diferença de poder estimado entre, por exemplo, $n = 30$, $b/C_v = 0.01$ e κ 's iguais a -0.30 (~ 0.20) e $+0.30$ (< 0.1). O fato de a distância entre as linhas de igual poder para $\kappa = -0.30$ possuírem menor distância entre si do que para $\kappa = +0.30$, para um mesmo valor de n , indica também que o poder decresce à medida que o valor do parâmetro de forma se eleva. A relação crescente da magnitude da tendência (b) e tamanho amostral (n) com o poder do teste de Mann-Kendall também pode ser vista na Figura 5.5 ao se deslocar no sentido positivo, respectivamente, de x para b e y para n .

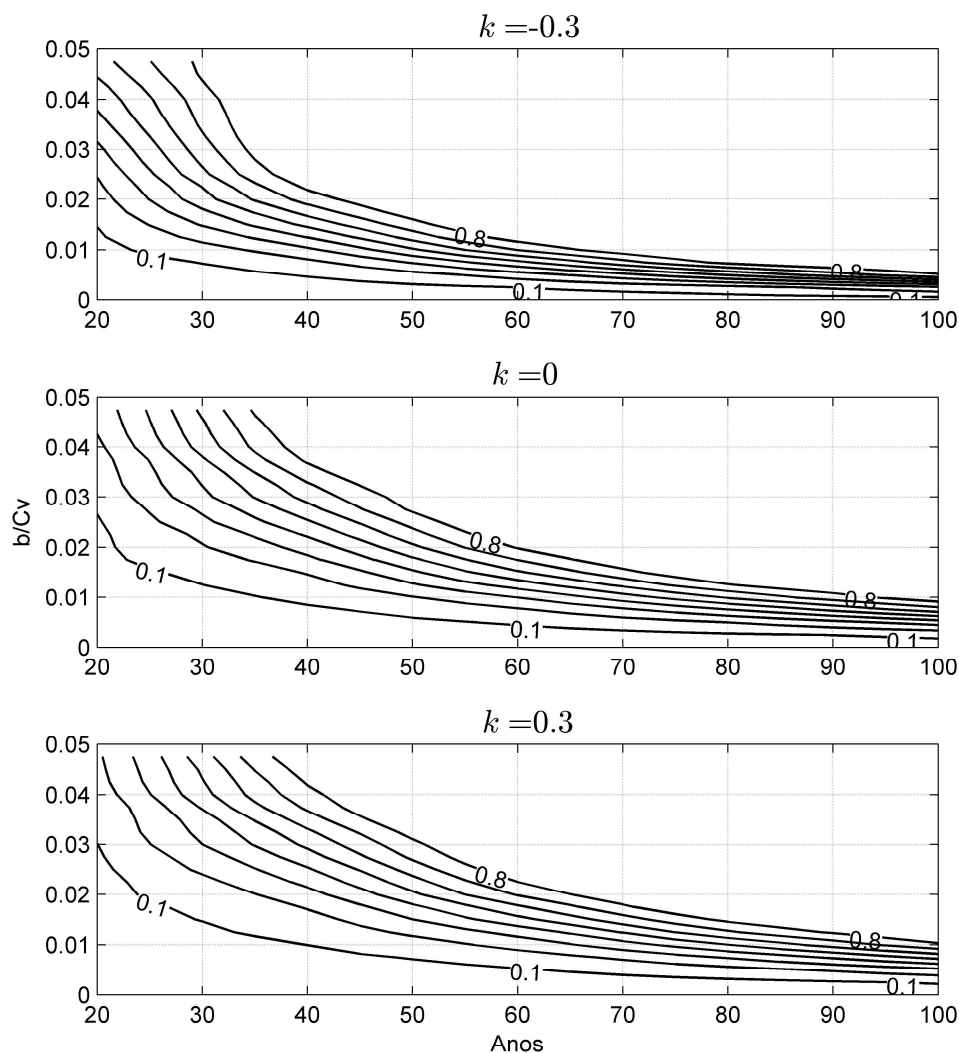


Figura 5.5 – Curvas de igual poder do teste de MK para diferentes relações sinal-ruído (b/C_v), tamanhos amostrais (n) e parâmetros de forma da GEV ($\kappa = -0.30, 0.0, +0.30$)

5.2. AVALIAÇÃO DE DESEMPENHO DO ESTIMADOR B_SEN APLICADO A SÉRIES DE EXTREMOS

Em conformidade com o exposto na seção 4.2.3 e com a meta de cumprir o que propõe o objetivo específico 02 deste trabalho, foram elaborados gráficos para apresentar e sintetizar os principais resultados advindos de Simulação Monte Carlo realizada para avaliar o estimador de magnitude de tendência proposto por Sen (1968), o B_SEN.

A Figura 5.6 mostra a dependência da probabilidade do Erro Tipo S para séries do tipo GEV ao se variar o tamanho amostral ($n = [20\ 40\ 60\ 80\ 100]$), coeficiente de variação ($Cv = [0.2\ 0.6\ 1.00]$) e a magnitude da tendência decenal relativa positiva ($b = [2.0\ 4.0\ 6.0\ 8.0\ 10.0\%]$), mantendo-se fixo apenas o parâmetro de forma $\kappa = -0.30$. Assim, pode-se constatar que a proporção de séries que tiveram o seu sinal estimado por B_SEN contrário àquele assumido durante o processo de geração serial pode ser expressiva em diversas situações, ficando claro que a probabilidade do Erro Tipo S é decrescente à medida que se eleva o tamanho da amostra (n). Em outras palavras, ao se gerar 10,000 séries do tipo GEV, via MC, para um dado κ , Cv , b fixos e diversos n , a taxa de equívocos de sinal estimados pela metodologia proposta por SEN será tão menor quanto maior o comprimento da série. Por exemplo, para $Cv = 1.00$ e $b = 2.00\%$, o caso mais extremo dentre os considerados neste trabalho, a probabilidade de se obter uma tendência com sinal trocado é de 36.4% para $n = 20$, 18% para $n = 40$, 8% para $n = 60$, 2.5% para $n = 80$ e menos de 1% para uma série de tamanho $n = 100$.

Naturalmente, os resultados apresentados acima estão relacionados com uma combinação extrema de Cv e b , sendo outras combinações possíveis de serem encontradas na prática. Por exemplo, um comportamento similar ao observado para o tamanho amostral nos casos mencionados acima ocorre com o aumento da magnitude da tendência, em que a relação negativa fica evidente ao se constatar que as barras da Figura 5.6 diminuem de magnitude da esquerda para a direita, ou seja, no sentido de crescimento da tendência relativa decenal. Cita-se a diminuição no valor da probabilidade do *Erro Tipo S* em amostras cujo $n = 40$ e $Cv = 1.0$, passando de 18.4%, quando $b = 2.00\%$, para 6.3% quando $b = 4.00\%$, o que mostra claramente que o problema tem em uma de suas causas a relação entre Cv e b .

Além disso, a Figura 5.6 evidencia o crescimento da probabilidade do Erro Tipo S com o aumento do coeficiente de variação, o que pode ser visto para amostras, por exemplo, de tamanho $n=20$ e $b = 2.00\%$ e que tem o Cv aumentado de 0.20 (0.076) para 0.60 (0.295) e, posteriormente para 1.00 (0.364). Intuitivamente, essa relação entre Cv e o erro em questão era esperada, dado que o aumento do ruído serial proporciona maior variabilidade à série, conferindo a possibilidade de aumento no número de equívocos do sentido de sinal de mudança. Ratificando a impressão anteriormente relatada, a Figura 5.7 mostra que, mesmo para diferentes valores do parâmetro de forma (κ), a relação positiva entre Cv e probabilidade do Erro Tipo S permanece, desde que os demais fatores intervenientes estudados sejam mantidos constantes. Diz-se isso ao observar que, para $n = 50$, $b=2.00\%$ e $k=0.3$, os valores das ordenadas para $Cv=0.2$ e 1.0 são, respectivamente, iguais a 0.010 e 0.285.

Em relação à Figura 5.7, nota-se que os efeitos de aumento da magnitude de b , de maneira análoga ao que ocorre para o tamanho amostral, resultam em decréscimos no valor da probabilidade do Erro Tipo S observado para diferentes Cv 's simulados, ou seja, a relação de dependência entre essas variáveis é negativa.

A relação positiva existente entre o parâmetro de forma da GEV (κ) e a probabilidade de se cometer equívocos na indicação do sinal de tendência estimada por B_SEN , já sugerida pela Figura 5.6, é facilmente constatada na Figura 5.7 ao se verificar que os menores e maiores valores de probabilidade do Erro Tipo S estão associados, respectivamente, ao menor (-0.3) e maior (0.3) κ simulados, para n , Cv e b fixos. Traduzindo em números, tem-se que a diferença observada entre os Erros do Tipo S para $\kappa = -0.30$ e $\kappa = 0.30$, considerando-se $n = 50$, $Cv = 0.60$, $b = 2.00\%$ fixos, é de 13.2%. As relações do parâmetro de forma e a taxa de erro em análise com a variação da magnitude da tendência (b) e o coeficiente de variação (Cv) são, respectivamente, crescentes e decrescentes.

Adicionalmente, o que chama a atenção é que, independente das variáveis intervenientes testadas, n , Cv , b e κ , existem combinações possíveis entre elas que resultam em percentuais de sinais com indicação equivocada em torno de 30%.

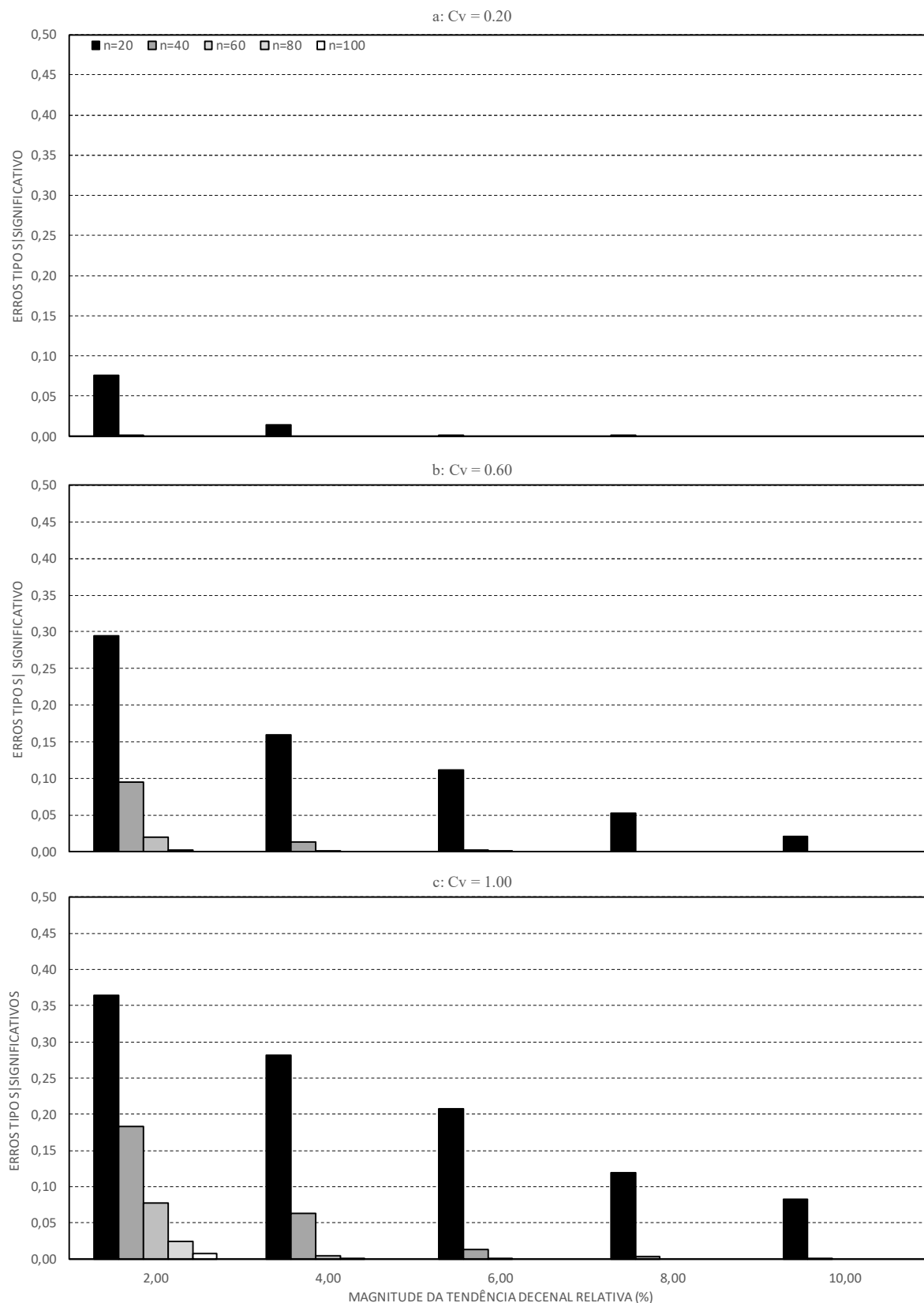


Figura 5.6 – Probabilidade do Erro Tipo S (sinal) dado que a estação foi declarada não-estacionária ($\alpha = 0.05$) para $\kappa = -0.30$ e diferentes valores de $n = [20\ 40\ 60\ 80\ 100]$ e $Cv = 0.2$ (a), 0.6 (b) e 1.0 (c)

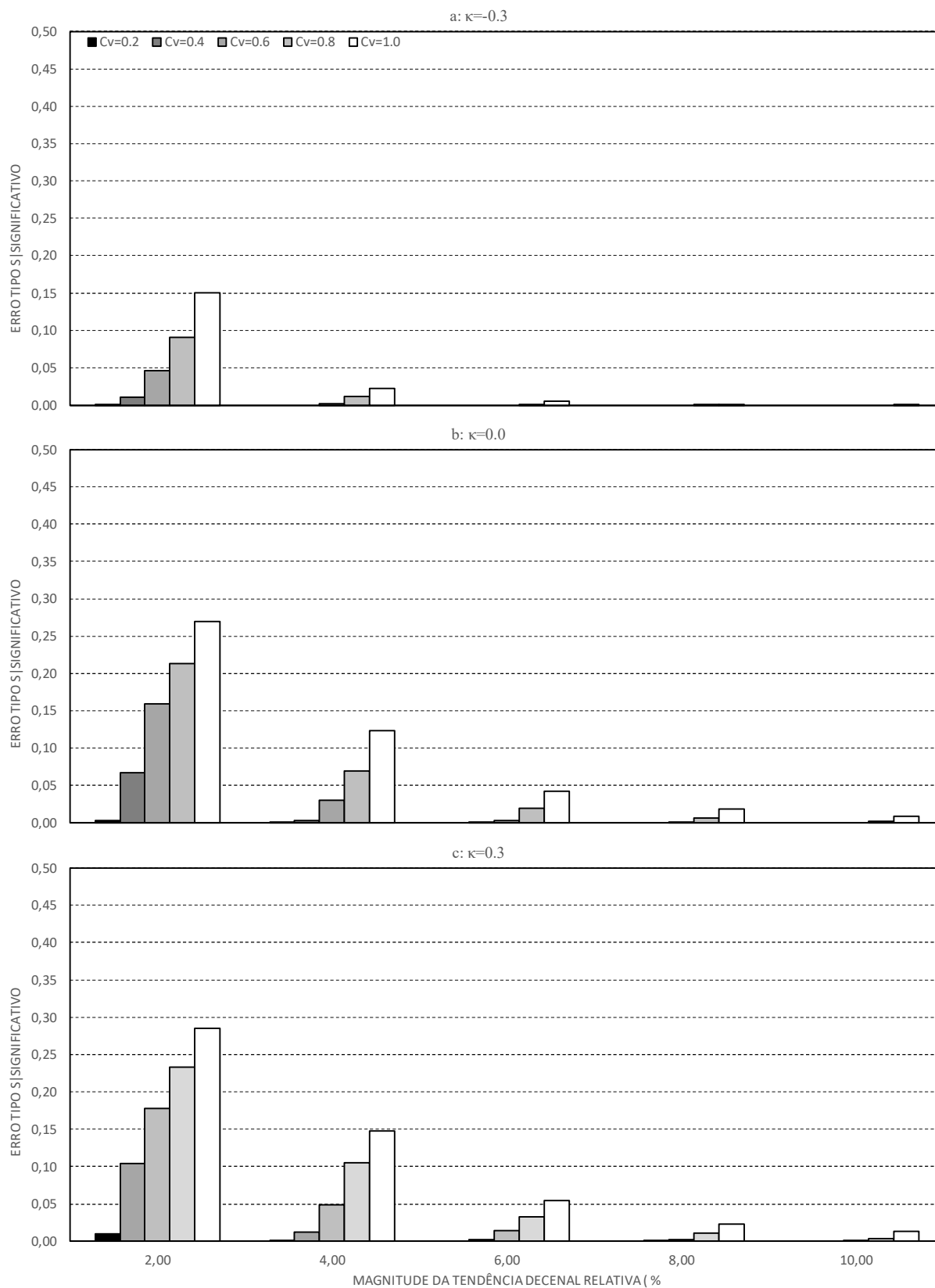


Figura 5.7 - Probabilidade do Erro Tipo S (sinal) dado que a estação foi declarada não-estacionária ($\alpha = 0.05$) para $n=50$ e diferentes valores de $Cv = [0.2 \ 0.4 \ 0.6 \ 0.8 \ 1.0]$ e $\kappa = -0.3$ (a), 0.0 (b) e $0.3.0$ (c)

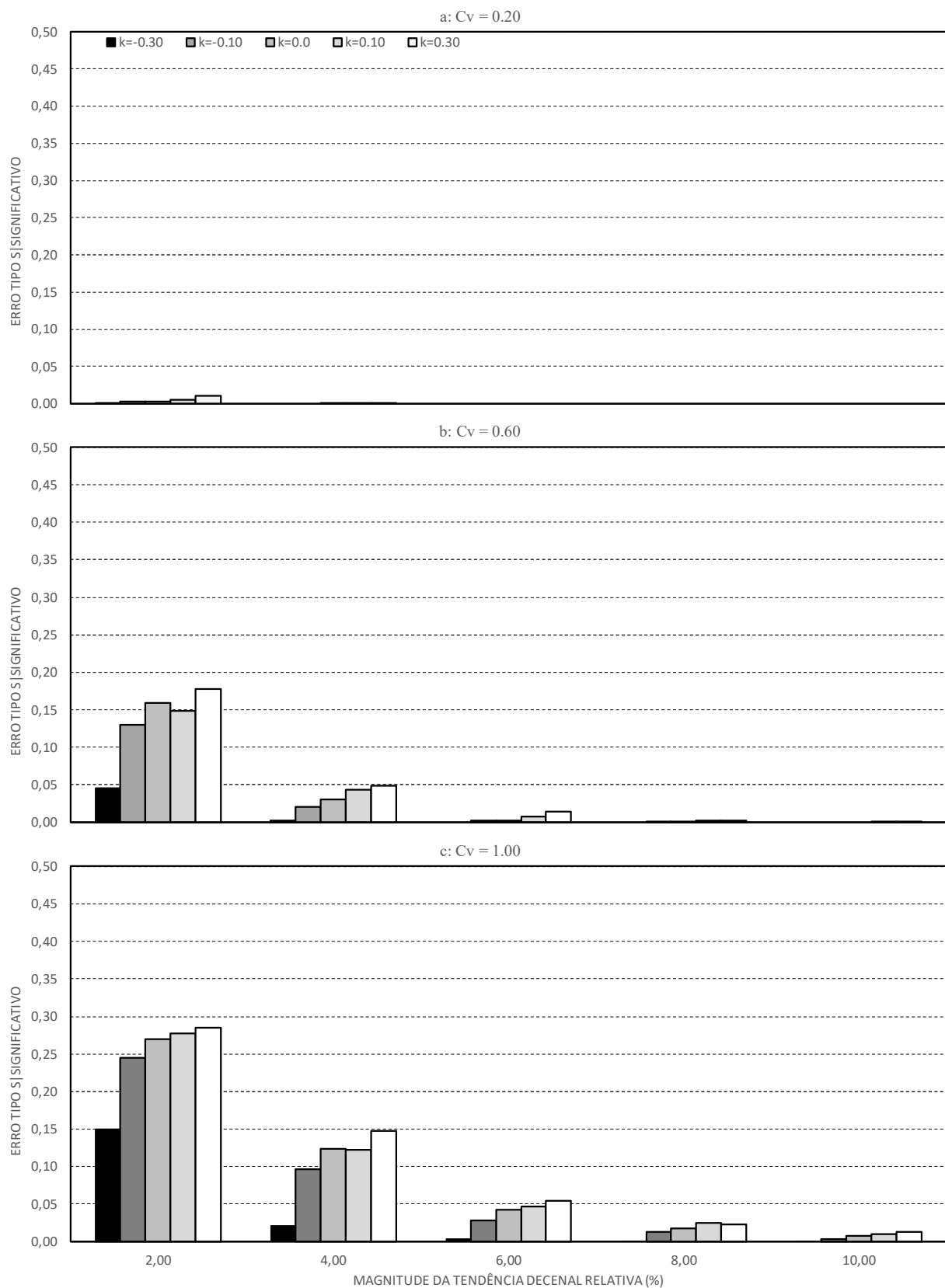


Figura 5.8 - Probabilidade do Erro Tipo S (sinal) dado que a estação foi declarada não-estacionária ($\alpha = 0.05$) para $n=50$ e diferentes valores de $\kappa = [-0.3 - 0.1 0 0.1 0.3]$ e $C_v = 0.2$ (a), 0.6 (b) e 1.0 (c)

Os resultados apresentados até aqui mostram que, em séries com alta variabilidade anual e com magnitude de tendência relativamente pequena – características que resultam em baixo poder de um teste estatístico (depende também do parâmetro de forma κ da GEV) –, a probabilidade de se obter uma estimativa da tendência com sinal trocado é relativamente alta para amostras com tamanhos geralmente constatados para as estações brasileiras, normalmente inferiores a 60 anos.

A Figura 5.9 mostra que, para poder do teste com valores inferiores a 0.15, a probabilidade do Erro Tipo S pode ser bastante elevada. Esse fato é um complicador em estudos de detecção, pois pode resultar, por exemplo, em situações em que duas estações próximas venham a ser identificadas como não-estacionárias, porém uma com tendência crescente e outra com tendência decrescente, levantando dúvidas sobre a análise realizada e causando dificuldades na interpretação dos resultados. Como será visto na sequência, esse problema pode ser exacerbado porque o estimador da magnitude da tendência tende a superestimar a real magnitude quando o poder do teste é baixo.

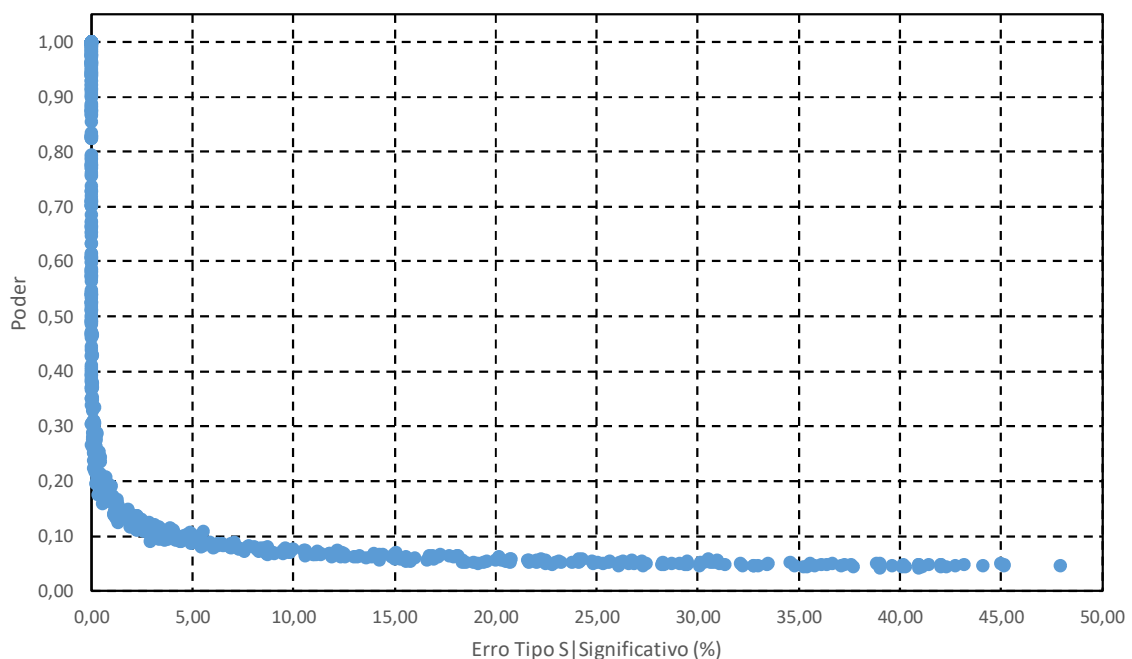


Figura 5.9 – Curva que expressa a relação entre o poder do teste e a probabilidade do Erro Tipo S | significativo (%)

Os resultados do estudo Monte Carlo permitiram avaliar também o grau de tendenciosidade e a variância do estimador B_SEN para diferentes características da série. A Figura 5.10 relaciona o poder do teste com a tendenciosidade relativa do estimador B_SEN quando a série é declarada não-estacionária pelo teste MK. Salienta-se que a tendenciosidade relativa é expressa pela razão entre o valor esperado do estimador e o real da magnitude, $\overline{B_SEN}/b$. Pode-se observar que o estimador tende a superestimar a magnitude real, especialmente em situações em que o poder do teste é baixo. Por exemplo, em casos em que o poder do teste é da ordem de 0.50, a estimativa da magnitude da tendência pode ser, em média, 40% maior do que a magnitude real, chegando à 100% (dobro) para valores de poder iguais a 0.20. Esclarece-se que distorções como essas podem comprometer significativamente a interpretação dos resultados e as decisões de gestão de recursos hídricos ou de dimensionamento de estruturas de engenharia que venham a ser baseadas em tais achados.

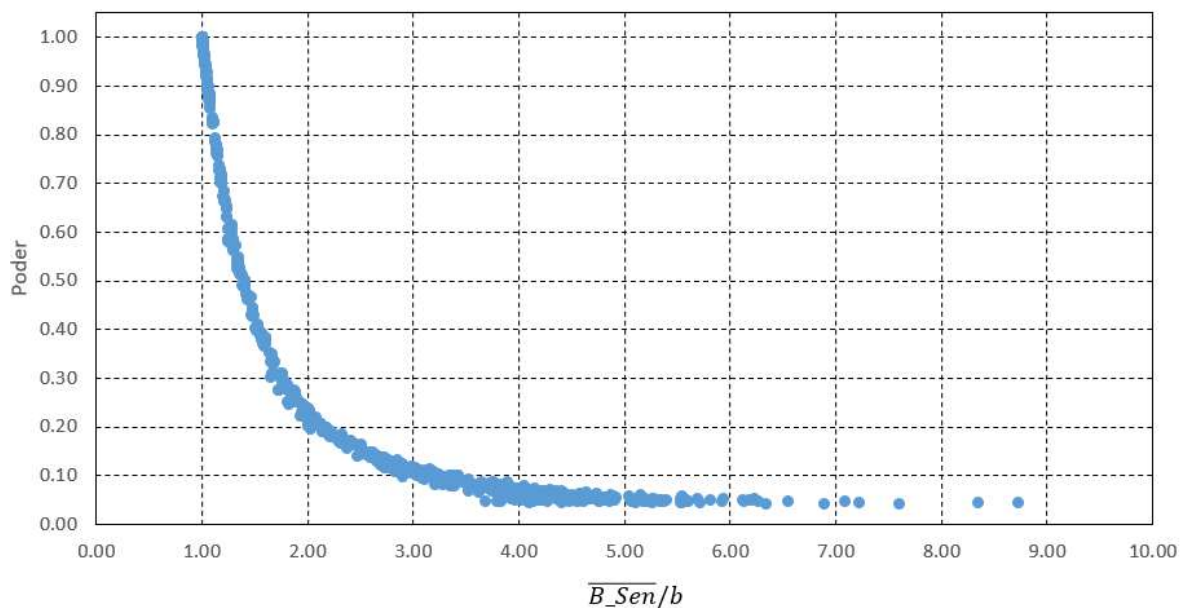


Figura 5.10 - Curva que expressa a relação entre o poder do teste e o valor médio estimado por B_SEN dividido por b ($\overline{B_SEN}/b$)

Além da tendenciosidade relativa do estimador, foi feita uma análise da variância do mesmo, utilizando para isso os percentis de 2.5% e 97.5% (IC) dos valores de B_SEN estimados para as 10,000 séries geradas para valores fixos de n , Cv e κ . Na verdade, os valores apresentados e discutidos a seguir consideram apenas os casos em que a estação foi declarada não-estacionária, utilizando-se como métrica de avaliação a denominada *Largura relativa do Erro Tipo M* | Significativas, que nada mais é do que a diferença entre os valores dos percentis

dividida pelo valor b para as séries significativas. Os resultados obtidos para essa avaliação são apresentados nas Figura 5.11 e Figura 5.12, em que se pode observar que, para algumas situações de baixo poder do teste, a Largura relativa do Erro Tipo M pode chegar a valores bastante elevados, 20, por exemplo.

Analisando a Figura 5.11, que apresenta a relação existente entre o tamanho amostral e a largura do intervalo de confiança obtida para um valor fixo de $\kappa = -0.30$ e $b = 0.006$ e diferentes magnitudes de coeficiente de variação ($Cv = [0.2 (0.2) 1.0]$), nota-se que, à medida que o número de anos de dados disponíveis para uma amostra se eleva, o intervalo de confiança se estreita. Já para o coeficiente de variação, observa-se um comportamento contrário: a elevação da sua magnitude resulta em alargamento do IC, ou seja, maior variabilidade nos valores estimados para a tendência pelo B_SEN. A influência de Cv sobre a largura de IC é mais pronunciada em amostras de tamanho reduzido ($n < 40$), visto que o aumento do coeficiente de variação, por exemplo, de 0.20 para 1.00, resulta em diferenças que chegam a quase 25 e 1,1 vez o valor de b simulado para, respectivamente, $n = 20$ e $n = 100$.

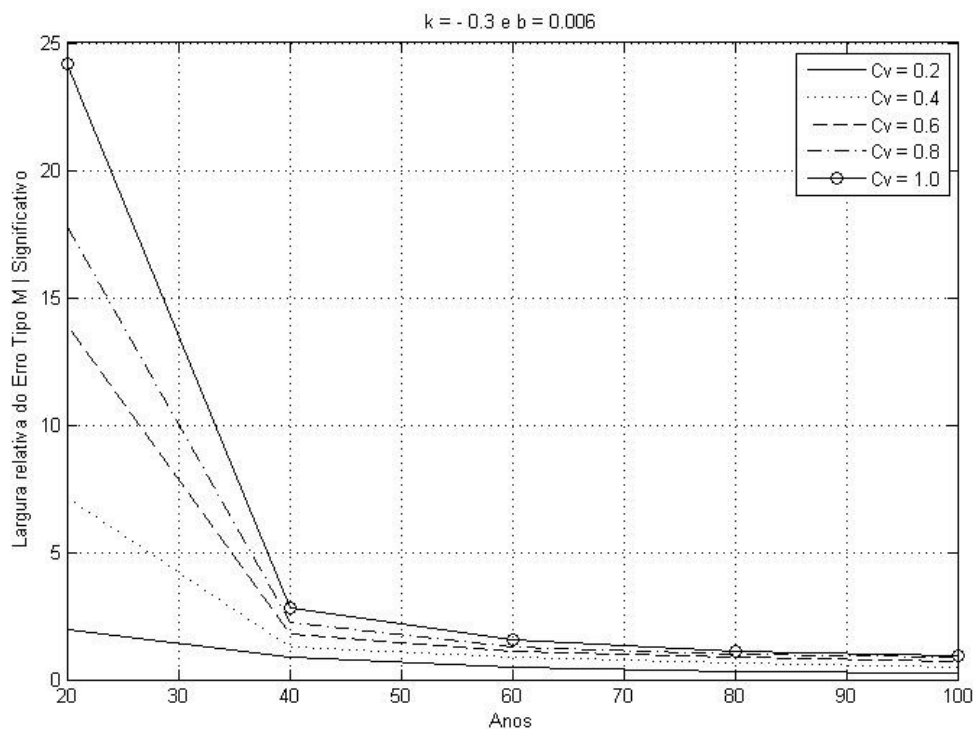


Figura 5.11 -Largura relativa do IC do Erro tipo M, dado que a estação foi declarada não-estacionária ($\alpha = 0.50$) para $\kappa = -0.3$, $b = 0.006$ e diferentes valores de $Cv = [0.2 (0.2) 1.0]$ e $n = [20 40 60 80 100]$.

A Figura 5.12 mostra a interdependência constatada entre o parâmetro de forma (κ) e a largura do IC da estimativa da magnitude da tendência por meio B_SEN, adotando-se um nível de confiança de 5%. Fixando-se os valores de $b = 0.006$ e do $C_v = 0.6$, construiu-se, para cada um dos κ avaliados ($[-0.3, -0.1, 0, +0.1, +0.3]$), curvas que expressam a relação entre as variáveis procuradas e plotadas nos pares de eixos coordenados. É interessante notar que, assim como ocorre para o C_v , o IC se estreita à medida que se aumenta o tamanho amostral, atingindo praticamente o valor esperado de b para séries com $n \geq 60$. Já o parâmetro de forma tem efeitos mais severos para amostras de tamanho reduzido ($n < 40$, para $k < -0.1$ e $n < 60$, para $k > 0$) e atua de maneira significativa na dispersão dos resultados de B_SEN obtidos, sendo, para alguns casos, superiores a 25 vezes o módulo de b .

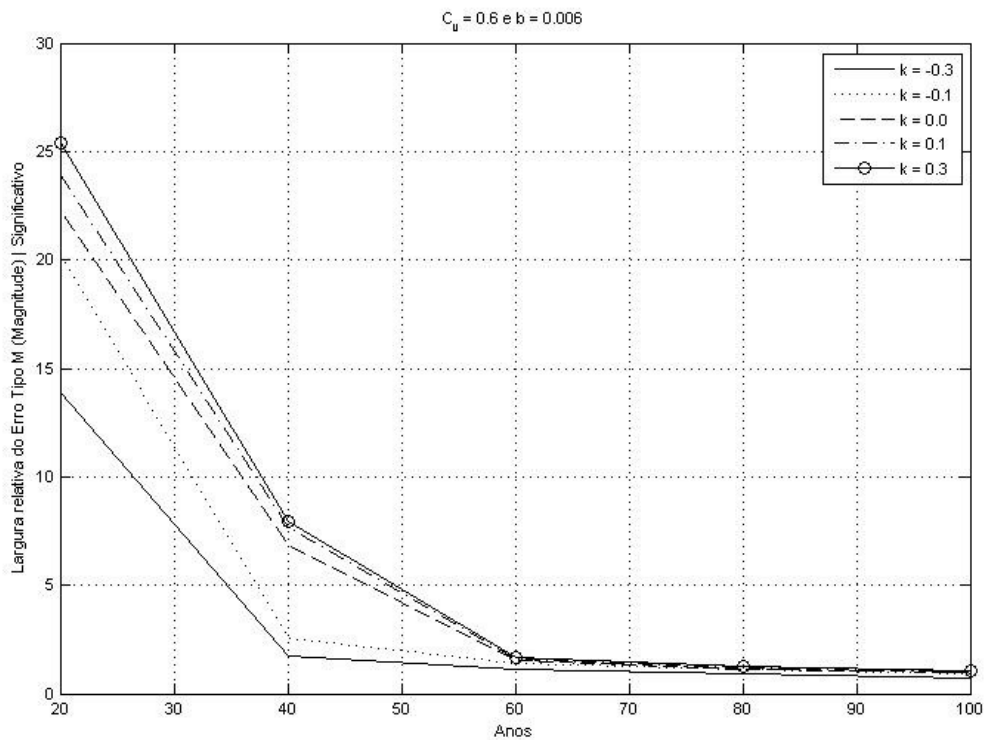


Figura 5.12 - Largura relativa do IC do Erro tipo M dado que a estação foi declarada não-estacionária ($\alpha = 0.50$) para $C_v = 0.6$, $b = 0.006$ e diferentes valores de $\kappa = [-0.3 -0.1 0 0.1 0.3]$ e $n = [20 40 60 80 100]$.

5.3. AVALIAÇÃO DE DESEMPENHO DAS TÉCNICAS QUE UTILIZAM O FDR COMO CONTROLE DE ERROS DO TIPO I – BH95 E BH_ADAP

No intuito de comparar o desempenho de testes tradicionais de detecção, que são baseados no controle da taxa de falsos positivos, do inglês *false positive rate (FPR)*, e aqueles que fazem uso da taxa de descobertas falsas (*false discovery rate - FDR*), foi proposta a geração de configurações distintas dos denominados campos sintéticos, a saber: (i) estações sem correlação espacial e sem tendência e (ii) com correlação espacial nula e com tendência.

Um campo sintético é composto por 179 estações fictícias que estão posicionadas nas mesmas coordenadas geográficas dos postos pluviométricos que compõem a base amostral utilizada no caso real. A cada uma dessas estações está associada uma série do tipo GEV, de média unitária, tamanho n , coeficiente de variação C_V , magnitude de tendência b e parâmetro de forma κ . O processo de geração desse campo segue os procedimentos descritos em 4.2.2, exceto pelo número vezes em que ele é repetido para cada uma das possíveis combinações de n , C_V , b e κ (2,000 vezes).

Cabe lembrar que três métricas de avaliação foram utilizadas nessa análise, sendo as mesmas assim definidas: (i) N_{TP} – expressa a razão entre o número de rejeições da hipótese nula de estacionariedade correta e o total de estações detentoras de tendência verdadeira, (ii) FDR – proporção entre o número de falsos positivos (n° erros do tipo I) e a totalidade de rejeições e (iii) $FNDR$ - resultante da divisão entre o número de falsos negativos (n° erros do tipo II) e a quantidade de hipótese aceitas pelas abordagens testadas.

Para evidenciar a relação existente entre o N_{TP} , FDR , $FNDR$ e o tamanho amostral, a Figura 5.13 exibe variação positiva na proporção de detecção correta de estações com tendência verdadeira (N_{TP}) à medida que os valores de n se elevam ($n = [40\ 60\ 80]$) e C_V (0.2), κ (-0,30), b (0.004) são mantidos fixos, com relativa estabilidade no valor esperado da proporção de falsa descoberta (do inglês, *false discovery proportion – FDP*) e redução no percentual de estações não-estacionárias não declaradas como tal. Esse comportamento se aplica a todas as metodologias de controle empregadas nessa avaliação: tradicional, BH95 e BH_Adap.

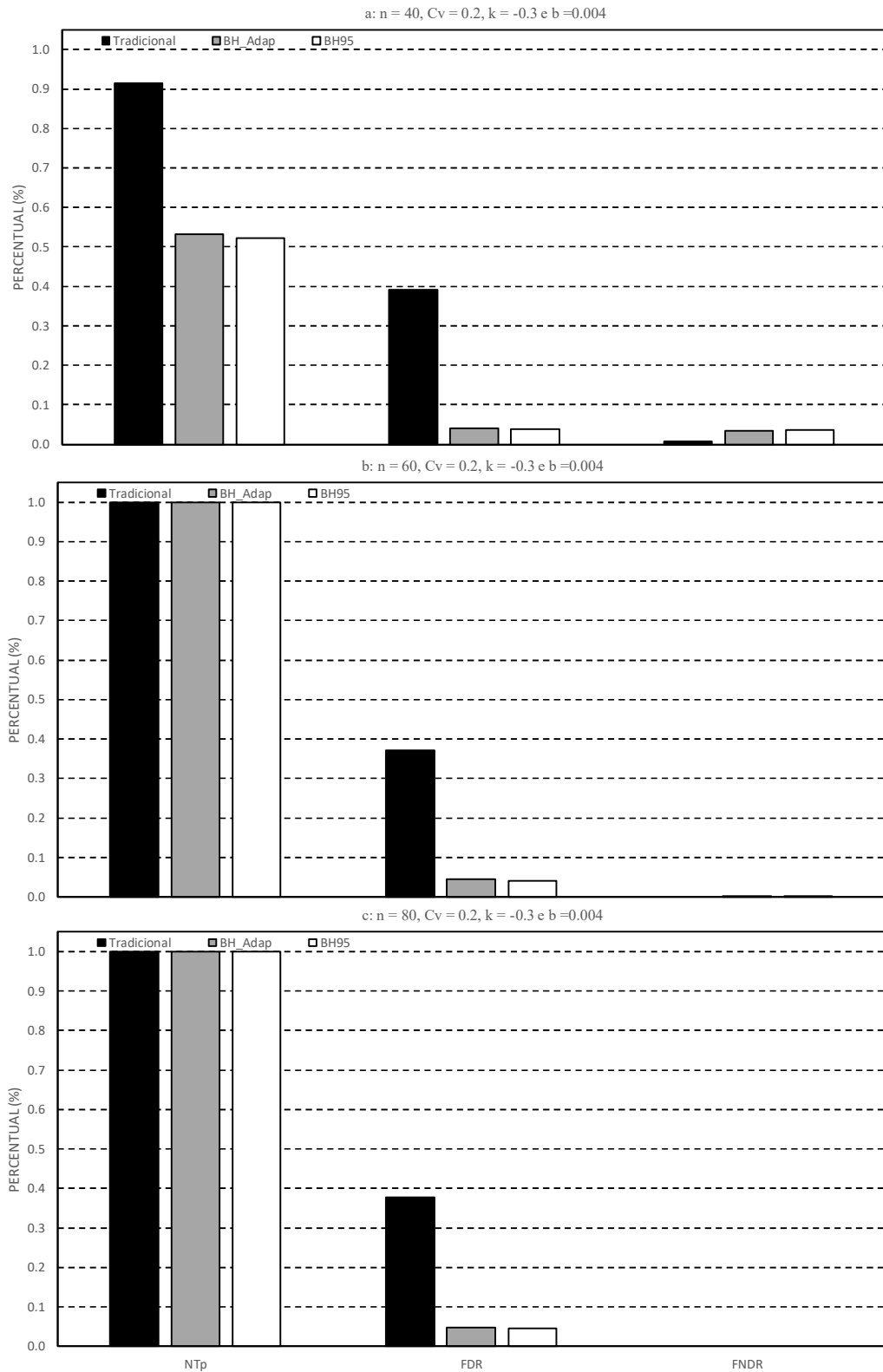


Figura 5.13 – Valores médios (%) de N_{TP} , FDR e $FNDR$ ao se aplicar as metodologias tradicionais, BH95 e BH_Adap à campos sintéticos cuja as séries possuem magnitude da tendência ($b = 0.002$), coeficiente de variação ($C_v = 0.2$), parâmetro de forma ($\kappa = -0.30$) e tamanhos amostrais ($n = 40, 60$ e 80)

Face aos resultados obtidos em seções anteriores, que indicam uma elevação de poder de detecção associado ao aumento do tamanho da amostra, a interdependência das variáveis mostrada na Figura 5.13 já era esperada. Entretanto, chama atenção a discrepância dos resultados obtidos para o N_{TP} e FDR entre as metodologias que utilizam como mecanismo de controle de erro o FPR e o FDR . A diferença observada chega a aproximadamente 45%, dado que os valores de N_{TP} são iguais a 0.90 e próximos a 0.55 para, respectivamente, o procedimento tradicional e BH95/BH_Adap, o que aponta para um cenário de maior poder de detecção daquela metodologia associada ao de maior percentual estimado. Isso se reflete em FDR , visto que a maior capacidade de detecção pode estar relacionada ao maior número de falsos positivos constatado para o dito tradicional. Quanto ao FNDR, nota-se uma aparente similaridade entre os percentuais obtidos para os três procedimentos testados, não sendo seu valor muito expressivo, já que o máximo, ocorrido para amostras de tamanho $n = 40$, é substancialmente inferior às demais métricas adotadas ($FNDR_{MAX} = 3.60\%$).

É sabido que um aumento no valor do coeficiente de variação amostral resulta em diminuição do poder de rejeição. Partindo-se dessa ideia é possível compreender a razão pela qual o percentual de rejeição correta (N_{TP}) se reduz ao se variar Cv de 0.40 (0.34) para 0.60 (0.11) e, posteriormente para 0.80 (0.08), mantidos n , κ e b constantes, conforme mostrada na Figura 5.14 para o procedimento tradicional. Em outras palavras, a relação entre o N_{TP} e o coeficiente de variação é decrescente, e o aumento em um deles implica na diminuição do outro. Destaca-se também a baixa capacidade de detecção do BH95 e BH_Adap, quando comparados ao controle tradicional, mesmo para situações de expressivo tamanho amostral ($n = 60$). Apenas para ilustrar, cita-se que, por exemplo, o BH95 e BH_Adap rejeitaram acertadamente, para $Cv = 0.40$, $b = 0.002$, $n = 60$ e $\kappa = -0.30$, apenas, em média, 1.53% do total de estações com tendência. Para a mesma configuração de campo, aponta-se que o procedimento tradicional obteve resultado bastante superior e igual a 33.8%.

A necessidade de uma metodologia que seja capaz de equilibrar poder e detecções equivocadas fica evidente na Figura 5.14. Isso porque, apesar do maior poder de detecção associado ao procedimento tradicional (33.8%), o seu percentual de falsos positivos declarado é igualmente relevante e, inclusive, superior àquele estimado para o N_{TP} (64,2%), e cresce à medida que os valores do coeficiente de variação também se elevam ($Cv = 0.60$

(81.4%) e $Cv = 0.80$ (84.4%). Já para aquelas que se baseiam no controle de erros pelo FDR , apesar da baixa capacidade de detecção, as taxas se mantêm inferiores à arbitrada (0.05) e sofrem flutuações mínimas com a oscilação de Cv . Quase imperceptíveis devido à escala adotada na Figura 5.14, as variações do $FNDR$ relacionam-se positivamente com o coeficiente de variação amostral, atingindo, portanto, seu valor máximo quando $Cv = 0.40$ (0.0725). As diferenças observadas entre os valores de $FNDR$ obtidos pelas abordagens tradicional, BH95 e BH_Adap, são negligenciáveis e não ultrapassam 1.00%.

Já a Figura 5.15 mostra a interdependência existente entre o parâmetro de forma da GEV (κ) e as métricas de avaliação adotadas. Assim, foram mantidos fixos o tamanho da amostral ($n = 60$), a magnitude da tendência ($b = 0.006$) e coeficiente de variação (0.40), enquanto a variável supracitada assumiu os valores -0.30, 0.00 e +0.30. Em uma primeira análise, nota-se o comportamento negativo de N_{TP} à medida que os valores de κ crescem, sendo mais pronunciado esse efeito para o BH95 e BH_Adap. Para ilustrar esse fato, observa-se que, ao se variar o κ de -0.30 para 0.0, a proporção de detecções corretas varia de 0.997, 0.948 e 0.951 para 0.578, 0.083 e 0.083, considerando-se, respectivamente, as abordagens tradicional, BH95 e BH_Adap. Seguindo o mesmo padrão de variação observado para o N_{TP} , o FDR se notabiliza por apresentar flutuações de magnitude menos abruptas para o BH95 e BH_Adap, sendo quase imperceptível na figura em questão. Entretanto, o mesmo não pode ser dito para o procedimento tradicional, já que a diferença na taxa de descobertas falsas varia entre 32 e 64%, ou seja, quase oito vezes a magnitude máxima visto nas demais abordagens testadas.

No caso do $FNDR$, a associação com κ é dita positiva. A incapacidade de detectar estações com tendências verdadeiras é próxima a zero quando o parâmetro de forma é igual a -0.30 e chega a quase 7.00% quando $\kappa = 0$, o que é um indicativo do seu poder, mesmo que em menor grau do que o observado para o N_{TP} . Apesar de diferentes, os valores de $FNDR$ obtidos para o procedimento tradicional e o BH95/BH_Adap podem ser apontados como similares, tendo em vista a reduzida diferença entre eles (inferior a 1.5%).

Acompanhando o crescimento do poder, no caso da Figura 5.16 sinalizado pelo aumento da magnitude da tendência ($b = [0.004 \ 0.006 \ 0.008]$), o N_{TP} se notabiliza por apresentar

valores de taxas de rejeição correta para o procedimento tradicional maiores do que as vistas para os demais testados. Isso fica evidente, por exemplo, no caso de $b = 0.004$, $n = 60$, $\kappa = -0.30$ e $Cv = 0.4$, em que as magnitudes de N_{TP} estimadas para BH95 (43%) e BH_Adap (44%) correspondem quase à metade daquela obtida para a abordagem tradicional. Por não exercer controle sobre o FDR, a proporção de número de erros do tipo I (falsos positivos) em relação à totalidade das estações rejeitadas estimada para o procedimento tradicional é tão menor quanto maior é o valor de b (poder) e é substancialmente diferente daquela obtida pelo uso do BH95 e BH_Adap, que mantém o valor do FDR inferior a q (0.05).

Seguindo o sentido oposto ao crescimento da magnitude de b , o $FNDR$ reduz-se a valores próximos de zero ao se elevar o módulo da tendência de 0.002 para 0.004, desde que mantidos as demais variáveis fixas e iguais $n = 60$, $\kappa = -0.30$ e $Cv = 0.4$, conforme mostra a Figura 5.16. O $FNDR$ permanece nulo para $b = 0.006$ e 0.008 e possui, em linhas gerais, o mesmo padrão comportamental observado para as demais variáveis analisadas, ou seja, a métrica em questão relaciona-se de maneira decrescente com a elevação do poder. Para o BH95 e BH_Adap, o $FNDR$, conforme previsto por Genovese e Wasserman (2004), esteve limitado superiormente ao valor obtido pela divisão entre a quantidade das hipóteses alternativas verdadeira e o número total de hipóteses testadas ($\sim 7.3\%$), não se importando com as magnitudes assumidas por n , Cv , b e κ .

Por fim, reforça-se que a geração dos campos sintéticos foi realizada adotando como premissa a inexistência de correlação espacial entre as estações ($\rho \neq 0$). Nesse caso, foi observada a existência de similaridade entre os valores de N_{TP} , FDR e $FNDR$ obtidos para o BH95 e BH_Adap independentemente da combinação entre n , Cv , κ e b . Apesar da diferença de poder entre essas metodologias relatada em seções anteriores, a simulação procedida mostrou, que para as condições escolhidas, as abordagens (b) e (c) apresentam desempenho similar.

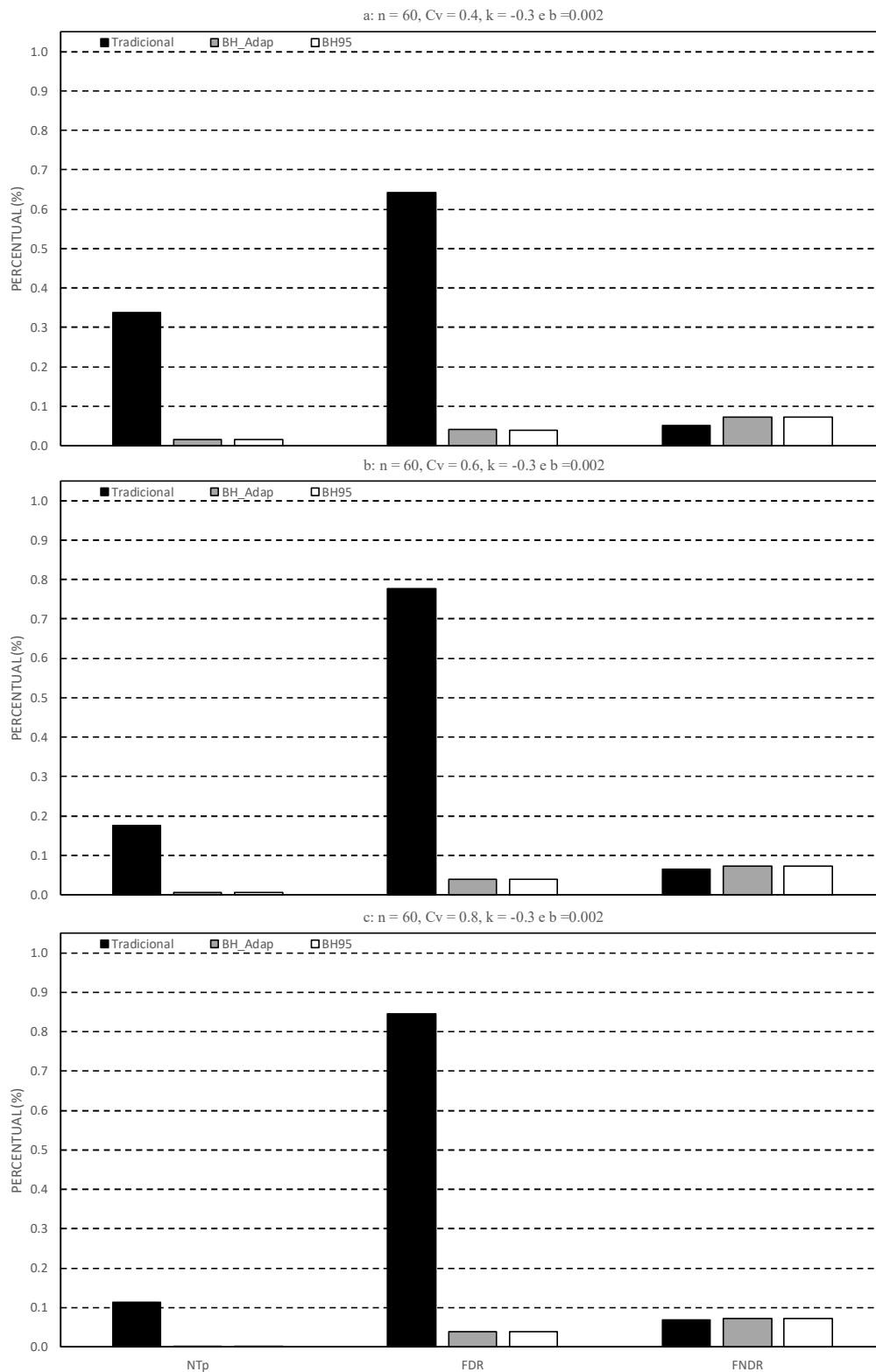


Figura 5.14 - Valores médios (%) de N_{TP} , FDR e $FNDR$ ao se aplicar as metodologias tradicionais, BH95 e BH_Adap à campos sintéticos cuja as séries possuem magnitude da tendência ($b = 0.002$), coeficiente de variação ($C_v = 0.4, 0.6, 0.8$), parâmetro de forma ($\kappa = -0.30$) e tamanhos amostrais ($n = 60$)

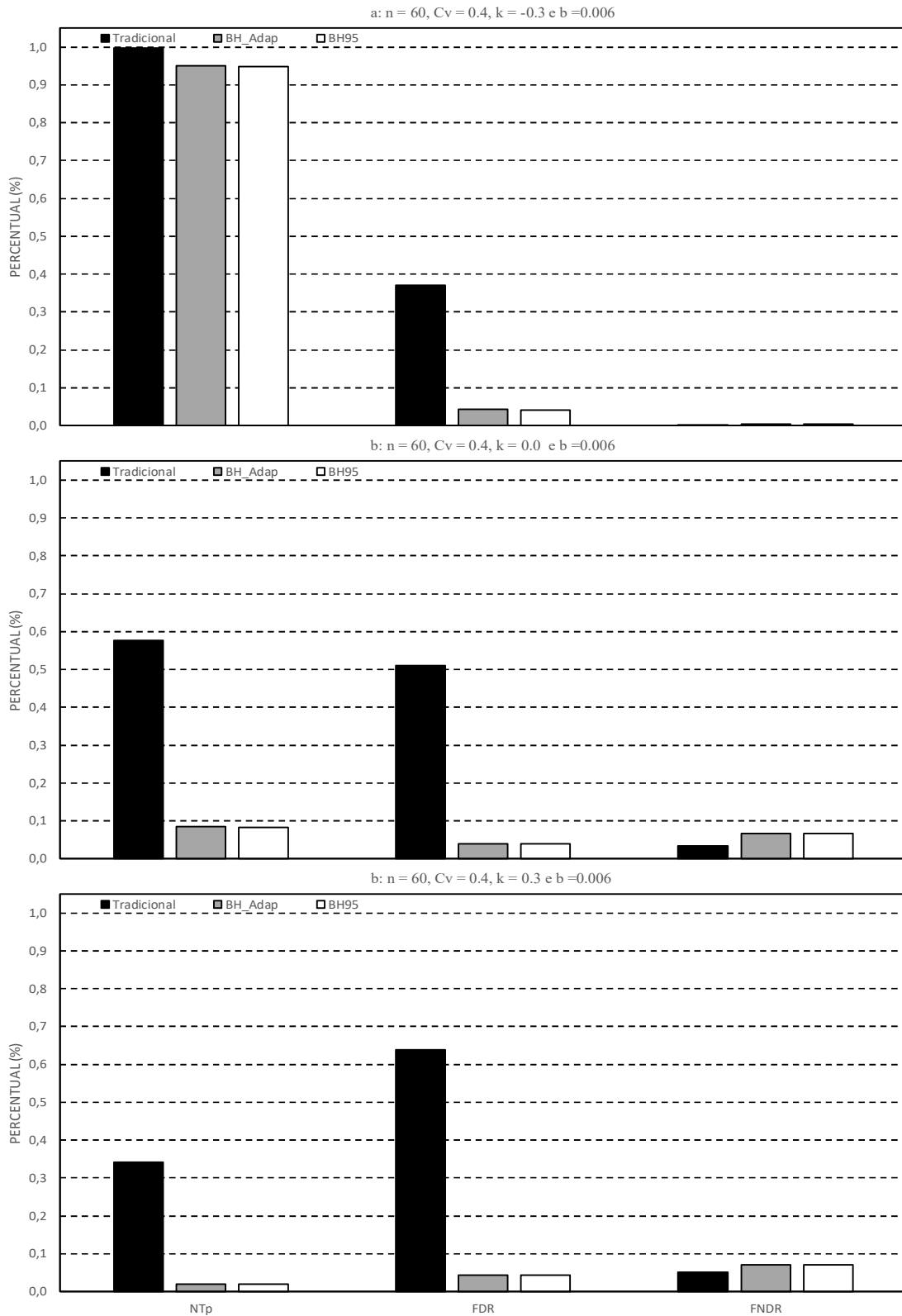


Figura 5.15 - Valores médios (%) de N_{TP} , FDR e $FNDR$ ao se aplicar as metodologias tradicionais, BH95 e BH_Adap à campos sintéticos cuja as séries possuem magnitude da tendência ($b = 0.006$), coeficiente de variação ($C_v = 0.4$), parâmetro de forma ($\kappa = -0.30, 0, +0.30$) e tamanhos amostrais ($n = 60$)

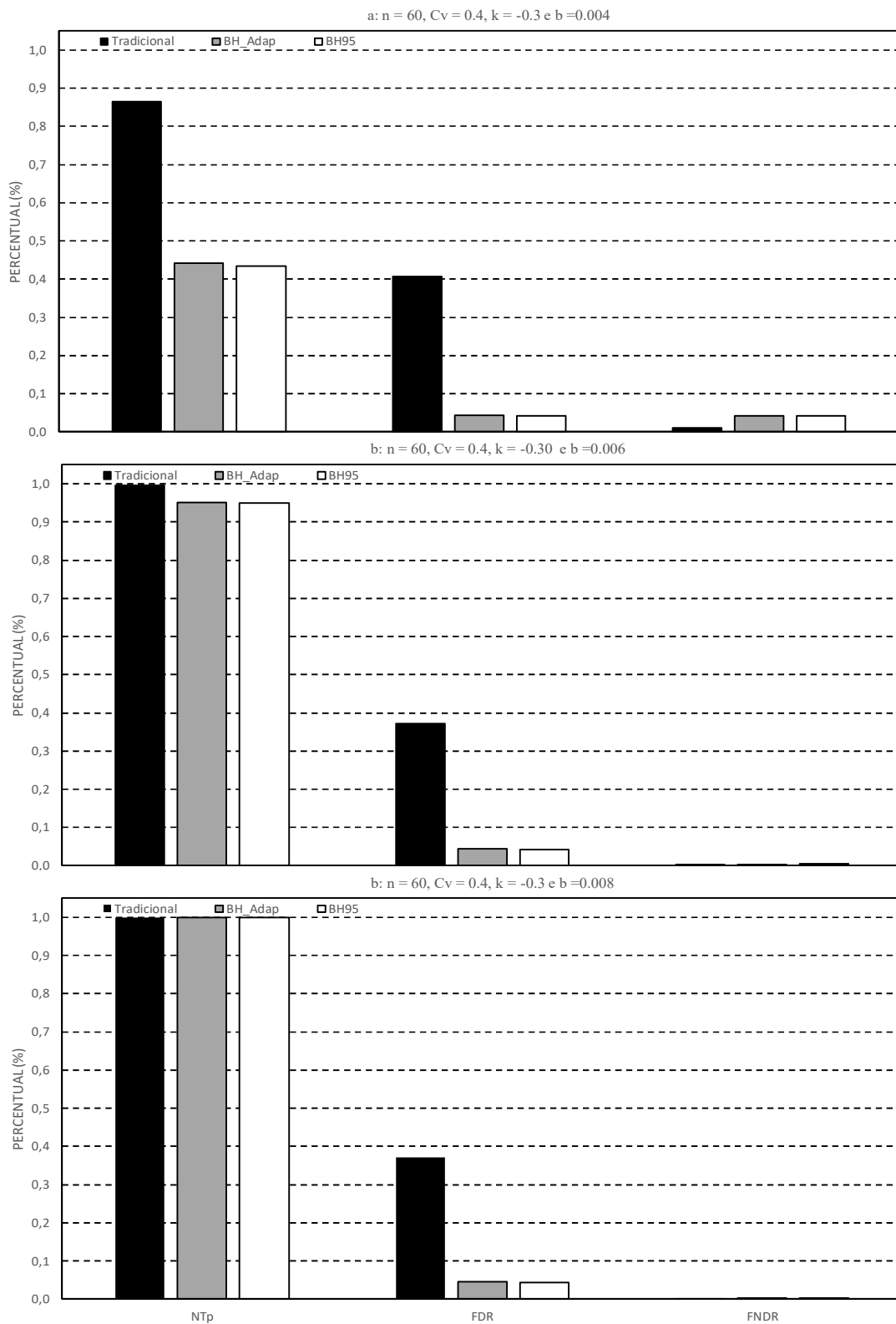


Figura 5.16 - Valores médios (%) de N_{TP} , FDR e $FNDR$ ao se aplicar as metodologias tradicionais, BH95 e BH_Adap à campos sintéticos cuja as séries possuem magnitude da tendência ($b = 0.004, 0.006, 0.008$), coeficiente de variação ($Cv = 0.4$), parâmetro de forma ($\kappa = -0.30$) e tamanhos amostrais ($n = 60$)

5.4. AVALIAÇÃO DE METODOLOGIAS DE DETECÇÃO DE TENDÊNCIAS EM SÉRIES AUTOCORRELACIONADAS

Visando exemplificar como a presença de autocorrelação afeta a análise de detecção de tendências e avaliar a existência de padrões regionais de mudança, crescentes ou decrescentes, na pluviometria e vazões das estações presentes na bacia do rio São Francisco, foram inicialmente adotadas 3 abordagens distintas para lidar com os efeitos da correlação serial estatisticamente significativa em uma dada base amostral: (a) simplesmente ignorar o problema da autocorrelação, (b) aplicar o PW e (c) o TFPW, ambos associados ao teste de Mann-Kendall.

Os resultados obtidos da aplicação dessas abordagens às séries das 179 estações pluviométricas e 100 fluviométricas selecionadas são apresentados nas Figura 5.17 e Figura 5.18, em que os pontos de monitoramento declarados como significativos foram ainda divididos em crescente (SC) e decrescente (SD), representados, respectivamente, pelas barras azuis e vermelhas.

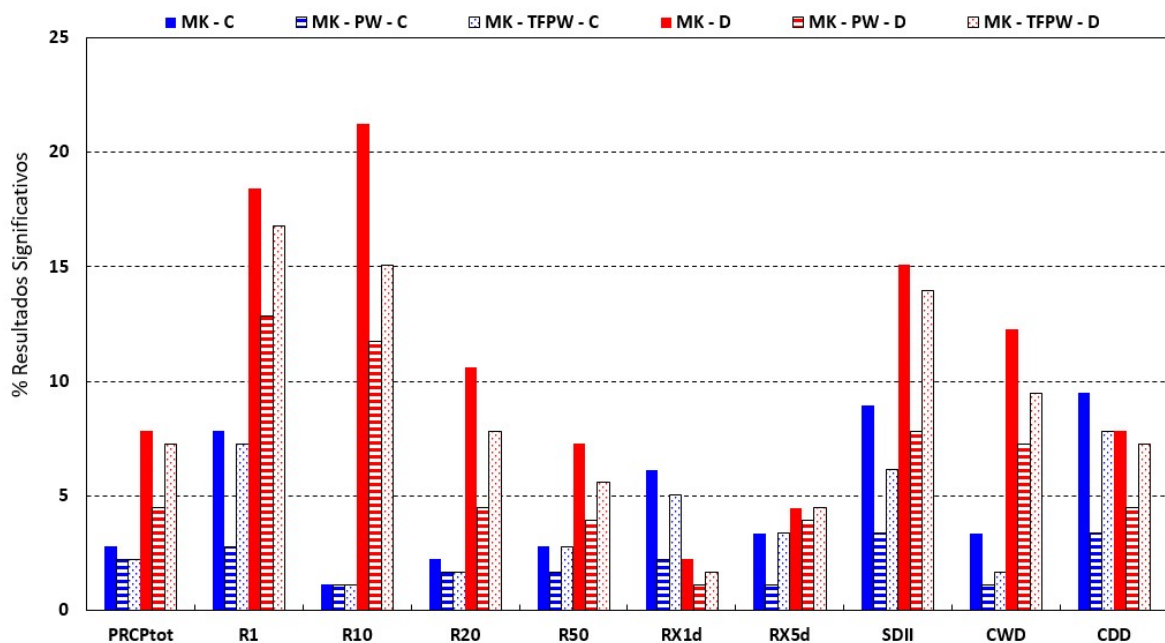


Figura 5.17 – Taxas de Rejeição das Abordagens (a: ignorar autocorrelação), (b: aplicar Pre_Whitening), (c: aplicar Trend-Free-Pre-Whitening) para os Índices de Precipitação

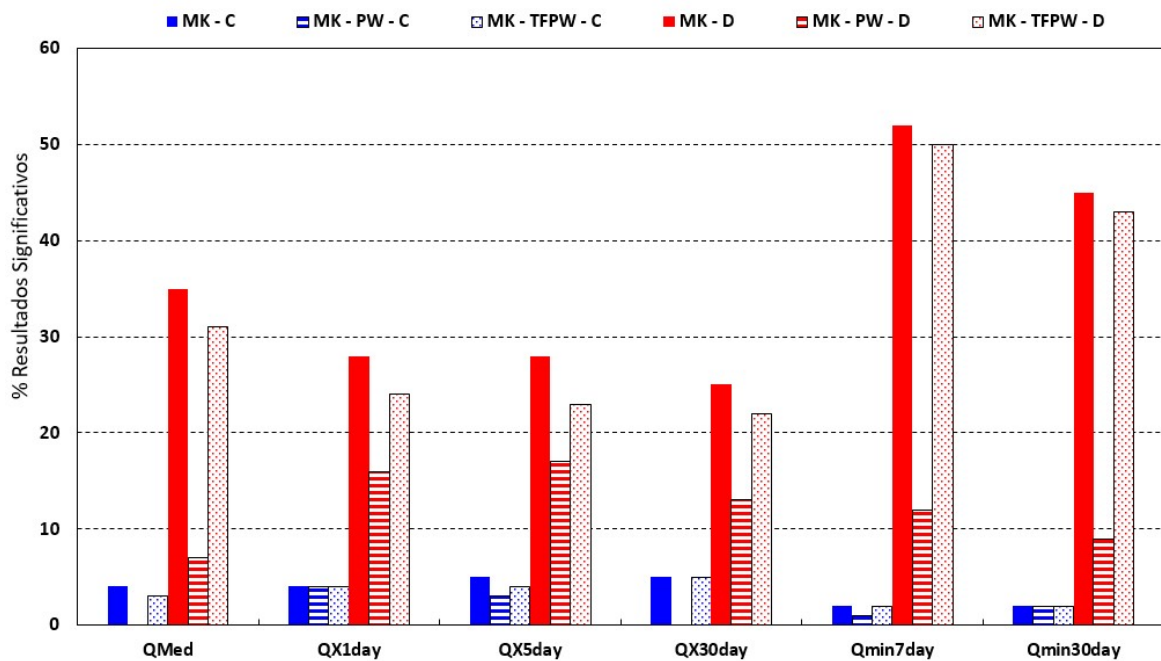


Figura 5.18 – Taxas de Rejeição das Abordagens (a: ignorar autocorrelação), (b: aplicar Pre-Whitening), (c: aplicar Trend-Free-Pre-Whitening) para os Índices de Vazão

Em linhas gerais, nota-se, dentre as séries cujos resultados foram significativos, a preponderância daquelas apontadas como decrescentes para as abordagens (a), (b) e (c). Isso ocorreu em 8 dos 10 índices extremos de precipitação analisados, excetuando-se o CDD e RX1d, e para todos os índices de vazão. Isso indica, em uma primeira análise, a possibilidade das condicionantes climáticas reinantes na localidade em estudo estarem contribuindo para uma condição de déficit hídrico na bacia do rio São Francisco, dado o decréscimo observado na grande maioria dos índices de precipitação e vazão, mesmo após a aplicação do MK-PW (b) e MK-TFPW (c).

Analisando as Figura 5.17 e Figura 5.18, é possível perceber que a proporção das estações declaradas como estatisticamente significativas é substancialmente inferior àquelas ditas sem tendência, no caso dos índices relacionados à pluviometria. Diz-se isso ao se constatar que, para o R1, índice com maior quantidade de estações avaliadas como não-estacionárias (26% – 47 estações), a proporção de pontos de monitoramento sem evidências de mudança é quase quatro vezes superior à estimada para aquelas declaradas como detentoras de tendência significativa (74%, 132 estações). A baixa proporção de estações em mudança fica ainda mais evidente quando se observa que sete dentre os dez índices extremos de pluviometria testados

possuem a soma das proporções de SD e SC inferiores a 20% (36 estações). Além de R1, constituem-se em exceção SDII e R10, que possuem, respectivamente, 22% (40 estações) e 24% (43 estações) como taxas de rejeição da hipótese nula.

O percentual de estações dotadas de tendência estatisticamente significantes para os índices extremos de precipitação, verificado para a abordagem (a), conforme esperado, se reduz quando a análise é feita sob a perspectiva de (b) e (c). O MK-PW e MK-TFPW se caracterizaram, respectivamente, pela redução mais pronunciada e não tão significativa das taxas de rejeição da hipótese nula para o caso em tela. Essa diferença no comportamento se deve à remoção de parte da tendência amostral juntamente com a tentativa de eliminação da correlação serial pelo MK-PW, o que, segundo Yue e Wang (2002), causa redução no número de detecções e possivelmente explica as discrepâncias entre as abordagens (b) e (c) constadas. Ademais, a fraca dependência temporal usualmente observada em extremos de pluviometria anuais (em média, o coeficiente de correlação lag-1 foi igual a 0,17) podem ter contribuído para a pequena diferença observada entre os valores de rejeição estimados para (a) e (c). Como exemplo do relatado, citam-se o SDII e o R1, que após a aplicação do MK-PW, tiveram a proporção de estações rejeitadas reduzida de, respectivamente, 22%(40 estações) para 16%(28 estações) e 26% (47 estações) para 16%(28 estações). Já no caso do MK-TFPW, as referidas taxas diminuem em menor grau em relação ao que ocorre para a abordagem (b), para 20% (36 estações) e 24% (43 estações).

O problema da autocorrelação na atividade de detecção fica mais evidente ao se constatar que os índices relacionados a pluviometria, que tiveram a maior redução absoluta e no percentual do número de estações apontadas como significativas após a aplicação do PW, foram também aqueles que possuem, em média, o maior coeficiente de correlação lag-1 (ρ) dentre aqueles testados, SDII (0.28) e R1 (0.34). Diante disso, os resultados aparentam ratificar a impossibilidade de uso indiscriminado do MK-PW relatado por Bayazit e Onoz (2007), sendo preferível o uso do MK-TFPW, principalmente, para situações de maior poder do teste: baixo coeficiente de variação, alta magnitude da tendência (Ver Figura 5.19) e grande tamanho amostral.

Mesmo para o caso em que a autocorrelação foi, em média, a mais baixa dentre todos os índices de precipitação testados, RX5d (0.11), o uso do MK-PW resultou comparativamente em diferenças nos resultados de significância. Nesse caso, o desempenho esperado para a abordagem (b) deveria ser o mais próximo possível daquele obtido para o MK-TFPW. Entretanto, por se tratar de média, a autocorrelação estimada para as séries variou dentro de um intervalo de valores (-0.27 – 0.32), o que, por fim, acabou por influenciar no número de estações apontadas como não-estacionárias pelas abordagens (b) e (c)

Na Figura 5.18, observa-se que as taxas de rejeição da hipótese nula associadas aos índices extremos de vazão são percentualmente superiores às estimadas para os de pluviometria. Nota-se, inclusive, que o percentual máximo de rejeição obtido para os índices ligados à chuva, R1 (26%), é inferior ao mínimo daqueles estimados para Qméd (39%), Qmáx (32%), Q7min (54%), Q30min (47%), QX5d (33%) e QX30d (30%). Conforme já mencionado, os índices extremos de vazão, majoritariamente negativos, se notabilizam pela baixa proporção de estações declaradas como crescentes, sejam elas significativas ou não, correspondendo a no máximo 5% (QX5day e QX30day) do total de rejeições.

No caso dos índices extremos relacionados à vazão, a diminuição no número de estações apontadas como significativas, após a aplicação do MK-PW, foi mais pronunciada do que a vista para aqueles associados à precipitação. Isso porque, por exemplo, a taxa de rejeição das estações declaradas como significativas decrescentes prevista na abordagem (a) para Q7min (52%) e Qméd (35%) diminuiu para, respectivamente, 12 e 7% quando o MK-PW foi aplicado. Comparativamente, nota-se que a referida diminuição é maior para Q7min (-40%) do que para Qméd (-28%), que por sua vez é superior a Qmáx (-12%). Esse aumento na diferença entre o pré e pós PW à medida que o valor da vazão se eleva possivelmente está associado à menor dependência temporal existente na série de máximos, sendo o efeito de sua remoção menos sentido na redução das taxas de rejeição do que para as médias e mínimas ($\rho_{(Q7min)} = 0.62 > \rho_{(Qmed)} = 0.55 > \rho_{(Qmax)} = 0.31$). Já para o MK-TFPW, a variação anteriormente relatada é mais sutil, não ultrapassando os 5% e 1% de variação para as estações ditas significativas, respectivamente, decrescentes e crescentes.

A diferença na redução das taxas de rejeições obtidas após o uso do MK-PW para os índices extremos de pluviometria e vazão parece realmente ser influenciada pela presença de correlação temporal. Via de regra, o maior valor, em média, de ρ (0.46) somado a maior magnitude da tendência estimada para as séries anuais de vazão conduzem a diminuições mais significativas no percentual de rejeição da hipótese nula. Esse efeito é menos pronunciado ao se utilizar o MK-TFPW, que reduz o valor de ρ estimado (0.28), ao prever a remoção da tendência amostral antes do cálculo do estimador do coeficiente de correlação amostral. A influência da presença de tendência no cálculo de ρ evidenciada pelos resultados citados já havia sido relatada por Douglas *et. al.* (2000): a existência de não-estacionariedade serial estatisticamente significativa modifica a magnitude da correlação serial estimada.

Assim, o emprego da abordagem de MK-PW para lidar com os efeitos da autocorrelação no processo de detecção de tendências parece limitado. Os resultados obtidos ao utilizá-la mostram que mesmo para situações de baixa correlação amostral, por exemplo o RX5d, a metodologia em questão não é capaz de manter as taxas de rejeição próximas as obtidas pelo MK-TFPW para um mesmo nível de correlação temporal. Reportado por Yue *et al.* (2002), Douglas *et. al.* (2000), Bayazit e Onoz (2007) e mostrado como suscetível a presença de tendências estatisticamente significantes na amostra, o MK-PW será negligenciado nas análises que serão realizadas na sequência deste trabalho. Esse comprometimento de desempenho relatado ainda é agravado pelo tamanho reduzido da maioria das séries testadas (< 30 anos), em que a superestimativa da magnitude da tendência, erro tipo S, poderá afetar mais significativamente o MK-PW do que o MK-TFPW.

Em linhas gerais, o panorama de incapacidade dos resultados de significância estatística obtidos para os índices de precipitação em explicar (ou pouco explicar) a mudança decrescente detectada para aqueles de vazão permanece, ao se analisar a magnitude das tendências relativas decenais previstas pelo estimador proposto por Sen (1968) para as séries de índices analisadas nas Figura 5.19 e Figura 5.20. Em outras palavras, situações como a incompatibilidade entre RX1d, RX5d, R10, R20 e R50, índices de chuva que estão relacionadas à eventos de cheia, e o Q_{máx}, ilustram a situação anteriormente relatada, dada a discrepância observada entre o percentual de rejeições da hipótese nula observado para os índices relacionados à chuva, máximo 24% (R10), e vazão máxima (32%).

As magnitudes de tendência decenais estimadas para as séries de índices de pluviometria e vazão foram subdivididas em três (3) faixas de valores, para que pudessem expressar o grau de relevância da mudança detectada e a necessidade de atenção dos gestores frente a um cenário de mudança, a saber: (i) 0 a 5%, (ii) 5 a 10% e (iii) >10%, que foram tratados aqui como, respectivamente, *sem importância*, *atenção* e *preocupante*, muito embora essa categorização não seja definitiva, podendo e devendo ser melhor discutida com aqueles responsáveis pela gestão dos recursos hídricos. Ainda quanto às Figura 5.19 e Figura 5.20, a magnitude da tendência foi estimada pelo $B_{SEN_Relativo}$, que nada mais é do uma adaptação do estimador proposto por Sen (1968) para expressar a mudança em termos decenais e médios:

$$B_{SEN_{Relativo}}(\%) = [(b \times 10/\bar{X}) \times 100] \quad (5.1)$$

Com o interesse naquelas faixas de maiores valores de magnitude de tendência, e, conseqüentemente, de potenciais mudanças verdadeiras mais significativas, o presente estudo foca em índices extremos cuja proporção de estações apontadas como não estacionárias seja relevante, ou seja, que expressiva parte daquelas declaradas como detentoras de tendência esteja concentrada nas faixas (ii) 5 – 10% e (iii) > 10%. A adoção desse critério se justifica pelo objetivo de buscar padrões regionais de mudança que sejam abrangentes e representativos, não sendo o foco deste trabalho indicar mudanças, mesmo que fortes, em uma única estação, por exemplo.

Assim, a Figura 5.19 mostra os percentuais de estações pluviométricas declaradas como significativas após a aplicação do MK-TFPW, subdivididas, de acordo com o $B_{SEN_Relativo}$ estimado, entre as três faixas de valores de magnitude de mudança decenais supracitadas. Os índices R1, R10 e SDII, que apresentaram o maior percentual de rejeição dentre os dez índices de pluviometria analisados, se notabilizam pelo predomínio de estações significativas de baixa magnitude de tendência ($R1_{|0-5\%|} = 11,20\%$, $R10_{|0-5\%|} = 7,30\%$ e $SDII_{|0-5\%|} = 10,10\%$), ou seja, para R1, R10 e SDII, os percentuais de estações significativas que se enquadram na faixa de tendências estimadas entre 0 e 5% são, respectivamente, 46,6%, 44,8% e 50,0%. Em contrapartida, os referidos índices estão também entre aqueles que possuem os maiores percentuais de rejeição agregados entre os intervalos 5 – 10% e >10%, respectivamente, 12,90%, 8,90% e 10,00%.

Diferentemente desses três índices, o RX1d e o RX5d possuem baixos percentuais de rejeição total, respectivamente, 6,70 e 5,60%, e apenas 2,2%(4) e 1,12%(2) das estações que foram declaradas como detentoras de tendência significativa se enquadram na faixa de mudanças decenais superiores a 10%. Em relação aos demais índices de pluviometria, PRCPTot, R20, R50, CWD e CDD, apesar de alguns deles indicarem a existência de mudanças que merecem atenção e são preocupantes na bacia do São Francisco ($CDD_{>5\%}$ (11,7%), $CWD_{>5\%}$ (9,50%)), entende-se que as alterações constatadas são de caráter localizado, dada a baixa proporção de estações rejeitadas (inferiores a 20%) para os mesmos.

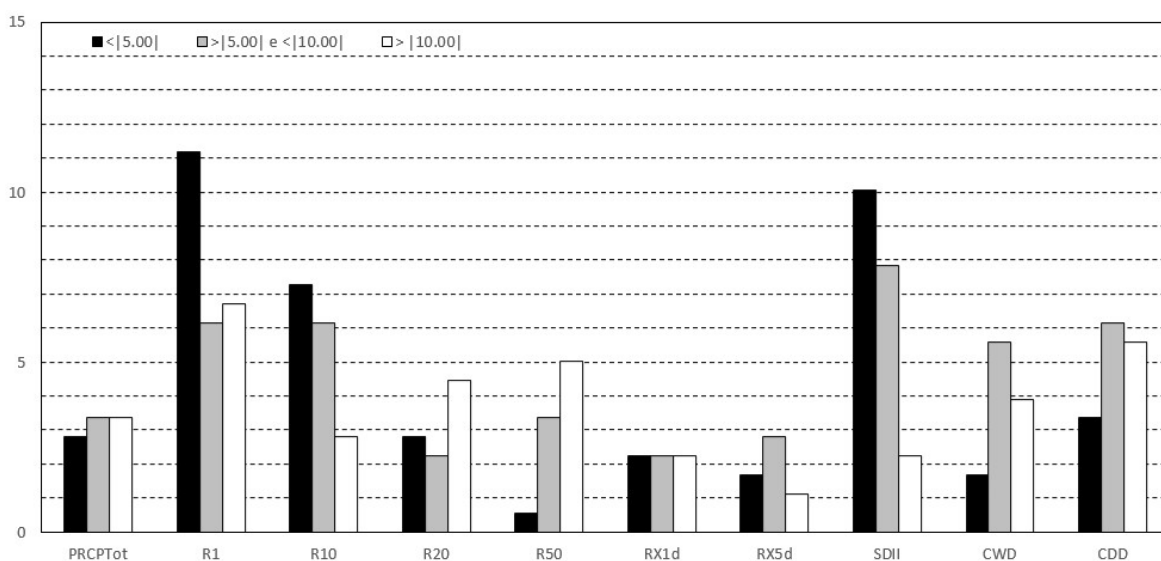


Figura 5.19 – Percentuais de estações pluviométricas de significância estatística declarada pelo MK-TFPW subdivididos em faixas de magnitude de tendência expressas pelo B_SEN_relativo (a): 0 – 5%, (b) 5 – 10% e (c) >10%.

Mostrando as mesmas informações apresentadas na Figura 5.19, só que para os seis (6) índices extremos associados à vazão, a Figura 5.20 evidencia que os maiores percentuais de rejeição da hipótese nula estão associados às estações cuja magnitude da mudança é superior a 10%. Esse comportamento é verificado para todos os índices de vazão testados e é mais pronunciado para as séries de mínimos, Q7min (31%) e Q30min (26%), do que para as máximas (Qmáx =18%) e médias, por exemplo. Chama a atenção que, diferentemente do constatado para os índices de pluviometria, a proporção de rejeitadas para o grupo de tendências entre 0 e 5% é a menor dentre os três utilizados, corroborando para a hipótese de que os resultados obtidos para os índices de chuva não são suficientes para explicar a totalidade das alterações no padrão de decréscimos da vazão, o que foi inicialmente

evidenciado pelo descompasso entre os números de estações apontados como detentoras de tendência para os índices extremos de chuva e vazão.

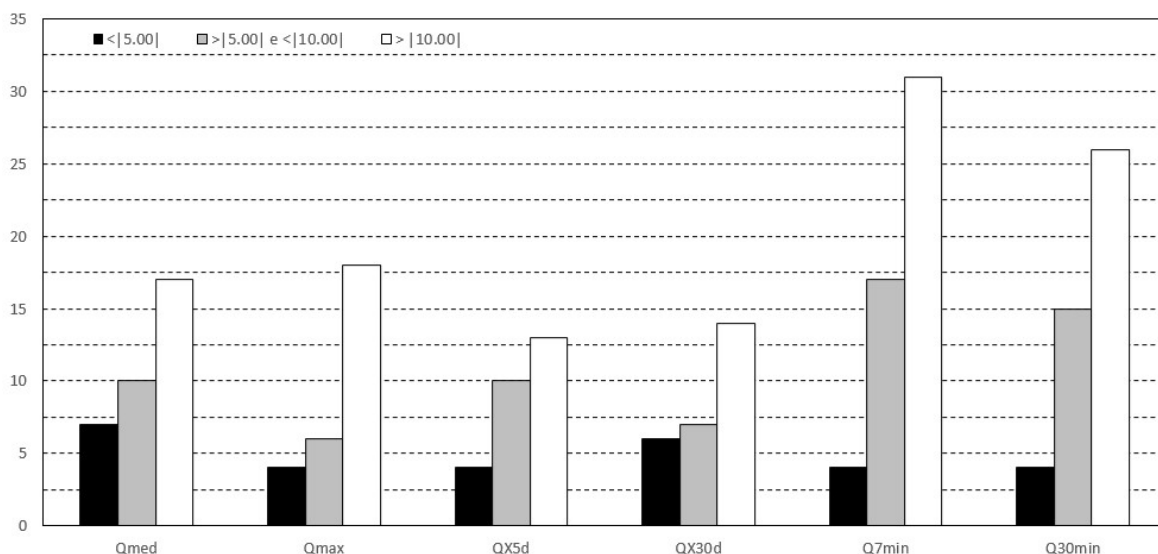


Figura 5.20 - Percentuais de estações pluviométricas de significância estatística declarada pelo MK-TFPW subdivididos em faixas de magnitude de tendência expressas pelo B_SEN_Relativo (a): 0 – 5%, (b) 5 – 10% e (c) >10%.

Tão importante quanto avaliar as mudanças ocorridas nas séries temporais dos índices extremos sob a perspectiva dos percentuais de rejeição e magnitudes das tendências até aqui realizada seja talvez analisá-las levando em consideração o tamanho dos registros históricos disponíveis das estações, os p -valores, além, é claro, da distribuição espacial das mesmas. A preocupação com o tamanho amostral se justifica pela influência que esse fator exerce no poder do teste de MK, bem como nos erros de sentido (tipo S) e estimativa da magnitude de mudança (tipo M) cometidos pelo B_SEN, o que já foi demonstrado em seções predecessoras deste trabalho.

Em termos práticos, a número de anos integrantes de um registro histórico se reflete na assertividade de declaração de tendência, que é crescente à medida que se eleva o tamanho amostral, e na percepção mais fidedigna dos efeitos da mudança (sentido e intensidade). Em outras palavras, as séries mais extensas consistem um bom panorama inicial para as análises de tendência regional, sendo a sua única desvantagem, no caso da área de estudo, o baixo quantitativo de estações cujo tamanho amostral é igual ou superior a 60 anos, o que levou à necessidade de utilizar séries menos extensas para ampliar a cobertura espacial da bacia do São Francisco.

Foram elaboradas figuras que procuram condensar os (3) três aspectos anteriormente citados: tamanho amostral, p_valores e distribuição espacial, permitindo que os padrões regionais de tendência possam ser visualizados com maior clareza. Compostas por quatro (4) gráficos em sua porção superior, cada um deles associado a uma das seguintes faixas de tamanhos serial: [30 39[, [40 49[, [50 60[e [60 +∞], as figuras contam ainda com mapa de toda a bacia, divididas em Alto, Médio, Médio-Baixo e Baixo São Francisco, contendo a marcação das estações apontadas como significativas decrescentes (vermelho) e crescentes (azul) pelo MK-TFPW. Quanto à simbologia adotada para o mapa, esclarece-se que ‘.’, ‘+’, ‘*’ e ‘x’ foram utilizados para expressar as faixas de tamanhos serial a que determinada estação pertence, sendo a menor expressa por “.” e a maior, por “x”. Adicionalmente, o símbolo adotado para representar uma estação qualquer será tão maior quanto mais próxima da faixa >10% estiver a tendência decenal estimada. Uma simbologia semelhante também foi utilizada para representar o tamanho amostral nos referidos gráficos, em que suas cores representam a sub-região de localização do ponto de monitoramento, a saber: vermelha (Alto), preta (Médio), verde (Médio-Baixo) e azul (Baixo).

A opção por apresentar os mapas (a), (b) e (c) na porção inferior das referidas figuras reflete-se na necessidade de comparação de resultados entre as metodologias tradicional, BH95 e BH_Adap. Ao colocá-las em uma mesma figura, a intenção é de possibilitar a visualização imediata das diferenças espaciais nas rejeições entre as três técnicas. No entanto, nesta seção, o foco restringe-se apenas a (a), ou seja, MK-TFPW, bem como o gráfico apresentados nas Figura 5.21 a Figura 5.23.

Na Figura 5.21 é possível perceber uma dificuldade de definição de um padrão de tendência regional para o número de dias de chuva (R1)no Alto São Francisco (ASF), dada a inexistência de zonas homogêneas de tendência única. A região do ASF é marcada pela coexistência entre estações de diferentes tamanhos amostrais e tendências que se alternam entre crescentes e decrescentes, sem configurar a existência de um padrão. Essa área, assim como a porção Sul do Médio São Francisco, pode ser vista como uma sub-região de mudança fraca, visto que a grande maioria das estações se situa na faixa de magnitudes decenais estimadas entre 0 e 5%, uma faixa de valores considerados negligenciáveis neste trabalho.

Chama a atenção a aglomeração de estações de R1 decrescentes posicionadas no Oeste Baiano, sendo que a grande maioria dos pontos de monitoramento está situada na faixa de tamanho amostral entre 30 e 39 anos e, portanto, sujeita às problemáticas inerentes a seu reduzido histórico de observação: baixo poder, erros tipo S e M, por exemplo. Nota-se ainda um aumento geral na magnitude das tendências negativas na porção inferior do Médio São Francisco (MSF) que se estende até o oeste do MBSF; esse comportamento, que é confirmado por estações de tamanhos amostrais mais representativos, parece realmente afetar a área de aglomeração de estações detectada no estado da Bahia, onde se verifica uma redução média de, aproximadamente, 10% no número de dias de chuva por década. Esse valor da magnitude de mudança é coerente com o que se verifica nas estações vizinhas e de maior comprimento serial, que possuem aproximadamente o mesmo módulo de tendência decenal visto para os pontos de monitoramento do oeste baiano. Diante disso, pode-se inferir que aparentemente os erros de magnitude advindos da estimativa da tendência pelo B_SEN para amostras de pequena extensão, não afetaram os resultados obtidos na região em questão.

Já o restante do MBSF aparenta ser ligeiramente positivo quanto a R1, comportamento esse que se mantém nas proximidades da foz. Apesar da existência de duas (2) estações de sinais opostos próximas uma da outra, de comprimento serial reduzido (30-39 anos) e forte magnitude de mudança ($>10\%$), entende-se que a região é de comportamento positivo, dada as evidências de existência de zona de transição decrescentes para crescente no Médio Baixo São Francisco. Estranha-se apenas a alta magnitude de tendência estimada para as estações de BSF, já que outras positivas, de maior tamanho serial e vizinhas às analisadas foram enquadradas na faixa de 0 a 5% de mudança. Dessa forma, a dualidade de tendências verificadas nessas duas estações e o aparente exagero na magnitude da mudança estimada podem ser creditadas, respectivamente, a erros do tipo S e M, associados à aplicação do B_SEN a séries de tamanho reduzido.

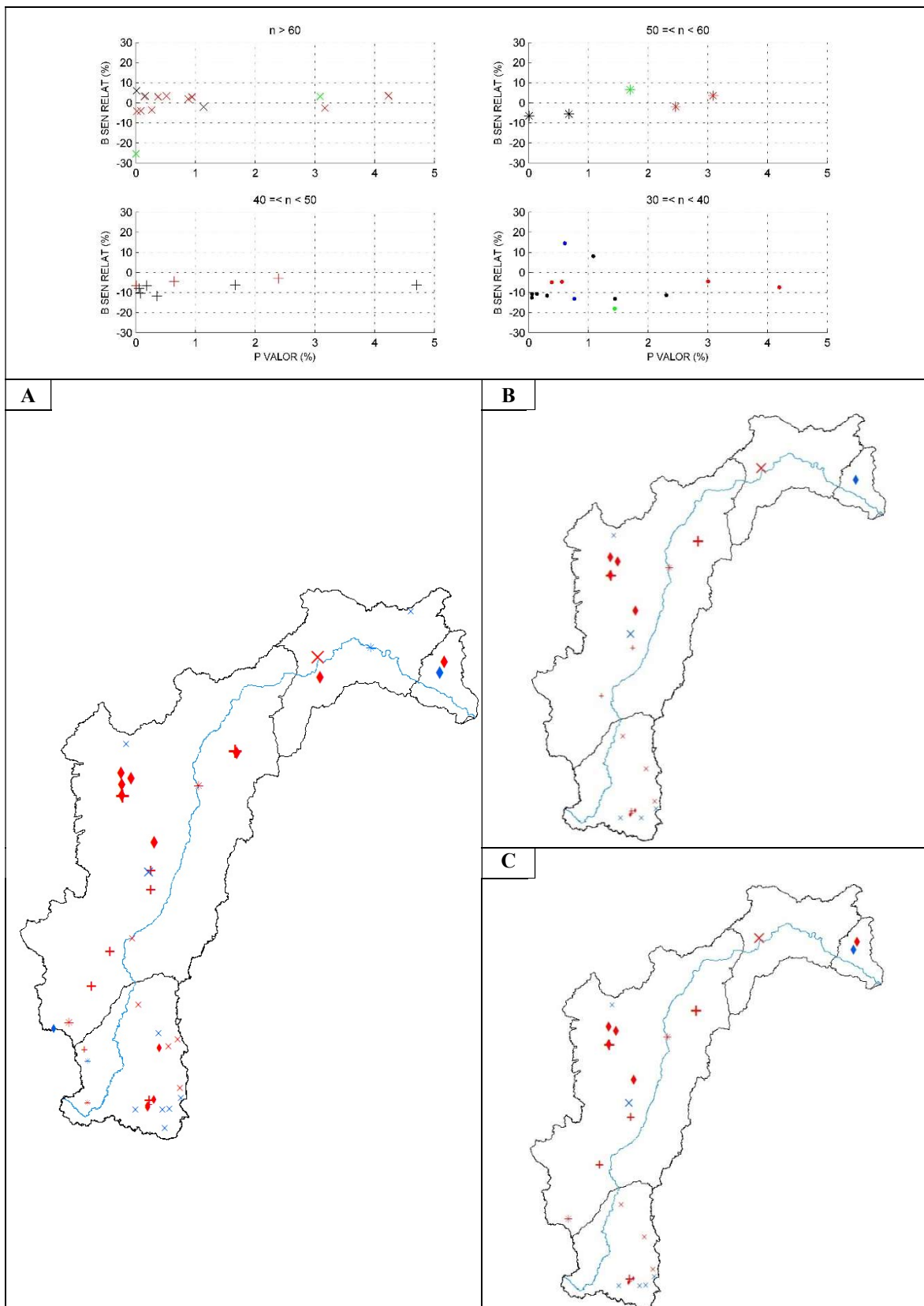


Figura 5.21 – Estações significativas para R1 após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

Na Figura 5.22, nota-se que, na região do ASF, há uma predominância de estações cujas magnitudes das tendências encontram-se enquadradas na menor dentre as três (3) faixas de valores utilizadas neste trabalho, 0 a 5%. O Alto São Francisco, que conta com onze (11) locais de monitoramento ditos não-estacionários decrescentes, a maioria representada por séries de tamanho superior a 60 anos (7), que coexistem com outros três (3) crescentes, pode ser apontado como de SDII em progressivo decréscimo e de lenta mudança. A negatividade da região em questão pode ser constatada pela diferença de percentual entre os pontos de monitoramento negativos e positivos, bem como pelo baixo sinal de mudança constatado nas estações crescentes (< 5%).

Por outro lado, o Oeste da Bahia apresenta três (3) estações de SDII positivo, situadas na mesma posição em que se constatou tendências negativas para o R1, de comprimentos amostrais entre 30-39 anos e magnitudes de mudanças decenais que chegam a quase 20%. É sabido que o SDII, conforme evidenciado pela equação (4.18), é a razão entre o total anual precipitado (PRCP_{tot}) e o número de dias de chuva superiores a 1mm (R1); portanto, reduções em R1 ou aumentos em PRCP_{tot} podem explicar o padrão crescente observado. No caso em questão, conforme já relatado, a forte redução no valor de R1 não é acompanhada por uma variação equivalente no total anual precipitado (Figura II.1– Apêndice II). Por fim, a região correspondente ao MBSF, que parece possuir evidências mais fortes de decréscimo do que as estimadas no ASF e MSF, apresenta sete (7) estações negativas e (2) duas positivas, sendo que o tamanho amostral das mesmas é heterogêneo, mas todas possuem séries de dados com extensão superior a 40 anos de registros. Os pontos de monitoramento em que se verificou tendência positiva apresentam magnitudes de tendências e p_valores semelhantes e parecem surgir em função de aspectos locais, a exemplo da orografia, já que erros de sinal de B_SEN_{relativo}, a priori, são menos prováveis de ocorrer pelo fato das estações crescentes em questão apresentarem registros que superam os 60 anos.

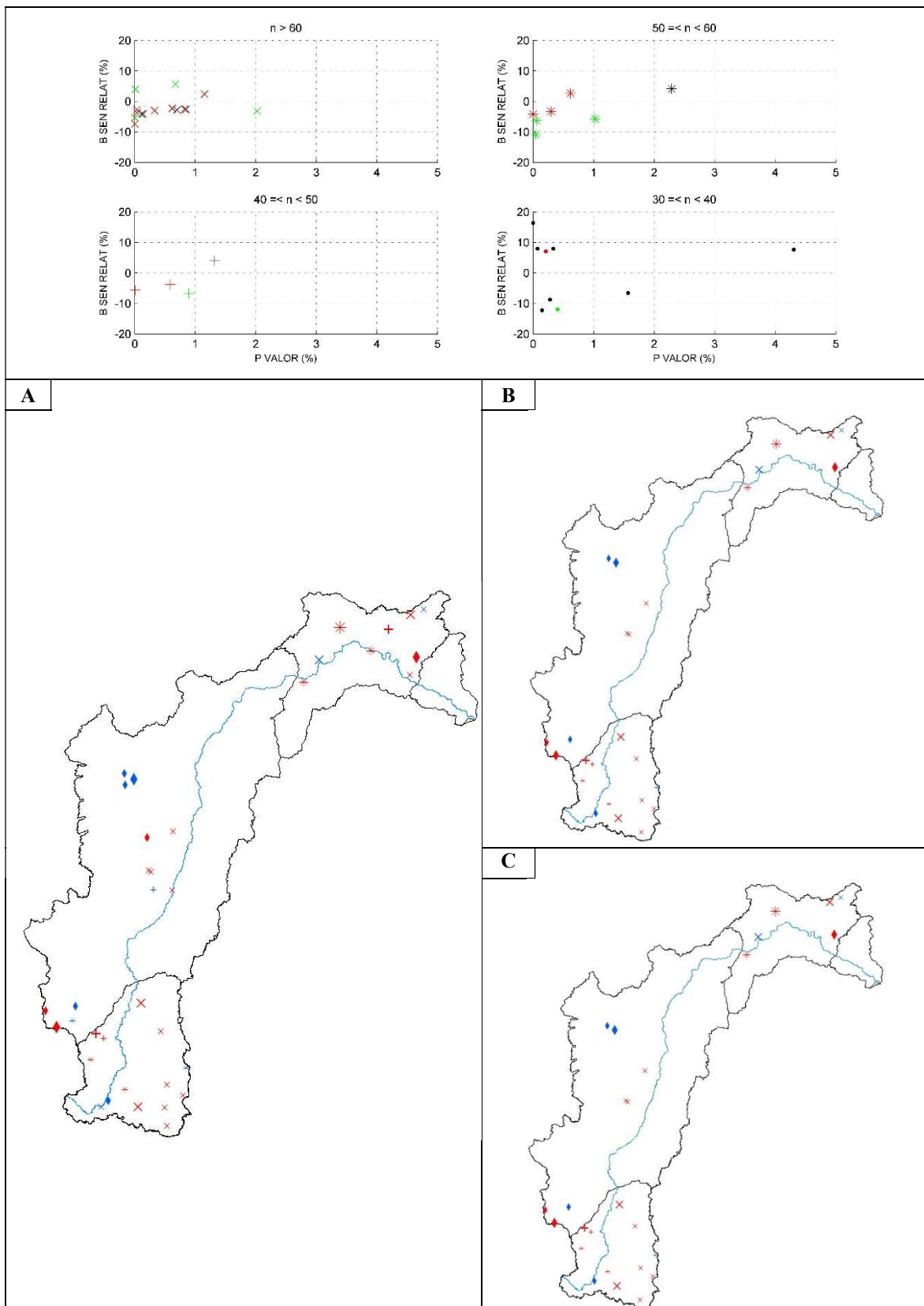


Figura 5.22 - Estações significativas para SDII após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

Para o R10, observa-se na Figura 5.23 que a totalidade da bacia do rio São Francisco encontra-se em decréscimo, à exceção da região do BSF. Verifica-se que a mudança expressa pelo B_SEN_relativo para o R10 é fraca na ASF e na porção superior do MSF e atinge valores máximos no MBSF, nas proximidades da divisa com a sub-região do Médio (>10%). De maneira geral, o R10 se comporta de maneira similar à vista para SDII e R1, ou seja, apresenta tendências negligenciáveis na porção superior da área de estudo, localizada no estado de Minas Gerais, e maiores magnitudes na Bahia (> 10%). Destaca-se que, no Oeste baiano, apenas um dos pontos de monitoramento apresenta magnitude de tendência entre 5 a 10% para R10, diferentemente do que ocorre para SDII e R1. Outro ponto de atenção é a presença de uma estação de tamanho superior a 60 anos, situada entre o Alto e Médio Alto São Francisco, cuja tendência estimada situa-se na faixa superior a 10%, o que pode ser um indicativo de uma possível zona de não-estacionariedade. Essa constatação é confirmada pelo fato de os índices R20 e R50 apresentarem mudança decrescente estatisticamente significativa para essa mesma região ainda maiores do que a estimada para R10. Em outras palavras, nota-se uma diminuição nas taxas de rejeição à medida que o valor da intensidade da chuva se eleva.

Em relação ao Oeste da Bahia, que é uma região de alta demanda hídrica e marcada por conflitos pelo uso da água (PRH – BHSF,2016), relata-se que o R50 e RX1d, além do SDII já relatado, são os únicos dentre os dez (10) índices de precipitação testados que possuem evidências de tendências crescentes, contrariando a lógica decrescente esperada para as variáveis em questão, face os resultados negativos obtidos para os índices de vazão.

O R50 no Oeste da Bahia é apontado como positivo para três (3) estações, códigos ANA: 1145004, 1144014 e 1344002, sendo que, para as duas primeiras, o sinal da tendência decenal é significativamente elevado (próximos a 50%). Há, para essa localidade, uma dificuldade em definir um padrão regional único de tendências para o R50, dada a coexistência dos pontos de monitoramento supracitados com outros quatro (4) de sinais contrários (Figura II.3 – Apêndice II).

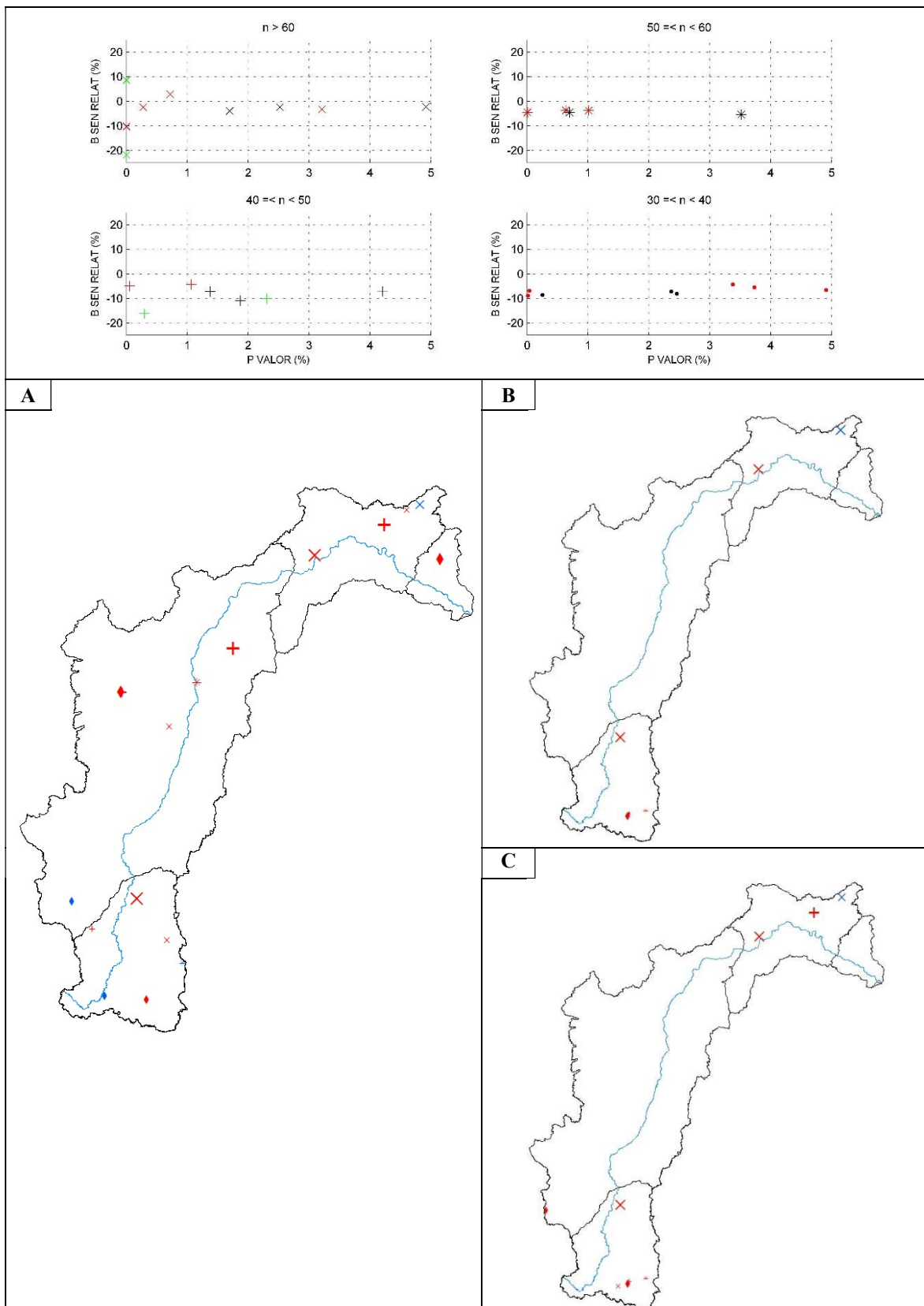


Figura 5.23-- Estações significativas para R10 após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

Diante do exposto, a Figura 5.24 mostra as séries observadas de R50 para as estações 1144014 e 1144005, cujos tamanhos seriais diferem em nove (9) anos (1144005:1963 – 2013 e 1144014: 1972-2013) e os sinais estatisticamente significantes de mudança são contrários, respectivamente, crescente e decrescente. O que chama a atenção é a existência de sucessivas falhas no período amostral considerado, principalmente para 1144014 (9 anos de falhas), dificultando a visualização de algum padrão regional. Entretanto, por compreender um período mais longo de observações, percebe-se no posto 1144005 uma aparente tendência à negatividade, graças a períodos de elevado valor de R50 na década de 60, que reduzem progressivamente até o ano de 2013. O mesmo não pode ser visto em 1144014, em que apesar do p_valor de $2.51E-04$ e $B_SEN_relativo$ de 47.38%, não é possível confirmar o aparente comportamento positivo, devido à não continuidade e ao tamanho reduzido da série analisada. Além disso, o aparente crescimento inferido em 1144014 é também observado em 1144005 entre os anos de 1972-2013, mas é suplantado pelo significativo decréscimo de R50 em período anterior ao que é comum entre as estações.

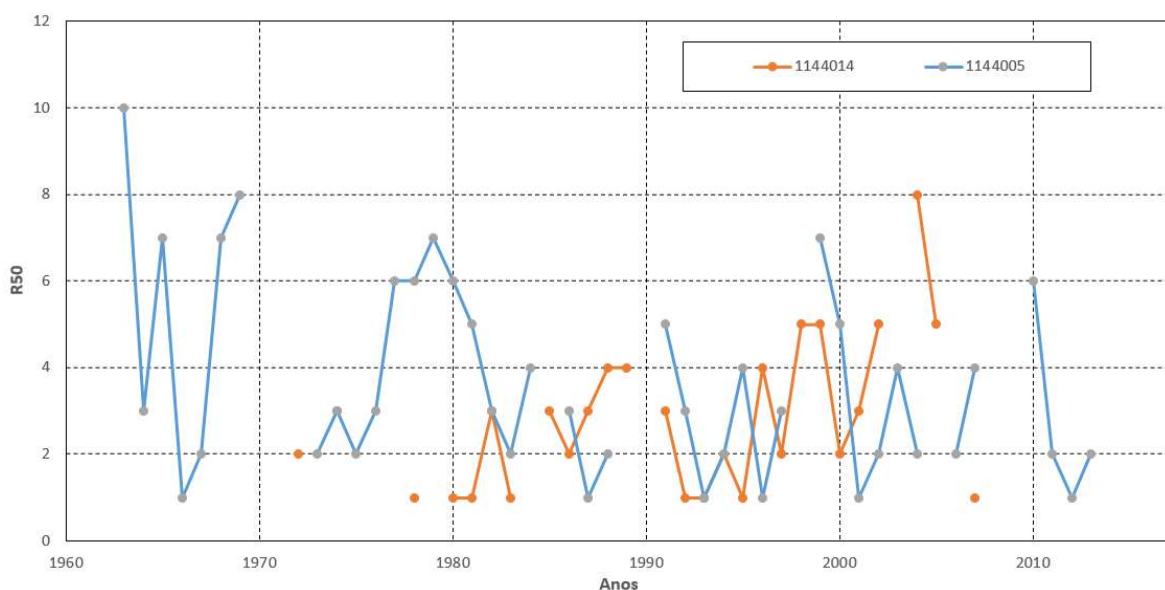


Figura 5.24 – Séries temporais de R50 para as estações 1144014 (laranja) e 1144005 (azul)

Mais a Oeste do MSF, problema similar ao relatado para as estações 1144014 e 1144005 ocorre entre 1344002 e 1343008, sendo a primeira significativa positiva e a segunda, negativa. Devido à grande extensão das séries associadas aos pontos de monitoramento em questão (1344002: 1945-2013 e 1343008: 1936-2012) e seus $p_valores$ estimados serem relativamente próximos (1344002: 0.0015 e 1343008: 0.016), a tarefa de definir a existência

de tendência significativa e seu sinal, sem a análise visual das séries propriamente ditas, torna-se difícil. Para facilitar essa avaliação, a Figura 5.25 mostra as séries temporais dessas estações, em que se evidenciam as grandes falhas existentes entre 1960-69 e 1973-76 em 1344002. Apesar da descontinuidade, é nítido que os mínimos e máximos locais observados para R50 ocorrem simultaneamente para as linhas laranja e azul, sendo invariavelmente maiores em magnitude para 1344002, o que indica uma compatibilidade de padrões entre as estações em análise, principalmente para o período pós-1976. Exclusivamente para 1343008, é possível se notar com clareza a redução nos valores de R50 à medida que se aproxima do ano de 2013. Entretanto, avaliações e constatações similares às anteriormente feitas para a estação decrescente são prejudicadas pela inexistência de maior base amostral e pela grande quantidade de falhas em sequência na série de 1344002. Em outras palavras, qualquer afirmação a respeito do comportamento da estação apontada como positiva e que abranja o período de 1936-2014 será baseada em conjecturas. Assim, assume-se que a região em questão apresenta tendência negativa por pelo menos dois motivos: aparentemente, nos anos em que ambas apresentam dados observados, essas duas estações se comportam de maneira similar, e, de maneira geral, há maior proporção de postos com tendência decrescente no MSF.

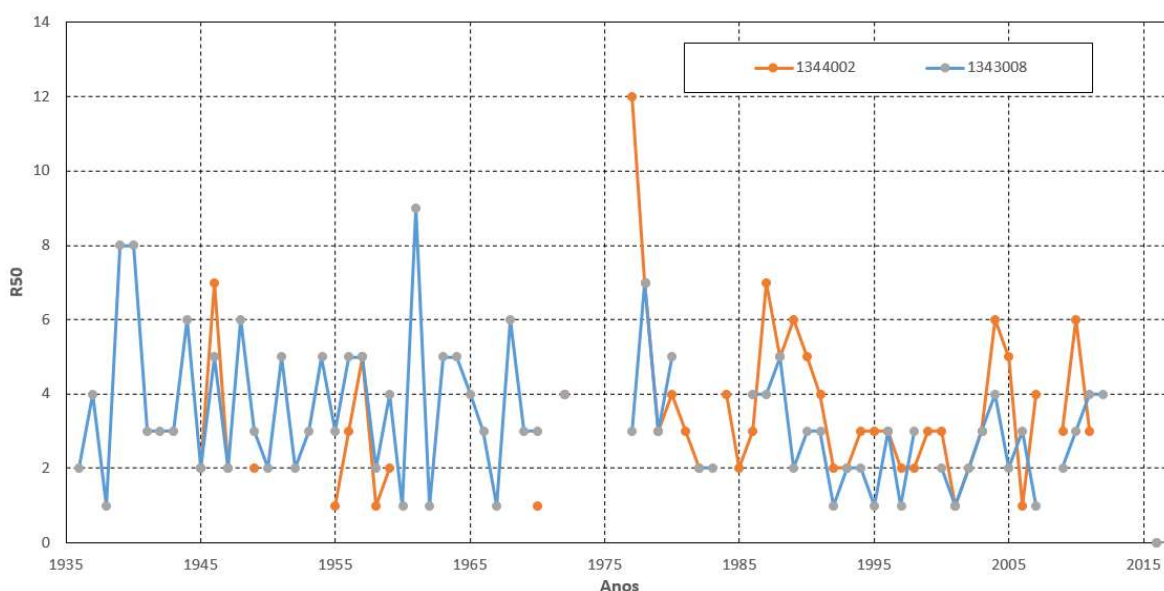


Figura 5.25 – Séries temporais de R50 para as estações 1344002 (laranja) e 1343008 (azul)

O ponto de monitoramento de maior comprimento serial localizado no Oeste Baiano se notabiliza por uma mudança crescente decenal de RX1d mais moderada e compreendida

entre 0 a 5% (1344002), diferentemente do evidenciado pelas duas estações de tamanhos amostrais na faixa de 30 a 39 anos, que apresentam tendências com valores superiores a 15%. O padrão observado para as estações de menor tamanho amostral não é compatível com os fortes decréscimos de vazão verificados para essa mesma região, o que sugere que o B_SEN superestimou a magnitude das mudanças positivas previstas para tais postos. No caso dos demais índices extremos associados à precipitação, não se nota a existência de tendências que ultrapassem o valor de B_SEN_Relativo de 5%, exceto o PRCPTot e CWD, que possuem, respectivamente duas e uma estação negativa no Oeste da Bahia. Visto que há poucos pontos de monitoramento com tendência decrescente para esses dois índices, não é possível afirmar que há um padrão comportamental para a região em análise. Uma situação análoga ocorre para CDD, R20 e RX5d, nos quais as mudanças são consideradas negligenciáveis.

Finalmente, quanto aos índices de vazão, observa-se nas Figura 5.26 e Figuras II.8 a II.12 - Apêndice II que todos eles apresentam alta concentração de estações decrescentes na região Centro-Oeste do Médio São Francisco. Para alguns dos índices, como no caso de Q7min (Figura 5.26) e Q30min (Figura II.10 – Apêndice II), existem estações significativas negativas espalhadas não apenas nessa localidade, como também pelo ASF e na porção superior da MSF. Outro aspecto que merece menção é que os índices associados a máximos de vazão, Qmáx, QX5d e QX30d, apresentam p_valores próximos de zero e magnitudes de mudança positiva decenais que se aproximam dos 10% no extremo Sul da ASF, região marcada pela existência de estações de comprimentos amostrais superiores a 60 anos. Em virtude do relatado, a região em questão pode ser declarada como positiva. Tal comportamento não-estacionário é compartilhado, em menor grau e abrangência, para o Qméd, mas não é verificado para Q7min e Q30min.

Conforme já relatado anteriormente, tanto em número de estações quanto em sinal de tendências estimadas pelo B_SEN_relativo, os índices associados à vazão mínima são aqueles que abrangem maior parte da área de estudo, seguidos pelo Qméd. No entanto, em virtude da inexistência de pontos de monitoramento que integram a base de dados no MBSF e BSF, não foi possível fazer uma análise conclusiva a respeito dessas regiões.

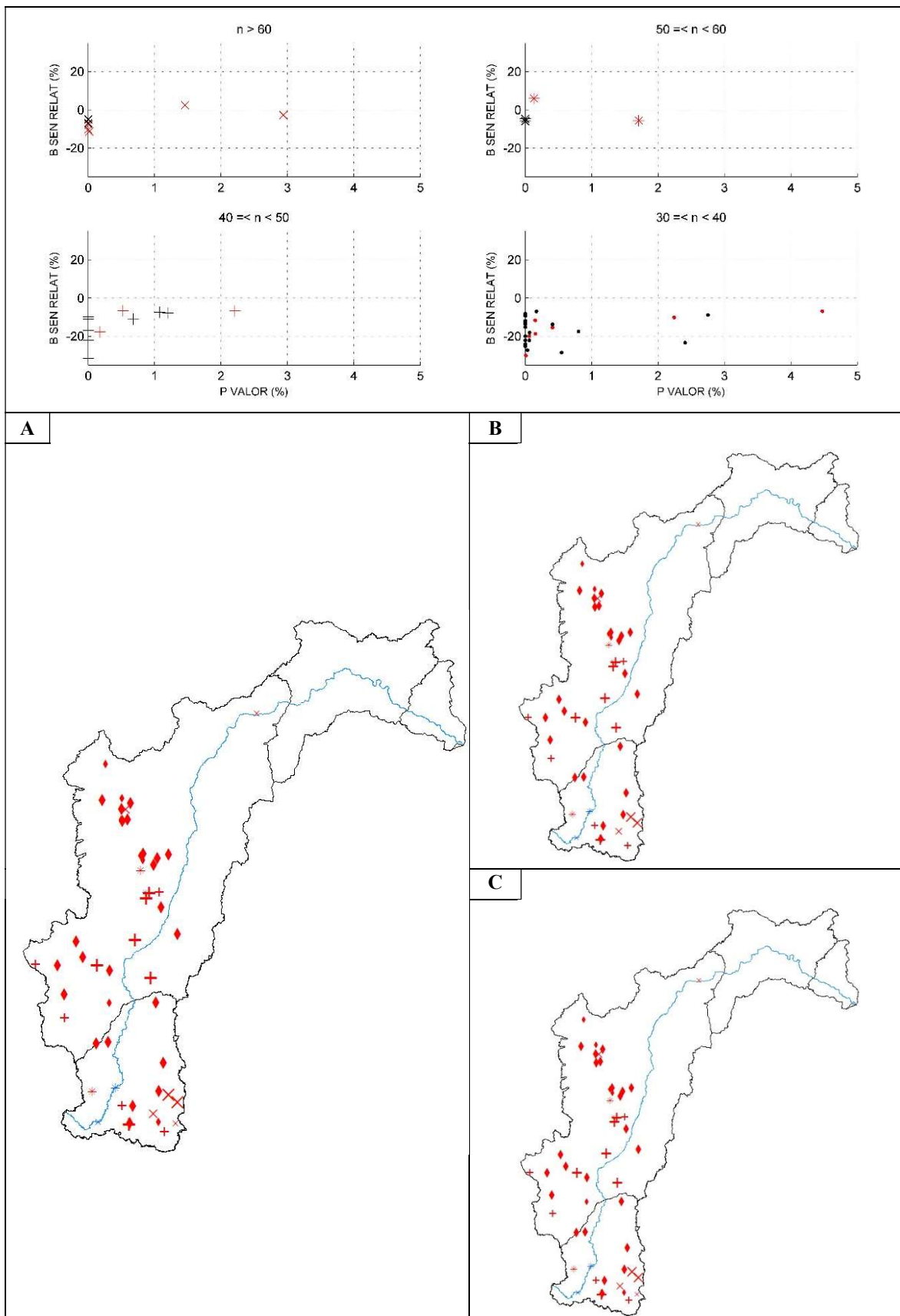


Figura 5.26- Estações significativas para Q7min após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

Apesar de, em alguns casos, o comportamento dos índices extremos associados à precipitação ser aderente ao decréscimo observado para aqueles de vazão, a análise procedida deixa claro que não se pode atribuir unicamente à diminuição das chuvas a redução generalizada observada para as descargas máximas, médias e mínimas.

Adicionalmente, é interessante notar que para todos os índices extremos testados, sejam eles de precipitação ou vazão, a dispersão dos valores observados para a magnitude de tendência de uma dada série temporal se reduz à medida que o tamanho amostral considerado diminui. Por fim, constatou-se que as estações que estão associadas a séries com maior disponibilidade de registros históricos são o ponto de partida para a análise da estacionariedade de uma região qualquer, já que (1) apresentam maior poder de detecção de mudança verdadeira, (2) são menos sujeitas às variações no cômputo da taxa de mudança pelo estimador proposto por Sen (1968) e (3) apresentam menores tamanhos de intervalo de confiança relativos (erros do tipo M).

5.5. EFEITOS DA MULTIPLICIDADE SOBRE A PROPORÇÃO DECLARADA DE ESTAÇÕES NÃO-ESTACIONÁRIAS

Com o objetivo de avaliar qual é o impacto de se ignorar o problema da multiplicidade ao se realizar uma análise de detecção de tendência em uma região dotada de mais de uma centena de estações, sejam elas pluviométricas ou fluviométricas, optou-se por adotar, para a bacia em estudo, três abordagens distintas de avaliação de estacionariedade, a saber: (a) simplesmente ignorá-la, (b) empregar o BH95 e (c) utilizar o BH_Adap.

A Figura 5.27 e a Figura 5.28 ilustram as diferenças dos resultados obtidos para as três metodologias, em termos do percentual de rejeições da hipótese nula para cada um dos índices extremos de precipitação (10) e vazão (6). Essas figuras permitem ainda identificar a prevalência do sinal de mudança decrescente em detrimento do crescente para a grande maioria das variáveis testadas. É importante ressaltar que os resultados apresentados nesta seção para (a), (b) e (c) contaram com a aplicação prévia do TFPW à base de dados, na tentativa de mitigar os efeitos da autocorrelação.

A questão da multiplicidade, ignorada em (a), além do menor poder de detecção associado às técnicas que utilizam o *FDR* como mecanismo de controle para os erros do tipo I, quando comparado com as tradicionais (vide seção 5.3), pode explicar, em parte, o menor número de rejeições observado pós aplicação do BH95 e BH_Adap, conforme a Figura 5.27 e a Figura 5.28. As diferenças entre as abordagens (a) e (b), ou (c), podem ser grandes, vide o exemplo de CDD e R10. Quando a multiplicidade é negligenciada, a proporção de estações apontadas como detentoras de tendências estatisticamente significantes para os dois índices anteriormente mencionados foi de 15-20%, enquanto os resultados obtidos ao se utilizar BH95 e BH_Adap indicaram a presença de mudança significativa em 0% e 4-6% do total de pontos testados para, respectivamente, CDD e R10.

Além de CDD e R10, R1 e CWD foram os índices de precipitação que tiveram reduções na proporção de estações apontadas como significativas superior à média (8,70%), considerando-se apenas os índices referentes à precipitação, ao se modificar a abordagem de estacionariedade empregada de (a) para (b) ou (c). Por exemplo, ao se utilizar a metodologia (b) ao invés de (a), nota-se que a diminuição no percentual de rejeições é de 11,20% e 8,90% para, respectivamente, R1 e CWD. Isso aponta que a evidência de não-estacionariedade contida nas séries de CDD, R10, R1 e CWD não são muito fortes, sendo que os *p*-valores associados às mesmas não são suficientemente pequenos para manter os resultados obtidos para (a). A redução no percentual médio das taxas de rejeição de Q_{\max} , $Q_{\text{méd}}$, $Q_{7\text{min}}$, $Q_{30\text{min}}$, Q_{X5d} e Q_{X30d} é de 6,0% e 3,50% ao se considerar a alteração entre o emprego exclusivo do MK+TFPW para, respectivamente, o BH95 e BH_Adap.

Apesar do relatado, as diferenças entre as abordagens tradicional e as que utilizam o *FDR* nem sempre são grandes, embora não negligenciáveis, como se pode notar para os resultados médios obtidos entre BH95 e BH_Adap para o PRCptot (7.8%), R20 (6.7%), R50 (6.4%), $Q_{\text{méd}}$ (7,6%) e Q_{X1d} (5,6%). Mesmo que as discrepâncias não sejam tão significativas, quanto aquelas obtidas para os índices discutidos no parágrafo anterior, a redução no número de detecções é a regra, à exceção do $Q_{7\text{min}}$, no qual a proporção de detecção foi exatamente a mesma para (a) e (c).

Assim, fica claro que a redução observada pelo emprego do BH95 e BH_Adap varia, índice a índice. Esse é um comportamento esperado, uma vez que a referida diminuição está estreitamente relacionada com a quantidade de evidências contida nos dados, ou seja, nos p_valores obtidos nos testes de hipóteses procedidos.

O BH_Adap, que difere da abordagem (b) apenas na técnica empregada para a estimativa do número de hipóteses nulas verdadeiras (m_0), padece das mesmas limitações metodológicas do BH95. No entanto, esse teste possui maior poder do que o seu predecessor. A diferença metodológica entre ambos, que se traduz por maiores taxas de rejeição da hipótese nula em (c) do que (b) para a maioria dos índices de vazão e precipitação testados (por exemplo, 5.6% para Qméd e 0.60% para R50), parece ser ligeiramente mais pronunciada, em média, na rejeição das séries associadas a estações fluviométricas (2,5%) do que pluviométricas (0.6%).

Adicionalmente, aponta-se que metade dos dez índices extremos associados à precipitação tiveram o mesmo percentual de rejeição ao se aplicar o BH95 e BH_Adap, a saber: PRCPTot, R20, RX1d, RX5d, CWD e CDD. Isso não ocorreu para aqueles relativos à vazão, já que a diferença mínima observada entre as abordagens (b) e (c) deste trabalho é de 0.8% (Q30min), chegando ao máximo de 5.6% (Qméd).

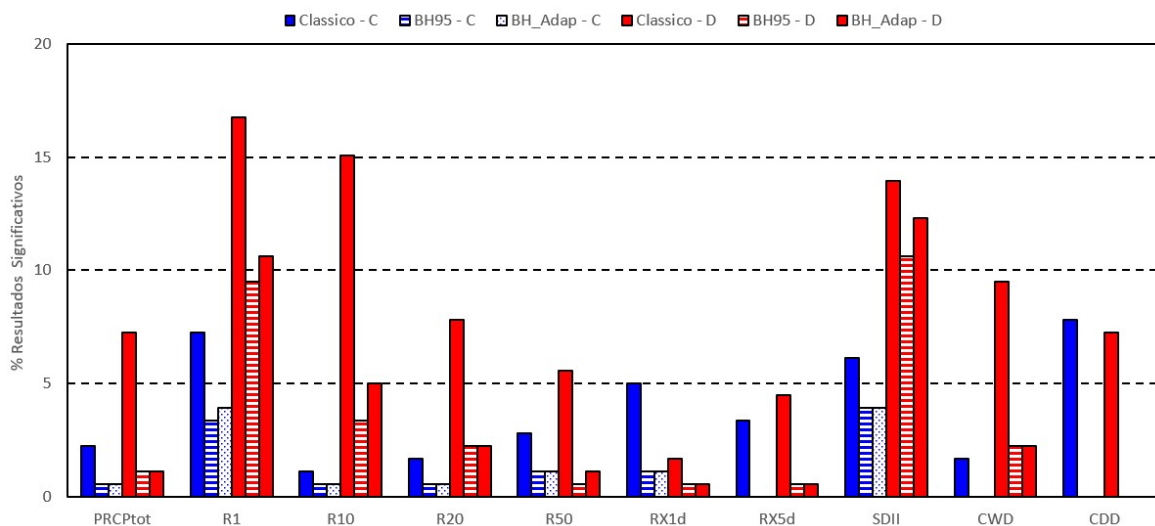


Figura 5.27: Proporção de estações declaradas como não-estacionárias ($\alpha=5\%$), utilizando-se procedimentos locais de análise (não considera a multiplicidade) e o BH95. Para situações em que a correlação temporal é significativa foi aplicado o TFP

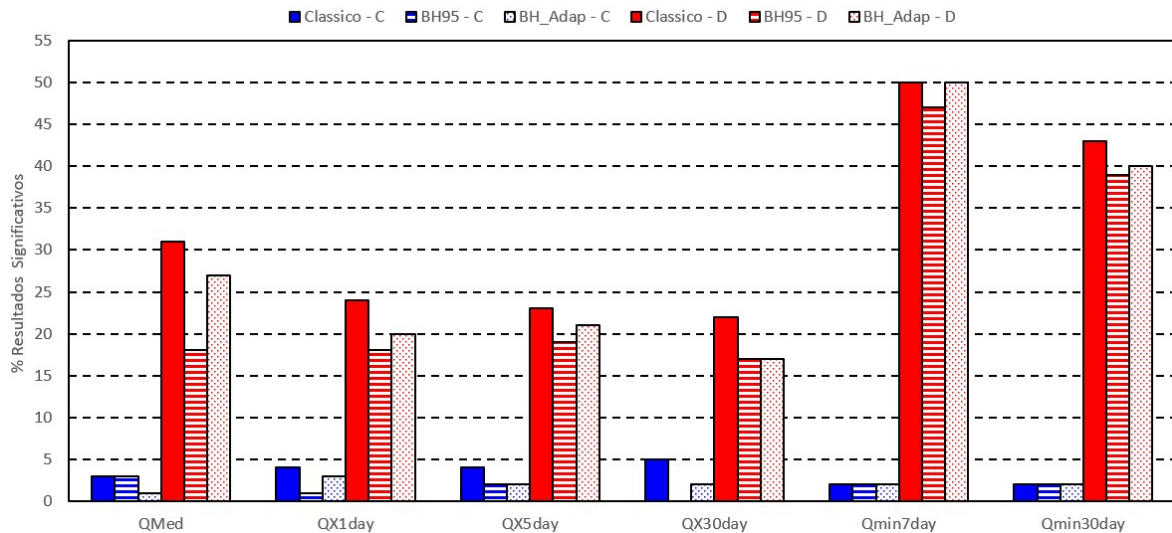


Figura 5.28: Proporção de estações declaradas como não-estacionárias ($\alpha=5\%$), utilizando-se procedimentos locais de análise (não considera a multiplicidade) e o BH95. Para situações em que a correlação temporal é significativa foi aplicado o TFP

Na Figura 5.28, nota-se que, mesmo após a tentativa de se mitigar o efeito da correlação serial pelo uso do TFPW e multiplicidade de testes pelo emprego do BH_95 e BH_Adap, a proporção de estações significativas para os índices de vazão avaliados permanece superior a 15%, chegando a valores próximos de 50% para o Q7min, que se distribuem pelo Alto e Médio São Francisco (Figura 5.26 (b) e (c)). Destaca-se ainda que o QX1d (Qmáx), Qméd, QX5d e Q30min tiveram declaradas como significativas negativas mais de 20% das estações fluviométricas testadas.

As Figura 5.29 e Figura 5.30 estão subdivididas em classes de magnitude do $B_SEN_relativo$ estimado e mostram a alteração do percentual das estações declaradas significativas pelo MK+TFPW após a aplicação das técnicas de controle de erros pelo FDR. Essas figuras indicam, em linhas gerais, uma ligeira diferença na média da variação das taxas de rejeição estimadas para os índices de precipitação ou vazão para os três (3) intervalos de mudança decenais previstos: (i) menores que 5,00%, (ii) maiores ou iguais a 5,00% e menores que 10,00% e (iii) superiores ou iguais a 10,00%. Salienta-se que, apesar de negligenciáveis na prática, essas diferenças são mais pronunciadas para o intervalo (ii), independentemente do emprego do BH95, BH_Adap ou da natureza do índice testado ($BH95_p = 3,13\%$ $BH95_q = 2,83\%$ e $BH_Adap_p = 2,96\%$ e $BH_Adap_q = 2,00\%$).

As variações nas taxas de rejeição constatadas pelo uso do BH95 e BH_Adap, que são relativamente equivalentes em média por classes de B_SEN_relativo, variam índice a índice, chegando a atingir valores máximos de 6,15% (R10) e mínimos de 0,00% (RX5d), no caso do BH95 para magnitudes de mudança classificadas, respectivamente, nos grupos (i) e (ii), por exemplo. Ainda quanto à Figura 5.29, nota-se que os índices associados ao maior número de estações declaradas como não-estacionárias, R1, R10 e SDII, experimentaram maior redução de suas taxas de rejeição na faixa de menor magnitude da tendência decenal, porém, não deixaram de ser aqueles que detém a maior proporção total de estações rejeitadas. Em contrapartida, os índices R50, CWD e CDD tiveram maiores mudanças percentuais nas estações integrantes das classes (ii) e (iii).

A sutileza das diferenças metodológicas existente entre BH95 e BH_Adap parece se refletir nas variações de rejeição obtidas entre as abordagens supracitadas, pelo menos no caso dos índices de precipitação. Coincidentemente, as reduções nas taxas de rejeição, que são idênticas em módulo para a maioria das classes de magnitudes de mudança das variáveis testadas, parecem seguir o seguinte padrão: se, para um determinado intervalo de B_SEN_relativo, a diferença na variação da rejeição for nula entre o BH95 e BH_Adap, as demais também o serão, e o contrário também verdadeiro.

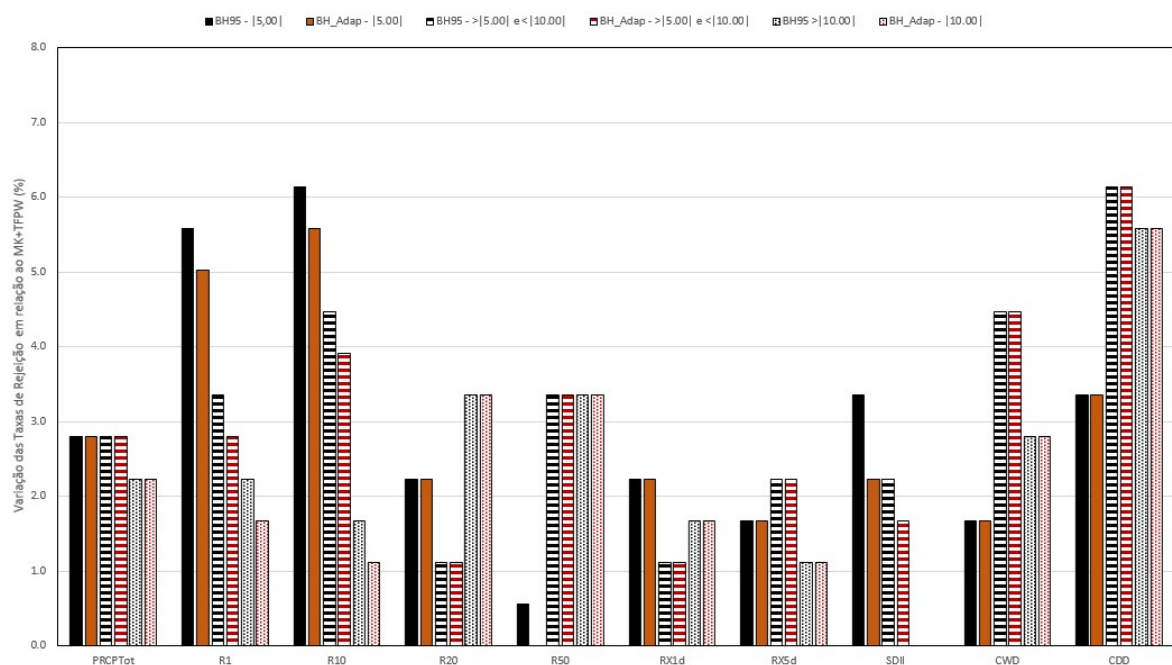


Figura 5.29 – Variação do percentual de rejeição dos índices extremos de pluviometria obtidos para o BH95 e BH_Adap em relação ao MK+TFPW (referência) por magnitudes de B_SEN_relativos

Na Figura 5.30, o comportamento anteriormente mencionado não se repete, já que um intervalo de B_SEN_relativo pode ter a mesma variação da taxa de rejeição para as técnicas que empregam o FDR e, ainda assim, variar de maneira diferente para os demais intervalos. Para exemplificar, considerando o QX5d, a variação das taxas de rejeição estimadas para o BH95 e BH_Adap são iguais para magnitudes de mudanças decenais inferiores a 10.00%, sendo diferentes para o intervalo (iii). Nesse caso, nota-se ainda que estações que haviam sido apontadas como não-significativas pelo uso exclusivo do TFPW associado ao MK (MK+TFPW) foram declaradas como não-estacionárias pelas abordagens (b) e (c), tornando negativa a diferença apresentada na Figura 5.30. É importante ressaltar que essa elevação no número de estações que integram a faixa de B_SEN_relativo superior ou igual a 10.00% para o BH95 (2,00%) e BH_Adap (4,00%), vista em QX5d, não alteram as constatações de que o número de rejeições total obtido pelo MK+TFPW é superior ao visto nas demais abordagens propostas nesta seção.

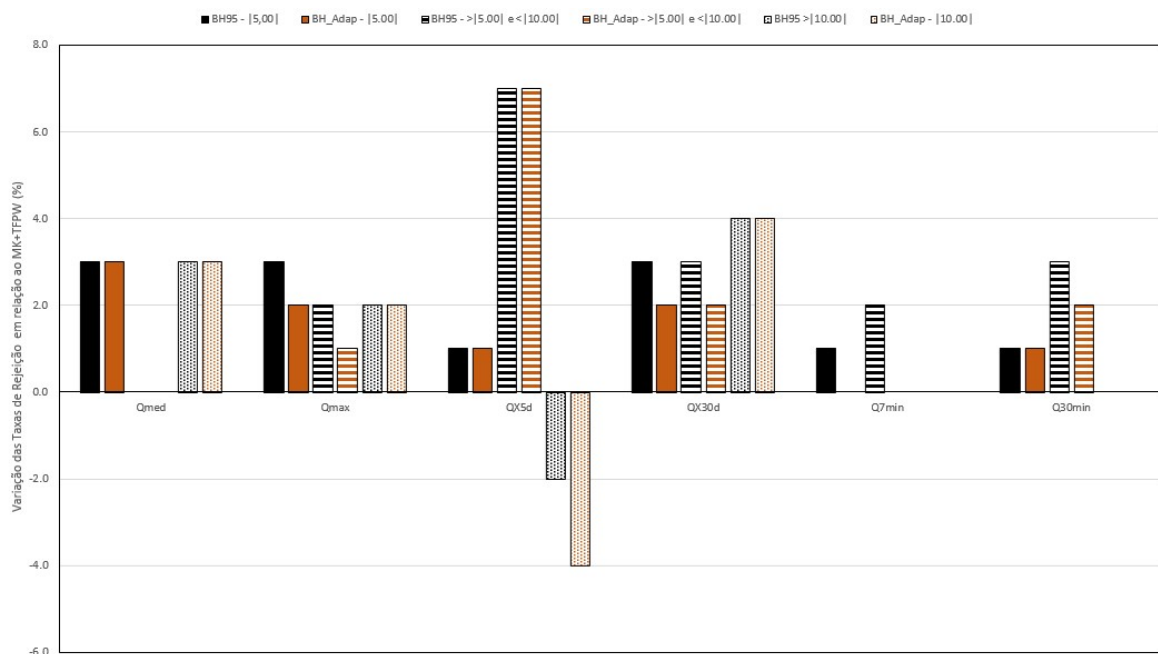


Figura 5.30 - Variação do percentual de rejeição dos índices extremos de pluviometria obtidos para o BH95 e BH_Adap em relação ao MK+TFPW (referência) por magnitudes de B_SEN_relativos

À exceção do que se verifica em QX5d, as variações observadas para os índices extremos de vazão, assim como constatado para aqueles relacionados à precipitação, são variáveis e de comportamento único. Isso porque, conforme ilustrado na Figura 5.30, alguns índices, tais como Qmáx e QX30d, apresentam certo equilíbrio nas magnitudes da variação das taxas

de rejeição estimadas para os diversos intervalos de $B_SEN_relativo$, enquanto que, para outros índices, a exemplo do $Qméd$ para a classe (ii), os valores obtidos pelo MK+TFPW após a aplicação do BH_95 e BH_Adap permanecem constantes, ou seja, o percentual de variação é igual a 0,00%.

O R1, R10 e SDII apresentaram maiores proporções de estações declaradas como detentoras de tendências dentre os 10 índices de precipitação avaliados. Para esses três índices, especialmente, nota-se uma maior redução no número de estações significativas localizadas no Alto e Médio São Francisco, principalmente, no oeste do estado da Bahia. Trata-se de um comportamento esperado, visto que essas duas regiões possuem o maior quantitativo de pontos de monitoramento dentre as quatro analisadas neste trabalho. A mesma justificativa se aplica à maior diminuição da taxa de rejeição associada às estações decrescentes, em que prevalece a máxima: quanto maior a proporção de estações de sinal de mudança negativa, maior será a variação observada na sua taxa de rejeição após a aplicação do BH95 e BH_Adap e vice-versa.

A utilização do BH95 e BH_Adap eliminou parte da quase sobreposição ou alternância de sinais de mudança entre as estações para o R1, permitindo que fossem estabelecidas zonas homogêneas de mudança no ASF. O mesmo não ocorreu para o caso da aplicação exclusiva do MK-TFPW, conforme a Figura 5.21(a). Apesar do sinal fraco da tendência nessa sub-região, o que se nota é que o Sul da ASF concentra estações significativas decrescentes, sendo o Leste e Norte negativo. Quanto às demais regiões da bacia do São Francisco, Médio, Médio-Baixo e Baixo, a utilização das técnicas que empregam o *FDR* não alterou significativamente os padrões de mudança relatados em 5.4. Ao contrário, o BH95 forneceu evidência adicional à existência de tendência positiva no BSF, antes sob suspeita graças à coexistência de estações com sinal positivo e negativo lado a lado; essa aparente incongruência desapareceu ao se levar em consideração a presença de correlação espacial por meio da abordagem (b) (Figura 5.21(b)).

Dentre os testados, o R1 é o único índice pluviométrico que, após o uso do BH95 e BH_Adap, permaneceu com estações declaradas como decrescente na região do oeste

baiano. Nessa mesma região, o R10, R50, CWD e CDD foram apontados como não significativas pelas abordagens (b) e (c).

Para o R10, a Figura 5.23 (b) e (c) mostra que não só a região Oeste da Bahia foi afetada, mas sim toda a área do Médio São Francisco. Onde antes, pelo uso exclusivo do MK+TFPW, foram detectadas onze (11) estações com tendência negativa estatística significativa, agora, nota-se a presença de 0 e 1 pontos de monitoramento significativos, respectivamente, para o BH95 e BH_Adap. Avaliando-se os p_valores das estações localizadas na MSF, observa-se que a maioria deles (8) são superiores a 1,00%, ou seja, a assertividade na declaração de estacionariedade não é tão grande quanto a vista para os índices de vazão, por exemplo. Elenca-se esse como um dos fatores que impactaram na diferença comportamental relatada entre as abordagens (a) e aquelas que empregam o *FDR*. Apesar de esperada, situações como a observada para R10 nos remetem aos resultados obtidos na seção 5.3 para o BH95 e BH_Adap, em que a não detecção observada pode estar relacionada, dentre outros fatores, à dificuldade dos testes na identificação de tendências em séries de baixa magnitude de tendência, pequeno tamanho amostral e alto coeficientes de variação. A despeito do que possa ter ocasionado essa dificuldade, a baixa magnitude de mudança expressa pelo B_SEN_relativo para as estações localizadas na MSF torna despropositada a discussão sobre a existência de tendência ou não para a sub-região em questão, já que, caso existente, a mesma pode ser interpretada como negligenciável, em termos de gestão, por este trabalho.

Diferente do observado para R1 e R10, o SDII experimentou redução quase equânime no número de estações declaradas significativas crescentes (4) e decrescentes (6), o que, em parte, contribuiu para a manutenção do cenário de tendências apontado pela seção 5.4 após a aplicação do BH95 e BH_Adap. Assim, permanecem as constatações de tendência regional majoritariamente decrescente e lenta na ASF e a mudança positivo de SDII no Oeste Baiano, decorrente da diminuição de R1. Antes sinalizado pelo MK+TFPW, a relação anteriormente descrita entre R1 e SDII pode mais claramente vista após o emprego das abordagens (b) e (c), que eliminaram a presença de pontos de monitoramento negativos de PRCPTot nessa região, o que evidenciou a dependência exclusiva de SDII e R1 para este caso.

Pelo fato de o BH95 e o BH_Adap reduzirem o total de estações declaradas como significativas, a avaliação de tendência para o R50 foi facilitada, se comparada à procedida na seção 5.4, que se baseou na análise de séries temporais dotadas de falhas e conjecturas sobre o comportamento dominante regional. Apesar dos fatores complicadores, a região foi apontada como negativa pelo MK+TFPW. Contrariando as impressões fornecidas pela abordagem (a), as técnicas que empregam o FDR como mecanismo de controle de erros tipo I reduziram o número de estações significativas no Oeste da Bahia a apenas duas positivas. Visto que o BH95 e o BH_Adap possuem menor poder, o que implica em detectar como não estacionárias aquelas estações de maior evidência de mudanças, os resultados para esses testes devem ser interpretados com parcimônia, tendo-se em mente que a exclusividade de tendência constatada em seus resultados pode mascarar mudanças existentes menos expressivas, mas relevantes para o processo decisório de gestão. Isso porque alterações associadas a séries de tamanho reduzido, alto coeficiente de variação e tendência, por exemplo, não foram detectadas por BH95 e BH_Adap, conforme mostram os resultados apresentados na seção 5.3. Assim, mesmo diante das evidências apontadas pela Figura II.3 (b) e (c) – Apêndice II, torna-se difícil negligenciar os resultados obtidos pelo MK+TFPW, onde fatores locais, tais como orografia e urbanização, por exemplo, podem causar alterações relevantes no sinal da mudança observado em determinadas estações.

Esclarece-se que, assim como na seção 5.4, os resultados obtidos para CDD, CWD, PRCPTot, RX1d e RX5d foram negligenciados nesta seção. Isso se deve à presença de mudança regional fraca e/ou à baixa proporção de estações apontadas como detentoras de tendência estatisticamente significantes associada a tais índices extremos.

Em contraste com o que se verificou para os índices de precipitação, os índices de vazão reduziram em menor proporção, ou seja, foram menos impactados pelo emprego das abordagens (b) e (c). Foi verificado que tais índices mantiveram o mesmo padrão espacial visto na seção 5.4. Via de regra, observa-se uma alta concentração de estações decrescentes no Centro-Oeste do Médio e Centro-Norte do ASF, variando apenas, índice a índice, a tendência dominante no extremo Sul da bacia em análise. A porção sul da ASF apresentou tendência decrescente para Q7min e Q30min, crescente para Qmáx, QX5d e QX30d e comportamento estacionário para Qméd. Essa área é também marcada pela presença de

crescimento de R1, e, em decorrência disso, de decréscimos no SDII. Diante desse cenário, sugere-se que o número de dias de chuva na ASF está aumentando no período úmido e se distribuindo de maneira menos uniforme ao longo do ano, tornando as estações secas e chuvosas mais severas. Isso justifica, em parte, o comportamento positivo de $Q_{m\acute{a}x}$, $QX5d$ e $QX30d$, bem como o decréscimo observado para os índices de vazão mínima, mas não explica a totalidade do ocorrido. Diz-se isso, já que a magnitude do sinal de mudança vista em $Q7min$ é superior à vista para o R1, por exemplo

6. CONCLUSÕES E RECOMENDAÇÕES

Esta dissertação possui três objetivos específicos. No primeiro, procura-se avaliar como o tamanho da série histórica (n), magnitude da tendência (b), o coeficiente de variação (C_v) e o coeficiente de assimetria da série histórica, este último função do parâmetro de forma (k) da distribuição Generalizada de Valores Extremos (GEV), influenciam, numa análise local, o poder do teste Mann-Kendall (MK). O segundo, foca na probabilidade de se obter uma estimativa da magnitude da tendência com sinal oposto ao real (erro tipo S) empregando o estimador sugerido por Sen (1968) e o grau de tendenciosidade dessa estimativa (erro tipo M), quando as variáveis n , b , C_v e κ variam. O terceiro, aborda a questão da análise regional de tendência, procurando avaliar o desempenho de metodologias baseadas no conceito do False Discovery Rate (FDR), sugeridas na literatura (Benjamini e Hockberg, 1995, 200) para amenizar o problema da multiplicidade de testes de hipóteses, que tendem a resultar em número exagerado de resultados falso positivos. Como estudo de caso, o presente trabalho ilustra o uso das referidas metodologias num estudo de detecção de tendências monotônicas em extremos de variáveis hidrometeorológicas na bacia do Rio São Francisco.

Os resultados baseados em Simulação Monte Carlo permitiram mapear, para séries geradas pela distribuição Generalizada de Valores Extremos (GEV), a relação entre o poder do teste MK e características da série histórica de extremos, tais como tamanho da amostra (n), coeficiente de variação (C_v), parâmetro de forma (κ) da distribuição GEV (relacionado com o grau de assimetria da série), e a magnitude da tendência monotônica (b).

Observou-se, corroborando trabalhos anteriores, que o aumento do tamanho da amostra ou a redução da razão sinal/ruído (b/C_v) resultam num aumento do poder do teste MK. Além disso, percebeu-se que à medida que κ se torna mais negativo (assimetrias positivas maiores), o poder do teste MK aumenta, ou em outras palavras, quanto maior o valor de κ , menor poder do teste MK. A quantificação desses resultados preenche uma lacuna importante, pois os trabalhos apresentados na literatura focam quase que exclusivamente em séries normalmente distribuídas, inadequadas para análise de extremos. Os poucos trabalhos que avaliam o poder do teste empregando outras distribuições teóricas de probabilidade não realizaram as análises com o grau de detalhamento realizado neste trabalho.

Esses resultados sugerem também que séries de máximos ou mínimos anuais com 30 anos de dados, valor esse geralmente indicado na literatura como adequado para estudos de detecção de tendência, podem não ser adequadas para identificar tendências que estejam presentes na série. Por exemplo, considerando uma série GEV com $C_v = 1$, valor alto, porém corriqueiro em séries de extremos, e uma tendência monotônica de 20% por década em relação à média da série, um valor bastante elevado e com repercussões práticas importantes, o teste MK irá detectar a tendência, em média, em apenas 45%, 20% e 17% do tempo, para valores de $\kappa = -0,30$, $\kappa = 0,0$ e $\kappa = +0,30$, respectivamente. Esses valores de poder do teste MK são relativamente baixos, fato possivelmente desconhecido da maioria dos usuários do teste MK. O aumento de poder do teste MK é significativo quando o tamanho da série aumenta de 30 para 50 anos. Por exemplo, mantidos os outros fatores constantes, quando $n = 50$, o teste MK passa a detectar a tendência, em média, em 90%, 60% e 50% do tempo, para valores de $\kappa = -0,30$, $\kappa = 0,0$ e $\kappa = +0,30$, respectivamente.

Essas análises com base em simulação Monte Carlo permitiram também concluir que em situações onde o poder do teste MK é baixo, a probabilidade de se estimar a magnitude da tendência com o sinal trocado (erro Tipo S) empregando o estimador proposto por Sen (1968), muito utilizado em análises desse tipo, não é desprezível. Verificou-se que a probabilidade do erro Tipo S pode variar de 5% a 45% para valores de poder do teste inferiores a 10%, valor esse observado, por exemplo, quando a série é distribuída por uma distribuição Gumbel ($\kappa = 0$), com 30 anos de dados, $C_v = 1$, e tendência monotônica de 10% por década em relação à média da série. A estimativa de magnitude de tendência com sinal trocado pode causar interpretações equivocadas sobre os resultados de análise de tendência e pode resultar em decisões equivocadas sobre investimentos em adaptação. Ainda em relação à magnitude da tendência, este trabalho concluiu que o estimador proposto por Sen (1968) tende a superestimar a magnitude (erro Tipo M) quando o poder do teste MK é baixo. O valor estimado da magnitude da tendência, em média, pode ser 1,5 vezes maior que o valor real quando o poder do teste é igual a 0,40, podendo ainda ser até 5 vezes maior do que o valor real, em média, quando o poder do teste é igual a 0,05. A quantificação dos erros Tipo S e Tipo M em estudos de Simulação Monte Carlos com séries GEV com características de extremos hidrológicos é uma inovação deste trabalho.

Um segundo conjunto de estudos com Simulação Monte Carlo procurou entender melhor a questão da multiplicidade de testes em análises de detecção de caráter regional, quando se realiza vários testes de detecção em várias estações localizadas na região, objetivando-se obter uma proporção de estações não-estacionárias. Foi possível com esses testes avaliar o desempenho de estratégias sugeridas na literatura baseadas no conceito do *False Discovery Rate* (FDR), denominadas aqui neste trabalho de BH95 e BH_Adap. Tais estratégias foram sugeridas fora da área de Hidrologia para controlar a taxa de falsa detecção, ou seja, a proporção de estações declaradas erroneamente como não-estacionárias, taxa essa geralmente bastante elevada se o teste MK for aplicado individualmente a cada estação sem a preocupação com a visão de conjunto da região.

Os resultados obtidos aqui, baseados em séries GEV, mostraram, como já era previsto, que as taxas de detecção falsas podem ser muito elevadas quando o teste MK é aplicado em cada estação de forma individualizada, sem preocupação com os erros cometidos para a região como um todo. Mas esses valores dependem de características das séries. Por exemplo, mesmo se a região tiver séries GEV ($\kappa = -0.3$) com 60 anos de dados, a análise individual das estações empregando o teste MK pode resultar em taxas de detecções falsas da ordem 80%, em situações em que o $C_v = 1$ e a magnitude da tendência for da ordem de 5% por década em relação à média da série. Isso significa que de cada 10 estações declaradas não-estacionárias, em média, 8 estações são na verdade estacionárias. Esse tipo de comportamento dos testes pode gerar distorções importantes na identificação de regiões que devam ser analisadas com mais detalhamento, contribuindo para um possível desperdício de recursos materiais e humanos em análises subsequentes do padrão de comportamento hidrometeorológicos na região. Os mesmos resultados mostraram que, de fato, as estratégias BH95 e BH_Adap foram capazes, independente das características da série, de controlar as taxas de falsas detecções no nível previamente definido como sendo aceitável, neste caso, de 5%. Mas o benefício de poder controlar as taxas de falsas detecções, advindo do uso das estratégias BH95 e BH_Adap, possui um custo importante, que é o de ter reduzida a capacidade de detectar as estações que são realmente não-estacionárias. A capacidade das estratégias BH95 a BH_Adap em detectar as estações que são realmente não-estacionárias varia de acordo com as características da série, mas é quase sempre inferior àquela obtida quando se emprega o teste MK de forma individualizada na região. Por exemplo, a

capacidade de BH95 e BH_Adap pode ser de aproximadamente 95%, um valor próximo daquele obtido pela estratégia de empregar o teste MK de forma individualizada na região (100%), no caso da região possuir séries GEV ($\kappa = -0.3$) com 60 anos de dados, $C_v = 1$, e tendência com magnitude de 15% por década em relação à média. Por outro lado, se a região possuir séries GEV ($\kappa = -0.3$) com 40 anos de dados, $C_v = 1$, e tendência com magnitude de 4% por década em relação à média, esses valores diminuem para aproximadamente 50%, um valor bem abaixo dos 90% obtidos pelo teste MK aplicado individualmente. A diferença é que a taxa de falsa detecção é mantida abaixo de 5% quando as estratégias BH95 e BH_Adap são empregadas, enquanto a mesma sobe para 39% quando MK é aplicado de forma individualizado.

Na análise de tendências realizada para a bacia do Rio São Francisco, além da aplicação das técnicas de MK, BH_95, BH_Adap e B_Sen, procurou-se, primeiramente, eliminar a presença da autocorrelação na base de dados de pluviometria e fluviometria empregada. A influência da autocorrelação serial no processo de detecção já foi amplamente discutida na literatura, levando à proposição de algumas abordagens para mitigar seus efeitos, a exemplo do Pre-Whitening (PW) e Trend-Free-Pre-Whitening (TFPW). O PW e o TFPW, associados ao teste de Mann-Kendall, resultaram em redução nas taxas de rejeições obtidas para o emprego exclusivo do MK, independentemente da natureza da variável avaliada. A redução percentual, que ocorreu de maneira diferente para cada uma das metodologias que lidam com a correlação temporal testadas, foi mais significativa para o PW, principalmente, nos índices climáticos associados à vazão mínima (Q7min) – coincidentemente, aquele cuja correlação serial esperada é a maior (0.62). Credita-se essa diferença de comportamento à remoção de parte da tendência da amostra, que ocorre juntamente com a tentativa de eliminação da correlação serial pelo MK-PW. Esse fato, segundo Yue e Wang (2002), é causador de redução no número de detecções, e possivelmente explica as discrepâncias entre o MK-PW e o MK-TFPW constadas.

Diante disso, o emprego da abordagem de MK-PW para lidar com os efeitos da autocorrelação no processo de detecção de tendências parece limitado. Os resultados obtidos ao utiliza-la mostram que mesmo para situações de baixa correlação amostral, por exemplo o RX5d (0.10), a metodologia em questão não é capaz de manter as taxas de rejeição próximas as obtidas pelo MK-TFPW para um mesmo nível de correlação temporal.

Reportado por Yue *et al.* (2002), Douglas *et al.* (2000), Bayazit e Onoz (2007) e mostrado como suscetível a presença de tendências estatisticamente significantes na amostra, principalmente de grande magnitude, o MK-PW deve ser preterido em relação ao MK-TFPW. O uso do MK-TFPW é preferível até em situações em que o tamanho amostral da série testada é reduzido, devido à possibilidade da superestimativa da magnitude da tendência, erro tipo S. do estimador proposto por Sen (1968).

Chama a atenção o elevado número de estações com tendências decrescentes nas séries de vazões mínimas (Q7min), com uma proporção de aproximadamente 50%. Na maioria dessas estações, a magnitude da tendência ultrapassa a taxa de 10% por década. Esses resultados foram obtidos com o emprego da técnica de Trend-Free Pre-Whitening (TFPW) para lidar com a autocorrelação temporal e com o uso dos métodos baseados no conceito de FDR para manter a proporção de falsos positivos menor do que 5%. O mesmo não pode ser dito para os índices associados à precipitação, já que apenas um (SDII) dentre os dez testados apresentaram taxas de rejeição decrescentes ligeiramente superiores a 10% após o uso do TFPW sucedido pelo BH_Adap.

Essa constatação se deve às diferenças significativas na magnitude de tendência decenal e na quantidade de pontos julgados como não-estacionários. A fraca dependência entre o padrão comportamental das tendências de chuva e de vazão pode ser vista em regiões específicas da bacia, tais como o extremo Sul do Alto São Francisco e Oeste da Bahia. A redução no número de dias de chuva (R1) foi evidenciada pela aplicação do MK-TFPW, mas ficou mais clara após o uso do BH_95 e BH_Adap. Essa diminuição contribuiu ainda para o crescimento de SDII, e, portanto, para o decréscimo dos índices de vazão de algumas das sub-regiões da área de estudo. Chama a atenção que, diferente da tendência negativa irrestrita vigente no Oeste baiano, o extremo sul do Alto São Francisco apresenta diferentes comportamentos para as descargas máximas, médias e mínimos – respectivamente, crescente, não-estacionário e decrescente.

Ventura *et al.* (2004) sugerem que o impacto da correlação espacial não afeta de maneira tão significativa os resultados obtidos pelas técnicas que empregam o *FDR*. Apesar de esse aspecto não ter sido abordado pelo presente trabalho, acredita-se que é necessário avaliar melhor a constatação desses autores. Aplicar diferentes metodologias baseadas no controle

pelo FDR à base de dados sintética correlacionada espacialmente, por exemplo, consiste em tarefa importante para melhor conhecimento de suas limitações técnicas e de desempenho, em que conhecer os pontos fracos e forte de determinada metodologia submetida às mais diversas condições é a diferença entre a obtenção de resultados coerentes e equivocados.

A tentativa de formar *clusters* de estações e, posterior, aplicação das técnicas de *FDR* aos grupos homogêneos formados é outra linha a ser estudada em trabalhos futuros. Apesar de Sun *et. al.* (2014) indicarem que essa proposição é potencialmente benéfica, a mesma carece de maiores definições que envolvem não só a escolha da técnica de clusterização a ser empregada, bem como qual das inúmeras metodologias de controle de *FDR* é a mais adequada para tal atividade.

O possível próximo passo a ser dado na continuidade da avaliação dos erros Tipo S e M na Hidrologia é estender a análise de *B_Sen* para outras técnicas de estimativa de magnitude de tendência, bem como para séries que seguem distribuições estatísticas diferentes, como a LP3, e a normal, por exemplo, ao invés da GEV utilizada. Comparar o desempenho de outras metodologias de estimativa de tendência e potenciais diferenças entre os resultados obtidos pelas mesmas, via simulação Monte Carlo (conforme realizado na seção 5.2), pode resultar na indicação de uso de uma técnica em detrimento de outra, quando se deseja evitar os erros tipo S e M. Espera-se também que o uso de diferentes distribuições possibilite uma maior generalização dos resultados até então obtidos, permitindo que as avaliações dos erros tipo S e M não se restrinjam a extremos, mas a situações em que o interesse seja, por exemplo, em variáveis hidrológicas de maior frequência e, portanto, assumidas como detentoras de outras distribuições que não à GEV.

Já para o caso da continuidade dos estudos desenvolvidos para a bacia do São Francisco, entende-se ser necessário avaliar a relação entre as mudanças detectadas e fenômenos climáticos de larga escala, por exemplo, ENSO (*El-Nino Southern Oscillation*), PDO (*Pacific Decadal Oscillation*) e ADO (*Atlantic Decadal Oscillation*), e aprofundar nas avaliações que considerem alterações no uso e ocupação do solo como causadoras da ruptura dos padrões estacionários.

REFERÊNCIAS BIBLIOGRÁFICAS

- Aires de Souza, S., Araújo, A. A., Troger, F. H. (2013). “Análise de Estacionariedade das Séries Hidrológicas no local do future aproveitamento hidrelétrico São Luiz do Tapajós. In: “XI Simpósio de Recursos Hídricos do Nordeste, 2012.
- Alves, B. C. C., Souza, F. A. S., Silveira, C. S. (2013). “Análise de Tendências e Padrões de Variação das Séries Históricas de Vazões do Operador Nacional do Sistema Elétrico (ONS)”. In: RBRH – Revista Brasileira de Recursos Hídricos, vol. 18, n. 4 (Out/Dez), 19 – 34.
- Alkama, R.; Marchand, L., Ribes, A., Decharme, B. (2013). “Detection global runoff changes: results from observations and CMIP5 experiments”. In: Hydrol. Earth Syst. Sci., 17, 2967 – 2979.
- Alpert, P., Ben_Gai, T., Baharad, A., Benjamini, Y., Yekutieli, D., Colacino, M., Diodato RC., Homar, V., Romero, R., Michaelides, S., Manes, A. (2002). “The paradoxical increase of Mediterranean extrem daily rainfall in spite of decrease in total values”. In: Geophys Res. Lett 29: 311-314.
- Araújo, A. A., Aires de Souza, S. (2014). “Análise de Estacionariedade nas series hidrometeorológicas da bacia do rio Grande aplicando a técnica de reamostragem bootstrap”. In: XII Simpósio de Recursos Hídricos do Nordeste, Natal.
- Bayazit, M. (2015). “Nonstationary of hydrological records and recent trends in trend analysis: A State-of-the-art review”. In: Environ. Process. 2, 527 – 542.
- Bayazit, M. e B. Onoz, (2007). “To prewhiten or not prewhiten in trend analysis?”. In: Hydrological Science Journal, v.52, p.611-624
- Benjamini, Y. e Y. Hochberg, (1995). “Controlling the false Discovery rate: A practical and powerful approach to multiple testing”. In: J. Roy. Stat. Soc., 57B, 289-300
- Benjamini, Y. e Y. Hochberg, (1997). “Multiple Hypotheses Testing with weights”. In: Scand. J. Statist. 24, 407-418
- Benjamini, Y. e Y. Hochberg, (2000). “The adaptive control of the false discovery rate in multiple hypotheses testing”. In: J. Behav. Educ. Statist., 25, 60-83
- Benjamini, Y. e D. Yekutieli, (2001). ” The control of the false discovery rate in a multiples testing under dependency”. In: Ann. Stat., 29, 1165-1188

- Benjamini, Y., Krieger, M. A. e Yekutieli, D. (2006) “Adaptive linear step-up procedures that control the false discovery rate”. In: *Biometrika*, 93, 491–507.
- Birkmann, J. e Welle, T. (2015). “Assessing the risk of loss and damage: exposure, vulnerability and risk to climate-related hazards for different country classifications.” In: *Int. J. Global Warming*, X, No. Y
- Blain, G.C. (2014).”Removing the influence of the serial correlation on the Mann-Kendall test”. In: *Revista Brasileira de Meteorologia*, v.29, n.2, 161-170.
- Brown, A.E., Zhang, L., McMahon, T.A., Western, A.W., Vertessy, R.A. (2005). “A review of paired catchment studies for determining changes in water yield resulting from alterations in vegetation.” In: *Journal of Hydrology*, 310 (1–4), 28–61.
- Cammalleri, C., Micale, F., Vogt, J.V. (2016). “Recent temporal trend in modelled soil water deficit over Europe driven by meteorological observations”. In: *Int. J. Climatology*. DOI: 10.1002/joc.4677
- Chiew, F. H. S., e McMahon, T. A. (1993). “Detection of trend and change in annual flows of Australian rivers”. In: *Int. J. of Climatology*, 13, 643 – 653.
- Detzel, D. H. M., Bessa, M. R. B., Vallejos, C.A.V., Santos, A.B., Thomsen, L.S., Mine, M. R. M., Bloot, M. L., Estrócio, J.P. (2011). “Estacionariedade das Afluências às Usinas Hidrelétricas Brasileiras”. In: *RBRH – Revista Brasileira de Recursos Hídricos*, vol. 16, n. 3 (Jul/Set), 95 – 111.
- Dooge, J.C.I. (1992). “Hydrologic models and climate change.” In: *Journal of Geophysical Research*, 97 (D3), 2677–2686.
- Douglas, E. M., Vogel, R. M. e Kroll, C. N. (2000). “Trends in flood and low flows in the United States: impact of spatial correlation”. In: *J. Hydrol.*, 240, 90 – 105.
- Efron, B. (2008). “Microarrays, empirical Bayes and the two groups model.” In: *Statist. Sci.*, 23, 1–22
- Fatichi, F., Caporali, E. (2009). “A comprehensive analysis of changes in precipitation regime in Tuscany”. In: *Int. J. Clymatol.*, 29, 1883-1893.
- Ganguli, P., Ganguly, A.R. (2015).”Space-time trends in U.S. meteorological droughts”. In: Google Scholar. ArXiv: 1512.08526.
- Gelman, A. e Carlin, J. (2014).”Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors”. In: *Perspectives on Psychological Science*, vol. 9(6), 641-651.

- Genovese, C. e Wasserman, L. (2002). “Operating characteristics and extensions of the false discovery rate procedure.” In: *J. R. Statist. Soc. B*, 64, 499–517.
- Gudmundsson, L., Seneviratne, S. I. (2015): “European drought trends”. In: *Proc. IAHS*, 369, 75–79, proc-iahs.net/369/75/2015/doi:10.5194/piahs-369-75-2015.
- Hamed, K.H. (2009). “Exact Distribution of the Mann–Kendall Trend Test Statistic for Persistent Data.” In: *Journal of Hydrology*, 365, 86–94.
- Hamed, K.H. e Hao, A.R. (1998). “A Modified Mann-Kendall Trend Test for Autocorrelated Data.” In: *Journal of Hydrology*, 294, 182-196.
- Hay J.E., Easterling D., Ebi K.L, Kitoh A., Parry M. (2015). “Introduction to the special issue: Observed and projected changes in weather and climate extremes”. In: *Weather Clim. Ext.*, 11 (2015), pp. 1–3
- Hockberg, Y., (1988) “A sharper Bonferroni procedure for multiple tests of significance”. In: *Biometrika*, 75, 800 – 802.
- Hochberg, Y. e Benjamini, Y. (1990) “More powerful procedures for multiple significance testing”. In: *Statist. Med.*, 9, 811–818.
- Holm, S., (1979). “A simple sequentially rejective multiple test procedure”. In: *Scand. J. Statist.*, 6, 65-70
- Hossain, F. (2014). “Paradox of peak flows in Changing Climate”. In: *J. Hydrol. Eng. ASCE*, 19(9): 02514001.
- Humphrey, V., Gudmundsson, L., Seneviratne, S. I., (2016). “Assessing global water storage variability from GRACE: trends, seasonal cycles, subseasonal anomalies e extremes”. In: *Surv Geophys*, 37, 357-395. doi 10.1007/s10712-016-9367-1
- Khaliq, M. N., Ouarda, T. B. M. J., Gachon, P., Sushama, L., St-Hilare, A., (2009). “Identification of hydrological trends in the presence of serial and cross correlations: A review of selected methods and their application to annual regimes of Canadian rivers”. In: *J. Hydrology*, 368, 117 – 130.
- Katz, R. W.,(2002). “Sir Gilbert Walker and a connection between El Niño and statistics”. In: *Stat. Sci.*, 17, 97-112
- Kendall, M. G. (1975). “Rank Correlation Methods”. Griffin, London, UK
- Kulkarni, A. e von Storch, H. (1995). “Monte Carlos Experiments on the Effect of Serial Correlation on the Mann-Kendall Test of Trend.” In: *Meteorologische Zeitschrift* 4(2):82-85, January.

- Kundzewicz, Z.W., Mata, L.J., Arnell, N.W., Döll, P., Kabat, P., Jiménez, B., Miller, K.A., Oki, T., Sen, Z., Shiklomanov, I.A. (2007). “Freshwater resources and their management.” In: Chapter in *Climate Change 2007: Impacts, Adaptation and Vulnerability*, M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden, and C.E. Hanson, eds. Cambridge University Press, Cambridge, UK, pp. 173–210.
- Krakauer, N. Y., e Fung, I. (2008). “Mapping and attribution of change in streamflow in the coterminous United States”. In: *Hydrol. Earth Syst. Sci.*, 12, 1111 – 1120.
- Lehmann, E.L. (1975). “Nonparametric, Statistical Methods Based on Ranks”. In: Holden-Day, Inc, California.
- Lettenmaier, D. P., Wood, E. F., e Wallis, J. R. (1994). “Hydro-climatological trends in the continental United States”. In: *Journal of Climate*, 7(4), 586 – 607.
- Livezey, R., e W. Chen (1983): “Statistical field significance and its determination by Monte Carlo techniques”. In: *Mon. Wea. Rev.*, 111, 46-59
- Madsen, H. Lawrence, D., Lang, M., Martinkova, M. Kjeldsen T. (2014). “Review of trend analysis and climate change projections of extreme precipitation and floods in Europe”. In: *J. Hydrol.*, 519, 3614 – 3625.
- Mallya, G., Mishra, V., Niyogi, D., Tripathi, S., Govindaraju, R.S., (2016). “Trends and variability of droughts over the Indian monsoon region”. In: *Weather and Climate Extremes*, 12, 43-68.
- Mann, H. B. (1945). “Nonparametric tests against trend.” *Econometrica*, 13, 245– 259, doi:10.2307/1907187.
- Marcus, R., Peritz, E., Gabriel, KR., (1976). “On closed testing procedure with special reference to ordered analysis of variance”. In: *Biometrika*, 63, 655 – 660.
- Martins, E. S e Stedinger, J. R. (2000). “Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data”. In: *Water Resources Research*, 36, no.3, 737 - 744
- Merz, B., Nguyen, V. D., Vorogushyn, S. (2016). “Temporal clustering of floods in Germany: Do flood-rich and flood-poor periods exist?” In: *J. Hydrol.*, 541, 824-838.
- Miao, C. Y., Shi, W., Chen, X. H., Yang, L. (2012). “Spatio-temporal variability of streamflow in the Yellow river: possible causes and implications”. In: *Hydrological Sciences Journal*, 57(7), 1355–1367

- Mueller, I.I, Kruger, C.M., Kaviski, E. “Análise de Estacionariedade de Séries Hidrológicas na bacia incremental de Itaipu”. In: RBRH – Revista Brasileira de Recursos Hídricos, vol. 3, n. 4 (Out/Dez), 51 – 71.
- Naghetini, Mauro e Pinto, Éber José de Andrada, (2007). “Hidrologia Estatística”. CPRM, Belo Horizonte
- Onoz, B., e Bayazit, M. (2012). “Block Bootstrap for Mann-Kendall Trend Test of Serially Dependent Data.” *Hydrological Processes*, 26(23), doi:10.1002/hyp.8438.
- Paciorek, C., Risbery, J., Ventura, V., e Rosen, R. (2002). “Multiple indices of Northern Hemisphere cyclone activity, winters 1949 – 1999”. In: *J. Climate*, 15, 1573 – 1590.
- Peterson, T.C., (2005). “Climate Change Indices”. In: *WMO Bulletin*, 54 (2), 83-86.
- Pittock, A. B., (1980). “Monitoring, causality and uncertainty in a stratospheric context”. In: *Pageoph*, vol. 118, 643 – 660.
- PRH – BSF: Plano de Recursos Hídricos da Bacia Hidrográfica do Rio São Francisco (2016 – 2025), vol.1, 237 p.
- Radziejewski, M. e Kundzewicz, Z. W., (2004), “Detectability of Changes in Hydrological records”. In: *Hydrol. Sci. J.*, 49 (1), 39 – 51.
- Renard, B., Lang, M., Bois, P., Dupreyat, A., Mestre, O., Niel, H., Sauquet, E., Prudhomme, C., Parey, S., Paquet, E., Neppel, L., Gailhard, J. (2008). “Regional methods for trend detection: Assessing field significance and regional consistency”. In: *Water Resources Research*, vol. 44, W08419, doi:10.1029/2007WR006268.
- Rosin, C., Amorim, R. S. S., Morais, T. S. T., (2015). “Análise de Tendências Hidrológicas na bacia do rio das Mortes”. In: – *Revista Brasileira de Recursos Hídricos – Versão on-line*, vol. 20, n. 4 (Out/Dez), 991 – 998.
- Scofield, G. B., Angelis, C. F., Souza, W. C. (2014). “Estudo das Tendências do Total de Precipitação e do Número de dias chuvosos no Litoral Norte de São Paulo”. In: RBRH – *Revista Brasileira de Recursos Hídricos*, vol. 19, n. 3 (Jul/Set), 19 – 31.
- Sen, P.K., (1968). “Estimates of regression coefficient based on Kendall's tau”. In: *J. Am. Stat.Assoc.*, 63, 1379-1389.
- Seneviratne, SI., Nicholls, N., Easterling, D., Goodess, CM., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McIness, K., Rahimi, M., Reichstein, M., Sorterberg, A., Vera, C., Zhang, X. (2012). “Changes in climate extremes and their impacts on the natural physical environment.” In: *Field CB, Barros, V., Stocker, TF, Qin, D., Dokken, D., Ebi, KL.*

- Mastrandea, MD., Mach, KJ., Plattner, G-K, Allen, SK., Tignor, M., Midgley, PM. (Eds) Intergovernmental Panel on Climate Change special report on managing the risks of extreme events and disasters to advance climate change adaptation. Cambridge University Press, Cambridge/New York
- Schweder, T. e Spjøtvoll, E. (1982). “Plots of p-values to evaluate many tests simultaneously”. In: *Biometrika*, 69, 493–502.
- Seeger, P. (1986). “A note on a method for the analysis of significance en masse”. In: *Technometrics*, 10, 586 – 593.
- Shafer, J. P. (1995). “Multiple Hypothesis Testing”. In: *Annu. Rev. Psychol*, 46, 561 – 584.
- Simes, R. J. (1986). “An improved Bonferroni procedure for multiple tests of significance”. In: *Biometrika*, 73, 751 – 754.
- Soriç, B. (1989) “Statistical “discoveries” and effect size estimation”. In: *J. Am. Statist. Ass.*, 84, 608–610.
- Storey, J. D. (2002) “A direct approach to false discovery rates.” In: *J. R. Statist. Soc. B*, 64, 479–498.
- Storey, J. D. e Tibshirani, R. (2003) “Statistical significance for genomewide studies”. In: *Proc. Natn. Acad. Sci. USA*, 100, 9440–9445.
- Storey, J. D., Taylor, J. E. and Siegmund, D. (2004) “Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach.” In: *J. R. Statist. Soc. B*, 66, 187–205.
- Storey, J. D. and Tibshirani, R. J. (2003) Statistical significance for genome-wide experiments. *Proc. Natn. Acad.*
- Sun, W., Reich, B. J., Cai, T. T., Guindani, M. e Schwartzman, A. (2014). “False Discovery Control in Large-Scale Spatial Multiple Testing”. In: *Journal of Royal Statistical Society: Series B (Statistical Methodology)*.
- Ventura, V., Paciorek, C.J., and Risbey, J.S. (2004). “Controlling the Proportion of Falsely Rejected Hypotheses when Conducting Multiple Tests with Climatological Data.” In: *Journal of Climate*, 17(22), 4343-4356.
- Vogel, R. M., Yaoundl, C. Walter, M. (2011). “Nonstationarity: flood magnification and recurrence reduction factors in the United States”. In: *J. Amer. Water Resources Assoc.*, 47, 464 – 474.

- Von Storch H (1995). “Misuses of statistical analysis in climate research. Analysis of Climate Variability: Applications of Statistical Techniques”. In: ed H.v.Storch, A.Navarra, Springer-Verlag, New York: 11–26
- WMO (World Meteorological Organization). (2009). Guidelines on analysis of extremes in a changing climate in support of informed decisions for adaptation, Climate Data and Monitoring, WCDMP-No. 72
- Yekutieli, D. e Y. Benjamini (1999). “Resampling –based false discovery rate controlling multiple test procedures for correlated test statistics”. In: J. Stat. Plan. Inf., 82, 171-196
- Yue, S. e Wang, C.Y. (2002). “Applicability of Prewhitening to Eliminate the Influence of Serial Correlation on the Mann-Kendall Test.” In: Water Resources Research, 38(6), 1068, 10.1029/2001WR000861.
- Yue, S., Pilon, P., and Cavadias, G. (2002a). “Power of Mann-Kendall and Spearman’s Rho Tests for Detecting Monotonic Trends in Hydrological Series.” In: Journal of Hydrology, 259, 254-271.
- Yue, S., Pilon, P., Phinney, B., and Cavadias, G. (2002b). “The Influence of Autocorrelation on the Ability to Detect Trend in Hydrological Series.” In: Hydrological Processes, 16, 1807-1829.
- Yue, S. e Pilon, P., (2004); “A Comparison of the Power of the t Test, Mann-Kendall and Bootstrap Tests for Trend Detection.” In: Hydrological Sciences–Journal–des Sciences Hydrologiques, 49(1), February.
- Zhang, X., Alexander, L., Hegerl, G.C., Jones, P., Tank, A.K., Peterson, T.C., Trewin, B. and Zweirs, F.W. (2011). “Indices for monitoring changes in extremes based on daily temperature and precipitation data.” In: WIREs Clim Change. doi: 10.1002/wcc.147.
- Zwiers, F. W., Alexander, L. V., Hegerl, G. C., Knutson, T. R., Kossin, J.P., Naveau, P., Nichols, N., Schär, C., Seneviratne, S. I. e Zhang, X. (2011). “Challenges in estimating and understanding recent changes in the frequency and intensity of extreme climate and weather events”. In: Proc. World Climate Research Programme Open Science Conf., Denver, CO, WCRP, 45 pp. [Available online at http://library.wmo.int/pmb_ged/wcrp_2011-zwiers.pdf.

**APÊNDICE I – ESTAÇÕES PLUVIOMÉTRICAS E
FLUVIOMÉTRICAS**

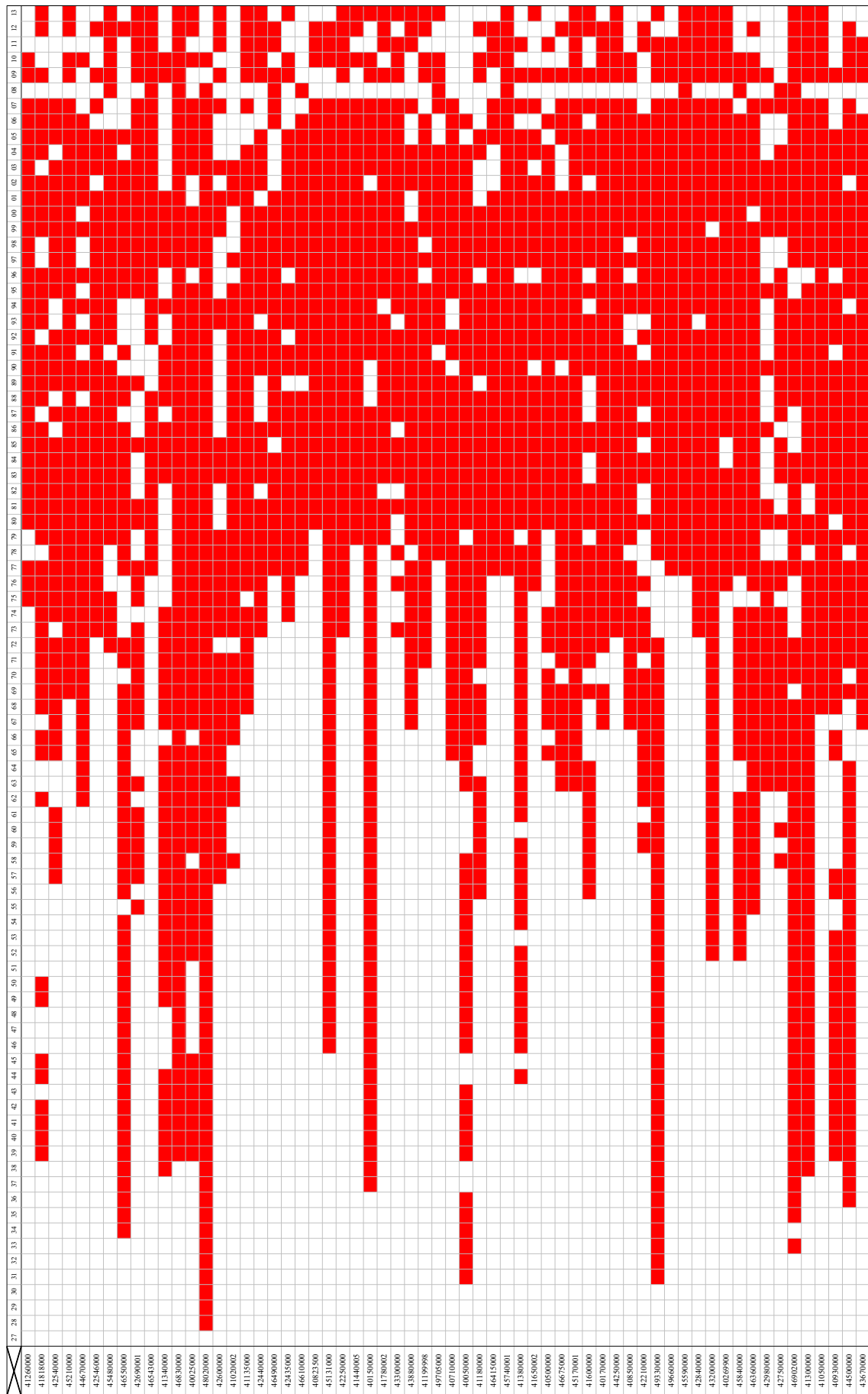
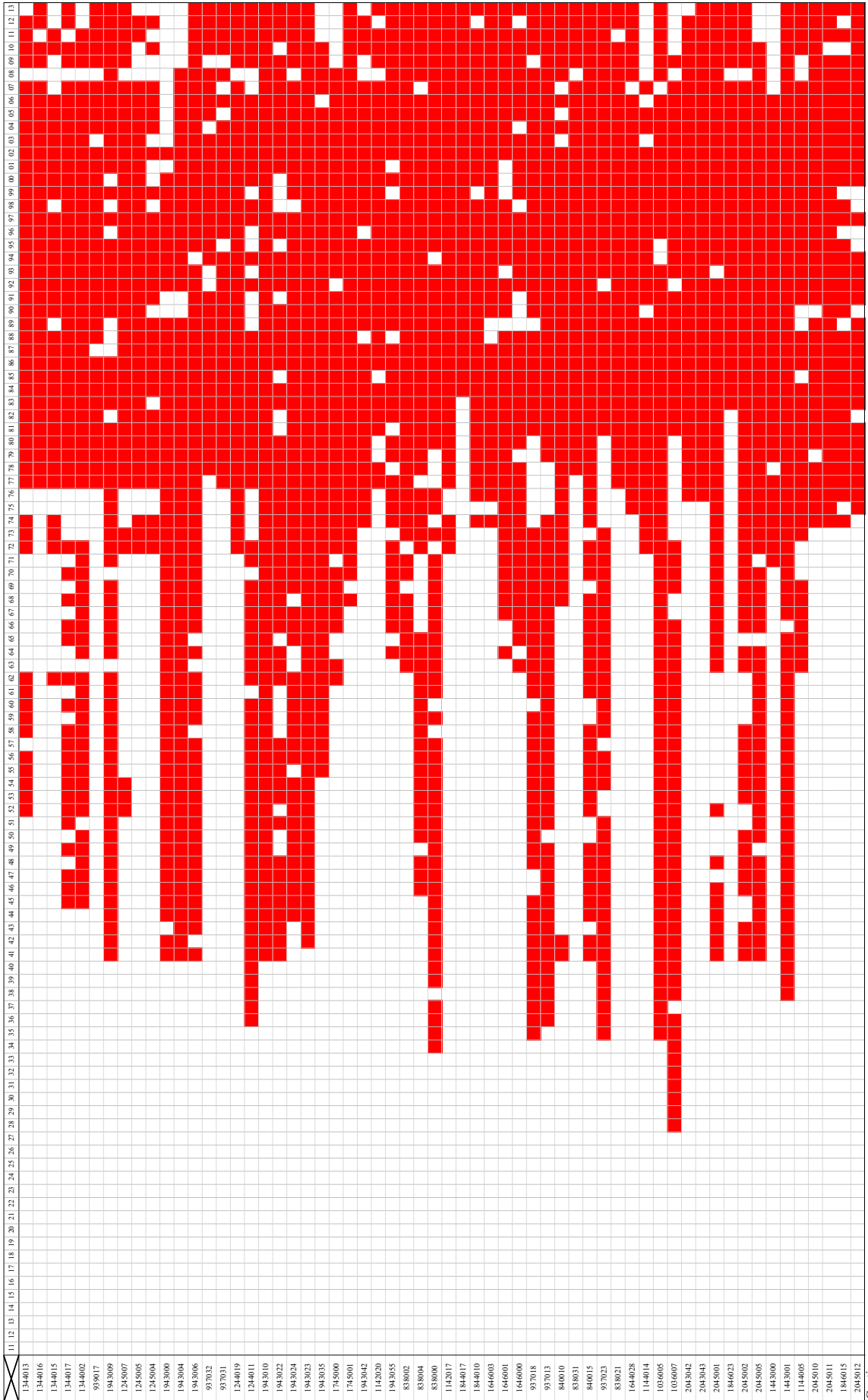


Figura 1.2 - Anos de Disponibilidade de Dados (vermelho) das Estações Pluviométricas que integram a Base Amostral utilizada. A 1ª linha refere-se ao período de anos (1911 - 2013) e a 1ª coluna são os códigos ANA/ONS dos postos
 * Estações excluídas da base amostral devido à presença de barragem



Continua

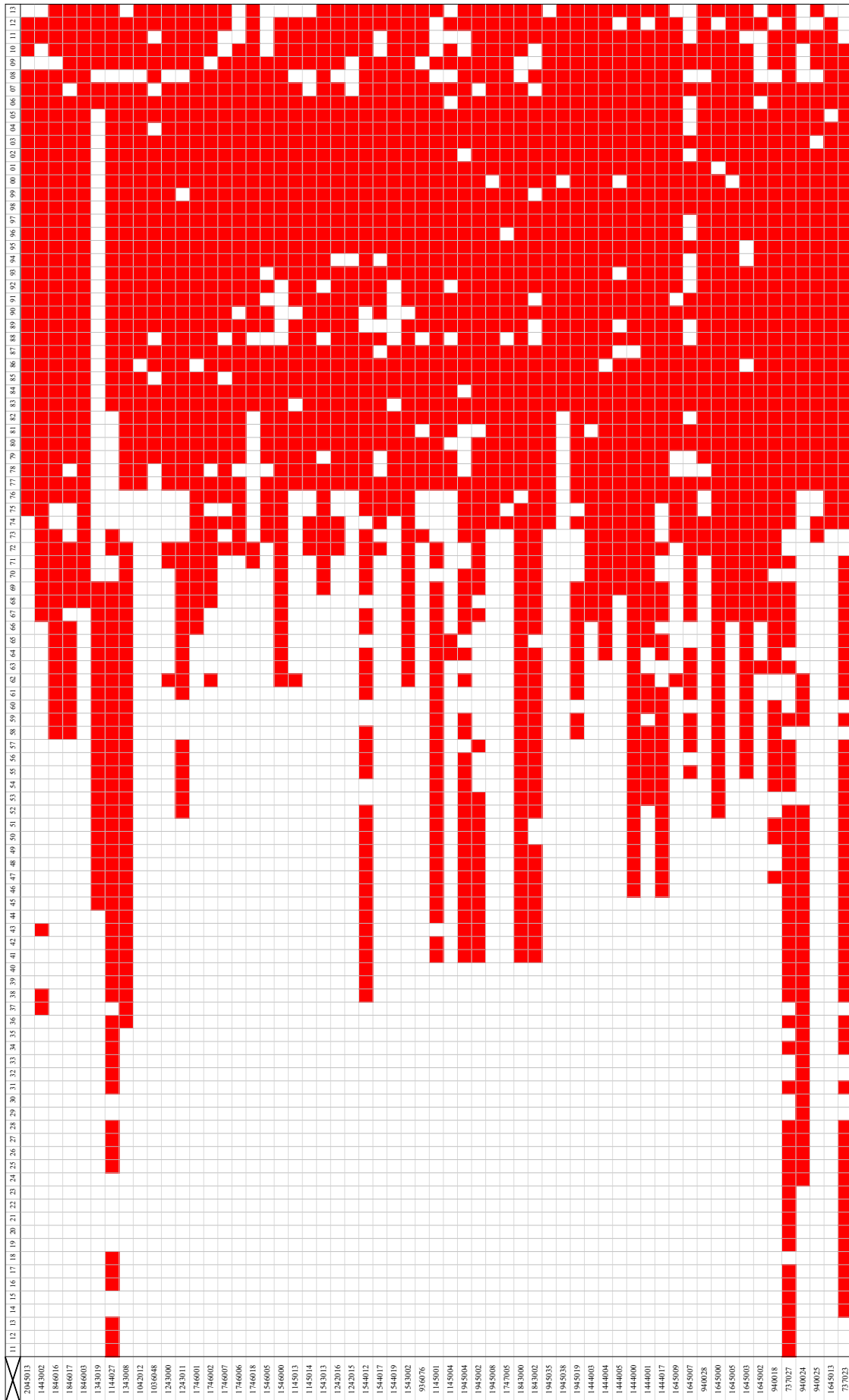


Figura 1.1 - Anos de Disponibilidade de Dados (vermelho) das Estações Pluviométricas que integram a Base Amostral utilizada. A 1ª linha refere-se ao período de anos (1911 - 2013) e a 1ª coluna são os códigos ANA/ONS dos postos

**APÊNDICE II – FIGURAS RESULTADOS MK-TFPW, BH95 E
BH_ADAP**

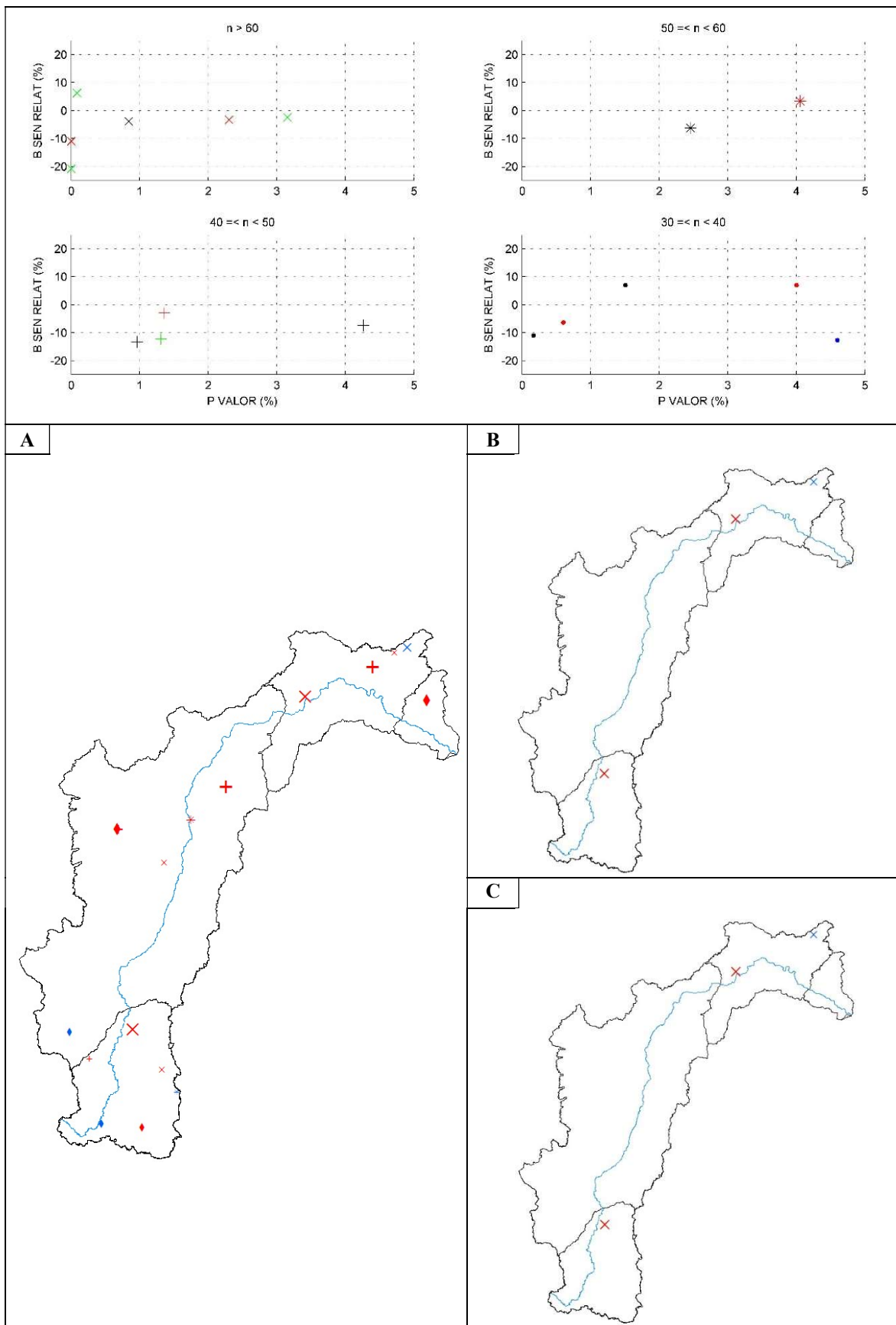


Figura II.1 - Estações significativas para PRCPTot após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

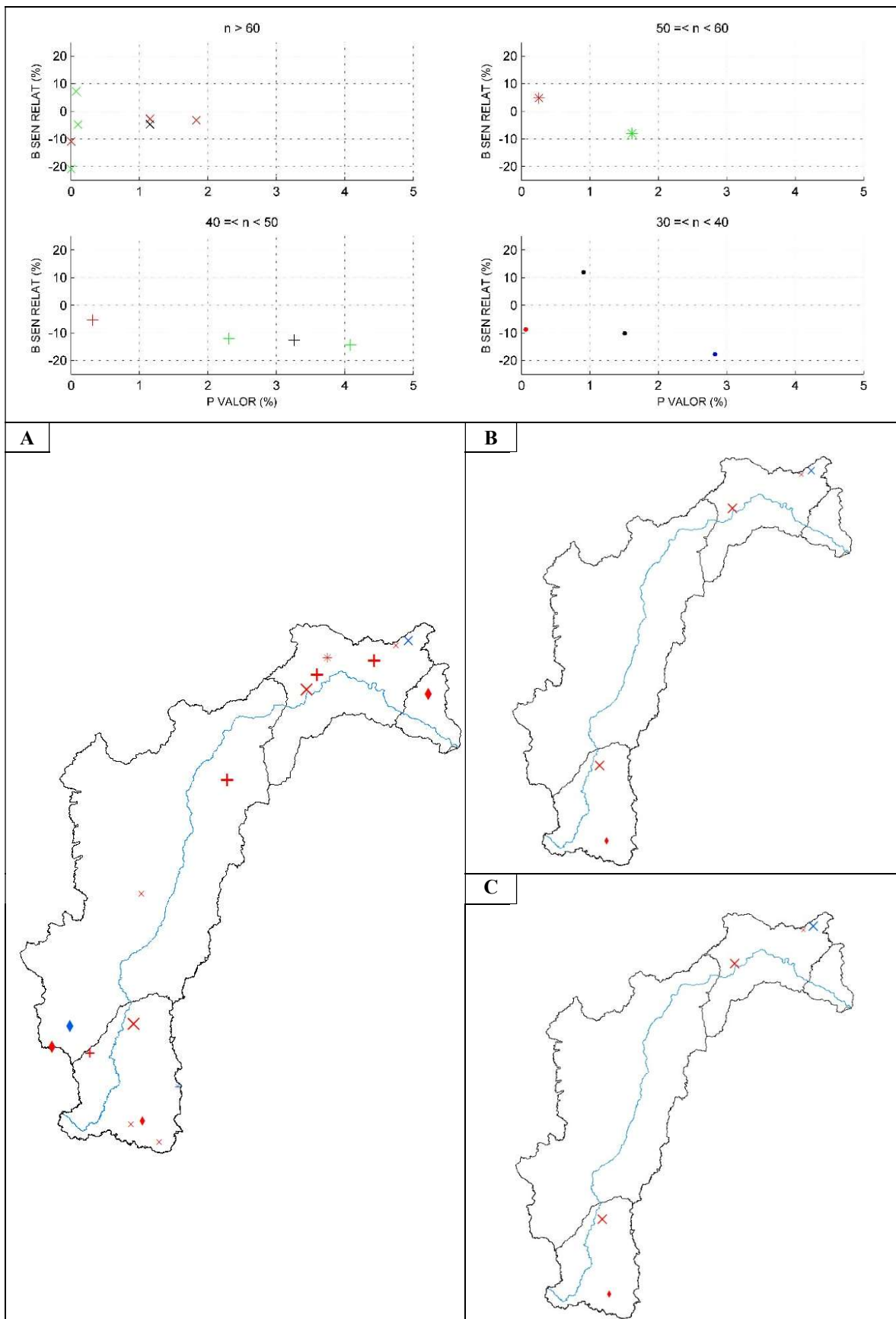


Figura II.2 - Estações significativas para R20 após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

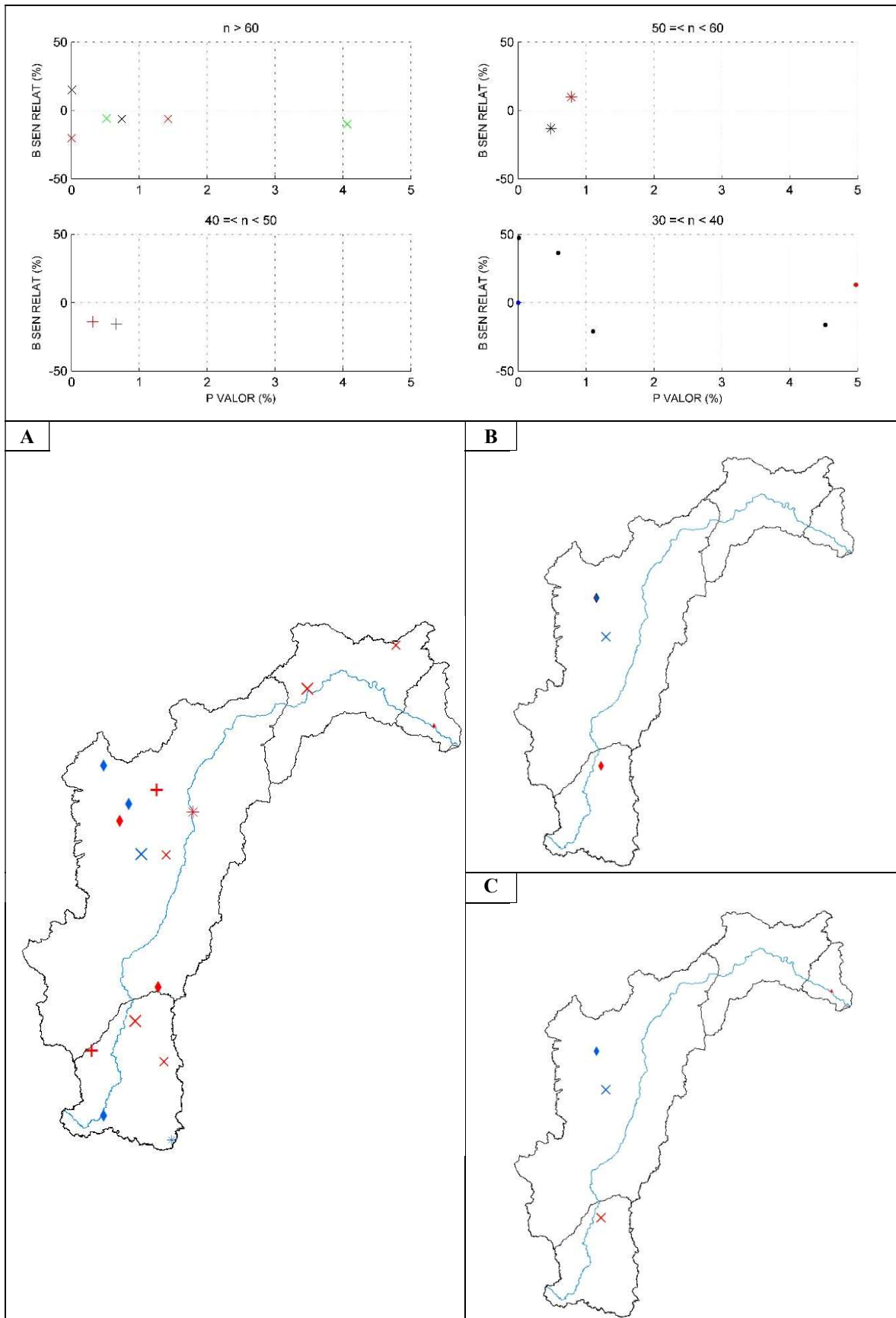


Figura II.3 - Estações significativas para R50 após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_Sen_relativos dos pontos de monitoramento julgadas significativas por (a)

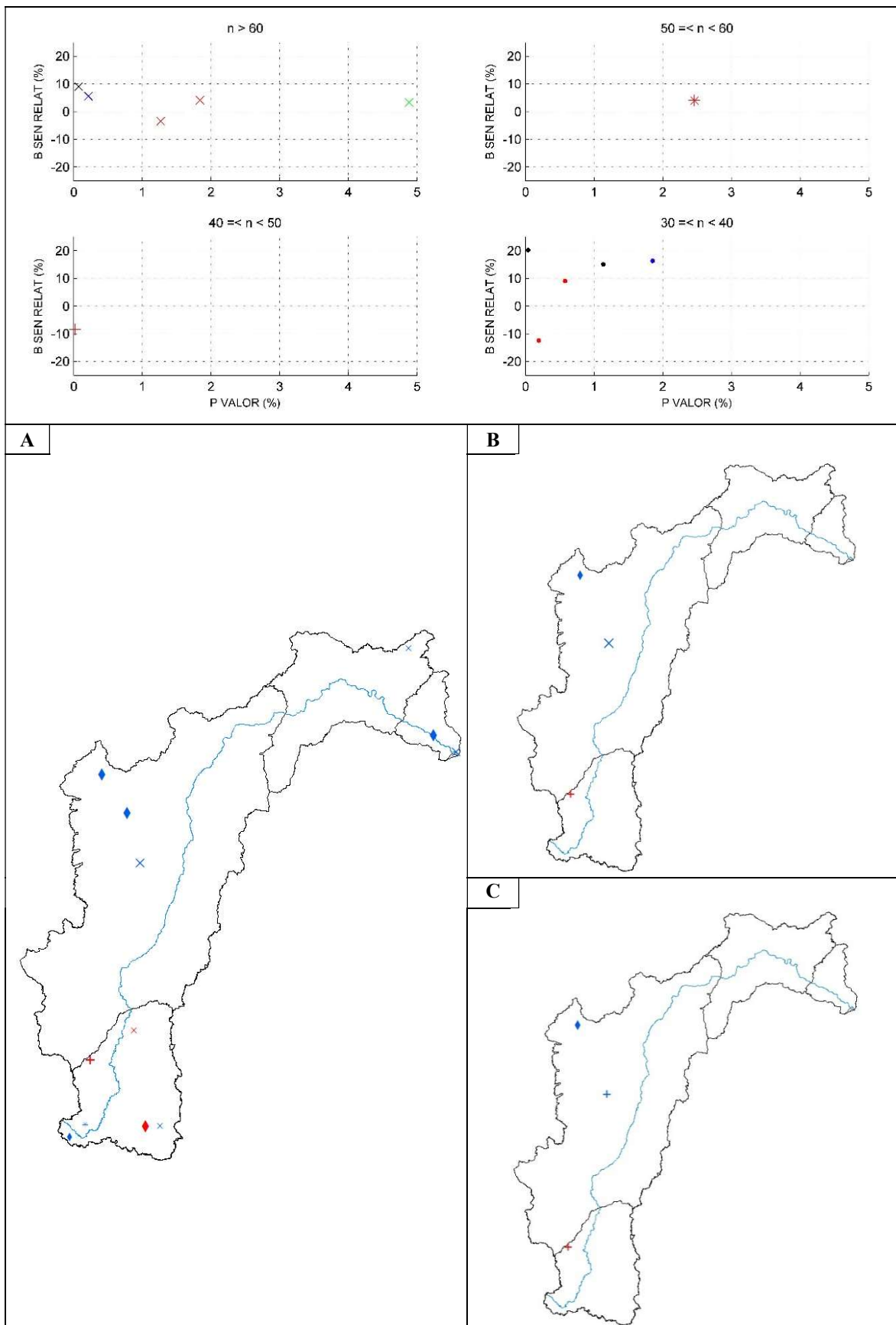


Figura II.1.4 - Estações significativas para RX1d após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

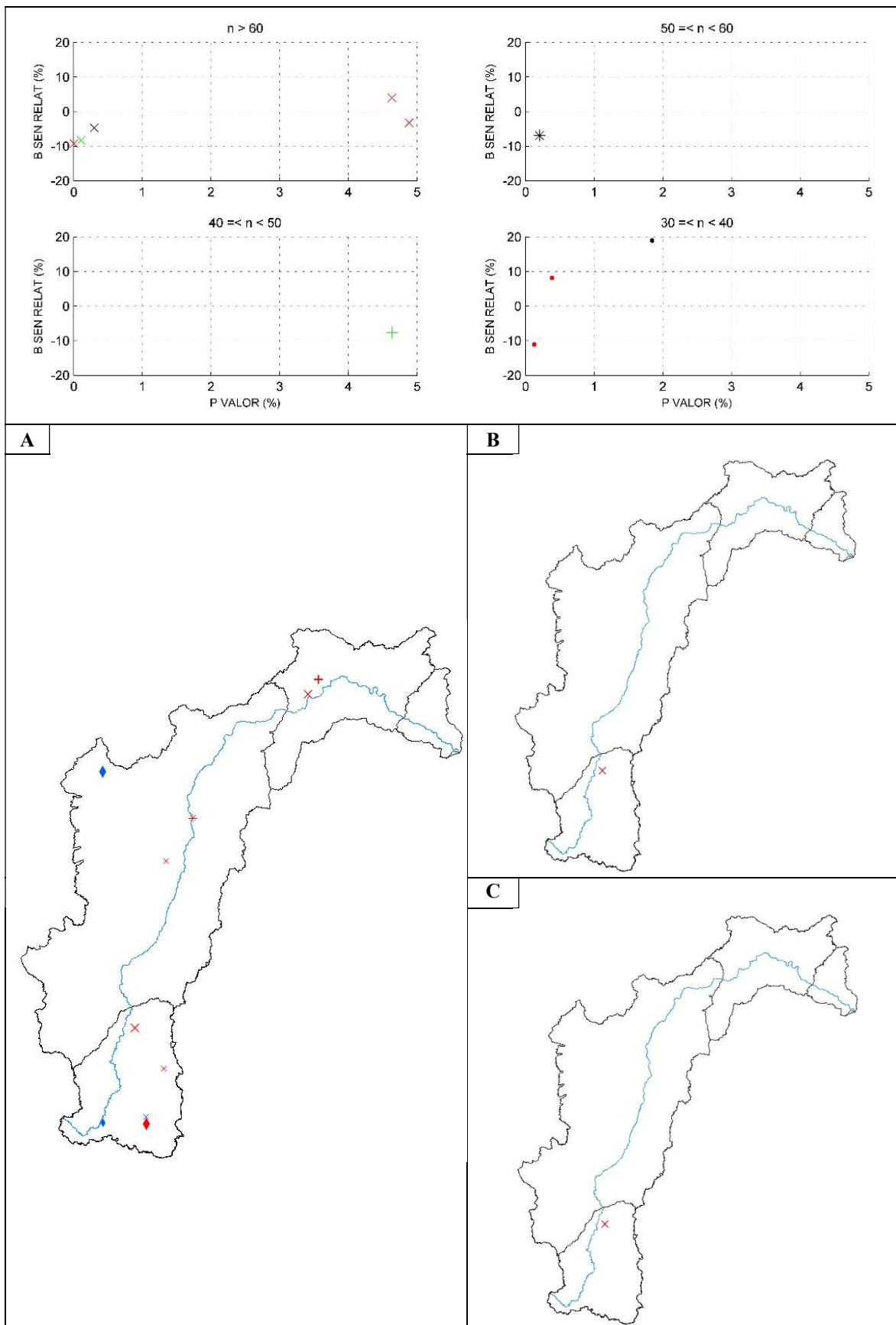


Figura II.5 - Estações significativas para RX5d após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

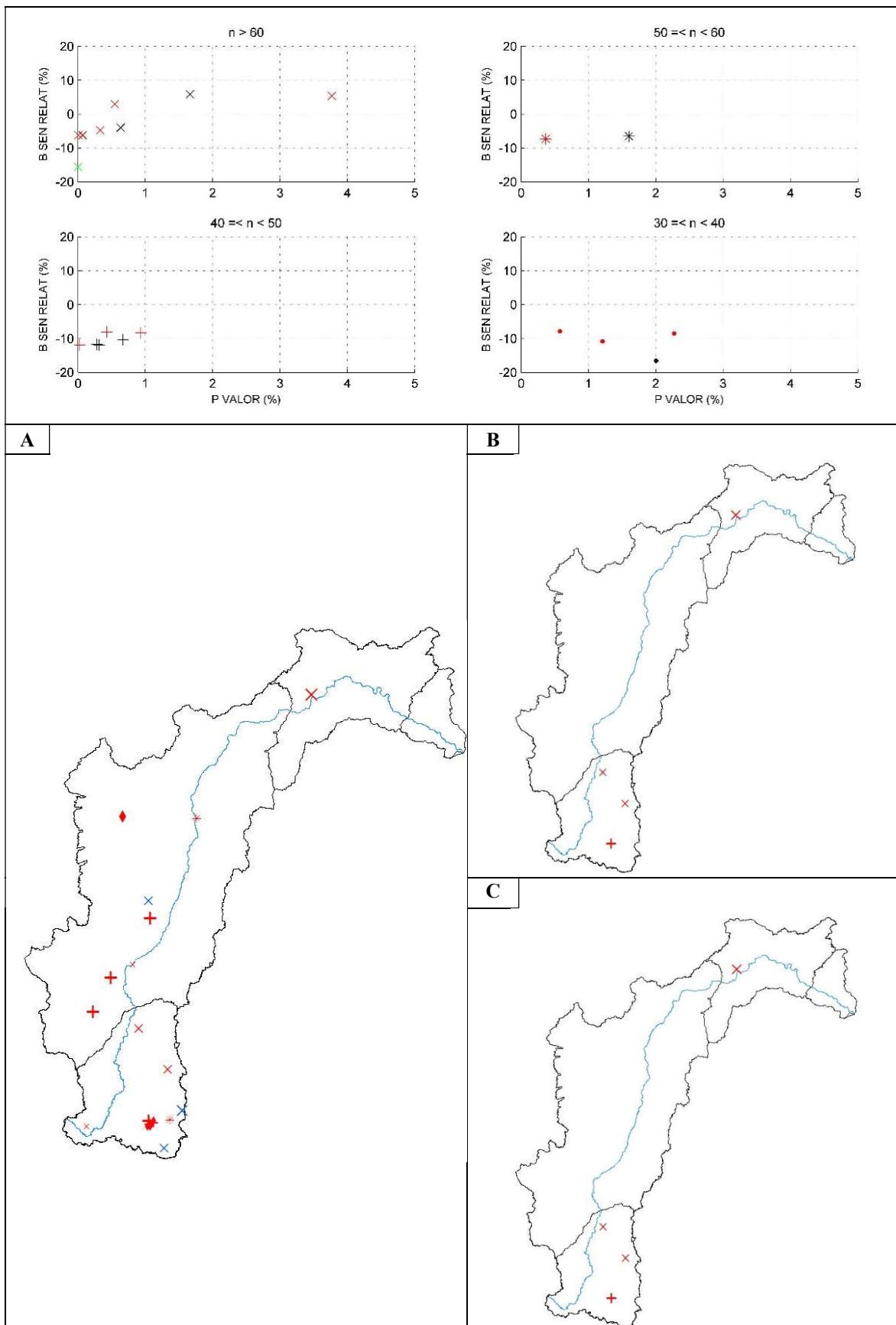


Figura II.6 - Estações significativas para CWD após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

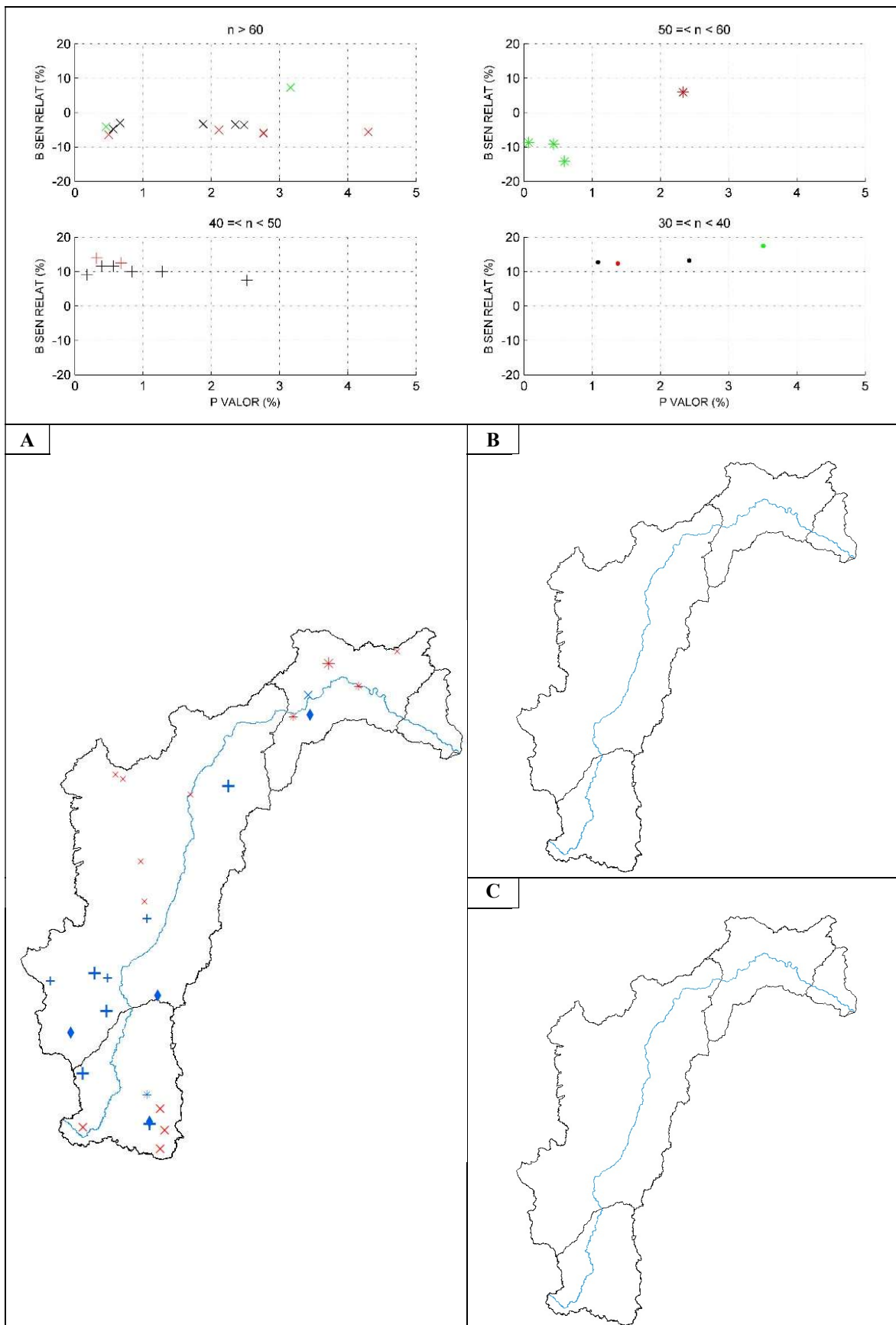


Figura II.7 - Estações significativas para CDD após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

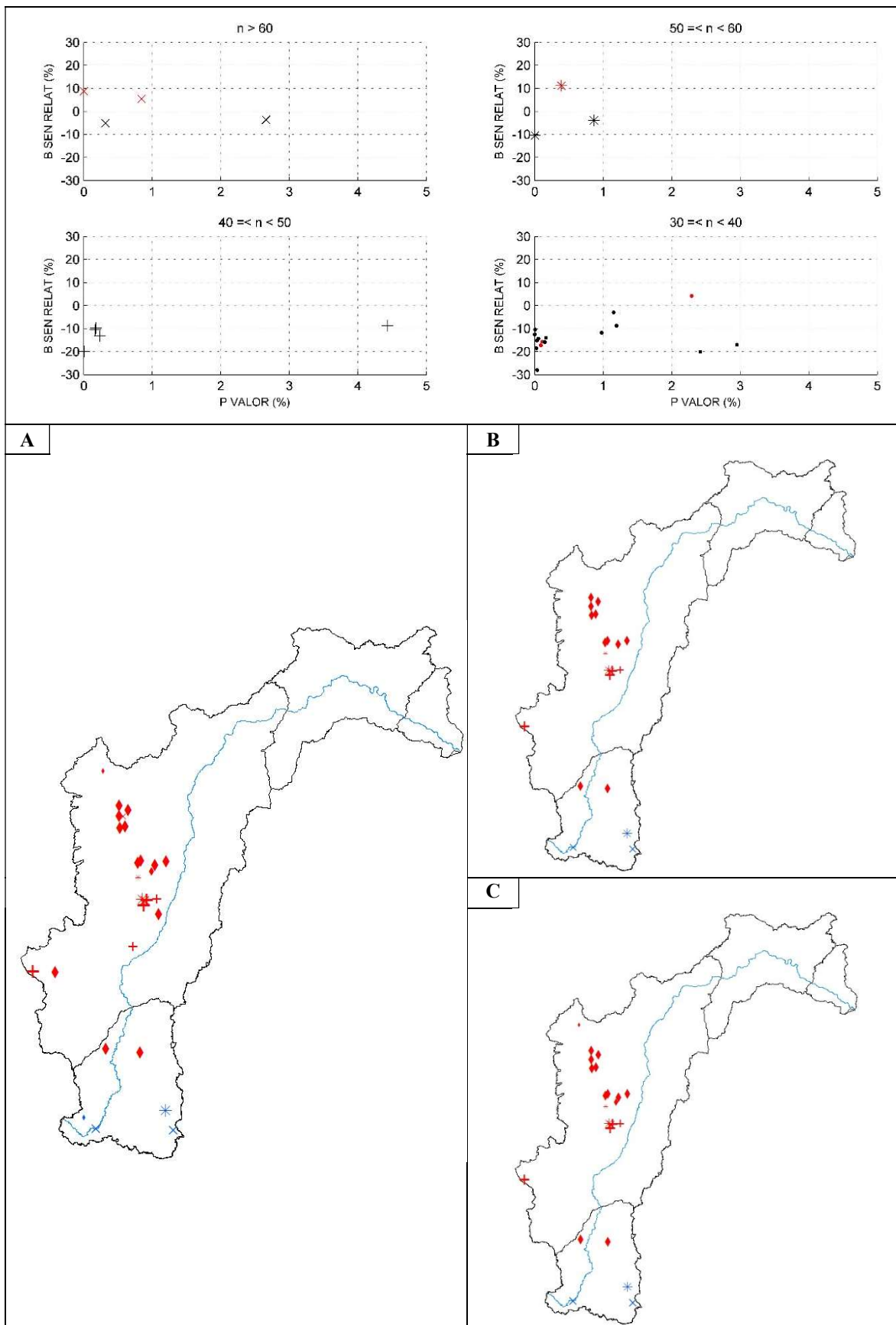


Figura II.8 - Estações significativas para Q_{\max} após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativos por (a)

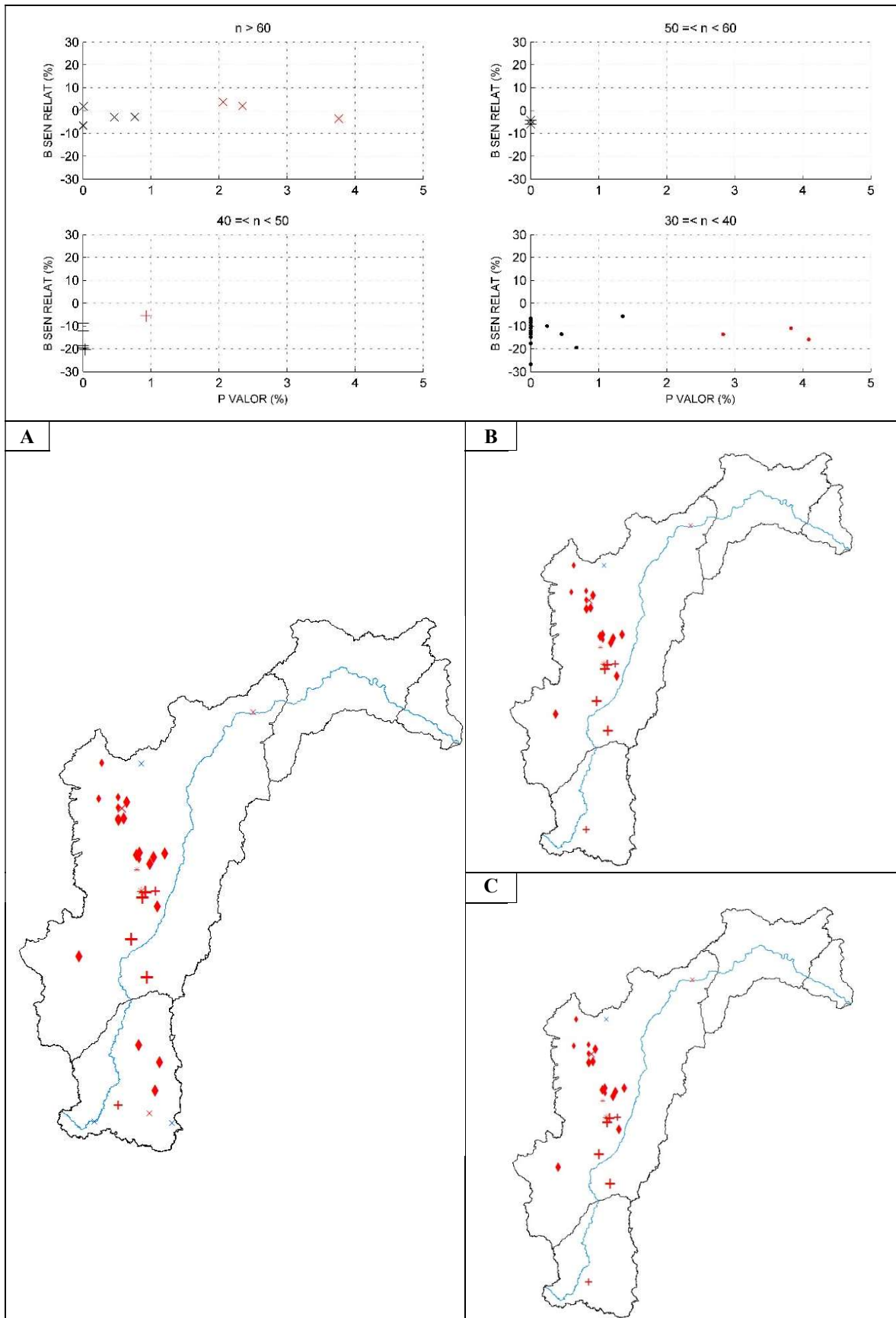


Figura II.9 - Estações significativas para Q_{med} após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

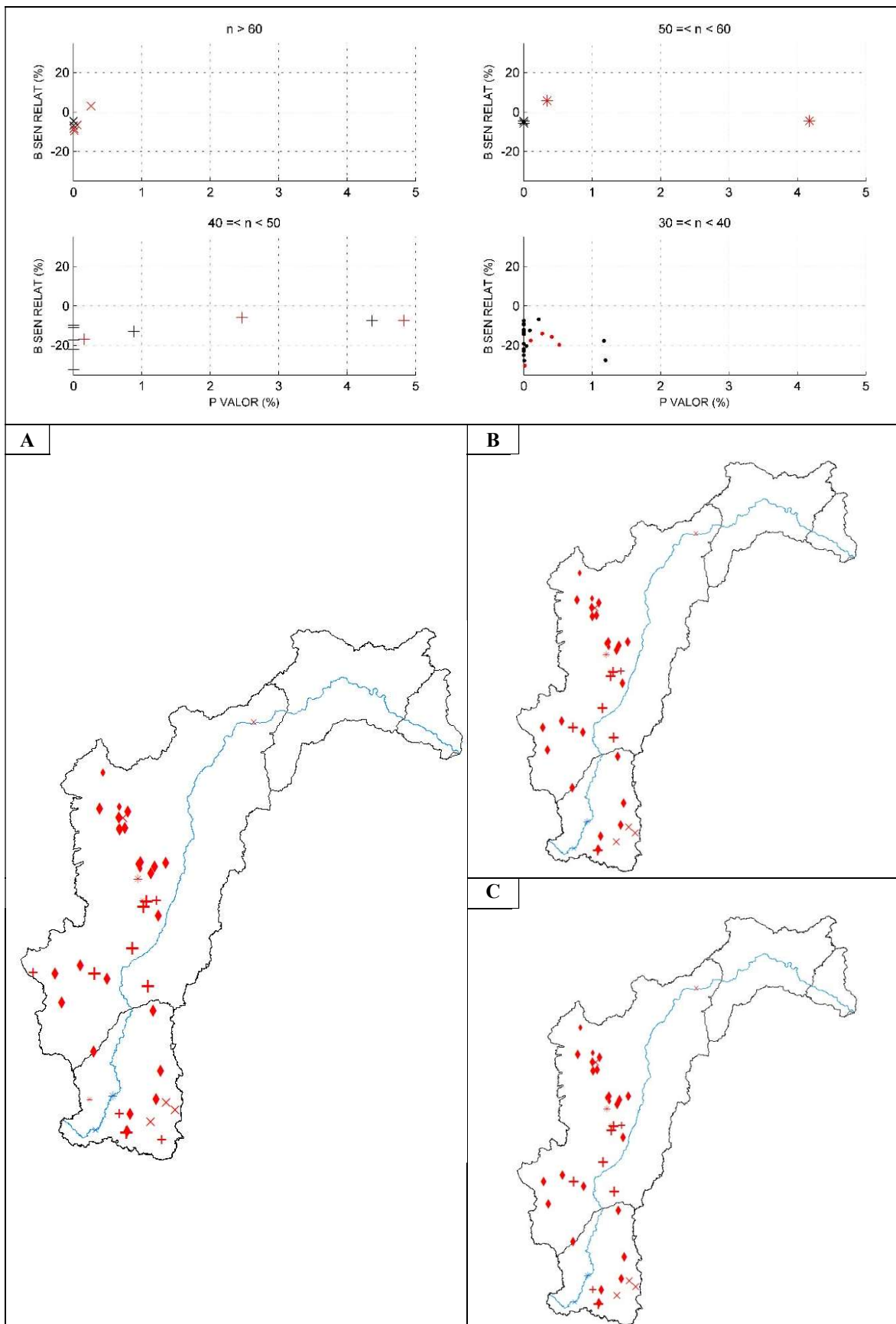


Figura II.10 - Estações significativas para Q30min após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

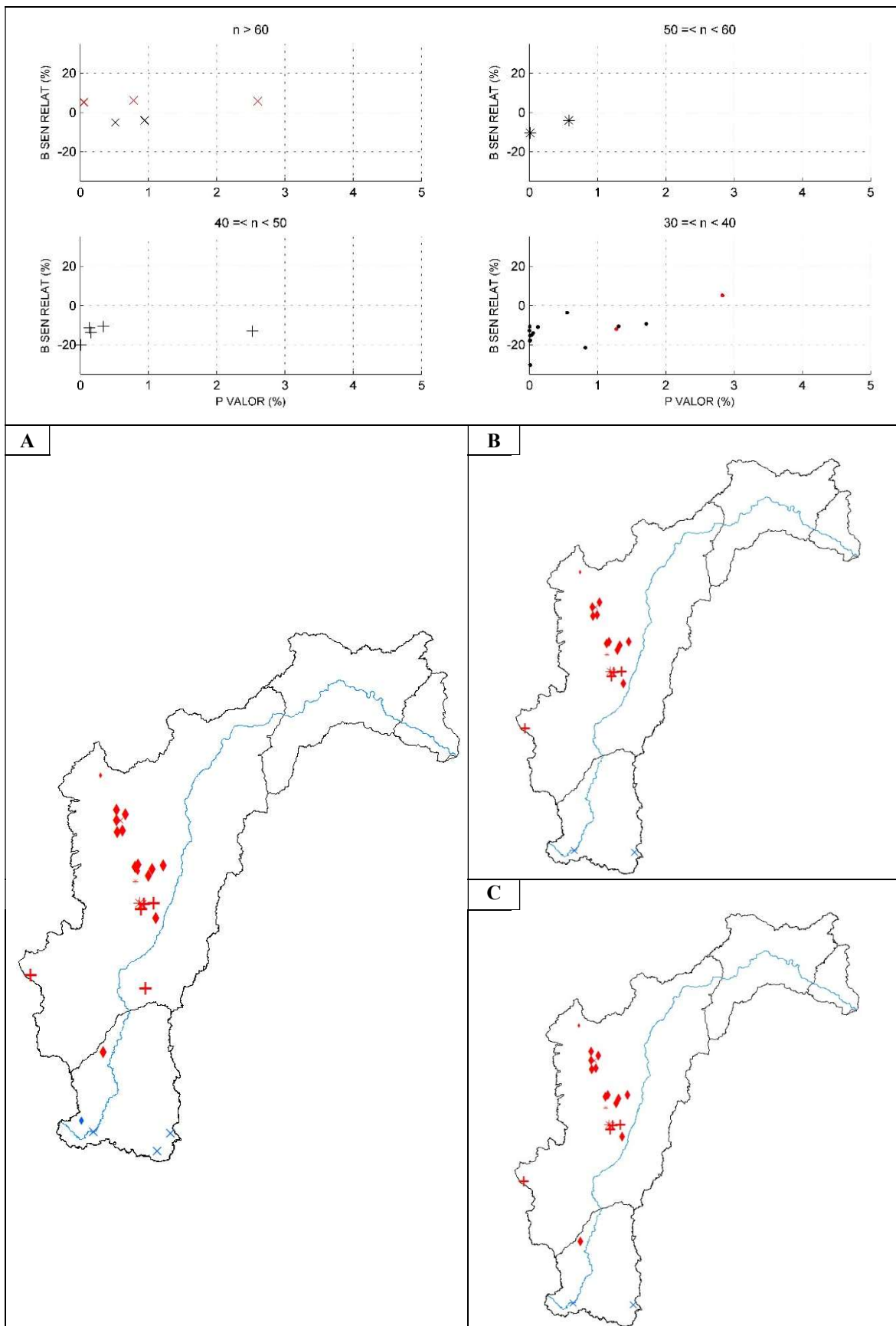


Figura II.11 - Estações significativas para QX5d após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)

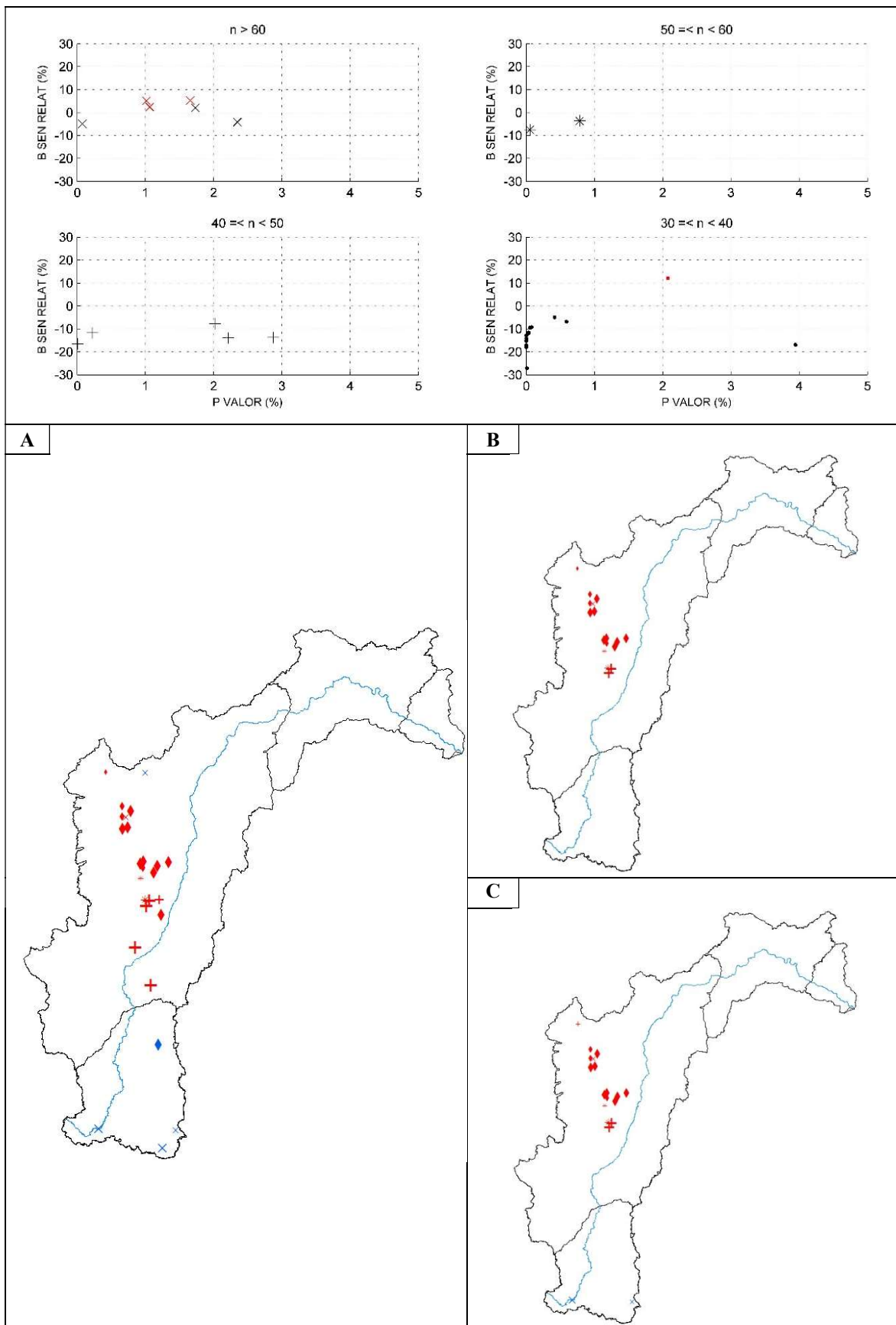


Figura II.12 - Estações significativas para QX30d após a aplicação (a) MK-TFPW, (b) BH95 e (c) BH_Adap. O gráfico mostra os p_valores e B_SEN_relativos dos pontos de monitoramento julgadas significativas por (a)