

UNIVERSIDADE FEDERAL DE MINAS GERAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM SANEAMENTO,
MEIO AMBIENTE E RECURSOS HÍDRICOS

**COMPARAÇÃO DE TÉCNICAS DE EXTENSÃO
DE SÉRIES HIDROLÓGICAS**

Lucas Filipe Lucena Jesus

Belo Horizonte

2016

COMPARAÇÃO DE TÉCNICAS DE EXTENSÃO DE SÉRIES HIDROLÓGICAS

Lucas Filipe Lucena Jesus

Lucas Filipe Lucena Jesus

COMPARAÇÃO DE TÉCNICAS DE EXTENSÃO DE SÉRIES HIDROLÓGICAS

Dissertação apresentada ao Programa de Pós-graduação em Saneamento, Meio Ambiente e Recursos Hídricos da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre em Saneamento, Meio Ambiente e Recursos Hídricos.

Área de concentração: Hidráulica e Recursos Hídricos

Linha de pesquisa: Modelagem de Processos Hidrológicos

Orientador: prof. Wilson Fernandes

Belo Horizonte
Escola de Engenharia da UFMG
2016

J58c Jesus, Lucas Filipe Lucena.
Comparação de técnicas de extensão de séries hidrológicas [manuscrito]
/ Lucas Filipe Lucena Jesus. - 2016.
xiii, 146 f.: il.

Orientador: Wilson dos Santos Fernandes.

Dissertação (mestrado) – Universidade Federal de Minas Gerais, Escola de Engenharia.

Apêndice: f. 142-146.

Bibliografia: f.136-141.

1. Engenharia sanitária- Teses. 2. Recursos hídricos - Desenvolvimento - Teses. 3. Hidrologia - Modelos matemáticos - Teses. I. Fernandes, Wilson dos Santos. II. Universidade Federal de Minas Gerais. Escola de Engenharia. III. Título.

CDU: 628(043)



FOLHA DE APROVAÇÃO

Comparação de técnicas de extensão de séries hidrológicas

LUCAS FILIPE LUCENA JESUS

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:


Prof. WILSON DOS SANTOS FERNANDES - orientador


Prof. VEBER AFONSO FIGUEIREDO COSTA


Prof. ÉBER JOSÉ DE ANDRADE PINTO


Prof. OTTO CORRÊA ROTUNNO FILHO

Aprovada pelo Colegiado do PG SMARH

Versão Final aprovada por

Prof. Nilo de Oliveira Nascimento
Coordenador

Prof. Wilson dos Santos Fernandes
Orientador

Belo Horizonte, 16 de dezembro de 2016.

*“Todo o problema do mundo é que os
tolos e os fanáticos estão sempre
cheios de certezas, enquanto os sábios
estão cheios de dúvidas.”*

Bertrand Russell

AGRADECIMENTOS

A toda a minha família, em especial aos meus pais, Eva e Tarcízio, por serem um referencial e um exemplo, pela ajuda que me deram nos momentos de necessidade e por sempre acreditarem em mim.

À Géssica, pelo apoio em todos os momentos. Por ter entendido as eventuais ausências e por fazer a vida valer ainda mais a pena.

Ao meu orientador, Wilson Fernandes, pela sugestão do tema, pelo apoio, pela condução tranquila do trabalho e por ter despertado em mim o interesse pela área de recursos hídricos.

Aos amigos da Pimenta de Ávila, com os quais aprendi tantas e tantas coisas.

Aos colegas do mestrado, em especial à Nathália Machado, pelo apoio principalmente durante as disciplinas.

Aos professores e funcionários do Departamento de Engenharia Hidráulica e de Recursos Hídricos da UFMG e do SMARH pela presteza.

RESUMO

A análise de frequência de variáveis hidrológicas é um tema extensivamente estudado na literatura científica, tanto pela responsabilidade dos seus resultados quanto pelas diversas incertezas inerentes à metodologia tradicionalmente empregada. Uma das fontes de incertezas é a pequena extensão das séries históricas hidrológicas reduzidas. Diante disso, diversas técnicas de extensão de séries hidrológicas foram desenvolvidas a partir de meados do século passado, como o grupo denominado Manutenção da Variância (MOVE1 a MOVE4) e as técnicas KTRL2 (Linha Robusta Kendall-Theil 2) e RLOC (Linha Robusta de Correlação). Essas técnicas foram estudadas em conjunto com os modelos estatísticos de regressão OLS (Mínimos Quadrados Ordinários), KTRL e GLM (Modelos Lineares Generalizados), com o objetivo de obter séries hidrológicas reduzidas estendidas mais representativas das medidas descritivas populacionais do que a série reduzida curta. Dessa forma, seriam proporcionadas melhorias na estimativa da frequência de eventos hidrológicos extremos. Todos os modelos estudados baseiam-se nas observações de uma estação fluviométrica ou pluviométrica de referência para extrapolar os registros de uma série hidrológica curta. A metodologia do estudo consistiu de simulações de Monte Carlo contemplando vários cenários de extensão das séries curta e de referência, além do nível de correlação linear entre elas. Também foram simulados cenários abrangendo dados modelados por diferentes distribuições bivariadas (Normal, Gumbel e Pearson 3 com assimetria positiva e negativa). Foram analisados o viés, a variância e o erro das estimativas da média, da variância, do coeficiente de assimetria, dos percentis e dos quantis populacionais, além do erro e do viés na estimativa dos registros individuais. Esses resultados foram utilizados para a elaboração de critérios de aplicação das técnicas baseados na teoria dos estimadores. Os critérios elaborados reúnem as técnicas que geraram séries estendidas com desempenho superior à série curta, no que tange à estimativa dos momentos populacionais analisados. Os resultados indicaram que os modelos de regressão KTRL e GLM não são indicados para a extensão das séries hidrológicas. O modelo OLS teve desempenho satisfatório apenas na estimativa dos registros individuais. As técnicas com melhor desempenho são aquelas pertencentes à família MOVE. O critério de aplicação das técnicas elaborado foi avaliado por meio da aplicação sobre dados reais. Concluiu-se que as técnicas de extensão de séries hidrológicas são capazes de gerar séries estendidas com medidas descritivas mais próximas dos respectivos valores populacionais do que a série curta.

Palavras-chave: Distribuição Temporal de Precipitação; Eventos Extremos; Hietogramas de Projeto; Hidrologia - Modelos Matemáticos.

ABSTRACT

Hydrologic frequency analysis is an extensively studied research topic in the literature, due to the relevant role it plays in the practice of water resources engineering and to several uncertainties inherent to the traditional methodology. One of the sources of these uncertainties is the short extension of the reduced hydrologic series used as input to frequency analysis. Several record extension techniques have been developed since the middle of the last century, such as the group named Maintenance of Variance Extension (MOVE1 to MOVE4) and the techniques KTRL2 (Kendall-Theil Robust Line 2) and RLOC (Robust Line of Organic Correlation). Along with regression models OLS (Ordinary Least Squares), KTRL and GLM (Generalized Linear Models), such techniques have been studied with the objective to extend reduced hydrologic series in order to obtain more accurate time series than the original ones with respect to the estimation of the population moments, leading to improvements in the estimates of frequency of extreme hydrologic events. All models studied are based on a reference time series (rainfall or streamflow records) used to extrapolate the observations of the short time series. The methodology consisted of Monte Carlo simulations encompassing several scenarios of short and long series extension, and the linear correlation between them. Also, scenarios covering the distribution of the data (bivariate Normal, Gumbel and Pearson 3 with positive and negative skewness) were simulated. Results evaluated were bias, variance and error in the estimation of the population mean, variance, skewness coefficient and quantiles, along with the error and the bias in the estimation of the individual records. These results were used for the definition of criteria for the practical application of record extension techniques, based on the estimators theory. The criteria developed pooled the techniques that generated extended records with superior performance than the short series with respect to the estimation of population moments. Results show that KTRL and GLM regression models are not suitable for the extension of the hydrologic time series. In the estimation of individual records, OLS outperformed the other techniques. In the other aspects, the MOVE techniques performed best. The developed criteria were tested by means of the application on real data. It was concluded that the record extension techniques are able to generate extended records with better estimates of the population descriptive measures than the short series.

Keywords: Time Distribution of Rainfall; Extreme Events; Design Storms; Hydrology – Mathematical Modeling.

SUMÁRIO

LISTA DE FIGURAS	V
LISTA DE TABELAS.....	VIII
LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS	X
1 INTRODUÇÃO.....	1
2 OBJETIVOS.....	3
2.1 OBJETIVO GERAL	3
2.2 OBJETIVOS ESPECÍFICOS.....	3
3 REVISÃO DA LITERATURA	4
3.1 CONCEITOS DE HIDROLOGIA ESTATÍSTICA	4
3.1.1 <i>Distribuições de probabilidades utilizadas na Engenharia de Recursos Hídricos</i>	4
3.1.2 <i>Estimação de parâmetros</i>	9
3.2 TÉCNICAS DE EXTENSÃO DE SÉRIES HIDROLÓGICAS	14
3.2.1 <i>OLS (Ordinary Least Squares)</i>	16
3.2.2 <i>RPN (Regression Plus Noise)</i>	22
3.2.3 <i>MOVE (Maintenance of Variance Extension)</i>	25
3.2.4 <i>KTRL (Kendall-Theil Robust Line)</i>	30
3.2.5 <i>KTRL2 (Kendall-Theil Robust Line 2)</i>	34
3.2.6 <i>RLOC (Robust Line of Organic Correlation)</i>	35
3.2.7 <i>GLM (Generalized Linear Models)</i>	37
3.2.8 <i>Técnicas que consideram a autocorrelação amostral</i>	43
3.2.9 <i>Comentários sobre as técnicas de extensão de séries hidrológicas</i>	45
4 MATERIAL E MÉTODOS	49
4.1 SEQUÊNCIA METODOLÓGICA.....	49
4.1.1 <i>Experimentos sobre amostras sintéticas</i>	49
4.1.2 <i>Aplicação sobre amostras reais</i>	51
4.2 SIMULAÇÃO DE MONTE CARLO	52
4.3 GERAÇÃO DE AMOSTRAS SINTÉTICAS.....	54
4.3.1 <i>Distribuição Normal Bivariada</i>	55
4.3.2 <i>Outras distribuições multivariadas</i>	57
4.4 TÉCNICAS DE EXTENSÃO DE SÉRIES HIDROLÓGICAS	60
4.5 DEFINIÇÃO DE CRITÉRIOS PARA A APLICAÇÃO DAS TÉCNICAS DE EXTENSÃO DE SÉRIES HIDROLÓGICAS SOBRE DADOS REAIS	63
4.6 APLICAÇÃO SOBRE AMOSTRAS REAIS.....	65
4.6.1 <i>Critérios para a escolha da região estudada</i>	65
4.6.2 <i>Região estudada</i>	65
4.6.3 <i>Aplicação das técnicas de extensão de séries hidrológicas</i>	72
4.7 AVALIAÇÃO DOS RESULTADOS	73
5 RESULTADOS E DISCUSSÃO	75
5.1 COMPARAÇÃO DAS TÉCNICAS DE EXTENSÃO DE SÉRIES HIDROLÓGICAS	75
5.1.1 <i>Momentos amostrais</i>	77
5.1.2 <i>Percentis</i>	88
5.1.3 <i>Registros individuais</i>	94
5.1.4 <i>Quantil associado ao TR = 100 anos</i>	97
5.2 CRITÉRIOS PARA A APLICAÇÃO SOBRE DADOS REAIS.....	103
5.3 APLICAÇÃO SOBRE DADOS REAIS	114
5.3.1 <i>Simulação histórica</i>	115
5.3.2 <i>Simulações bootstrap</i>	128
6 CONCLUSÕES E RECOMENDAÇÕES	131
REFERÊNCIAS	136
APÊNDICE A	142

LISTA DE FIGURAS

FIGURA 3.1 – CLASSIFICAÇÃO DAS DISTRIBUIÇÕES COMUMENTE UTILIZADAS NA ENGENHARIA DE RECURSOS HÍDRICOS, EM RELAÇÃO AO COMPORTAMENTO DA CAUDA	7
FIGURA 3.2 - HIPÓTESE DE NORMALIDADE DAS ESTIMATIVAS DA REGRESSÃO LINEAR	18
FIGURA 3.3 - VIÉS INSERIDO NOS VALORES PREVISTOS DE Y POR MEIO DA APLICAÇÃO DA TÉCNICA OLS	22
FIGURA 3.4 - DETERMINAÇÃO DA DECLIVIDADE DA RETA DA TÉCNICA MOVE1	25
FIGURA 3.5 - COMPARAÇÃO DA RESISTÊNCIA A <i>OUTLIERS</i> ENTRE AS TÉCNICAS OLS E KTRL ...	31
FIGURA 3.6 - DETERMINAÇÃO DA DECLIVIDADE DA RETA DE REGRESSÃO DA TÉCNICA KTRL..	32
FIGURA 3.7 - RESULTADO GRÁFICO DO EXPERIMENTO DE HIRSCH <i>ET AL.</i> (1991): VANTAGEM DA TÉCNICA KTRL EM RELAÇÃO À TÉCNICA OLS PARA AMOSTRAS QUE NÃO SEGUEM A DISTRIBUIÇÃO NORMAL	34
FIGURA 3.8 - RESULTADO GRÁFICO DO EXPERIMENTO DE HIRSCH <i>ET AL.</i> (1991): VANTAGEM DA TÉCNICA RLOC EM RELAÇÃO À TÉCNICA MOVE1 PARA AMOSTRAS QUE NÃO SEGUEM A DISTRIBUIÇÃO NORMAL	36
FIGURA 3.9 – FORMATO DE ALGUMAS DAS FUNÇÕES DE LIGAÇÃO NORMALMENTE UTILIZADAS NOS MODELOS LINEARES GENERALIZADOS	38
FIGURA 3.10 – RESULTADOS DA APLICAÇÃO DA TÉCNICA MOVE1 NO ESTUDO DE REINFELDS <i>ET AL.</i> (2006).....	46
FIGURA 4.1 - RESUMO DA METODOLOGIA UTILIZADA NO EXPERIMENTO SOBRE AMOSTRAS SINTÉTICAS	50
FIGURA 4.2 – DISTRIBUIÇÕES MARGINAIS UTILIZADAS NAS AMOSTRAGENS	60
FIGURA 4.3 – ESTAÇÕES PLUVIOMÉTRICAS DENTRO E NAS ADJACÊNCIAS DA BACIA HIDROGRÁFICA EM ESTUDO	67
FIGURA 4.4 - ESTAÇÕES FLUVIOMÉTRICAS NA BACIA HIDROGRÁFICA EM ESTUDO.....	68
FIGURA 4.5 – GRÁFICOS DE DUPLA MASSA DAS ESTAÇÕES SELECIONADAS.....	71
FIGURA 5.1 – RMSE NA ESTIMATIVA DA MÉDIA POPULACIONAL – DADOS PERTENCENTES À DISTRIBUIÇÃO NORMAL	77
FIGURA 5.2 – RMSE NA ESTIMATIVA DA MÉDIA POPULACIONAL – INFLUÊNCIA DA DENSIDADE DE PROBABILIDADE DOS DADOS	78
FIGURA 5.3 – VIÉS NA ESTIMATIVA DA MÉDIA POPULACIONAL – INFLUÊNCIA DA DENSIDADE DE PROBABILIDADE DOS DADOS	79
FIGURA 5.4 – RMSE NA ESTIMATIVA DA VARIÂNCIA POPULACIONAL – DADOS PERTENCENTES À DISTRIBUIÇÃO NORMAL.....	80
FIGURA 5.5 – RMSE NA ESTIMATIVA DA VARIÂNCIA POPULACIONAL – INFLUÊNCIA DA DENSIDADE DE PROBABILIDADE DOS DADOS	82
FIGURA 5.6 – VIÉS NA ESTIMATIVA DA VARIÂNCIA POPULACIONAL – RESULTADOS PARA DADOS GUM.....	83
FIGURA 5.7 – VIÉS NA ESTIMATIVA DA VARIÂNCIA POPULACIONAL – INFLUÊNCIA DA DISTRIBUIÇÃO DE PROBABILIDADES DOS DADOS	85

FIGURA 5.8 – VARIÂNCIA NA ESTIMATIVA DA VARIÂNCIA POPULACIONAL.....	85
FIGURA 5.9 – RMSE NA ESTIMATIVA DO COEFICIENTE DE ASSIMETRIA POPULACIONAL – RESULTADOS PARA DADOS NOR	86
FIGURA 5.10 – VIÉS NA ESTIMATIVA DO COEFICIENTE DE ASSIMETRIA POPULACIONAL – RESULTADOS PARA DADOS GUM, PE3+ E PE3-	87
FIGURA 5.11 – VARIÂNCIA NA ESTIMATIVA DO COEFICIENTE DE ASSIMETRIA POPULACIONAL – RESULTADOS PARA DADOS NOR	88
FIGURA 5.12 – RMSE E VIÉS NA ESTIMATIVA DOS PERCENTIS – RESULTADOS PARA DADOS NOR	90
FIGURA 5.13 – RMSE E VIÉS NA ESTIMATIVA DOS PERCENTIS – RESULTADOS PARA DADOS GUM	92
FIGURA 5.14 – RMSE NA ESTIMATIVA DOS PERCENTIS – INFLUÊNCIA DA ASSIMETRIA DOS DADOS	93
FIGURA 5.15 – VIÉS NA ESTIMATIVA DOS PERCENTIS – INFLUÊNCIA DA ASSIMETRIA DOS DADOS	95
FIGURA 5.16 – ERRO NA ESTIMATIVA DOS REGISTROS INDIVIDUAIS	95
FIGURA 5.17 – VIÉS NA ESTIMATIVA DOS REGISTROS INDIVIDUAIS – INFLUÊNCIA DA DISTRIBUIÇÃO DOS DADOS	96
FIGURA 5.18 – AJUSTE DAS TÉCNICAS DE EXTENSÃO PARA O CASO $n_1= 25$, $n_2= 25$ E $P = 0,7$..	97
FIGURA 5.19 – RMSE NA ESTIMATIVA DO QUANTIL POPULACIONAL ASSOCIADO AO TR = 100 ANOS.....	98
FIGURA 5.20 – VIÉS NA ESTIMATIVA DO QUANTIL POPULACIONAL ASSOCIADO AO TR = 100 ANOS.....	100
FIGURA 5.21 – VARIÂNCIA NA ESTIMATIVA DO QUANTIL POPULACIONAL ASSOCIADO AO TR = 100 ANOS.....	102
FIGURA 5.22 – RESULTADOS DA ANÁLISE DO VIÉS, DA VARIÂNCIA E DO ERRO NA ESTIMATIVA DE QUANTIS EM COMPARAÇÃO COM A SÉRIE CURTA – DADOS COM DISTRIBUIÇÃO NORMAL.....	103
FIGURA 5.23 – RESULTADOS DA ANÁLISE DO VIÉS, DA VARIÂNCIA E DO ERRO NA ESTIMATIVA DE QUANTIS EM COMPARAÇÃO COM A SÉRIE CURTA – DADOS COM DISTRIBUIÇÃO GUMBEL	104
FIGURA 5.24 – RESULTADOS DA ANÁLISE DO VIÉS, DA VARIÂNCIA E DO ERRO NA ESTIMATIVA DE QUANTIS EM COMPARAÇÃO COM A SÉRIE CURTA – DADOS COM DISTRIBUIÇÃO PEARSON 3, ASSIMETRIA POSITIVA	105
FIGURA 5.25 – RESULTADOS DA ANÁLISE DO VIÉS, DA VARIÂNCIA E DO ERRO NA ESTIMATIVA DE QUANTIS EM COMPARAÇÃO COM A SÉRIE CURTA – DADOS COM DISTRIBUIÇÃO PEARSON 3, ASSIMETRIA NEGATIVA	106
FIGURA 5.26 – PROPORÇÃO DAS AMOSTRAS NA QUAL HOVE REDUÇÃO DO ERRO ABSOLUTO NA ESTIMATIVA DO QUANTIL ASSOCIADO AO TR = 100 ANOS	110
FIGURA 5.27 – COEFICIENTES DE CORRELAÇÃO ENTRE AS SÉRIES CURTA E DE REFERÊNCIA DAS SIMULAÇÕES NAS QUAIS HOVE REDUÇÃO DO ERRO NA ESTIMATIVA DOS QUANTIS APÓS A EXTENSÃO	112
FIGURA 5.28 – AJUSTE DAS TÉCNICAS MOVE4 E OLS – CENÁRIO 1.1	117

FIGURA 5.29 – CENÁRIO 1.2 – MEDIDAS DESCRITIVAS DAS SÉRIES HISTÓRICAS COMPLETAS - ESTAÇÕES PRÓXIMAS À ESTAÇÃO PLUVIOMÉTRICA PIRANGA.....	120
FIGURA 5.30 – CENÁRIO 1.2 – ESTIMATIVA DOS PERCENTIS E REGISTROS INDIVIDUAIS DA ESTAÇÃO PLUVIOMÉTRICA PIRANGA.....	121
FIGURA 5.31 – CENÁRIO 1.2 – MEDIDAS DESCRITIVAS DAS SÉRIES HISTÓRICAS COMPLETAS - ESTAÇÕES À MONTANTE DA ESTAÇÃO FLUVIOMÉTRICA PORTO FIRME.....	125
FIGURA 5.32 – CENÁRIO 2.2 – ESTIMATIVA DOS PERCENTIS E REGISTROS INDIVIDUAIS DA ESTAÇÃO FLUVIOMÉTRICA PORTO FIRME	128
FIGURA 5.33 – RESULTADOS DAS SIMULAÇÕES DE <i>BOOTSTRAP</i>	129

LISTA DE TABELAS

TABELA 3.1 - ALGUMAS VANTAGENS E DESVANTAGENS DE CADA MÉTODO DE ESTIMAÇÃO DE PARÂMETROS	13
TABELA 3.2 - PRINCIPAIS EXPRESSÕES UTILIZADAS NA APLICAÇÃO DA REGRESSÃO LINEAR SIMPLES (TÉCNICA OLS).....	18
TABELA 3.3 - PRINCIPAIS EXPRESSÕES UTILIZADAS NA APLICAÇÃO DA REGRESSÃO LINEAR MÚLTIPLA (TÉCNICA OLS).....	20
TABELA 3.4 - HIPÓTESES DA REGRESSÃO LINEAR SIMPLES (TÉCNICA OLS).....	21
TABELA 3.5 – VALORES CRÍTICOS DO COEFICIENTE DE CORRELAÇÃO POPULACIONAL PARA A EFICIÊNCIA DOS ESTIMADORES DA MÉDIA E DA VARIÂNCIA POPULACIONAL – TÉCNICAS OLS E RPN	24
TABELA 3.6 – VALORES DE $E[s_Y^2/\sigma_Y^2]$ – TÉCNICAS OLS E MOVE1	26
TABELA 3.7 – ALGUMAS DAS FUNÇÕES DE LIGAÇÃO MAIS COMUNS E SUAS INVERSAS	38
TABELA 3.8 - FUNÇÕES DE LIGAÇÃO CANÔNICAS ASSOCIADAS A ALGUMAS DISTRIBUIÇÕES DAS FAMÍLIAS EXPONENCIAIS	42
TABELA 4.1 - ESTAÇÕES PLUVIOMÉTRICAS PRÉ-SELECIONADAS.....	69
TABELA 4.2 - ESTAÇÕES FLUVIOMÉTRICAS PRÉ-SELECIONADAS NA BACIA DO RIO PIRANGA.....	69
TABELA 4.3 - MATRIZ DE CORRELAÇÃO ENTRE AS SÉRIES DE MÁXIMOS DIÁRIOS	70
TABELA 4.4 - MATRIZ DE CORRELAÇÃO ENTRE AS SÉRIES DE PRECIPITAÇÃO ACUMULADA ANUAL / VAZÕES MÉDIAS ANUAIS	70
TABELA 4.5 - SIMULAÇÕES NA APLICAÇÃO DAS TÉCNICAS DE EXTENSÃO SOBRE DADOS REAIS .	73
TABELA 5.1 – TÉCNICAS APLICÁVEIS NA EXTENSÃO DE SÉRIES HIDROLÓGICAS – DADOS COM DISTRIBUIÇÃO NORMAL.....	107
TABELA 5.2 – TÉCNICAS APLICÁVEIS NA EXTENSÃO DE SÉRIES HIDROLÓGICAS – DADOS COM DISTRIBUIÇÃO GUMBEL	108
TABELA 5.3 – TÉCNICAS APLICÁVEIS NA EXTENSÃO DE SÉRIES HIDROLÓGICAS – DADOS COM DISTRIBUIÇÃO PEARSON 3, ASSIMETRIA POSITIVA	109
TABELA 5.4 – TÉCNICAS APLICÁVEIS NA EXTENSÃO DE SÉRIES HIDROLÓGICAS – DADOS COM DISTRIBUIÇÃO PEARSON 3, ASSIMETRIA NEGATIVA	109
TABELA 5.5 – INFORMAÇÕES SOBRE OS CENÁRIOS CONSIDERADOS E SOBRE AS SÉRIES HIDROLÓGICAS CONSIDERADAS NA APLICAÇÃO SOBRE DADOS REAIS	115
TABELA 5.6 – RESULTADOS DA APLICAÇÃO DAS TÉCNICAS DE EXTENSÃO DE SÉRIES HIDROLÓGICAS – CENÁRIO 1.1	116
TABELA 5.7 – RESULTADOS DA APLICAÇÃO DAS TÉCNICAS DE EXTENSÃO DE SÉRIES HIDROLÓGICAS – CENÁRIO 1.2	118
TABELA 5.8 – RESULTADOS DA APLICAÇÃO DAS TÉCNICAS DE EXTENSÃO DE SÉRIES HIDROLÓGICAS – CENÁRIO 2.1	122
TABELA 5.9 – RESULTADOS DA APLICAÇÃO DAS TÉCNICAS DE EXTENSÃO DE SÉRIES HIDROLÓGICAS – CENÁRIO 2.1 (ESTAÇÃO PIRANGA COMO REFERÊNCIA)	124

TABELA 5.10 – RESULTADOS DA APLICAÇÃO DAS TÉCNICAS DE EXTENSÃO DE SÉRIES HIDROLÓGICAS – CENÁRIO 2.2	126
--	-----

LISTA DE ABREVIATURAS, SIGLAS E SÍMBOLOS

2P	Distribuição de probabilidades com 2 parâmetros
3P	Distribuição de probabilidades com 3 parâmetros
a	Intercepto em um modelo de regressão / extensão
AAS	Amostra Aleatória Simples
AIC	Amplitude Inter-Quartil
AIC	Critério de Informação de Akaike
ANA	Agência Nacional de Águas
ANN	Redes Neurais Artificiais
b	Coefficiente angular em um modelo de regressão / extensão
BrXII	Distribuição Burr tipo XII
CV	Coefficiente de variação
e	Resíduo de um modelo de regressão
$E[...]$	Valor esperado de uma expressão
EV (1/2/3)	Distribuição de Valor Extremo
f_X	Função densidade de probabilidades da variável aleatória X
f_{X_1, X_2}	Função densidade de probabilidade conjunta das variáveis aleatórias X_1 e X_2
F_X	Função acumulada de probabilidades da variável aleatória X
FAP	Função Acumulada de Probabilidades
FDP	Função Densidade de Probabilidades
FDPC	Função Densidade de Probabilidades Conjunta
FLU	Estação Fluviométrica
FMP	Função Massa de Probabilidades
GEV	Distribuição Generalizada de Valores Extremos

GG	Distribuição Gama Generalizada
GLM	Modelos Lineares Generalizados
GMOVE	Manutenção da Variância Generalizada
GUM	Dados distribuídos segundo a distribuição Gumbel
IID	Independentes e Identicamente Distribuídas
IWLS	Mínimos Quadrados Ponderado Iterativo
KTRL	Linha Robusta Kendall-Theil
LN2	Distribuição Log-Normal com 2 parâmetros
LN3	Distribuição Log-Normal com 3 parâmetros
LOC	Linha de Correlação Orgânica
LP3	Distribuição Log-Pearson 3
m_j	Momento amostral de ordem j
ML	Método da Máxima Verossimilhança
MLR	Regressão Linear Múltipla
MOM	Método dos Momentos
MOVE	Manutenção da Variância
n	Extensão de uma série hidrológica
$n_1, N1$	Extensão da série curta
$n_2, N2$	Diferença entre a extensão da série longa e a extensão da série curta
NOR	Dados distribuídos segundo a distribuição Normal
OLS	Mínimos Quadrados Ordinário
$P[...]$	Probabilidade de um evento
PE3	Distribuição Pearson 3
PE3-	Dados distribuídos segundo a distribuição Pearson 3 com assimetria negativa de -1,4

PE3+	Dados distribuídos segundo a distribuição Pearson 3 com assimetria positiva de 1,4
PLU	Estação Pluviométrica
Q _{7,10}	Vazão mínima com 7 dias de duração e tempo de retorno de 10 anos
Q ₉₅	Vazão com permanência em 95% do tempo
r	Coefficiente de correlação amostral de Pearson
R^2	Coefficiente de determinação
RLOC	Linha Robusta de Correlação
RMSE	Erro Quadrático Médio
RPN	Regressão Mais Ruído
s	Desvio-padrão amostral
s_X^2	Variância amostral da variável aleatória X
SHT	Temperatura Média do Hemisfério Sul
TR	Tempo de Retorno
USBR	<i>United States Bureau of Reclamation</i>
USWRC	<i>United States Water Resources Council</i>
VA	Variável aleatória
\bar{x}	Média amostral de x
$x_{0,99}$	Estimativa do quantil associado ao TR = 100 anos da variável aleatória X
X	Variável aleatória independente em um modelo de regressão / extensão
y_i	i -ésimo valor amostral da variável aleatória Y
\hat{y}_i	Estimativa da i -ésima observação da variável aleatória Y
Y	Variável aleatória dependente em um modelo de regressão / extensão
α	Parâmetro de um modelo populacional de regressão

β	Parâmetro de um modelo populacional de regressão
γ	Coefficiente de assimetria populacional
θ	Parâmetro
$\hat{\theta}$	Estimativa de um parâmetro
μ_j	Momento populacional de ordem j
ρ	Coefficiente de correlação populacional de Pearson
σ_X^2	Variância populacional da variável aleatória X
τ	Coefficiente de correlação populacional de Kendall
∞	Infinito

1 INTRODUÇÃO

A análise de frequência de variáveis hidrológicas é um dos temas de grande relevância da Engenharia de Recursos Hídricos, dado que, por meio dela, são obtidas as chuvas e vazões de referência que são utilizadas em estudos e projetos. Por se tratarem de grandezas às quais são atribuídas grande responsabilidade e por se constituírem em valores com baixa probabilidade de ocorrência, com várias incertezas em sua avaliação, a estimativa dos quantis de projeto têm sido objeto de estudo na literatura científica. Uma das grandes dificuldades na análise de frequência é estimar, com a maior confiabilidade possível, os quantis de projeto a partir de registros com extensão pequena.

No estudo de Benson (1960), por exemplo, concluiu-se que, para estimar a vazão associada ao período de retorno de 50 anos, com o erro situando-se em uma faixa de 24% do quantil populacional, em 95% das amostras, são necessários pelo menos 39 anos de observações. Já Naghettini e Pinto (2007) fazem referência à consideração prática de que, a partir de uma série hidrológica com extensão de n anos, pode-se estimar, com alguma confiabilidade, quantis associados a períodos de retorno de até $2n$ anos.

A Agência Nacional de Águas (ANA), responsável sobre a rede hidrometeorológica brasileira, divulgou, em 2007, um documento consolidando a evolução desta desde a década de 1920 (ANA, 2007). Analisando este e outros documentos (ANA, 2009a; ANA, 2009b), fica evidente que algumas das estações de monitoramento possuem uma extensão muito pequena da série de dados registrados. Além disso, a qualidade dos dados monitorados nessas estações não é a ideal, o que fica evidenciado pela presença de falhas, por exemplo.

Existem diversas técnicas para preenchimento de falhas, entre as quais estão a regressão linear simples, ponderação regional, redes neurais artificiais e vetor regional (ANA, 2012). Tais técnicas foram estudadas por Oliveira *et al.* (2010) no contexto do preenchimento de falhas de precipitações anuais. Um problema dessas técnicas é que elas se propõem a minimizar o erro das estimativas dos registros individuais que não foram monitorados. Ao utilizar essa abordagem, aspectos estatísticos da série hidrológica como um todo são deixados de lado, como sua capacidade de aumentar a confiabilidade das estimativas dos percentis populacionais, além da média, variância e outros momentos estatísticos.

As técnicas de extensão de séries hidrológicas ocupam-se da tarefa de estimar valores não monitorados em uma série hidrológica, com base em uma série hidrológica de referência, para obter uma série estendida com determinadas propriedades estatísticas desejadas (como média e variâncias mais próximas dos valores populacionais, por exemplo). Esse pressuposto torna as técnicas de extensão de séries hidrológicas potencialmente adequadas para utilização anterior à etapa de análise de frequência.

Apesar disso, após uma pesquisa na literatura científica mundial, constatou-se que ainda não foi desenvolvido qualquer estudo, que seja de conhecimento do autor, dedicado à aplicação prática das técnicas de extensão de séries hidrológicas.

Esta dissertação está organizada em capítulos, sendo que o primeiro capítulo apresenta uma breve introdução ao tema estudado, com foco na relevância desta pesquisa, e o segundo capítulo traz os objetivos do trabalho.

O terceiro capítulo contempla a revisão da literatura, abordando conceitos de estatística que serão relevantes no decorrer da pesquisa e apresentando as técnicas de extensão de séries hidrológicas. São feitos comentários sobre os avanços obtidos até o momento sobre o tema.

O quarto capítulo apresenta a metodologia do trabalho, enfocando basicamente nos cenários de simulação de Monte Carlo adotados e nas premissas adotadas em cada etapa dos estudos. Também é apresentada a região onde será estudada a aplicação prática das técnicas de extensão de séries hidrológicas;

O quinto capítulo apresenta os resultados da pesquisa, com a definição dos critérios de aplicação das técnicas e a própria aplicação sobre as séries hidrológicas registradas na região estudada.

O sexto capítulo apresenta as conclusões, recomendações e considerações finais do estudo, seguido pelas referências bibliográficas utilizadas neste trabalho. Ao final do documento, são apresentados, em apêndice, os *scripts* utilizados para o ajuste das técnicas de extensão.

2 OBJETIVOS

2.1 Objetivo geral

O objetivo geral desta pesquisa é comparar o desempenho de técnicas de extensão de séries hidrológicas reduzidas, com foco na estimativa de quantis máximos de precipitação e/ou vazão com baixas probabilidades de ocorrência.

2.2 Objetivos específicos

Os objetivos específicos da pesquisa são:

1. Comparar as técnicas desenvolvidas para uso exclusivo na extensão das séries hidrológicas com técnicas estatísticas de regressão mais complexas, como os modelos lineares generalizados;
2. Avaliar o desempenho das técnicas de extensão de séries hidrológicas na estimativa dos momentos populacionais (média, variância e coeficiente de assimetria);
3. Avaliar a aplicabilidade das técnicas de extensão de séries hidrológicas sobre variáveis aleatórias descritas por distribuições diferentes da Normal; e
4. Elaborar critérios práticos para a aplicação das técnicas de extensão sobre dados hidrológicos reais.

3 REVISÃO DA LITERATURA

3.1 Conceitos de Hidrologia Estatística

3.1.1 Distribuições de probabilidades utilizadas na Engenharia de Recursos Hídricos

A escolha de distribuições de probabilidades para modelar extremos hidrológicos tem sido continuamente estudada na literatura científica (PAPALEXIOU *et al.*, 2013; GOTTSCHALK *et al.*, 2013; SERINALDI & KILSBY, 2014; CAVANAUGH *et al.*, 2015; YALCIN *et al.*, 2016). De fato, a análise estatística das precipitações é um problema da maior importância ambiental (YALCIN *et al.*, 2016), e, por consequência, também o é a análise estatística das vazões dos cursos de água naturais, seja com foco nas vazões máximas ou nas vazões mínimas.

Malamud e Turcotte (2006) definiram, para a análise de frequência de vazões máximas, quatro grupos de distribuições, a saber: i) família normal (Normal, Log-Normal 2P e Log-Normal 3P); ii) família generalizada de valores extremos (GEV, Gumbel, Fréchet e Weibull reversa); iii) família Pearson 3 (Pearson 3 e Log-Pearson 3); e iv) distribuição Generalizada de Pareto. Tais distribuições também podem ser aplicadas para a análise de frequência de precipitações máximas, tendo em vista, por exemplo, a relação física de causa e consequência entre as chuvas e as vazões.

À análise de frequência de vazões mínimas, têm sido dedicados um número menor de estudos em comparação à análise de frequência de vazões máximas. Atribui-se essa disparidade aos tempos de retorno relativamente baixos utilizados como critério de projeto das vazões mínimas, tipicamente menores que 50 anos (GUSTARD *et al.*, 1989 *apud* GOTTSCHALK *et al.*, 2013). De fato, entre as vazões mínimas de referência utilizadas em projetos no Brasil, estão as vazões $Q_{7,10}$ e Q_{95} (ANA, 2011). A primeira caracteriza-se pelo baixo tempo de retorno, enquanto que a segunda, desvinculada de tempo de retorno, é estimada por meio da curva de permanência de toda a série fluviométrica registrada. Apesar disso, Smakhtin (2001) aponta que, entre as distribuições mais relacionadas às vazões mínimas, estão Weibull, Gumbel (mínimos), Pearson 3 e Log Normal.

De modo mais preciso, quando se possuem N observações independentes e identicamente distribuídas (IID) da variável aleatória X , distribuída de acordo com a função acumulada de

probabilidades F_X , a distribuição das probabilidades acumuladas de $Y = \text{máximo}(X_1, X_2, \dots, X_N)$ é dada por (NAGHETTINI & PINTO, 2007)

$$G_Y(x) = (F_X(x))^N \quad (3.1)$$

Para o caso assintótico em que $N \rightarrow \infty$, caso a Equação 3.1 seja desenvolvida sem demais tratamentos, obtêm-se como resultados distribuições limitantes degeneradas. Fisher e Tippet (1928) propuseram que a Equação 3.1 fosse alterada para contemplar o caso de uma sequência de valores transformados linearmente de X , definidos por $a_N + b_N X$. Considera-se que a_N e b_N dependem apenas de N , e não de X . Desta forma, a Equação 3.1 é reescrita como

$$G_Y(x) = \lim_{N \rightarrow \infty} F_X(a_N + b_N x)^N \quad (3.2)$$

Ao adotar valores apropriados para a_N e b_N na solução da Equação 3.2, obtêm-se as distribuições acumuladas assintóticas de máximos de Gumbel, Fréchet e Weibull (máximos), apresentadas respectivamente nas equações (PAPALEXIOU & KOUTSOYIANNIS, 2013)

$$G(x) = \exp \left[- \exp \left(- \frac{x-\beta}{\alpha} \right) \right], x \in \mathbb{R}, \beta \in \mathbb{R}, \alpha > 0 \quad (3.3)$$

$$G(x) = \exp \left[- \left(\frac{x-\beta}{\alpha} \right)^{-\frac{1}{\gamma}} \right], x \geq \beta, \beta \in \mathbb{R}, \alpha > 0, \gamma > 0 \quad (3.4)$$

$$G(x) = \exp \left[- \left(- \frac{x-\beta}{\alpha} \right)^{\frac{1}{\gamma}} \right], x \leq \beta, \beta \in \mathbb{R}, \alpha > 0, \gamma > 0 \quad (3.5)$$

em que

α representa o parâmetro de escala;

β representa o parâmetro de posição; e

γ representa o parâmetro de forma.

A distribuição dada pela Equação 3.5 é conhecida como Weibull inversa, pois é limitada superiormente, diferente das distribuições Fréchet e Gumbel (PAPALEXIOU & KOUTSOYIANNIS, 2013). A distribuição Gumbel é amplamente utilizada, sendo inclusive indicada por Eletrobrás (2000), juntamente com a distribuição exponencial de 2 parâmetros, para a análise de frequência de vazões máximas. A distribuição Fréchet chegou a ser recentemente recomendada para a análise de frequência de precipitações diárias máximas anuais, em detrimento da distribuição Gumbel, devido ao comportamento polinomial de sua

cauda superior (PAPALEXIOU & KOUTSOYIANNIS, 2013). A distribuição Fréchet também é conhecida como Log-Gumbel, denominação que se justifica pelo fato de que, se uma dada variável aleatória (VA) se distribuir conforme a Gumbel, o logaritmo dessa VA se distribuirá segundo uma distribuição Fréchet (NAGHETTINI & PINTO, 2007).

As distribuições de máximos apresentadas nas Equações 3.3, 3.4 e 3.5 podem ser sintetizadas em uma única distribuição, denominada GEV, ou Generalizada de Valores Extremos, cuja função acumulada de probabilidades é apresentada na Equação 3.6 (JENKINSON, 1955).

$$G(x) = \exp \left[- \left(1 - \gamma \frac{x - \beta}{\alpha} \right)^{\frac{1}{\gamma}} \right] \quad (3.6)$$

A família de distribuições GEV pode ser dividida em três classes, associadas a valores distintos do parâmetro de forma γ : i) $\gamma < 0$ corresponde à distribuição Fréchet (EV2); ii) $\gamma = 0$ corresponde à distribuição Gumbel (EV1); e iii) $\gamma > 0$ corresponde à distribuição Weibull inversa (EV3). Na prática, os valores de γ situam-se usualmente no intervalo $[-0,5; 0,5]$ (ADLOUNI *et al.*, 2008).

As distribuições teóricas assintóticas de máximos não são as únicas utilizadas na análise de frequência de máximos. De fato, em seu estudo, Brooks e Carruthers (1953) *apud* Adlouni *et al.* (2008) concluíram que a distribuição Gumbel tende a subestimar a altura das precipitações menos frequentes, em função das características de sua cauda superior. Dada a subjetividade da escolha da distribuição representativa das vazões médias diárias máximas anuais e as possíveis inconsistências decorrentes do uso de distribuições diferentes em um mesmo contexto, Benson (1968) avaliou a metodologia mais apropriada para a análise de frequência. Entre as distribuições analisadas estavam Gumbel, Fréchet, Log-Normal e Log-Pearson 3. A distribuição Log-Pearson 3 foi considerada a que melhor aderiu aos dados e, portanto, foi recomendada, juntamente com a distribuição Log-Normal (caso particular da distribuição LP3 para quando a assimetria do logaritmo dos dados é nula). Essa conclusão tornou-se diretriz para a análise de frequência de vazões máximas nos Estados Unidos, onde padronizou-se a utilização da distribuição LP3 para tanto (US WATER RESOURCES COUNCIL, 1981).

Poucos estudos foram desenvolvidos no Brasil com o objetivo de definir as distribuições aplicáveis às vazões médias diárias máximas anuais. Destaca-se o trabalho de Costa (2014), no qual foram estudadas séries hidrológicas originadas de 1.253 postos fluviométricos de todo

o país. As distribuições candidatas consideradas foram a LN2, Exponencial, Gama, Gumbel, PE3, GEV, LP3, LN3 e Generalizada de Pareto. Pelos critérios definidos na pesquisa, foram indicadas, para a modelagem das vazões máximas, as distribuições Log-Normal de 2 e 3 parâmetros, juntamente com as distribuições Log-Pearson 3 e Gumbel. Apenas a distribuição Generalizada de Pareto foi considerada inadequada para aplicação sobre os registros fluviométricos brasileiros.

A estimativa de quantis de referência, seja de eventos máximos ou mínimos, é diretamente influenciada pela cauda da distribuição utilizada para representar as observações das variáveis hidrológicas. Papalexiou *et al.* (2013) apontam, de modo simplificado, que as distribuições de probabilidades podem ser classificadas de duas formas: aquelas com a cauda pesada e aquelas com cauda leve. As distribuições com cauda pesada apresentam função densidade de probabilidade com decaimento lento (seja na cauda superior e/ou na cauda inferior). Assim, a eventos máximos, por exemplo, distribuições com cauda pesada associam frequências maiores do que as distribuições com cauda leve. A Figura 3.1 apresenta a classificação de algumas das distribuições de eventos máximos mais utilizadas na Engenharia de Recursos Hídricos em relação ao “peso” da cauda. Apesar de não indicado explicitamente, existem métodos criteriosos para classificar as caudas das distribuições, como aquele apresentado por Ouarda *et al.* (1994) *apud* Adlouni *et al.* (2008).

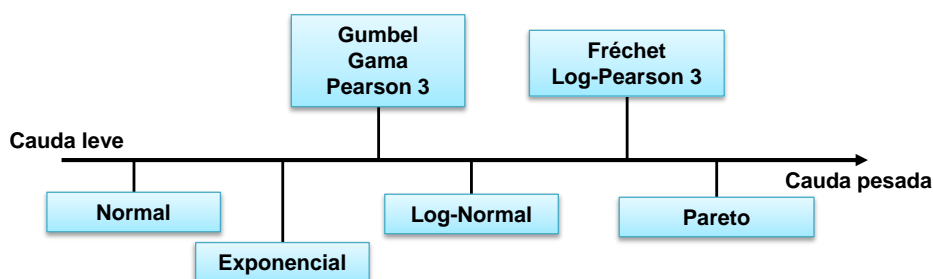


Figura 3.1 – Classificação das distribuições comumente utilizadas na Engenharia de Recursos Hídricos, em relação ao comportamento da cauda
 Fonte: Adaptado de Adlouni *et al.* (2008)

Indícios de que os eventos hidrológicos máximos são governadas por distribuições de probabilidades com cauda pesada são apontados por diversos autores (KATZ *et al.*, 2002; BOWERS *et al.*, 2012; PAPALEXIOU *et al.*, 2013). Papalexiou e Koutsoyiannis (2013) obtiveram as distribuições máximas exatas para precipitações diárias reais, considerando cenários em que as alturas pluviométricas maiores que zero seguem a distribuição Burr tipo XII (BrXII) e a distribuição Gama Generalizada (GG). As distribuições resultantes estiveram

dentro do domínio de atração das distribuições Fréchet e Gumbel, respectivamente. Essas duas distribuições foram ajustadas tanto para os dados BrXII quanto para os dados GG. Os quantis populacionais foram obtidos por meio da aplicação direta da Equação 3.1 sobre a função acumulada de probabilidades das distribuições BrXII e GG. Independente da distribuição parental dos dados, a distribuição Fréchet, com cauda mais pesada do que a distribuição Gumbel, foi a que mais se aproximou dos quantis populacionais extremos.

De modo análogo às distribuições teóricas de extremos máximos, podem ser obtidas as distribuições teóricas de mínimos. Considerando a variável aleatória $Z = \text{mínimo}(X_1, X_2, \dots, X_N)$, sua distribuição acumulada de probabilidades é dada por (NAGHETTINI & PINTO, 2007)

$$H_Z(x) = 1 - (1 - F_X(x))^N \quad (3.7)$$

O desenvolvimento da Equação 3.7 retorna distribuições teóricas também denominadas Gumbel, Fréchet e Weibull, porém em suas versões para eventos mínimos. As Equações 3.8 a 3.10 apresentam, respectivamente, tais distribuições na forma acumulada (GOTTSCHALK *et al.*, 2013).

$$H(x) = 1 - \exp\left[-\exp\left(\frac{x-\beta}{\alpha}\right)\right], x \in \mathbb{R}, \beta \in \mathbb{R}, \alpha > 0 \quad (3.8)$$

$$H(x) = 1 - \exp\left[-\left(-\frac{x-\beta}{\alpha}\right)^{\frac{1}{\gamma}}\right], \beta \geq x, \beta \in \mathbb{R}, \alpha > 0, \gamma > 0 \quad (3.9)$$

$$H(x) = 1 - \exp\left[-\left(\frac{x-\beta}{\alpha}\right)^{\frac{1}{\gamma}}\right], x \geq \beta, \beta \in \mathbb{R}, \alpha > 0, \gamma > 0 \quad (3.10)$$

Dentre as distribuições apresentadas nas Equações 3.8 a 3.10, a única considerada aplicável para representar vazões mínimas é a distribuição Weibull, pois somente ela é limitada inferiormente (GOTTSCHALK *et al.*, 2013).

Outras variáveis aleatórias são mais bem representadas por distribuições diferentes destas aqui apresentadas para eventos máximos. Com relação às vazões sazonais (em períodos secos e chuvosos), Bowers *et al.* (2012) apontam que a distribuição Log-Normal se ajusta bem às observações. Já para as precipitações mensais, sazonais e anuais, Yue e Hashino (2007) apontaram as distribuições Log-Pearson 3, Pearson 3, GEV e Log-Normal de três parâmetros.

Diversos outros estudos analisam a distribuição probabilística de certos eventos hidrológicos em várias regiões do planeta, destacando-se, em todos eles, as distribuições aqui citadas.

A escolha da distribuição mais adequada deve ser feita caso a caso, utilizando as ferramentas disponíveis, como os testes de aderência. Apesar de haver distribuições mais utilizadas em determinadas situações, garantir que o modelo adotado é o mais adequado não é uma tarefa simples.

As distribuições de probabilidades aqui citadas fornecem modelos do comportamento aleatório das variáveis frequentemente estudadas na Hidrologia. Contudo, a representatividade desses modelos fica condicionada a vários outros aspectos, entre os quais está a eficácia da estimativa dos seus parâmetros em cada caso estudado. A seguir, apresenta-se uma breve revisão das técnicas estatísticas de estimação de parâmetros.

3.1.2 Estimação de parâmetros

Os parâmetros de determinados modelos probabilísticos, como distribuições de probabilidades e modelos de regressão, são estimados em função dos dados amostrais disponíveis. Um conjunto de dados amostrais x com extensão n (x_1, x_2, \dots, x_n) é entendido como uma das infinitas realizações possíveis da variável aleatória X com extensão n . Como o valor populacional de um parâmetro θ não pode ser determinado sem o conhecimento de toda a população de X , são feitas estimativas diferentes desse parâmetro ($\hat{\theta}$) para cada realização disponível de X . Ainda, é possível utilizar várias expressões matemáticas, denominadas estimadores, para determinar um valor de $\hat{\theta}$, dada a amostra de X . A qualidade dos estimadores pode ser determinada por meio de quatro propriedades desejáveis (NAGHETTINI e PINTO, 2007):

- Ausência de viés: um estimador é considerado não enviesado se seu valor esperado, para várias realizações de X , for igual ao valor populacional do parâmetro estudado. Essa característica é definida matematicamente pela expressão

$$E[\hat{\theta}] = \theta, \tag{3.11}$$

na qual $E[\hat{\theta}]$ denota o valor esperado das estimativas de θ ;

- Consistência: um estimador é definido como consistente se a seguinte expressão é válida:

$$\lim_{n \rightarrow \infty} P[|\hat{\theta} - \theta| \leq \varepsilon] = 1, \quad (3.12)$$

na qual ε é qualquer número positivo.

A Equação 3.12 pode ser entendida como a condição de que, para uma amostra de extensão tendendo ao infinito, a probabilidade de $\hat{\theta}$ se igualar a θ é unitária;

- Eficiência: um estimador não enviesado é considerado o mais eficiente entre todos os outros estimadores se a variância das suas estimativas, obtida para várias realizações da variável aleatória, for menor ou igual à variância das estimativas obtidas por meio de qualquer outro estimador não enviesado; e
- Suficiência: um estimador é considerado suficiente se utiliza todas as informações sobre o parâmetro estimado fornecidas pelos dados amostrais, de tal forma que nenhuma outra informação possa ser acrescentada por qualquer outro estimador.

Outras propriedades dos estimadores também são importantes, como o erro das estimativas de modelos de regressão, de preenchimento de falhas ou de extensão de séries hidrológicas, por exemplo. Como apontado por Van Der Heijden *et al.* (2004), o erro das estimativas corresponde à soma entre o viés e uma parcela aleatória. Já a estabilidade é definida como a capacidade de os estimadores serem insensíveis a pequenos erros aleatórios nas observações. Por fim, a robustez consiste na propriedade de o estimador ser insensível a grandes erros de observação em alguns registros (*outliers*).

Entre os métodos de estimação de parâmetros das distribuições de probabilidades utilizadas na Engenharia de Recursos Hídricos estão: i) o método dos momentos (MOM); ii) o método da máxima verossimilhança (ML, do inglês *Maximum Likelihood*); iii) o método dos mínimos quadrados (OLS); iv) o método dos momentos mistos; vi) o método da máxima entropia; vii) o método generalizado dos momentos; e viii) o método dos momentos-L (NAGHETTINI e PINTO, 2007). Desses métodos, os mais utilizados na prática da engenharia e mais estudados na literatura científica, especialmente no contexto de sua aplicação na Engenharia de Recursos Hídricos, são os métodos i), ii) e viii).

O método dos momentos (MOM) consiste em igualar os momentos populacionais aos momentos amostrais da distribuição de probabilidades. Para uma amostra aleatória simples

(AAS), constituída de n observações IID da variável x (x_1, x_2, \dots, x_n), obtida por meio de uma realização da variável aleatória distribuída de acordo com $f_X(x; \theta_1, \theta_2, \dots, \theta_p)$ com p parâmetros, o método MOM é definido pela expressão (NAGHETTINI e PINTO, 2007)

$$\mu_j(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p) = m_j \quad \text{com } j = 1, 2, \dots, p \quad (3.13)$$

na qual μ_j e m_j representam, respectivamente, os momentos populacionais e amostrais de ordem j . Os parâmetros são obtidos por meio da solução do sistema de equações resultante da Equação 3.13.

O método da máxima verossimilhança (ML) apresenta-se como uma técnica capaz de fornecer soluções otimizadas para diversos problemas paramétricos, não apenas a estimação de parâmetros das distribuições de probabilidades. Os primeiros usos documentados do método ML, ainda sem tal denominação, remontam ao século XVIII (STIGLER, 2007).

Considerando o mesmo conjunto de dados de x apresentado acima, a função densidade de probabilidade conjunta (FDPC) da amostra é dada por $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f_X(x_1)f_X(x_2) \dots f_X(x_n)$. Essa FDPC, denominada função de verossimilhança, é proporcional à probabilidade de que a AAS tenha sido extraída da população descrita pela FDP $f_X(x; \theta_1, \theta_2, \dots, \theta_p)$. A função de verossimilhança é dada, portanto, por (NAGHETTINI e PINTO, 2007):

$$L(\theta_1, \theta_2, \dots, \theta_p) = \prod_{i=1}^n f_X(x_i; \theta_1, \theta_2, \dots, \theta_p) \quad (3.14)$$

O processo de estimação dos parâmetros pelo método ML consiste em maximizar a função de máxima verossimilhança para o mesmo número de parâmetros p , por meio do sistema de equações formado por (NAGHETTINI e PINTO, 2007)

$$\frac{\partial L(\theta_1, \theta_2, \dots, \theta_p)}{\partial \theta_j} = 0, \quad j = 1, 2, \dots, p \quad (3.15)$$

O método dos momentos-L consiste em igualar os momentos-L populacionais aos momentos-L amostrais. Os momentos-L, também conhecidos como momentos ponderados por

probabilidades, foram introduzidos por Greenwood *et al.* (1979), sendo definidos pela expressão

$$M_{p,r,s} = E[X^p [F_x(x)]^r [1 - F_x(x)]^s] = \int_0^1 [x(F)]^p F^r (1 - F)^s dF \quad (3.16)$$

em que

p , r e s são números reais; e

$x(F)$ denota a função de quantis da distribuição com FAP igual a F .

Quando r e s são iguais a zero, e p é um número não negativo, os momentos $M_{p,0,0}$ são iguais aos momentos convencionais de ordem p em relação à origem (μ'_p). Os momentos $M_{1,0,s}$ e $M_{1,r,0}$ são os mais frequentemente utilizados no ajuste de distribuições de probabilidades (NAGHETTINI e PINTO, 2007).

Cada um dos métodos aqui apresentados possui vantagens e desvantagens, algumas das quais estão sintetizadas na Tabela 3.1.

Tabela 3.1 - Algumas vantagens e desvantagens de cada método de estimação de parâmetros

Método	Vantagens	Desvantagens
Momentos (MOM)	<ul style="list-style-type: none"> • Simplicidade de cálculo; e • Aplicabilidade direta a todas as distribuições utilizadas na Engenharia de Recursos Hídricos. 	<ul style="list-style-type: none"> • Não inclui os estimadores mais eficientes; • Tem desempenho inferior aos demais métodos de estimação; • Caso utilize o estimador tradicional do coeficiente de assimetria, pode levar a erros significativos devido à sensibilidade deste estimador a valores extremos, que são mais detalhadas por Griffis e Stedinger (2009) e Brys <i>et al.</i> (2004); e • É mais difícil de ser generalizado (para o caso de dados dependentes, dados não identicamente distribuídos, etc.) (VAN GELDER, 2004).
Máxima Verossimilhança (ML)	<ul style="list-style-type: none"> • Fornece o estimador não enviesado com menor erro quadrático médio à medida que a extensão da amostra tende a infinito (VAN GELDER, 2004); • A função de máxima verossimilhança pode ser utilizada para, além de determinar os estimadores pontuais ML, fornecer um intervalo que contém valores plausíveis para o estimador (VAN GELDER, 2004); • Os estimadores ML são normalmente distribuídos (no caso assintótico) (VAN GELDER, 2004); e • Usualmente são de implementação simples em computador (VAN GELDER, 2004). 	<ul style="list-style-type: none"> • O uso do método ML pode gerar grandes erros ao analisar amostras pequenas (KATZ <i>et al.</i>, 2002); • Suas vantagens pressupõem os casos assintóticos (grandes amostras, por exemplo, caso que não é usual na engenharia de recursos hídricos) (KATZ <i>et al.</i>, 2002; VAN GELDER, 2004); e • Os estimadores ML podem não existir em algumas situações (VAN GELDER, 2004).
Momentos-L	<ul style="list-style-type: none"> • Fornece estimadores relativamente resistentes a <i>outliers</i> (VAN GELDER, 2004); • Simplicidade computacional (KATZ <i>et al.</i>, 2002); e • Bom desempenho para amostras pequenas (KATZ <i>et al.</i>, 2002). 	<ul style="list-style-type: none"> • Suas vantagens podem não se refletir em funções matemáticas dos momentos L (como as funções quantis) (VAN GELDER, 2004); • Para algumas distribuições, outros métodos de estimação podem ser superiores em termos do erro quadrático médio dos estimadores dos quantis (HOSKING e WALLIS, 1987; ROSBJERG <i>et al.</i>, 1992); e • A “resistência a <i>outliers</i>” pode levar à minoração da importância dos dados amostrais que trazem informação sobre quantis extremos (VAN GELDER, 2004).

Os métodos de estimação de parâmetros aqui apresentados aplicam-se ao ajuste de distribuições de probabilidades. Outra aplicação da estimação de parâmetros consiste em ajustar modelos de regressão entre variáveis aleatórias, que constituem a base das técnicas de extensão das séries hidrológicas apresentadas a seguir.

3.2 Técnicas de extensão de séries hidrológicas

As técnicas de extensão de séries hidrológicas podem ser definidas como métodos de amostragem indireta de séries temporais que utilizam, como referência, as observações de outra série hidrológica, obtendo séries estendidas com determinadas propriedades estatísticas. Diversas técnicas de extensão foram desenvolvidas especialmente com o objetivo de estender as séries hidrológicas. Outras são adaptações de modelos que possuem, originalmente, propósitos distantes da extensão das séries hidrológicas, como modelos estatísticos de regressão e técnicas de preenchimento de falhas.

Para a apresentação das técnicas de extensão de séries hidrológicas, será considerada a seguinte convenção: a série hidrológica curta, variável dependente nas expressões das técnicas, é composta pelos componentes y_i , totalizando, num primeiro momento, n_1 registros. A série de referência, utilizada como variável independente nas formulações, é composta pelos elementos x_i e possui $n_1 + n_2$ membros. Os registros de y_1 a y_{n_1} e de x_1 a x_{n_1} formam um período de observações comuns, sobre o qual é feito o ajuste das técnicas de extensão. Nesse período, a correlação linear amostral entre x e y é igual a r , que é representativa do coeficiente de correlação populacional ρ , entre X e Y . Não existe correlação serial entre os registros, de forma que não é necessário que as observações sejam consecutivas. É imprescindível, no entanto, que x_i e y_i representem observações dentro um mesmo intervalo de tempo. Esta configuração é mais facilmente entendida quando apresentada no formato

$$\begin{array}{ll} \text{Série longa:} & x_1, x_2, x_3, \dots, x_{n_1}, x_{n_1+1}, x_{n_1+2}, \dots, x_{n_1+n_2} \\ \text{Série curta:} & y_1, y_2, y_3, \dots, y_{n_1} \end{array}$$

A extensão da série curta consiste na estimativa dos valores de y_{n_1+1} a $y_{n_1+n_2}$, de forma a obter, para todo o conjunto dos $n_1 + n_2$ valores da série estendida, medidas descritivas tão próximas quanto possível das medidas descritivas populacionais da variável aleatória Y .

O desenvolvimento das técnicas de extensão de séries hidrológicas derivou de avanços estatísticos obtidos antes mesmo de meados do século passado, para o tratamento de falhas em séries temporais. O cenário estudado por Wilks (1932) consistiu em estimar os parâmetros de uma distribuição normal bivariada a partir de registros das variáveis aleatórias X e Y , com um período de observações comuns e falhas em cada uma das amostras. Concluiu-se que as estimativas são mais confiáveis quando todo o universo de observações, tanto de X quanto de

Y , é considerado nas estimativas, em relação a estimar os parâmetros populacionais de X ou de Y apenas com base nas observações de cada variável aleatória.

O estudo de Matthai (1951) trouxe colaborações, em estudos de amostragem, para o caso em que um fenômeno associado a determinada variável aleatória não pode ser observado. Por meio da aplicação de técnicas estatísticas de preenchimento de falhas, e tomando como base informações de variáveis observáveis, foram obtidos avanços no conhecimento da distribuição de probabilidades que descreve esse fenômeno. Foi estudado predominantemente o caso da distribuição normal multivariada envolvendo três variáveis.

Alguns estudos, como aquele desenvolvido por Lord (1955), utilizaram o indicador denominado razão de informação relativa (*relative-information ratio*), designado por I , para avaliar o desempenho das técnicas de extensão na estimativa das medidas populacionais. Esse indicador é definido de tal forma que, caso seja maior que 1, a variância da estimativa de um momento populacional considerando a série original (curta) é superior à variância da estimativa feita utilizando também os demais valores estimados pelas técnicas de extensão. Por outro lado, caso $I < 1$, a estimativa por meio série curta, apenas, é mais eficiente do que após a extensão. É desejável, após a extensão da série original, que I seja maior que 1. As expressões de I^{-1} , para a estimativa da média e da variância de y , são apresentadas nas Equações 3.17 e 3.18, respectivamente.

$$I_{\mu}^{-1} = 1 - \frac{n_1}{n_1 + n_2} \rho^2 \quad (3.17)$$

$$I_{\sigma}^{-1} = 1 - \frac{n_1}{n_1 + n_2} \rho^4 \quad (3.18)$$

Utilizando os estimadores do método da máxima verossimilhança para extrair informações populacionais sobre a distribuição normal multivariada, Lord (1955) concluiu que, desde que n_1 e n_2 fossem maiores que zero, o indicador I seria maior ou igual a 1 para a estimativa da média e da variância populacionais, quaisquer fossem os valores das demais variáveis envolvidas. Portanto, independente do coeficiente de correlação entre as séries estudadas, não haveria prejuízo nas estimativas dos momentos populacionais, ao utilizar as séries estendidas em detrimento das séries originais. Contudo, como apresentado na Tabela 3.1, o método da máxima verossimilhança e, portanto, as conclusões de Lord (1955), são válidas apenas para o caso assintótico em que a extensão da série curta tende para valores elevados.

Conclusões para casos práticos começaram a ser obtidas por Langbein e Hardison (1955) *apud* Fiering (1963), tendo-se chegado à conclusão de que só há ganho de informação para a estimativa da média populacional, após a extensão da série original de y , quando o coeficiente de correlação obedece à relação

$$|\rho| > \frac{1}{\sqrt{n_1 - 2}} \quad (3.19)$$

O resultado da Equação 3.19 é válido caso as seguintes premissas sejam atendidas (FIERING, 1963):

- existência de um regime hidrológico estável na região onde as observações de X e Y foram feitas, de forma que um nível significativo de correlação entre os registros seja esperado;
- as variáveis aleatórias estudadas podem ser modeladas pela distribuição Normal;
- ausência de autocorrelação entre os registros;
- existência de um período de observações comuns entre X e Y , com n_1 elementos, e X possui n_2 observações adicionais; e
- as observações de X e Y podem ser relacionadas por meio da equação de uma reta.

A primeira premissa baseia-se no fato de que os registros das séries hidrológicas que se prestam à aplicação das técnicas de extensão são representativos dos fenômenos hidrológicos observados na região de estudo. A estabilidade do regime hidrológico traduz-se na constância da correlação populacional entre X e Y ao longo do tempo. Já a última premissa diz respeito à técnica de extensão utilizada por Fiering (1963), que foi a OLS.

3.2.1 OLS (*Ordinary Least Squares*)

A técnica OLS (*Ordinary Least Squares* ou Mínimos Quadrados Ordinário) pode ser apresentada nos formatos

$$\hat{y}_i = a + bx_i, \text{ e} \quad (3.20)$$

$$\hat{y}_i = \bar{y}_C + r(s_{y_C}/s_{x_C})(x_i - \bar{x}_C), \quad (3.21)$$

nos quais

\hat{y}_i e x_i denotam os valores estimado e observado dos elementos das séries históricas curta e longa, respectivamente;

a e b são parâmetros do modelo de regressão;

\bar{y}_C e \bar{x}_C denotam as médias aritméticas de y_C e x_C , respectivamente, sendo que y_C e x_C representam o período de observações comuns das séries históricas curta e longa, respectivamente, para $i = 1, 2, 3, \dots, n_1$;

s_{y_C} e s_{x_C} denotam os valores do desvio-padrão de y_C e x_C , respectivamente; e

r denota o coeficiente de correlação de Pearson entre y_C e x_C .

A técnica OLS possui o caso simples, para a regressão entre duas variáveis aleatórias, e a regressão múltipla, para três ou mais. As Equações 3.20 e 3.21 representam o primeiro caso.

As Equações 3.20 e 3.21 apresentam o ajuste da técnica OLS aos dados amostrais. Na verdade, os parâmetros a e b são estimativas dos parâmetros α e β que descrevem a relação linear populacional entre as variáveis aleatórias X e Y em questão, representada pela equação (HELSEL & HIRSCH, 2002)

$$Y = \alpha + \beta X + \varepsilon \quad (3.22)$$

em que ε denota o resíduo do modelo, uma variável aleatória normal com média 0 e desvio padrão σ_e .

Entendendo que a técnica OLS fornece meios para a estimativa de y_i , dado um valor particular de x_i , o resíduo dessa estimativa pertence à variável aleatória ε , cuja estimativa populacional é designada por e . Tradicionalmente, assume-se a hipótese de que as estimativas de y são variáveis aleatórias normalmente distribuídas, cuja média é obtida utilizando a Equação 3.20 ou 3.21 (HELSEL e HIRSCH, 2002). Essa interpretação é apresentada visualmente na Figura 3.2.

O método de cálculo dos coeficientes da regressão linear simples mais utilizado é o Mínimos Quadrados Ordinário, que dá o nome à técnica OLS. Por meio da minimização da equação do quadrado do resíduo, $e_i^2 = (y_i - a - bx_i)^2$, para todo i , obtém-se um sistema de equações normais, a partir do qual são obtidos estimadores dos coeficientes a e b (NAGHETTINI e PINTO, 2007). Conforme comprovado pelo teorema Gauss Markov, a técnica OLS é a mais eficiente para estimativa dos parâmetros de a e b , ao serem atendidas as premissas de: i) linearidade entre Y e X ; e ii) independência e homoscedasticidade dos resíduos (WOOD,

2006). A Tabela 3.2 apresenta um resumo das principais expressões derivadas da determinação dos coeficientes de regressão por meio da técnica OLS.

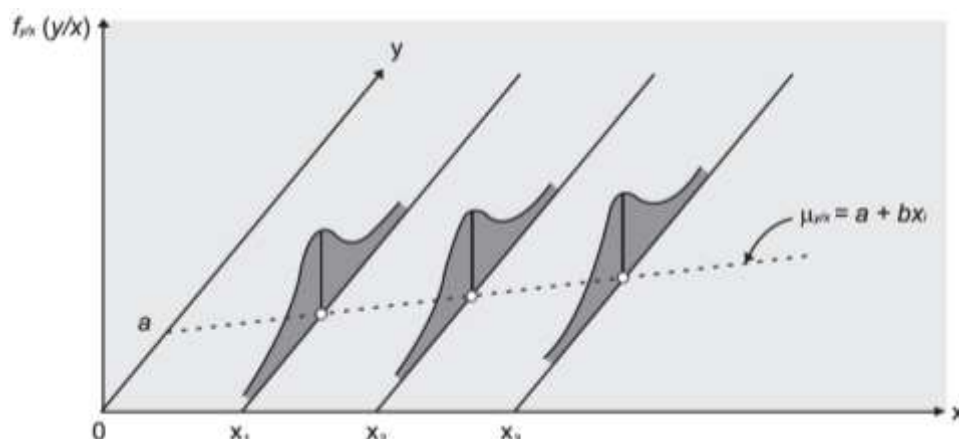


Figura 3.2 - Hipótese de normalidade das estimativas da regressão linear

Fonte: NAGHETTINI e PINTO, 2007.

Tabela 3.2 - Principais expressões utilizadas na aplicação da regressão linear simples (técnica OLS)

Variável	Equação
Coeficiente angular da reta de regressão	$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \quad (3.23)$
Intercepto da reta de regressão	$a = \bar{y} - b\bar{x} \quad (3.24)$
Estimativa de y em função de x	$\hat{y}_i = a + bx_i \quad (3.25)$
Erro-padrão ou estimativa da variância dos resíduos	$\hat{\sigma}_e^2 = s_e^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2} \quad (3.26)$
Coeficiente de determinação	$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = b^2 \frac{s_x^2}{s_y^2} \quad (3.27)$

Fonte: Adaptado de HELSEL e HIRSCH, 2002, e NAGHETTINI e PINTO, 2007.

O coeficiente de determinação, R^2 , é definido como a proporção da variância total da variável dependente que é explicada pelo modelo de regressão, e, portanto, situa-se entre 0 e 1. O coeficiente de correlação amostral r de Pearson relaciona-se com o coeficiente de determinação do método OLS por meio da equação (NAGHETTINI e PINTO, 2007)

$$r = \pm \sqrt{R^2} \quad (3.28)$$

sendo que o coeficiente de correlação linear r de Pearson pode ser obtido diretamente dos dados amostrais a partir da equação

$$r = \frac{s_{X,Y}}{s_X s_Y} = \frac{\sum_{i=1}^n [(x_i - \bar{x})(y_i - \bar{y})]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.29)$$

em que $s_{X,Y}$ denota a estimativa da covariância entre as variáveis X e Y .

As medidas de correlação fornecem uma medida do grau de associação entre duas variáveis. Uma forte correlação não indica necessariamente uma relação causal entre elas, apesar de poder indicar que as variáveis compartilham uma ou mais causas. Uma relação causa-efeito está fora do escopo da análise estatística e só poderá ser determinada por meio do conhecimento físico do processo envolvido (HELSEL e HIRSCH, 2002).

As medidas de correlação ρ são adimensionais e situam-se na faixa $-1 \leq \rho \leq 1$. A ausência de correlação é indicada por $\rho = 0$ e ρ possui sinal positivo quando X e Y possuem mesmo comportamento (caracterizam uma relação crescente). ρ possui sinal negativo quando as variáveis possuem comportamento oposto, caracterizando uma relação decrescente. Quando uma das variáveis descreve tempo ou espaço, ρ torna-se uma medida de tendência temporal ou espacial (HELSEL e HIRSCH, 2002).

O coeficiente de correlação r de Pearson é aplicável a relações monotônicas lineares. A condição de linearidade torna r mais sensível à presença de *outliers*, dado que seu cálculo utiliza medidas pouco robustas, como médias e desvios-padrão. Cabe ressaltar, ainda, que r é invariante em relação a mudanças de escala, uma vez que é medido por meio da distância entre os termos das séries de dados em relação à média normalizada pelo desvio padrão (HELSEL e HIRSCH, 2002).

A regressão linear múltipla (MLR, do inglês *Multiple Linear Regression*) é a extensão da regressão linear simples para o caso em que há mais de uma variável explicativa. A evidência de que variáveis explicativas adicionais são necessárias para o modelo podem vir do entendimento científico do fenômeno analisado ou mesmo da análise dos resíduos da regressão linear simples. Os resíduos podem indicar tendências temporais, espaciais, sazonalidade, etc. Nesses casos, os resíduos do modelo de regressão linear simples podem ser plotados em função de muitas variáveis, até que a decisão sobre qual variável adicional deveria ser incorporada ao modelo original seja tomada (HELSEL e HIRSCH, 2002).

O modelo de regressão linear múltipla é representado pela expressão

$$Y = \beta_0 X_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K + \varepsilon \quad (3.30)$$

em que

β_0 é o intercepto do modelo, com $X_0 = 1$; e

β_K é o coeficiente angular associado à variável X_K , com $K > 0$

A estimativa dos parâmetros de um modelo MLR é feito para a estrutura apresentada na equação

$$\mathbf{Y} = \mathbf{X} \times \boldsymbol{\beta} \quad (3.31)$$

em que

\mathbf{Y} é um vetor ($n \times 1$) das n observações da variável dependente;

$\boldsymbol{\beta}$ é um vetor ($K \times 1$) dos coeficientes das K respectivas variáveis explicativas; e

\mathbf{X} é uma matriz ($n \times K$) com as n observações de cada variável K .

Um resumo das expressões utilizadas para ajuste da regressão MLR e para a determinação de outras grandezas relevantes é apresentado na Tabela 3.3.

Tabela 3.3 - Principais expressões utilizadas na aplicação da regressão linear múltipla (técnica OLS)

Nome	Equação	
Vetor dos coeficientes do modelo	$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$	(3.32)
Somatório dos quadrados do modelo de regressão	$SQ\ Reg = \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y} - n \bar{Y}^2$	(3.33)
Somatório dos quadrados dos resíduos do modelo de regressão	$SQ\ Res = \mathbf{Y}^T \mathbf{Y} - \hat{\boldsymbol{\beta}}^T \mathbf{X}^T \mathbf{Y}$	(3.34)
Somatório dos quadrados totais	$SQT = \mathbf{Y}^T \mathbf{Y} - n \bar{Y}^2$	(3.35)
Graus de liberdade do modelo de regressão	K	(3.36)
Graus de liberdade dos resíduos	$n - K - 1$	(3.37)
Quadrado médio do modelo de regressão	$QM\ Reg = \frac{SQ\ Reg}{K}$	(3.38)
Quadrado médio dos resíduos, erro-padrão ou estimativa da variância dos resíduos	$QM\ Res = s_e = \frac{SQ\ Res}{n - K - 1}$	(3.39)
Coefficiente de determinação	$R^2 = \frac{SQ\ Reg}{SQT}$	(3.40)

Fonte: adaptado de NAGHETTINI e PINTO, 2007.

As hipóteses da regressão linear simples, que corresponde ao caso abordado nesta pesquisa, são apresentadas na Tabela 3.4.

Tabela 3.4 - Hipóteses da regressão linear simples (técnica OLS)

Hipótese	Objetivo da aplicação da técnica OLS			
	Estimar Y em função de X	Estimar Y e a variância da estimativa	Obter o melhor estimador não enviesado de Y	Testar hipóteses e estimar intervalos de confiança
1) Y é realmente uma função linear de X	Sim	Sim	Sim	Sim
2) Os dados utilizados para construir a reta de regressão são representativos	Sim	Sim	Sim	Sim
3) Homoscedasticidade (variância dos resíduos é constante em função de X)			Sim	Sim
4) Os resíduos são independentes				Sim
5) Os resíduos pertencem a uma distribuição normal				Sim

Fonte: Adaptado de HELSEL e HIRSCH, 2002.

A aplicação da técnica OLS implica que (HELSEL e HIRSCH, 2002):

- Se as hipóteses 1 a 4 da Tabela 3.4 forem atendidas, os estimadores a e b são os estimadores não enviesados com menor variância de α e β ;
- A média dos resíduos (e_i) é exatamente zero;
- A média das previsões (\hat{y}_i) é igual à média dos valores amostrais y_i ;
- A linha de regressão passa pelo centróide dos dados amostrais (\bar{x}, \bar{y}); e
- A variância das previsões (\hat{y}_i) é menor que variância dos dados amostrais (exceto quando $R^2 = 1$ e a variância das previsões iguala-se à variância dos dados amostrais).

Entretanto, como mostra a Tabela 3.4, a simples estimativa de Y não requer premissas relacionadas aos resíduos. Também é importante ressaltar que não são feitas hipóteses a respeito das distribuições das variáveis independentes ou dependentes, conclusão útil ao lidar com séries hidrológicas, frequentemente caracterizadas como assimétricas (HELSEL & HIRSCH, 2002). A hipótese de homoscedasticidade e normalidade dos resíduos só é necessária para a aplicação de testes de hipóteses, para a estimativa dos intervalos de confiança e para garantir que a técnica OLS fornece o melhor estimador não enviesado de cada registro. Para a extensão das séries hidrológicas, essas premissas não precisam ser necessariamente verdadeiras.

O fato de as previsões possuírem variância inferior à variância dos dados amostrais é de extrema importância na aplicação da técnica OLS como técnica de extensão de séries hidrológicas. Matalas & Jacobs (1964) observaram que, apesar dos estimadores de μ_Y obtidos para a série estendida serem não viesados, o uso da técnica OLS para a extensão da série histórica de y leva à obtenção de valores viesados de σ_Y^2 , uma vez que os valores previstos situam-se exatamente sobre a reta de regressão, como mostra a Figura 3.3. Na estimativa da média populacional, concluiu-se que o uso da técnica OLS na extensão da série curta conduz a um estimador mais eficiente do que a média da série original se a relação estabelecida pela Equação 3.19 for satisfeita. Para resolver o problema do viés da variância das estimativas de y , Matalas e Jacobs (1964) propuseram incorporar o resíduo à técnica de extensão, como descrito no item a seguir.

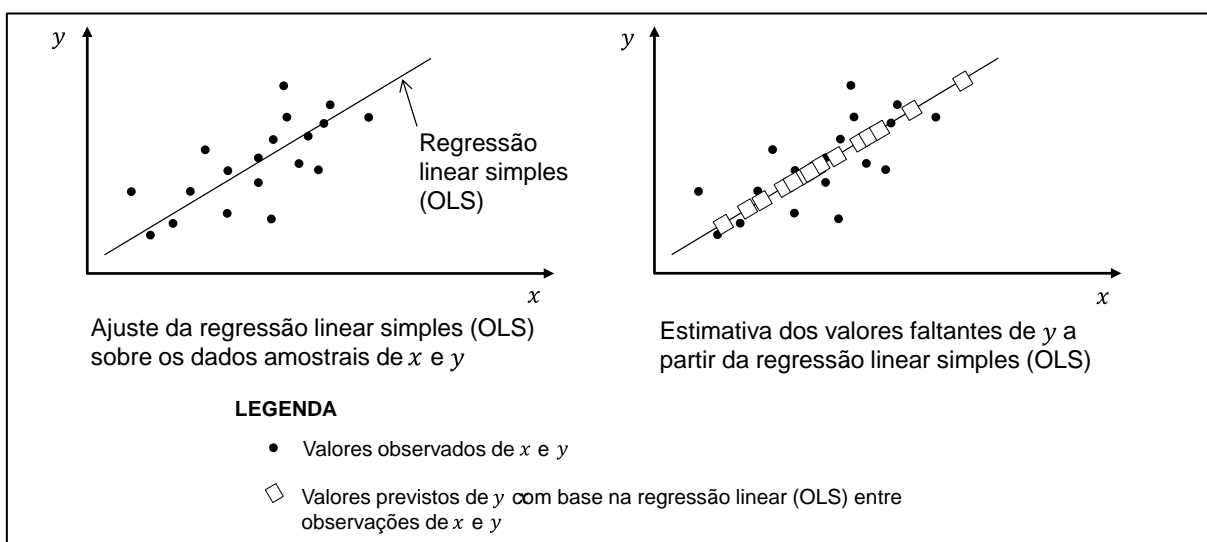


Figura 3.3 - Viés inserido nos valores previstos de y por meio da aplicação da técnica OLS

3.2.2 RPN (*Regression Plus Noise*)

Com o objetivo de eliminar o viés na estimativa da variância populacional e obter um novo estimador por meio da extensão da série histórica de y , mantendo a variabilidade da série original, Matalas & Jacobs (1964) propuseram a técnica *Regression Plus Noise* (Regressão Mais Ruído, ou simplesmente RPN), que consiste na incorporação de uma parcela de resíduo ao modelo da técnica OLS. A técnica RPN foi definida por meio da expressão

$$\hat{y}_i = \bar{y}_c + r \left(\frac{s_{y_c}}{s_{x_c}} \right) (x_i - \bar{x}_c) + \delta \sqrt{1 - r^2} s_{y_c} e_i \quad (3.41)$$

Na qual e_i denota um valor aleatório da distribuição normal padrão e

$$\delta^2 = \frac{n_2(n_1 - 4)(n_1 - 1)}{(n_2 - 1)(n_1 - 3)(n_1 - 2)} \quad (3.42)$$

Os estudos de Matalas e Jacobs (1964) mostraram que a média e a variância da série estendida por meio da utilização da técnica RPN são estimadores não enviesados dos momentos populacionais. Além disso, os autores também mostraram que esse estimador é mais eficiente do que a série curta, na estimativa da variância, se

$$|\rho| > \left[\frac{-B \pm \sqrt{B^2 - 4AC}}{2A} \right]^{\frac{1}{2}} \quad (3.43)$$

Em que

$$A = \left[\frac{(n_2 + 2)(n_1 - 6)(n_1 - 8)}{(n_1 - 3)(n_1 - 5)} - \frac{8(n_1 - 4)}{(n_1 - 3)} - \frac{2n_2(n_1 - 4)^2}{(n_1 - 3)^2} + \frac{n_1 n_2 (n_1 - 4)^2}{(n_1 - 3)^2 (n_1 - 2)} \right] \rho^4 \quad (3.44)$$

$$B = \left[\frac{6(n_2 + 2)(n_1 - 6)}{(n_1 - 3)(n_1 - 5)} + \frac{2(n_1^2 - n_1 - 14)}{(n_1 - 3)} + \frac{2n_2(n_1 - 4)(n_1 - 5)}{(n_1 - 3)^2} - \frac{2(n_1 - 4)(n_1 + 3)}{(n_1 - 3)} - \frac{2n_1 n_2 (n_1 - 4)^2}{(n_1 - 3)^2 (n_1 - 2)} \right] \rho^2 \quad (3.45)$$

$$C = \left[\frac{2(n_1 + 1)}{(n_1 - 3)} + \frac{3(n_2 + 2)}{(n_1 - 3)(n_1 - 5)} - \frac{(n_1 + 1)(2n_1 + n_2 - 2)}{(n_1 - 1)} + \frac{2n_2(n_1 - 4)}{(n_1 - 3)^2} + \frac{2(n_1 - 4)(n_1 + 1)}{(n_1 - 3)} + \frac{n_1 n_2 (n_1 - 4)^2}{(n_1 - 3)^2 (n_1 - 2)} \right] \quad (3.46)$$

A Tabela 3.5 sintetiza os valores mínimos do coeficiente de correlação populacional a partir dos quais os estimadores da média e da variância, para a série estendida, são mais eficientes do que os estimadores para a série original, considerando as técnicas OLS e RPN. É importante ratificar aqui que a estimativa da variância populacional, por meio da extensão utilizando a técnica OLS, é enviesada.

Matalas & Jacobs (1964) ainda fazem os seguintes comentários com relação às premissas assumidas no estudo:

- A estimativa do coeficiente de correlação populacional crítico é diretamente afetada pelo valor de n_1 . Para melhorar essa estimativa, o valor de n_1 deve ser o maior possível;

Tabela 3.5 – Valores críticos do coeficiente de correlação populacional para a eficiência dos estimadores da média e da variância populacional – técnicas OLS e RPN

Estimativa da média populacional – valores mínimos de ρ					
Técnica	$n_1 = 10$	$n_1 = 15$	$n_1 = 20$	$n_1 = 25$	$n_1 = 30$
OLS	0,35	0,28	0,24	0,21	0,19
RPN	0,35	0,28	0,24	0,21	0,19
Estimativa da variância populacional – valores mínimos de ρ					
Técnica OLS					
n_2	$n_1 = 10$	$n_1 = 15$	$n_1 = 20$	$n_1 = 25$	$n_1 = 30$
10	0,73	0,63	0,70	0,76	0,76
15	0,75	0,77	0,79	0,80	0,80
20	0,76	0,79	0,81	0,81	0,82
25	0,78	0,80	0,84	0,83	0,81
30	0,77	0,80	0,82	0,83	0,84
Técnica RPN					
n_2	$n_1 = 10$	$n_1 = 15$	$n_1 = 20$	$n_1 = 25$	$n_1 = 30$
10	0,65	0,54	0,52	0,42	0,38
15	0,65	0,54	0,51	0,42	0,39
20	0,65	0,54	0,51	0,42	0,39
25	0,65	0,54	0,50	0,42	0,39
30	0,65	0,54	0,50	0,42	0,39

Fonte: adaptado de Hirsch (1982)

- A maioria das variáveis hidrológicas apresenta assimetria e, portanto, não podem ser consideradas como normalmente distribuídas. A utilização de transformações pode auxiliar na normalização dos dados, bem como na linearização da relação entre as variáveis aleatórias X e Y ; e
- Algumas variáveis aleatórias estudadas em hidrologia não podem ser consideradas como independentes. Precipitações acumuladas ou vazões médias mensais tendem a ter valores elevados em sequência e valores baixos em sequência, em virtude da sazonalidade do regime hidrológico anual. A análise de variáveis aleatórias hidrológicas próprias de intervalos de tempo maiores, como precipitações acumuladas ou vazões médias anuais, reduz fortemente a dependência entre valores consecutivos da série de dados.

Hirsch (1982) aponta a grande desvantagem da técnica RPN, que se resume no fato de utilizar uma parcela aleatória na estimativa dos elementos da série estendida. Pode não ser o ideal obter várias realizações da série estendida para aplicações práticas. Além disso, basear decisões importantes de engenharia em apenas uma única realização da técnica RPN pode ser problemático. Uma alternativa à técnica RPN, com foco na melhoria da estimativa da variância da série estendida, é dada pelo conjunto de técnicas MOVE.

3.2.3 MOVE (*Maintenance of Variance Extension*)

3.2.3.1 MOVE1

A denominação MOVE refere-se a uma família de técnicas (MOVE1 a MOVE4). A técnica MOVE1 também é conhecida como LOC (*Line of Organic Correlation*), e teve suas primeiras aplicações na Engenharia de Recursos Hídricos na publicação de Kritskiy e Menkel (1968) *apud* Helsel e Hirsch (2002). As principais vantagens da técnica MOVE1 em relação à técnica OLS são (HELSEL e HIRSCH, 2002):

- A técnica MOVE1 minimiza erros nas variáveis X e Y , o que implica que a reta de regressão estimada é a mesma independente de qual variável, X ou Y , é considerada a variável explicativa; e
- Os valores estimados para a série estendida possuem propriedades que são mais similares às propriedades dos registros que se deseja representar, enquanto a aplicação da técnica OLS leva à estimativa de registros com variância subestimada (HIRSCH, 1982).

Os coeficientes do modelo de regressão MOVE1 são calculados para minimizar a soma das áreas dos triângulos formados pelas linhas verticais e horizontais originadas na linha de regressão (Figura 3.4). A expressão do modelo de regressão é dada por

$$\hat{y}_i = \bar{y}_C + \text{sign}[r] \frac{S_{yC}}{S_{xC}} (x_i - \bar{x}_C) \quad (3.47)$$

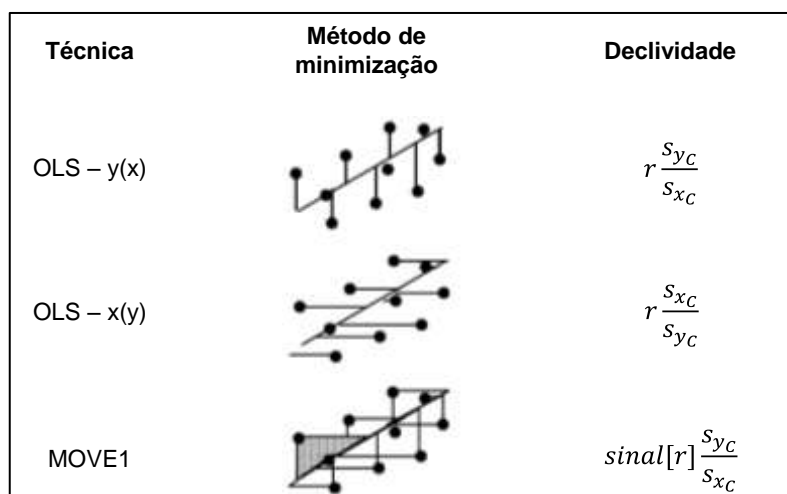


Figura 3.4 - Determinação da declividade da reta da técnica MOVE1

Fonte: Adaptado de HELSEL e HIRSCH, 2002.

Por fornecer uma relação única entre X e Y , independente de qual é a variável explicativa, a técnica MOVE1 pode ser entendida como uma forma de estabelecer a relação “média” entre as variáveis, com erros na estimativa de ambas. Outra virtude da técnica MOVE1 é produzir estimativas de y a partir de valores de x cuja distribuição de probabilidades é semelhante àquela que se espera que y possua. Sua aplicação é maior, portanto, quando se deseja obter não apenas estimativas individuais, como a média ou o desvio padrão de y , mas também outras características de y , como seus percentis (HELSEL e HIRSCH, 2002). A desvantagem da técnica MOVE1 é a sensibilidade a valores extremos (KHALIL e ADAMOWSKI, 2012).

A Tabela 3.6 apresenta os resultados da relação entre o valor esperado da variância amostral e a variância populacional ($E[s_y^2/\sigma_y^2]$), para as técnicas OLS e MOVE1, como apresentado por Hirsch (1982). Quanto mais próxima de 1, melhor é a estimativa da variância populacional. Valores abaixo e acima de 1 indicam um viés de subestimar e superestimar, respectivamente, a variância populacional.

Tabela 3.6 – Valores de $E[s_y^2/\sigma_y^2]$ – técnicas OLS e MOVE1

Técnica OLS				
n_1	n_2	$\rho = 0,5$	$\rho = 0,7$	$\rho = 0,9$
10	10	0,66	0,77	0,91
20	20	0,64	0,75	0,91
30	30	0,63	0,75	0,91
50	10	0,88	0,92	0,97
10	50	0,46	0,63	0,86
Técnica MOVE1				
n_1	n_2	$\rho = 0,5$	$\rho = 0,7$	$\rho = 0,9$
10	10	1,19	1,08	1,03
20	20	1,05	1,03	1,03
30	30	1,03	1,02	1,01
50	10	1,01	1,00	1,00
10	50	1,18	1,12	1,05

Fonte: adaptado de Hirsch (1982)

Analisando a Tabela 3.6, pode-se concluir que, enquanto a técnica OLS leva a séries estendidas com variância menor do que a variância populacional, a técnica MOVE1 superestima a variância. Apesar disso, a técnica MOVE1 se aproxima muito mais do valor populacional do que a técnica OLS (HIRSCH, 1982).

3.2.3.2 MOVE2

Enquanto o ajuste da técnica MOVE1 utiliza apenas informações do período de observações comuns de x e y , a técnica MOVE2, proposta por Hirsch (1982), utiliza também as

informações de x para as quais não há observações correspondentes de y . Para tanto, são utilizados os estimadores não enviesados da média ($\hat{\mu}_Y$) e da variância ($\hat{\sigma}_Y$) propostos por Matalas e Jacobs (1964). Pela proposição original da técnica MOVE2, a média e a variância da série estendida seriam iguais a $\hat{\mu}_Y$ e $\hat{\sigma}_Y$. A técnica MOVE2 é definida pelas equações

$$\hat{y}_i = \hat{\mu}_Y + \frac{\hat{\sigma}_Y}{s_{x_c}}(x_i - \bar{x}) \quad (3.48)$$

$$\hat{\mu}_Y = \bar{y}_c + \frac{n_2}{n_1 + n_2} r \frac{s_{x_c}}{s_{y_c}} (\bar{x}_E - \bar{x}_c) \quad (3.49)$$

$$\hat{\sigma}_Y^2 = \frac{1}{n_1 + n_2 - 1} \left[(n_1 - 1)s_{y_c}^2 + (n_2 - 1) \left(r \frac{s_{x_c}}{s_{y_c}} \right)^2 s_{x_E}^2 + (n_2 - 1)\delta^2(1 - r^2)s_{y_c}^2 + \frac{n_1 n_2}{n_1 + n_2} \left(r \frac{s_{x_c}}{s_{y_c}} \right)^2 (\bar{x}_E - \bar{x}_c)^2 \right] \quad (3.50)$$

nas quais \bar{x}_E denota média dos valores de x no período no qual não há observações comuns ($i = n_1 + 1, n_1 + 2, \dots, n_2$).

Para a aplicação da técnica MOVE2, de forma que os momentos da série estendida sejam realmente aqueles de Matalas e Jacobs (1964), deveriam ser gerados valores de y_i para $i = 1, 2, 3, \dots, n_1 + n_2$. Entretanto, como apontado por Vogel e Stedinger (1985), o procedimento prático consiste em gerar valores de y_i para $i = n_1 + 1, n_1 + 2, n_1 + 3, \dots, n_1 + n_2$, ou seja, a técnica MOVE2 proposta por Hirsch (1982) não atinge seu propósito original.

Segundo Hirsch (1982), a técnica MOVE2 não apresenta viés para a estimativa da variância populacional, apesar de não ser capaz de preservar a variabilidade em eventos hidrológicos extremos. Para Vogel e Stedinger (1985), as técnicas MOVE1 e MOVE2 devem ser utilizadas em situações nas quais as populações de X e Y não diferem significativamente entre si, em termos de distribuição de probabilidades, autocorrelação ou sazonalidade. As séries estendidas por meio dessas técnicas possuem propriedades muito mais próximas da série de referência. Para a verificação da validade da aplicação das técnicas OLS, MOVE1 e MOVE2, podem ser consideradas as premissas adotadas por Hirsch (1982), de séries hidrológicas de x e y estacionárias, sem autocorrelação e distribuídas segundo uma distribuição normal bivariada.

3.2.3.3 MOVE3

As técnicas MOVE3 e MOVE4 (que será detalhada a frente) foram propostas por Vogel e Stedinger (1985), motivadas por novos estimadores não enviesados da média e da variância populacional que são mais eficientes do que aqueles definidos por Matalas e Jacobs (1964). Vogel e Stedinger (1985) utilizaram premissas semelhantes às de Hirsch (1982): i) séries de x e y provenientes de uma distribuição normal bivariada (ou submetidas a alguma transformação para normalizar os dados); e ii) séries sem autocorrelação.

A técnica MOVE3 foi definida de tal forma que, após ser aplicada somente para os valores extrapolados de y (\hat{y}_i , para $i = n_1 + 1, n_1 + 2, \dots, n_1 + n_2$), toda a sequência de $n_1 + n_2$ registros possuísse média ($\hat{\mu}_Y$) e variância ($\hat{\sigma}_Y$) iguais aos valores propostos por Matalas e Jacobs (1964).

Os valores de \hat{y}_i da série estendida pela técnica MOVE3 são calculados por meio das equações

$$\hat{y}_i = a + b(x_i - \bar{x}_E) \quad (3.51)$$

$$a = \frac{(n_1 + n_2)\hat{\mu}_Y - n_1\bar{y}_C}{n_2} \quad (3.52)$$

$$b = \sqrt{\frac{(n_1 + n_2 - 1)\hat{\sigma}_Y^2 - (n_1 - 1)s_{y_C}^2 - n_1(\bar{y}_C - \hat{\mu}_Y)^2 - n_2(a - \hat{\mu}_Y)^2}{(n_2 - 1)s_{x_E}^2}} \quad (3.53)$$

Nas quais $s_{x_E}^2$ denota a variância amostral de x_i para $i = n_1 + 1, n_1 + 2, \dots, n_1 + n_2$.

3.2.3.4 MOVE4

Vogel e Stedinger (1985) também obtiveram estimadores não enviesados otimizados para a média ($\hat{\mu}_Y^*$) e para a variância ($\hat{\sigma}_Y^{*2}$) da série estendida de y por meio da introdução de duas variáveis, θ_1 e θ_2 . O estimador da média ($\hat{\mu}_Y^*$) mostrou-se mais eficiente do que aquele proposto por Matalas e Jacobs (1964) para valores pequenos de n_1 ($n_1 < 9$ para $n_2 = 60$ e $\rho = 0,7$, por exemplo). O estimador da variância ($\hat{\sigma}_Y^{*2}$) mostrou-se mais eficiente independente do valor de ρ e de n_1 em relação a n_2 . Para menores valores de ρ , maior foi a eficiência de $\hat{\sigma}_Y^{*2}$ em relação a $\hat{\sigma}_Y^2$.

Os valores de $\hat{\mu}_Y^*$ e $\hat{\sigma}_Y^{*2}$ são dados pelas equações

$$\hat{\mu}_Y^* = \bar{y}_C + \theta_1 \frac{n_2}{n_1 + n_2} r \frac{s_{x_C}}{s_{y_C}} (\bar{x}_E - \bar{x}_C) \quad (3.54)$$

$$\begin{aligned} \hat{\sigma}_Y^{*2} = & (1 - \theta_2) s_{y_C}^2 \\ & + \theta_2 \left[\left(r \frac{s_{x_C}}{s_{y_C}} \right)^2 (\bar{x}_E - \bar{x}_C)^2 s_x^2 \right. \\ & \left. + \left[1 - \frac{n_1 + n_2 - 3}{(n_1 - 3)(n_1 + n_2 - 1)} \right] \frac{n_1 - 1}{n_1 - 2} \left(s_{y_C}^2 - r \frac{s_{x_C}}{s_{y_C}} s_{x_C}^2 \right) \right] \end{aligned} \quad (3.55)$$

nas quais

$$\theta_1 = \frac{(n_1 - 3)\rho^2}{(n_1 - 4)\rho^2 + 1} \quad (3.56)$$

$$\theta_2 = \frac{(n_1 - 4)\rho^2}{(n_1 - 8,5)\rho^2 + 4,5} \quad (3.57)$$

A técnica MOVE4, também proposta por Vogel e Stedinger (1985), foi formulada para produzir valores de \hat{y}_i , com $i = n_1 + 1, n_1 + 2, \dots, n_1 + n_2$, de tal forma que a série histórica resultante, de $n_1 + n_2$ elementos ($y_1, y_2, \dots, y_{n_1}, \hat{y}_{n_1+1}, \hat{y}_{n_1+2}, \dots, \hat{y}_{n_1+n_2}$), possuísse média e variância dados por $\hat{\mu}_Y^*$ e $\hat{\sigma}_Y^{*2}$, respectivamente.

A expressão para estimativa dos valores de y_i é dada pela Equação 3.51, com coeficientes definidos pelas equações (VOGEL e STEDINGER, 1985)

$$a = \frac{(n_1 + n_2)\hat{\mu}_Y^* - n_1\bar{y}_C}{n_2} \quad (3.58)$$

$$b = \sqrt{\frac{(n_1 + n_2 - 1)\hat{\sigma}_Y^{*2} - (n_1 - 1)s_{y_C}^2 - n_1(\bar{y}_C - \hat{\mu}_Y^*)^2 - n_2(a - \hat{\mu}_Y^*)^2}{(n_2 - 1)s_{x_E}^2}} \quad (3.59)$$

As técnicas MOVE podem ser agrupadas em: i) MOVE1 e MOVE2, que mantêm o foco nos dois primeiros momentos estatísticos, e cujas formulações foram desenvolvidas para o caso em que toda a série histórica de y é estimada; e ii) MOVE3 e MOVE4, que têm o mesmo foco, apesar de, nas suas formulações, considerar que apenas os valores faltantes de y são estimados, em função de x (KHALIL e ADAMOWSKI, 2012).

As técnicas MOVE foram obtidas a partir de adaptações da técnica OLS. Um outro modelo de regressão, apresentado por Helsel e Hirsch (2002), que também já foi aplicado como técnica

de extensão de séries hidrológicas, é o KTRL (*Kendall-Theil Robust Line*, ou Linha Robusta Kendall Theil).

3.2.4 KTRL (*Kendall-Theil Robust Line*)

Na literatura, a linha de regressão KTRL também é conhecida como “*Sen slope*”. Assim como a técnica OLS está diretamente relacionada ao coeficiente de correlação r de Pearson, a técnica KTRL está relacionada ao coeficiente de correlação τ de Kendall (HELSEL e HIRSCH, 2002).

O coeficiente de correlação de Kendall fornece uma medida da relação monotônica entre X e Y . Por depender da classificação dos termos das séries de dados de X e Y , e não de seus valores, o coeficiente de correlação de Kendall possui resistência a *outliers* e pode ser utilizado se valores muito baixos ou muito elevados forem categorizados (como precipitações abaixo de determinado limiar). Em comparação com os demais coeficientes de correlação, τ destaca-se por apresentar valores de correlação numericamente mais baixos e ser capaz de mensurar correlações monotônicas lineares ou não (HELSEL e HIRSCH, 2002).

O coeficiente de correlação τ é calculado por meio da equação

$$\tau = \frac{P - M}{n(n - 1)/2} \quad (3.60)$$

Na qual

P denota o número de pares concordantes (Y aumenta à medida que X aumenta), ou seja, $y_i < y_j$ para todo $i < j$ para $i = 1, 2, 3, \dots, (n - 1)$ e $j = (i + 1), \dots, n$

M denota o número de pares discordantes (Y diminui à medida que X aumenta), ou seja, $y_i > y_j$ para todo $i < j$ para $i = 1, 2, 3, \dots, (n - 1)$ e $j = (i + 1), \dots, n$

Para propósitos de regressão, a técnica KTRL apresenta maior resistência a *outliers* do que a técnica OLS, como mostra a Figura 3.5, e não adota a hipótese de resíduos distribuídos segundo uma distribuição normal. Tendo em vista que as variáveis hidrológicas possuem erros de monitoramento associados à magnitude dos valores medidos, é de se esperar que a variância dos resíduos não seja constante para todo X , o que é uma hipótese da técnica OLS (GRANATO, 2006). Apesar de essas hipóteses serem mais generalistas do que as hipóteses da

técnica OLS, elas não necessariamente se apresentam como vantajosas no contexto da extensão de séries hidrológicas.

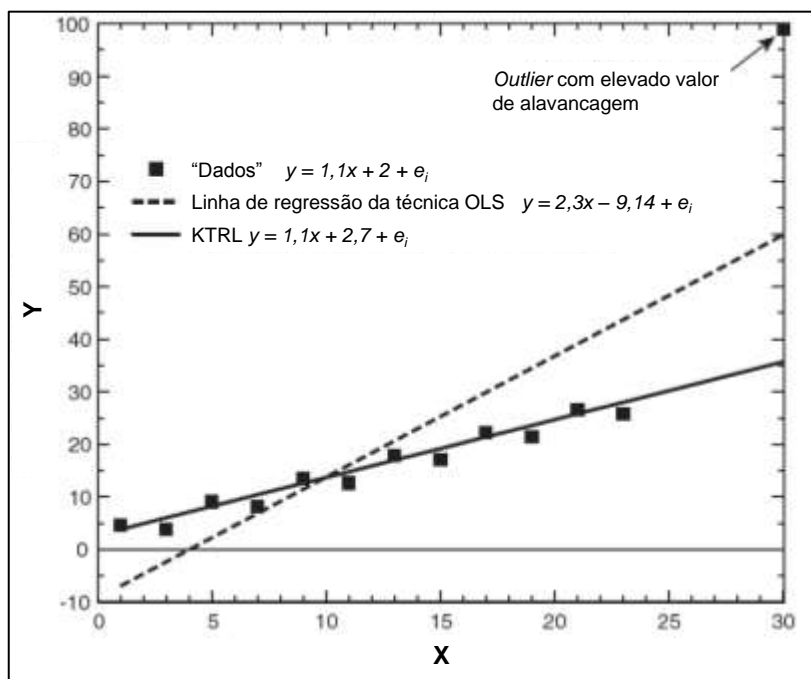


Figura 3.5 - Comparação da resistência a *outliers* entre as técnicas OLS e KTRL
 Fonte: adaptado de GRANATO, 2006.

A reta de regressão calculada por meio da técnica KTRL é identificada como a mediana das declividades das retas obtidas pela ligação de cada par de pontos (SEN, 1968), como mostra a Figura 3.6. Duas estimativas do intercepto da reta da técnica KTRL foram propostas na literatura. A primeira corresponde à mediana dos interceptos obtidos pelo prolongamento das retas da Figura 3.6-A (THEIL, 1950). A segunda é obtida por meio da passagem da reta de regressão pela mediana dos dados de x e y (CONOVER, 1980). Helsel e Hirsch (2002) apontam que a metodologia apresentada por Conover (1980) é mais vantajosa, por ser mais robusta, eficiente, pela sua facilidade de cálculo e por ser análoga à técnica OLS (na qual a reta de regressão passa pelos valores médios de X e Y).

A expressão geral da técnica KTRL é dada por

$$\hat{y}_i = a + bx_i \tag{3.61}$$

com declividade da reta dada pela equação

$$b = \text{mediana} \left[\frac{y_j - y_i}{x_j - x_i} \right] \quad \forall i < j \quad i = 1, 2, \dots, n_1 - 1 \quad e \quad j = 2, 3, \dots, n_1 \tag{3.62}$$

O intercepto, calculado conforme metodologia proposta por Conover (1980), é dado por

$$a = \text{mediana}[y_c] - b \times \text{mediana}[x_c] \quad (3.63)$$

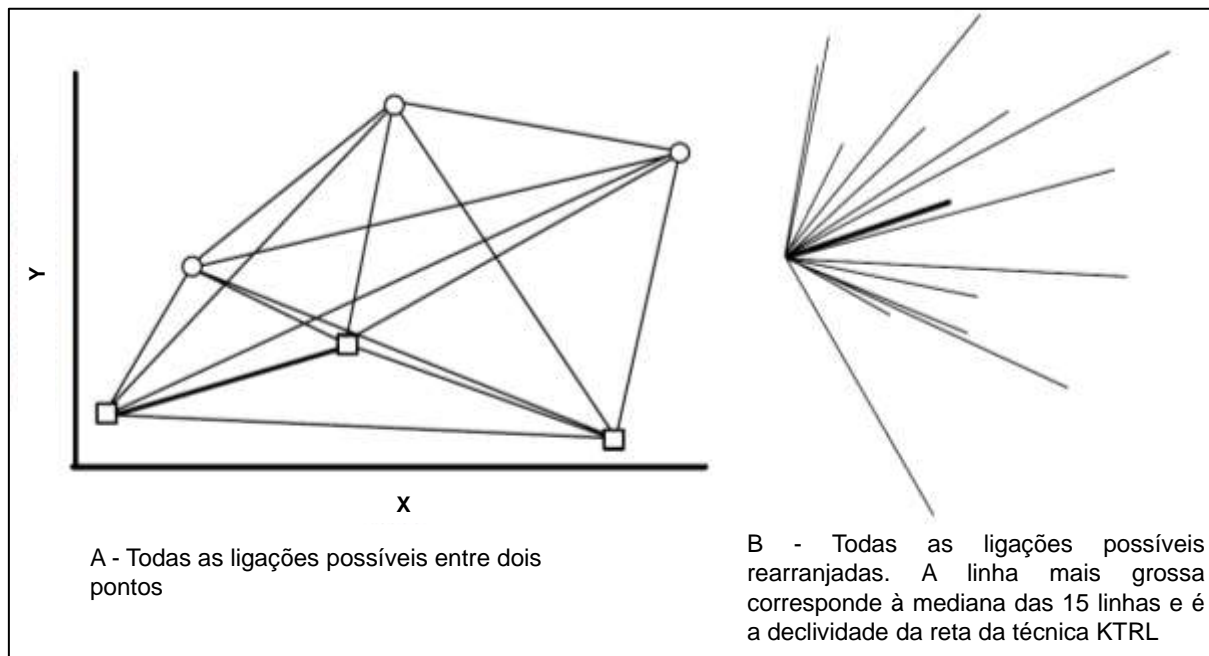


Figura 3.6 - Determinação da declividade da reta de regressão da técnica KTRL

Fonte: adaptado de HELSEL e HIRSCH, 2002.

Um problema na estimativa da declividade da reta de regressão surge quando uma amostra apresenta dois valores iguais de x , e o coeficiente angular calculado por meio da Equação 3.62 é enviesado. Para resolver esse problema, Granato (2006) não computou, na mediana, valores iguais de x , obtendo resultados satisfatórios.

Ambos os estimadores do coeficiente angular das técnicas OLS e KTRL são não enviesados. Contudo, o estimador do coeficiente angular da técnica OLS é ligeiramente mais eficiente do que o estimador da técnica KTRL. Para propósitos de regressão, e não necessariamente para a extensão de séries hidrológicas, a técnica KTRL apresenta-se com propriedades muito interessantes para uso por possuir estimadores dos coeficientes de regressão quase tão eficientes quanto os da técnica OLS, quando a hipótese de normalidade é atendida, e muito melhores quando essa hipóteses não é atendida (HELSEL e HIRSCH, 2002).

Estudos já foram desenvolvidos para comparar os resultados dos ajustes das técnicas OLS e KTRL, entre os quais se destaca a pesquisa de Hirsch *et al.* (1991), aplicada à qualidade da água. Com o intuito de avaliar a partir de qual grau de ausência de normalidade a técnica

KTRL passa a ser mais adequada, foi feita uma amostragem, por meio de simulação de Monte Carlo, de valores aleatórios provenientes de duas distribuições normais: a primeira, com média 10 e desvio padrão de 1, e a segunda, com média 11 e desvio padrão de 3. A proporção de mistura de cada amostragem variou de 80% a 100% da primeira distribuição e, portanto, de 0% a 20% da segunda distribuição. Foram considerados, ainda, dois cenários que simulam séries com extensão de 6 e 36 dados amostrados. Os autores relatam que a não normalidade das amostras geradas não é facilmente percebida. Os valores amostrados constituíram a variável dependente (Y), enquanto a variável independente considerada foi o tempo. Como Y independe do tempo (variável gerada aleatoriamente), a declividade populacional da reta de regressão é, teoricamente, nula. A medida da eficiência de cada técnica foi feita por meio da relação entre os indicadores RMSE (*Root Mean Square Error* ou Erro Quadrático Médio) obtidos para o ajuste das técnicas OLS e KTRL. Foi considerado que ocorre erro caso $b \neq 0$. Os resultados obtidos, apresentados na Figura 3.7, indicam que: i) a técnica OLS teve melhores resultados quando aplicada a uma série curta do que quando aplicada a uma série longa; ii) a técnica KTRL apresentou erros menores quando o percentual de mistura foi superior a 6% (amostra gerada de 6 valores) ou 2% (amostra gerada de 36 valores); e iii) quando os dados gerados se prestam à regressão linear, apresentam variância constante e normalidade dos resíduos, as duas técnicas apresentam resultados semelhantes. Baseados nesse estudo, Helsel e Hirsch (2002) indicam que pequenos desvios na normalidade dos dados, nem sempre detectáveis por meio de testes e histogramas, já favorecem o uso da técnica KTRL.

As vantagens mencionadas da técnica KTRL, para dados não normais, referem-se apenas aos coeficientes do modelo de regressão, e não necessariamente são mantidas para os momentos dos valores estimados. De fato, Khalil e Adamowski (2014a) apontam a não manutenção da variância da série curta na série estendida como grande deficiência da técnica KTRL. Na tentativa de solucionar esse problema, Khalil *et al.* (2012) propuseram uma nova técnica, baseada na técnica KTRL, denominada KTRL2.

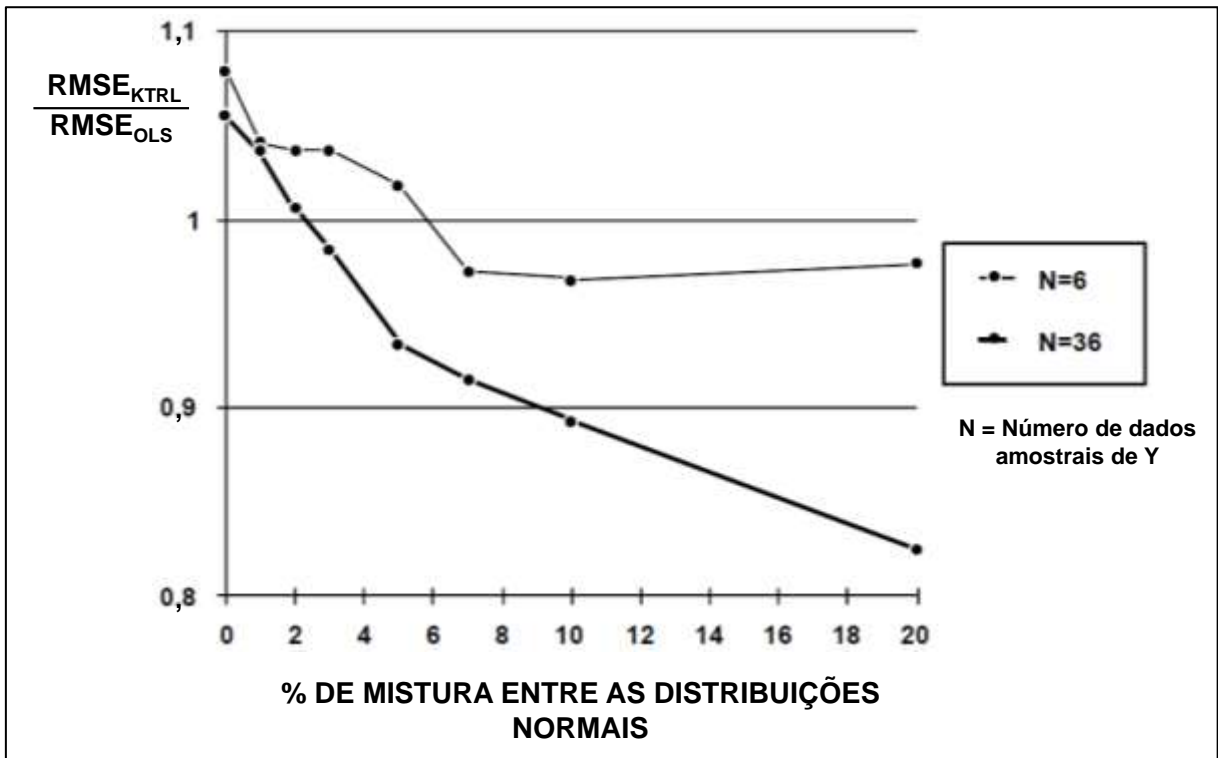


Figura 3.7 - Resultado gráfico do experimento de Hirsch *et al.* (1991): vantagem da técnica KTRL em relação à técnica OLS para amostras que não seguem a distribuição normal
 Fonte: adaptado de HIRSCH, 1991.

3.2.5 KTRL2 (Kendall-Theil Robust Line 2)

A técnica KTRL2 foi proposta por Khalil *et al.* (2012), com o intuito de manter a variabilidade da série curta na série estendida e a robustez da técnica KTRL na presença de *outliers*. Buscando obter uma série estendida com FDP próxima da FDP da série original, foram propostos novos estimadores dos coeficientes a e b do modelo de regressão (Equação 3.61), dados pelas expressões:

$$b = \text{mediana} \left[\frac{y_j - y_i}{x_j - x_i} \right] \quad \forall i < j \quad i = 5^{\circ}, 10^{\circ}, \dots, 90^{\circ} \quad e \quad j = 10^{\circ}, 15^{\circ}, \dots, 95^{\circ} \quad (3.64)$$

$$a = \text{mediana}[y_c] - b \times \text{mediana}[x_c] \quad (3.65)$$

A declividade da reta de regressão da técnica KTRL2 é estimada para minimizar o erro na estimativa dos percentis da série estendida (KHALIL & ADAMOWSKI, 2014a). Uma desvantagem da técnica KTRL2 é a necessidade de se estimar os percentis de ordem 5 a 95, o que pode se tornar uma dificuldade para séries curtas.

De forma geral, as técnicas MOVE apresentam grande sensibilidade a *outliers*, enquanto a técnica KTRL leva à estimativa de valores da série estendida com variância reduzida (KHALIL e ADAMOWSKI, 2012). Entendendo que a técnica KTRL2, proposta por Khalil *et al.* (2012), tem a finalidade de corrigir o viés na variância dos valores estendidos por meio da técnica KTRL, Khalil e Adamowski (2012) propuseram uma modificação da técnica MOVE1 (também denominada LOC) para reduzir sua sensibilidade a *outliers*. Essa técnica foi denominada RLOC (*Robust Line of Organic Correlation*).

3.2.6 RLOC (*Robust Line of Organic Correlation*)

A presença de *outliers*, superiores ou inferiores, possui mais efeito na estimativa da média do que na mediana (percentil de 50%) de uma amostra. Por outro lado, a variância amostral é afetada pelos desvios da amostra ao quadrado, e é mais sensível à presença de *outliers*. A amplitude interquartil (AIQ) é uma medida mais robusta da variabilidade amostral, dado que é obtida por meio da diferença entre os percentis de ordem 75 e 25 (considerando que os *outliers* são encontrados acima do percentil 75 e abaixo do percentil 25, e, portanto, não interfeririam no cômputo da AIQ) (KHALIL e ADAMOWSKI, 2012).

Para ser resistente a *outliers*, o coeficiente angular da técnica RLOC (*Robust Line of Organic Correlation*, ou Linha Robusta de Correlação) é estimado pela relação entre as amplitudes interquartis de y e x , ou seja (KHALIL e ADAMOWSKI, 2012)

$$b = \text{sinal}[r] \frac{y_{C(75)} - y_{C(25)}}{x_{C(75)} - x_{C(25)}} \quad (3.66)$$

na qual

$y_{C(p)}$ denota o percentil de ordem p da amostra y no período de observações comuns de y e x ;

e

$x_{C(p)}$ denota o percentil de ordem p da amostra x do período de observações comuns de y e x .

O intercepto da reta de regressão é análogo àquele fornecido por Conover (1980), adaptado para o novo coeficiente angular:

$$a = \text{mediana}[y_C] - b \times \text{mediana}[x_C] \quad (3.67)$$

A expressão geral da técnica RLOC é dada pela Equação 3.61.

Experimento semelhante àquele realizado por Hirsch *et al.* (1991), e descrito no item 3.2.4, foi realizado por Khalil e Adamowski (2012) para avaliar as vantagens do uso da técnica RLOC em amostras que não seguem a distribuição normal. Após realizar uma amostragem da variável dependente a partir de duas distribuições normais distintas, sendo a amostra da variável independente dada por números aleatórios da distribuição uniforme, foi calculado o coeficiente angular (teoricamente igual a zero, dada a ausência de relação entre x e y) por meio das técnicas MOVE1 e RLOC. A amostragem da variável dependente foi feita para um conteúdo de 80 % a 20% de valores pertencentes à distribuição normal principal. A medida do desempenho de cada técnica foi feita pela relação entre o RMSE calculado para as técnicas RLOC e MOVE1, de modo que uma relação inferior à unidade significa melhor desempenho da técnica RLOC. O resultado, sumarizado na Figura 3.8, mostra que a técnica RLOC é mais eficiente quando a distribuição das amostras não segue a distribuição normal.

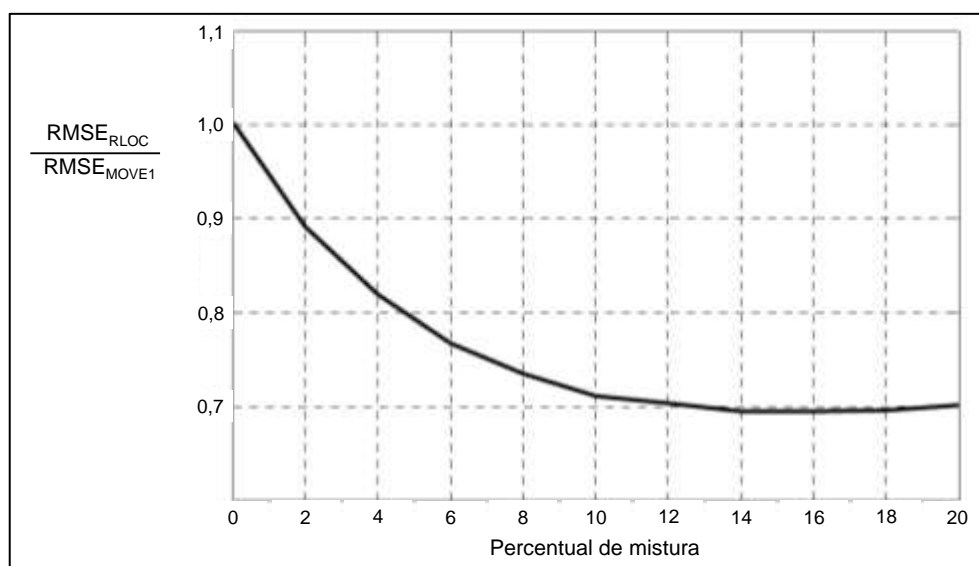


Figura 3.8 - Resultado gráfico do experimento de Hirsch *et al.* (1991): vantagem da técnica RLOC em relação à técnica MOVE1 para amostras que não seguem a distribuição normal
Fonte: adaptado de KHALIL e ADAMOWSKI, 2012.

Assim como comentado no item 3.2.4, um desempenho melhor da técnica RLOC na estimativa dos parâmetros populacionais de um modelo de regressão, em comparação com a técnica OLS não implicam, necessariamente, um desempenho melhor na extensão das séries hidrológicas. Afinal, nessa última, o desempenho é avaliado em função da estimativa dos momentos populacionais.

As técnicas apresentadas até o momento tratam do ajuste de uma reta às observações de X e Y . Outra abordagem refere-se ao ajuste de modelos capazes de introduzir alguma não

linearidade na relação entre as observações, com a subsequente avaliação da preservação da variabilidade dos dados. Um dos modelos capaz de ajustar diferentes formatos de regressão é apresentado a seguir.

3.2.7 GLM (*Generalized Linear Models*)

Os Modelos Lineares Generalizados, também conhecidos pela sua sigla em inglês (GLM, de *Generalized Linear Models*), foram propostos por Nelder e Wedderburn (1972). As vantagens dos GLM são permitir que a distribuição da variável resposta seja diferente da distribuição normal e permitir um certo grau de não-linearidade na estrutura do modelo (WOOD, 2006). A estrutura básica do modelo é dada pela equação

$$g(\mu_i) = \mathbf{X}_i\boldsymbol{\beta} \quad (3.68)$$

na qual

$$\mu_i = E(Y_i);$$

g é denominada função de ligação, monotônica e diferenciável, que relaciona o valor esperado da variável dependente ($E[Y_i]$) com o preditor linear; e

$\mathbf{X}_i\boldsymbol{\beta}$ é o preditor linear, que é a função linear de regressão propriamente dita. \mathbf{X}_i é a i -ésima linha da matriz de dados da variável independente e $\boldsymbol{\beta}$ é um vetor de parâmetros, com valores inicialmente desconhecidos. Na sua forma original, o preditor linear é o mesmo da regressão linear MLR.

Um modelo linear generalizado assume as premissas de que os valores Y_i são independentes e que a distribuição condicional da variável dependente Y_i pertence a uma família exponencial. A família exponencial inclui várias distribuições de probabilidades com aplicação prática, entre as quais estão a Poisson, Binomial, Gama e Normal (WOOD, 2006).

As funções de ligação mais comuns são apresentadas na Tabela 3.7, enquanto a Figura 3.9 apresenta o formato de algumas dessas funções para a relação $y = \mu_i = g^{-1}(a + bx_i)$.

Uma distribuição de probabilidades pertence à família exponencial se sua FDP ou FMP puder ser escrita na forma (FOX, 2015)

$$f(y; \theta, \phi) = \exp \left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right] \quad (3.69)$$

na qual

a , b e c são funções conhecidas que variam de acordo com a distribuição;

ϕ é um parâmetro de dispersão, correspondente a um valor fixo em algumas distribuições, e podendo ser um valor desconhecido em outras. Neste caso, é estimado em função de θ ; e

θ é o parâmetro canônico da distribuição, que, no contexto dos GLM, dependerá apenas dos parâmetros do modelo (β).

Tabela 3.7 – Algumas das funções de ligação mais comuns e suas inversas

Função de ligação	$X_i\beta_i = g^{-1}(\mu_i)$		$y_i = \mu_i = g^{-1}(X_i\beta)$	
Identidade	μ_i	(3.70)	$X_i\beta$	(3.71)
Logarítmica	$\ln(\mu_i)$	(3.72)	$e^{X_i\beta}$	(3.73)
Inversa	μ_i^{-1}	(3.74)	$(X_i\beta)^{-1}$	(3.75)
Inversa ao quadrado	μ_i^{-2}	(3.76)	$(X_i\beta)^{-1/2}$	(3.77)
Raiz quadrada	$\sqrt{\mu_i}$	(3.78)	$(X_i\beta)^2$	(3.79)
Logit	$\ln\left(\frac{\mu_i}{1-\mu_i}\right)$	(3.80)	$\frac{1}{1+e^{-X_i\beta}}$	(3.81)
Probit	$\Phi^{-1}(\mu_i)$	(3.82)	$\Phi(X_i\beta)$	(3.83)
Log-log	$-\ln[-\ln(\mu_i)]$	(3.84)	$\exp[-\exp(-X_i\beta)]$	(3.85)
Complemento log-log	$-\ln[-\ln(1-\mu_i)]$	(3.86)	$1 - \exp[-\exp(X_i\beta)]$	(3.87)

Fonte: adaptado de FOX, 2015.

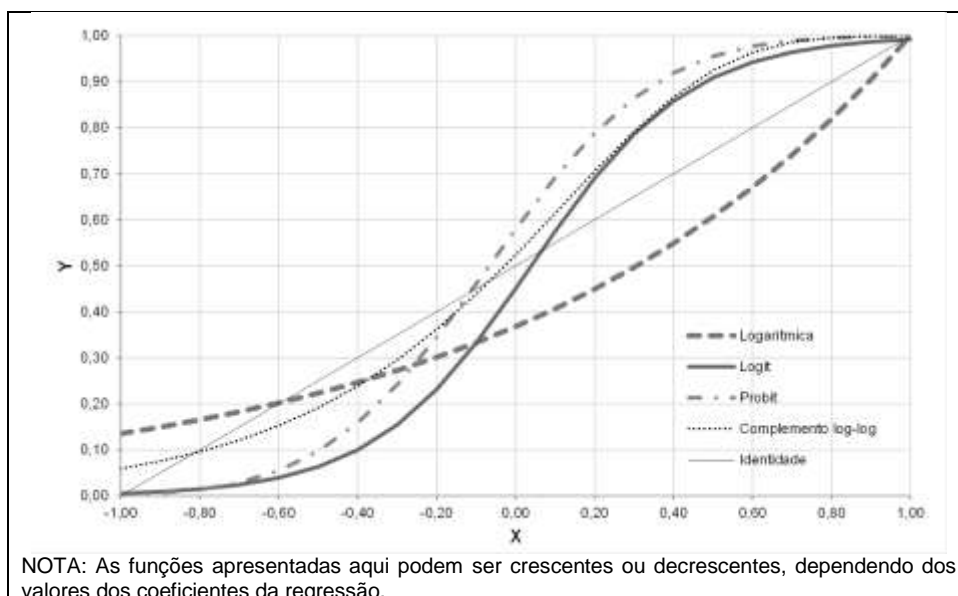


Figura 3.9 – Formato de algumas das funções de ligação normalmente utilizadas nos modelos lineares generalizados

A distribuição normal, por exemplo, pode ser escrita na forma

$$f(y; \mu, \sigma) = \exp \left[\frac{y\mu - \mu^2/2}{\sigma^2} + \frac{y^2}{2\sigma^2} - \log(\sigma\sqrt{2\pi}) \right] \quad (3.88)$$

na qual

$$\theta = \mu;$$

$$b(\theta) = \frac{\theta^2}{2} = \frac{\mu^2}{2};$$

$$a(\phi) = \phi = \sigma^2; \text{ e}$$

$$c(\phi, y) = -\frac{y^2}{2\phi} - \log(\sqrt{\phi 2\pi}) = -\frac{y^2}{2\sigma^2} - \log(\sigma\sqrt{2\pi}).$$

É possível obter expressões gerais para a média e a variância de distribuições da família exponencial em termos de a , b e ϕ . A função log-verossimilhança de θ , dado um valor particular de y , é definida por $\log[f_\theta(y; \theta, \phi)]$, considerando f_θ uma função de θ . Dessa forma, tem-se (WOOD, 2006):

$$l(\theta) = \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \quad (3.89)$$

na qual l é a função log-verossimilhança.

A derivada da função log-verossimilhança retorna

$$\frac{dl}{d\theta} = \frac{y - b'(\theta)}{a(\phi)} \quad (3.90)$$

Tomando l como uma variável aleatória, e substituindo as observações y pela variável aleatória Y , o valor esperado de $dl/d\theta$ pode ser calculado por

$$E\left(\frac{dl}{d\theta}\right) = \frac{E(Y) - b'(\theta)}{a(\phi)} \quad (3.91)$$

Adotando, para a máxima log-verossimilhança, que $E(dl/d\theta) = 0$ implica que

$$E(Y) = b'(\theta) \quad (3.92)$$

e pode-se concluir que a média de qualquer variável aleatória de uma distribuição pertencente à família exponencial é dada pela derivada de b em relação a θ , sendo que b depende da distribuição da VA (WOOD, 2006).

Derivando a função log-verossimilhança na segunda ordem, é possível obter

$$Var(Y) = b''(\theta)a(\phi) \quad (3.93)$$

Para o caso da distribuição normal, por exemplo, as equações 3.92 e 3.93 resultam em $E(Y) = \mu$ e $Var(Y) = \sigma^2$.

Os modelos lineares generalizados podem ser ajustados aos dados amostrais utilizando o método da máxima verossimilhança, por meio do qual podem ser estimados não apenas os parâmetros dos coeficientes de regressão, mas também o erro-padrão dos coeficientes, no caso assintótico. Quando a distribuição pertence à família exponencial, e resulta na situação em que ϕ é desconhecido, é comum seu valor ser estimado por meio do método dos momentos, embora possa ser estimado, a princípio, também pelo método ML (FOX, 2015).

O desvio residual, que é análogo à soma dos quadrados dos resíduos na técnica de regressão MLR, é estimado por (FOX, 2015)

$$D_m = 2(\ln(L_s) - \ln(L_m)) \quad (3.94)$$

em que

L_m é a máxima verossimilhança do modelo em estudo; e

L_s é a máxima verossimilhança de um modelo saturado, que consiste em um parâmetro para cada observação e, portanto, se ajusta às observações da melhor forma possível. É obtida por meio do cálculo da verossimilhança considerando $\hat{\mu} = y$.

Também de forma análoga a outras técnicas de regressão, a proporção da variabilidade de y explicada pelo modelo pode ser estimada por (FOX, 2015)

$$R^2 = 1 - \frac{D_1}{D_0} \quad (3.95)$$

em que

D_1 denota o desvio residual para o modelo em estudo; e

D_0 denota o desvio residual para o modelo que contém apenas uma constante (desvio residual nulo).

A função log-verossimilhança dos GLM, aplicada para n observações independentes, é dada por (FOX, 2015)

$$\ln L(\boldsymbol{\theta}, \phi; \mathbf{y}) = \sum_{i=1}^n \frac{y_i \theta_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi) \quad (3.96)$$

Supondo que a função de ligação do GLM é dada por (FOX, 2015)

$$g(\mu_i) = \eta_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_K x_{i,K} \quad (3.97)$$

Para estimar os parâmetros $\boldsymbol{\beta}$, é necessário diferenciar a função log-verossimilhança em relação a cada parâmetro. Aplicando a regra da cadeia e rearranjando os resultados, obtém-se as equações da máxima verossimilhança para os GLM (FOX, 2015)

$$\sum_{i=1}^n \frac{y_i - \mu_i}{a_i v(\mu_i)} \cdot \frac{d\mu_i}{d\eta_i} \cdot x_{i,j} = 0, \quad \text{para } j = 0, 1, \dots, K \quad (3.98)$$

em que

$v(\mu_i)$ representa a variância estimada de μ_i ; e

a_i é uma função do parâmetro de dispersão ϕ , que é constante ao longo das observações.

Algumas simplificações adicionais são possíveis quando a função de ligação g é canônica, resultando em (FOX, 2015).

$$\sum_{i=1}^n \frac{y_i x_{i,j}}{a_i} = \sum_{i=1}^n \frac{\mu_i x_{i,j}}{a_i} \quad (3.99)$$

Uma função de ligação é caracterizada como canônica quando o preditor linear η é igual a θ , o parâmetro canônico da família exponencial em questão. A Tabela 3.8 apresenta as funções de ligação canônicas para algumas distribuições da família exponencial (FOX, 2015).

Mesmo utilizando funções de ligação canônicas, chega-se à situação das equações de estimação de parâmetros envolverem funções não lineares, sendo necessário, portanto, utilizar métodos iterativos (FOX, 2015). Um dos métodos mais utilizados para tanto é o IWLS (*Iterative Weighted Least Squares*, ou Mínimos Quadrados Ponderado Iterativo).

Tabela 3.8 - Funções de ligação canônicas associadas a algumas distribuições das famílias exponenciais

Distribuição	Função de ligação canônica
Gaussiana	Identidade
Binomial	Logit
Poisson	Log
Gama	Inversa
Gaussiana Inversa	Inversa ao quadrado

Fonte: adaptado de FOX, 2015.

O método IWLS baseia-se na solução iterativa da equação (FOX, 2015)

$$\hat{\beta}^{(m+1)} = (\mathbf{X}^T \mathbf{W}^{(m)} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^{(m)} \mathbf{z}^{(m)} \quad (3.100)$$

na qual

$\hat{\beta}^{(m+1)}$ é o vetor dos coeficientes de regressão da iteração atual;

\mathbf{X} é a matriz contendo as observações da variável explicativa;

\mathbf{W} é a matriz diagonal dos pesos;

\mathbf{z} é o vetor das variáveis de ajuste na m -ésima iteração; e

m e $m + 1$ representam as iterações (anterior e atual, respectivamente) do método IWLS.

O método IWLS, aplicado para a função de ligação canônica, é equivalente ao método Newton-Raphson para a estimativa de raízes de uma equação (FOX, 2015).

Existem diversos critérios para a seleção do melhor GLM, entre os quais está o Critério de Informação de Akaike (AIC). Segundo Wood (2006), o valor do AIC para os modelos lineares generalizados é dado por

$$\text{AIC} = 2[-l(\hat{\beta}) + p] \quad (3.101)$$

em que p denota o número de parâmetros identificáveis no modelo, incluindo ϕ , que, caso não seja conhecido, deve ser estimado, e, então, o termo p torna-se $p + 1$.

Como mostra a Equação 3.101, utilizar o menor valor de AIC equivale a escolher o modelo com maior valor da log-verossimilhança (WOOD, 2006).

As técnicas apresentadas até aqui presumem que as observações de y são independentes. A seguir, são apresentadas, brevemente, técnicas de extensão de séries hidrológicas que consideram a autocorrelação amostral em sua formulação.

3.2.8 Técnicas que consideram a autocorrelação amostral

Uma outra técnica de extensão da família MOVE foi proposta por Grygier *et al.* (1989). Essa técnica, denominada de MOVE Generalizada (GMOVE), foi desenvolvida com o intuito de estender as séries hidrológicas dentro do contexto do *Central Valley Project*, desenvolvido pelo USBR (*United States Bureau of Reclamation*). Em uma das vertentes desse projeto, foi necessário prever vazões mensais com base nas informações históricas de precipitação e neve.

Na nomenclatura utilizada no desenvolvimento da técnica GMOVE, tem-se as seguintes variáveis (GRYGIER *et al.*, 1989):

- q_i é uma vazão utilizada como referência, no tempo i ; e
- m_i é o valor anterior da variável q_i (q_{i-1}); e
- y_i é o valor previsto da série de vazões estudada no tempo i .

Para o caso univariado, a técnica GMOVE pode ser escrita como (GRYGIER *et al.*, 1989)

$$\hat{y}_i = \bar{y}_C + b(m_i - \bar{m}_C) + c(q_i - \bar{q}_C) \quad (3.102)$$

em que

$$b^2 = \frac{\left(s_{y_1}^2 - \frac{s_{qy}^2}{s_{q_1}^2} \right)}{\left(s_{m_1}^2 - \frac{s_{mq}^2}{s_{q_1}^2} \right)} \quad (3.103)$$

$$c = (s_{qy} - bs_{mq})/s_{q_1}^2 \quad (3.104)$$

$s_{m_1}^2$, $s_{y_1}^2$ e $s_{q_1}^2$ denotam a variância amostral no período de observações comuns de m , y e q ; e s_{ij} denota a covariância entre as variáveis i e j .

É possível notar que, quando o termo m_i não é considerado, a técnica GMOVE (caso univariado) reduz-se para a técnica MOVE1. Cabe ressaltar que, ao considerar o termo m_i , a técnica GMOVE qualifica-se como uma técnica de previsão com a inclusão de um processo

estocástico, e, diferente das demais técnicas de extensão, a ordem na qual os eventos foram registrados passa a ter relevância.

Para o caso multivariado, a técnica GMOVE completa apresenta uma formulação matricial complexa. Para o caso reduzido em que o processo estocástico é removido e há k séries curtas a serem estendidas e h séries de referência, a técnica GMOVE multivariada é dada por (GRYGIER *et al.*, 1989)

$$\mathbf{y}_i = \bar{\mathbf{y}} + \mathbf{C}(\mathbf{q}_i - \bar{\mathbf{q}}) \quad (3.105)$$

$$\mathbf{C} = \mathbf{S}_{yq} \mathbf{S}_{qq}^{-1} \quad (3.106)$$

em que

\mathbf{y}_i é o vetor dos valores estimados, com comprimento $k \times 1$;

$\bar{\mathbf{y}}$ é o vetor das médias de cada série curta, com comprimento $k \times 1$;

\mathbf{C} é a matriz de parâmetros associados à série de referência \mathbf{q} , dimensões $k \times h$;

\mathbf{q}_i é o vetor dos valores de referência, com comprimento $h \times 1$;

$\bar{\mathbf{q}}$ é o vetor das médias de cada série de referência, com comprimento $h \times 1$;

\mathbf{S}_{yq} a matriz de covariância entre y e q ; e

\mathbf{S}_{qq} a matriz de covariância entre q e q ; e

Tarawneh (2011) também propôs uma nova técnica baseada na autocorrelação amostral, cuja formulação é apresentada nas Equações 3.107 a 3.109.

$$\hat{y}_i = \bar{y} + a(y_{i-1} - \bar{y}) + b(x_i - \bar{x}), \text{ para } i = n_1 + 1, n_1 + 2, \dots, n_1 + n_2 \quad (3.107)$$

$$a = \frac{C_{xy}C_{xy-1} - C_{xx}C_{yy-1}}{C_{xy-1}^2 - C_{xx}C_{yy}} \quad (3.108)$$

$$b = \frac{C_{xy} - aC_{xy-1}}{C_{xx}} \quad (3.109)$$

em que

C_{xx} e C_{yy} denotam a variância de x e y , respectivamente;

C_{xy} denota a covariância entre x e y ;

C_{yy-1} denota a covariância entre y_i e y_{i-1} ; e

C_{xy-1} denota a covariância entre x_i e y_{i-1} .

Uma vez apresentadas diversas técnicas de extensão de séries hidrológicas, são feitos, no item a seguir, diversos comentários sobre as técnicas e sua aplicação.

3.2.9 Comentários sobre as técnicas de extensão de séries hidrológicas

As técnicas aqui apresentadas tem tido diversas aplicações na Engenharia de Recursos Hídricos reportadas na literatura científica. A técnica OLS, como já mencionado, é uma das técnicas mais difundidas e sua aplicação tem-se feito em inúmeros aspectos, entre os quais está o preenchimento de falhas (OLIVEIRA, 2010).

As técnicas MOVE tem sido as mais utilizadas na literatura científica. Jia e Culver (2006) utilizaram a técnica MOVE2 para estender séries de vazões com o objetivo de gerar séries sintéticas a partir dos registros estimados. Além da MOVE2, foram utilizados, também, uma adaptação da técnica Redes Neurais Artificiais (*Artificial Neural Networks*, ou ANN) e um método envolvendo as áreas de drenagem em um posto fluviométrico e referência e no ponto em estudo. A técnica MOVE2 teve seu desempenho superado pela técnica ANN adaptada em vários quesitos, como a estimativa de vazões mínimas e máximas.

Duan *et al.* (2013) utilizaram a técnica MOVE3 para estimar as vazões no rio Ishikari, no Japão, de 1985 a 2010, e, a partir delas, utilizar o modelo específico para a estimativa de nutrientes e sólidos suspensos. A aplicação da técnica foi feita sobre vazões médias diárias, com correlações de Pearson entre as séries curta e longa variando de 0,79 a 0,96. Os resultados indicaram que a técnica MOVE3 teve melhor desempenho nas situações em que a correlação envolvida foi mais elevada.

A técnica MOVE1 foi utilizada no estudo de Reinfelds *et al.* (2006) para quantificar o impacto do uso da água para irrigação nas vazões diárias do rio Bega-Bemboka, na Austrália. Ao obter a série estendida pela técnica MOVE1, no posto Kanoona, tem-se o comportamento “natural” esperado para as vazões, sem a interferência de variáveis que não influenciam na série de referência. Confrontando a série estendida com os registros monitorados, nota-se uma grande diferença, como mostra a Figura 3.10. Tal diferença é atribuída, na pesquisa, ao uso da água para irrigação.

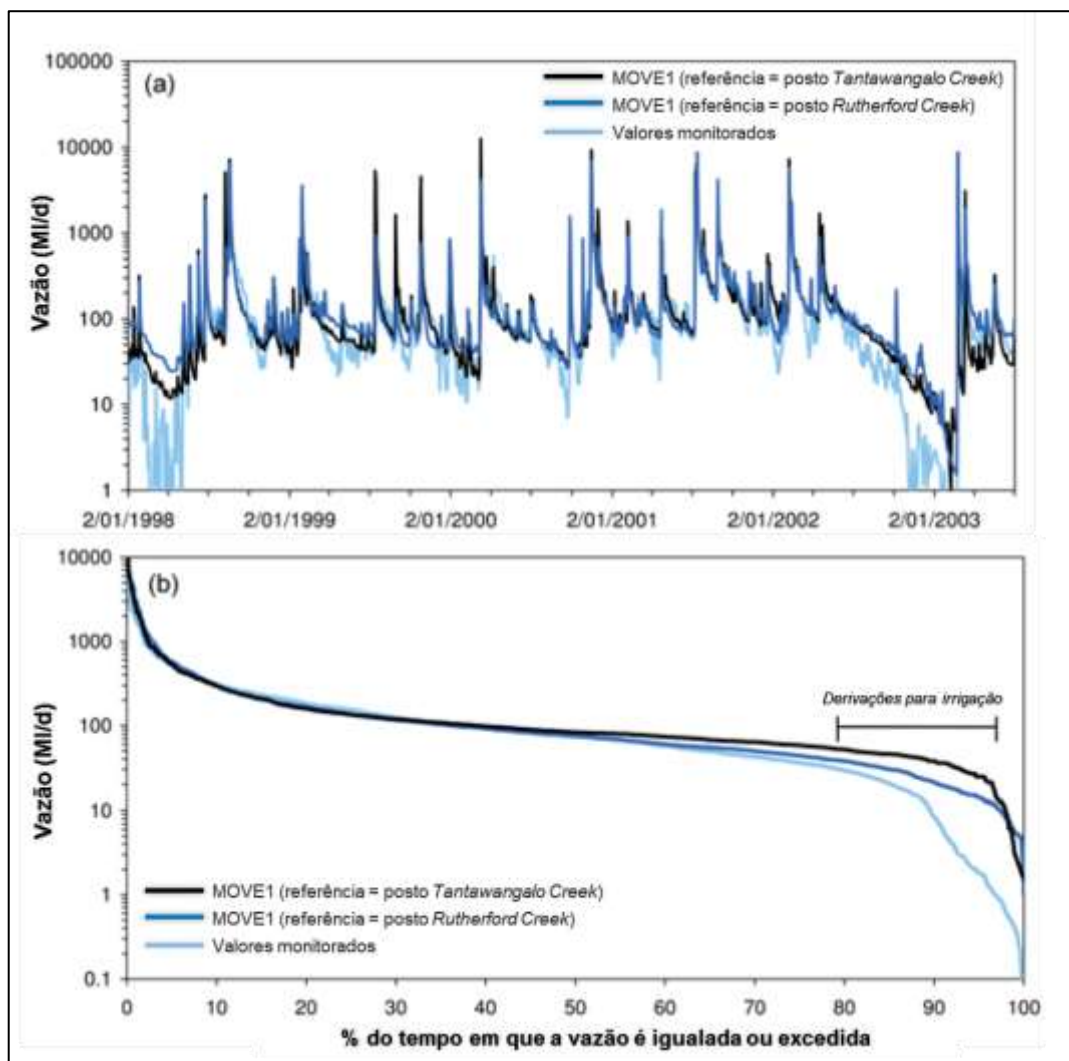


Figura 3.10 – Resultados da aplicação da técnica MOVE1 no estudo de Reinfelds *et al.* (2006)

Fonte: adaptado de REINFELDS *et al.*, 2006.

Hernandez-Henríquez *et al.* (2010) também utilizaram a técnica MOVE1 para reconstruir as vazões em um curso de água regulado por diversas restrições (operação de reservatórios, derivações, etc.). Sobre a série hidrológica obtida, foi feita uma análise da tendência utilizando duas metodologias, entre as quais está a técnica KTRL.

A técnica KTRL também foi aplicada por Albek (2003) para reconstruir a série temporal de indicadores de qualidade da água. Como apontado no estudo, a técnica KTRL teve desempenho superior à técnica OLS nessa tarefa devido à presença de *outliers* nos registros. Olsson *et al.* (2010) utilizaram a técnica KTRL para determinar tendências de redução da vazão média anual na bacia do rio Zerafshan, na Ásia Central.

Os Modelos Lineares Generalizados constituem uma técnica de regressão amplamente utilizada. Yang *et al.* (2005) utilizaram os GLM em uma das etapas de sua pesquisa para gerar séries sintéticas de precipitação espacialmente, com as séries mantendo algumas propriedades desejadas, como médias, variâncias, estrutura de autocorrelação, proporção de dias chuvosos, etc. O modelo adotado é composto de duas partes: i) modelagem da ocorrência ou não da precipitação, utilizando regressão logística, tomando como variáveis explicativas o mês do ano, as chuvas nos dias anteriores, indicadores topográficos, etc.; e ii) modelagem das alturas de chuva.

Outro exemplo de aplicação de GLM é dado por Ambrosino *et al.* (2011). Em seu estudo, os Modelos Lineares Generalizados foram utilizados para relacionar (temporal e espacialmente) as precipitações mensais no sul da África com indicadores climáticos, como umidade relativa do ar, índice SHT (temperatura média do hemisfério sul) e a temperatura do mar.

Uma grande quantidade de outros estudos utilizou as técnicas de extensão de séries hidrológicas aqui apresentadas, entre os quais estão os trabalhos de Eng *et al.* (2011), Déry *et al.* (2011) e Stahl *et al.* (2010). Não foram encontrados na literatura científica estudos que utilizaram as técnicas RPN, MOVE4, KTRL2, e RLOC, bem como as técnicas que consideram a autocorrelação amostral.

Como já mencionado, alguns estudos utilizaram os valores brutos das séries hidrológicas para proceder à aplicação das técnicas de extensão, enquanto, em outros, as técnicas foram aplicadas sobre os registros transformados, por exemplo, por meio da transformação logarítmica. Hirsch (1979) lidou com essa questão, tendo concluído que a aplicação da transformação logarítmica é recomendada quando o foco do estudo são as vazões mais baixas. Para o caso em que as vazões médias ou volumes escoados em um grande período são estudados, não recomenda-se que sejam feitas transformações.

Hirsch (1982) aponta que uma vantagem de utilizar a transformação logarítmica reside no fato de obter melhores resultados na extensão quando a relação entre as séries curta e de referência configura uma distribuição com assimetria mais próxima de ser claramente positiva do que nula. A distribuição acumulada da série estendida aproxima-se mais da distribuição acumulada das observações. Uma consequência indesejada desse procedimento é que a média amostral da série estendida (que é uma estimativa não enviesada da média dos logaritmos) será um estimador enviesado da média populacional.

Poucos estudos foram desenvolvidos na literatura para comparar o desempenho de técnicas de extensão de séries hidrológicas. Entre estes, destaca-se o trabalho de Khalil e Adamowski (2014a), no qual foi avaliado o comportamento das técnicas sob diferentes cenários de contaminação por *outliers*, de correlação e de extensão das séries curtas e de referência. Algumas conclusões obtidas foram:

- As técnicas OLS, MOVE3 e MOVE4 destacaram-se na estimativa dos registros individuais, no cenário sem contaminação por *outliers*. No cenário de contaminação por *outliers*, a técnica KTRL teve o melhor desempenho;
- As técnicas MOVE, KTRL2 e RLOC têm desempenho superior na preservação das características da distribuição populacional, em relação às técnicas OLS e KTRL; e
- Quando a série hidrológica possui *outliers*, as técnicas KTRL2 e RLOC são superiores às demais técnicas no que diz respeito à estimativa das medidas descritivas populacionais da série estendida.

Apesar de terem sido apresentadas aqui uma grande quantidade de técnicas de extensão de séries hidrológicas, existem várias outras técnicas que podem ser utilizadas neste contexto, entre as quais estão: Ponderação Regional e Vetor Regional (OLIVEIRA *et al.*, 2010); Redes Neurais Artificiais (DEPINÉ *et al.*, 2014; KHALIL & ADAMOWSKI, 2014b); Método dos Mínimos Quadrados Generalizado (STEDINGER & TASKER, 1985) e Ponderados (TASKER & STEDINGER, 1986); Regressão Baseada em Momentos-L (KHALIL *et al.*, 2016); e muitas outras. Provavelmente outras técnicas com potencial para a extensão de séries hidrológicas ainda surgirão. Ressalta-se, portanto, que não foi objetivo desta revisão da literatura e, sobretudo, desta pesquisa, esgotar a avaliação das técnicas de extensão de séries hidrológicas.

4 MATERIAL E MÉTODOS

4.1 Sequência metodológica

Com o intuito de facilitar o entendimento da metodologia adotada nesta pesquisa, é apresentada, neste item, uma síntese da sequência dos procedimentos utilizados, com indicação dos cenários simulados.

A pesquisa foi estruturada em duas etapas:

- Experimento sobre amostras sintéticas: foi realizado um experimento de Monte Carlo com o intuito de possibilitar a caracterização e a comparação das técnicas de extensão. Os resultados do experimento também foram utilizados para definir os critérios de aplicação das técnicas sobre séries hidrológicas reais; e
- Aplicação sobre amostras reais: as técnicas foram aplicadas a séries hidrológicas obtidas a partir do monitoramento da rede hidrometeorológica brasileira em uma região de estudo. Foram elaborados exemplos de aplicação das técnicas sobre dados reais, com o objetivo de validar os critérios de aplicação definidos na etapa anterior. Além disso, também foi possível avaliar a influência da relação física entre as séries hidrológicas estudadas no desempenho das técnicas de extensão. Esta etapa limitou-se ao estudo de séries hidrológicas de vazão e precipitação.

4.1.1 Experimentos sobre amostras sintéticas

Para a caracterização e comparação do desempenho das técnicas, foi desenvolvido um experimento de Monte Carlo baseado na extensão de séries temporais sintéticas. A principal vantagem desse experimento é tornar possível a avaliação do desempenho das técnicas em diversos aspectos, como a estimativa dos momentos e dos quantis populacionais. A avaliação do desempenho das técnicas foi feita por meio da comparação direta com as medidas descritivas populacionais, que estão relacionadas aos valores considerados como parâmetros para a geração das amostras sintéticas (média, variância e coeficiente de assimetria, nesta pesquisa). Essa conveniência não seria possível caso a pesquisa tivesse sido desenvolvida exclusivamente sobre séries hidrológicas reais, que, além de estarem sujeitas a erros de monitoramento e outros problemas, não possibilitam a obtenção exata das medidas descritivas populacionais.

Uma síntese dos cenários definidos para este experimento é apresentada na Figura 4.1. Alguns aspectos já apresentados serão abordados com maiores detalhes posteriormente neste capítulo.

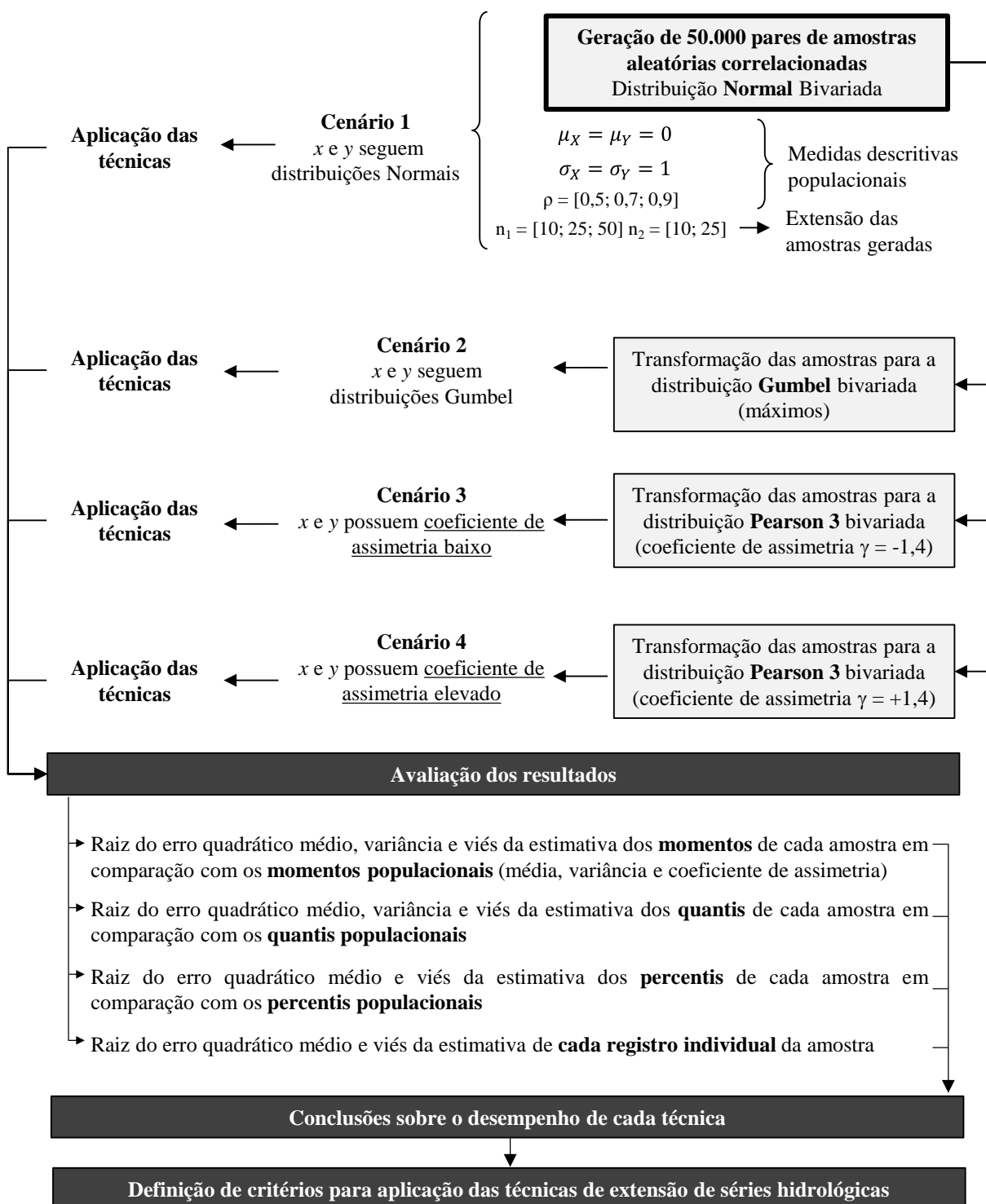


Figura 4.1 - Resumo da metodologia utilizada no experimento sobre amostras sintéticas

Como se nota na Figura 4.1, foram simulados 3 cenários de correlação populacional entre a série curta e a série de referência, 3 cenários de n_1 , 2 cenários de n_2 e 4 cenários de

distribuição das amostras. Dessa forma, totalizam-se 72 cenários, sendo que em cada um deles 50.000 pares de séries sintéticas aleatórias correlacionadas foram geradas. Em cada par de séries sintéticas, uma delas teve alguns valores desprezados, e foi estendida utilizando as técnicas de extensão.

4.1.2 Aplicação sobre amostras reais

Para a avaliação do desempenho das técnicas quando aplicadas sobre amostras reais de dados hidrológicos, optou-se por selecionar uma região de estudo monitorada por estações pluviométricas e fluviométricas. Os critérios para escolha e própria apresentação da região de estudo serão apresentados no decorrer do texto. Foram estudados dois cenários, a saber:

- Cenário 1: Extensão de uma série de precipitações tomando como referência uma outra série de precipitações; e
- Cenário 2: Extensão de uma série de vazões tomando como referência uma outra série de vazões.

As variáveis hidrológicas analisadas em cada um dos dois cenários foram: i) precipitação diária máxima anual e vazão média diária máxima anual, nos Cenários 1 e 2, respectivamente; e ii) precipitação anual e vazão média anual, nos Cenários 1 e 2, respectivamente. Tendo em vista que as técnicas de extensão de séries hidrológicas baseiam-se em princípios estatísticos, os postos de monitoramento pluviométrico e fluviométrico analisados em cada cenário foram escolhidos com base no coeficiente de correlação linear entre as séries registradas, com a abrangência de valores mais baixos, inferiores a 0,50, até valores elevados, próximos de 1. Foi escolhido um cenário único de extensão da série curta e do período estendido: $n_1 = 10$ e n_2 definido em função da extensão da série de referência. Cabe ratificar que a avaliação da influência de n_1 , n_2 e ρ no desempenho das técnicas de extensão foi avaliada no experimento de Monte Carlo.

A aplicação das técnicas sobre dados reais dividiu-se em duas etapas: i) aplicação obedecendo a ordem cronológica das observações, com análises mais detalhadas das particularidades de cada caso; e ii) aplicação em simulação de *bootstrap*, com o objetivo de avaliar a similaridade entre os resultados das simulações de Monte Carlo, empreendidas com base em amostras sintéticas, e aplicação semelhante, desenvolvida com base em amostras reais.

A seguir, cada uma das etapas metodológicas mencionadas é detalhada.

4.2 Simulação de Monte Carlo

Segundo Decker (1991), uma simulação de Monte Carlo pode ser informalmente definida como qualquer simulação que utilize números aleatórios para resolver determinado problema. Outra definição, mais precisa, é dada por Mun (2006): é o método no qual são calculados inúmeros cenários de um modelo por meio do sorteio repetitivo de valores pertencentes a uma distribuição de probabilidades, os quais representam uma variável com significativo grau de incerteza. Cada um desses cenários produz um resultado distinto. Sobre o conjunto de resultados formados, podem ser tomadas decisões baseadas ou não em medidas descritivas desse conjunto de resultados. Não raro, os resultados das simulações de Monte Carlo são apresentados como quantis pertencentes a distribuições de probabilidade. Como observado por James (1980), a aplicabilidade do método de Monte Carlo independe da natureza estocástica do sistema em estudo, mas sim na habilidade de formular o problema de forma que números aleatórios sejam utilizados para obter a solução.

Mais formalmente, os cálculos empreendidos no método de Monte Carlo equivalem à solução numérica de integrais. Supondo, por simplicidade, o caso em que F é uma função de números aleatórios uniformemente distribuídos x_i entre 0 e 1, o valor esperado do resultado F da simulação de Monte Carlo é um estimador não enviesado da integral múltipla (JAMES, 1980)

$$I = \int_0^1 \dots \int_0^1 F(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n \quad (4.1)$$

Dessa forma, as estimativas numéricas dos valores esperados, variâncias, erros, etc. das séries estendidas podem ser mais facilmente calculadas do que analiticamente, por meio das expressões de cada técnica de extensão.

A simulação de Monte Carlo é caracterizada como paramétrica por utilizar parâmetros de distribuições de probabilidade antes do início da simulação. Uma alternativa a essa abordagem é a simulação não paramétrica, na qual a amostragem de registros históricos é utilizada como dados de entrada na simulação e, portanto, não é necessário fornecer parâmetros de distribuições de probabilidades para iniciar a simulação. As simulações não paramétricas de *bootstrap*, por exemplo, são ferramentas robustas, porém aplicáveis apenas quando da disponibilidade de dados. Além disso, quanto maior é o conjunto de dados

históricos disponível, maior será o nível de precisão e confiança nos resultados da simulação (MUN, 2006).

Com relação à simulação de Monte Carlo, é importante destacar, especialmente no contexto da engenharia de recursos hídricos, a diferença entre a geração de variáveis aleatórias pertencentes a uma dada distribuição e a geração de séries hidrológicas sintéticas. Segundo Rubinstein (1981), na simulação de Monte Carlo, o tempo não possui papel necessariamente relevante nas análises. Além disso, nas simulações de Monte Carlo, os valores gerados, seja para entrada ou para a saída de dados, são independentes, o que não é observado nas séries temporais de diversas variáveis aleatórias hidrológicas. Nas simulações realizadas no contexto desta pesquisa, supõe-se que não existe dependência entre os números aleatórios gerados, condição que é premissa da metodologia tradicional de análise de frequência.

Os estudos sobre técnicas de extensão de séries hidrológicas desenvolvidos até a data desta pesquisa utilizam-se, em geral, de simulações de Monte Carlo para gerar uma elevada quantidade de amostras de duas variáveis aleatórias correlacionadas entre si, seguindo uma determinada distribuição de probabilidades. Um procedimento distinto foi adotado por Khalil e Adamowski (2014a), que utilizaram a técnica de *bootstrap* para gerar 5.000 pares de amostras correlacionadas com extensão entre 48 e 240 valores a partir de um universo de 1.000 pares de registros gerados sinteticamente.

Considerando o objetivo específico do presente trabalho, relacionado à aplicação das técnicas de extensão sobre dados hidrológicos reais, uma alternativa natural para o desenvolvimento metodológico seria utilizar registros de precipitação e vazão monitorados na rede hidrometeorológica brasileira e disponibilizados no sistema Hidroweb para efetuar simulações baseadas em técnicas não paramétricas. Contudo, esse procedimento encontra alguns empecilhos, entre os quais estão:

- Dificuldade em agregar um grande volume de dados: Além da complexidade de coletar todos os registros da rede hidrometeorológica brasileira, existem problemas como registros inconsistentes, falhas de monitoramento, erros sistemáticos, assincronismo entre períodos chuvosos, entre outros;
- Diferenças de escala: comparação entre séries hidrológicas com registros de magnitudes significativamente diferentes, especialmente no caso de vazões;

- O desconhecimento das medidas descritivas populacionais: a própria extensão reduzida das séries hidrológicas, que motivou o desenvolvimento desta pesquisa, reduz a confiabilidade das estimativas das medidas descritivas populacionais.

Os pontos acima mencionados indicam que uma eventual metodologia desenvolvida para a avaliação do desempenho das técnicas de extensão teria limitações óbvias. Com isso em mente, optou-se pela abordagem paramétrica, por meio da geração de séries sintéticas, para o desenvolvimento da pesquisa, obtendo, também, concordância com os demais trabalhos realizados sobre extensão de séries hidrológicas.

O número de realizações adotado nas simulações de Monte Carlo pode ser definido considerando o erro das estimativas das diversas medidas descritivas que são extraídas do modelo. No caso desta pesquisa, uma estimativa precisa da magnitude dos erros das simulações é de difícil obtenção, tendo em vista os quatro cenários considerados, as diversas técnicas estudadas e as várias medidas descritivas avaliadas. Levando em consideração também o tempo computacional dispendido nas simulações e as metodologias dos demais estudos relacionados às técnicas de extensão de séries hidrológicas, fixou-se o número de realizações em 50.000. O ponto de partida das simulações foi a geração das amostras sintéticas das séries curta e de referência.

4.3 Geração de amostras sintéticas

Segundo Genest e Favre (2007), a dependência de variáveis hidrológicas vem tradicionalmente sendo representada por meio de famílias de distribuições bivariadas, sendo as mais comuns a Normal, a Log-Normal, a Gama e as distribuições de valores extremos. Dando prosseguimento a essa abordagem, foram adotadas distribuições de probabilidade bivariadas também nesta pesquisa.

Para se atingir o objetivo de comparar o desempenho das técnicas e avaliar sua aplicabilidade para diversas variáveis aleatórias hidrológicas, adotou-se a premissa de utilizar distribuições de probabilidades que representassem tais variáveis. Por exemplo, ao amostrar quantis aleatórios da distribuição Gumbel bivariada, estaria sendo avaliado o desempenho das técnicas na extensão de variáveis aleatórias de máximos (como precipitações diárias máximas anuais ou vazões médias diárias máximas anuais). Também por esse motivo, e pelo fato de os demais estudos sobre as técnicas de extensão terem sido realizados com base em amostras

normais bivariadas, também foram considerados, nesta pesquisa, quantis aleatórios da distribuição Normal bivariada na amostragem.

Para avaliar a influência do coeficiente de assimetria no desempenho das técnicas, foram adotados dois outros cenários, correspondentes às distribuições Pearson 3 com coeficiente de assimetria positivo e negativo. Tais distribuições foram escolhidas por convergirem para a distribuição normal quando o coeficiente de assimetria tende a zero.

Totalizaram-se, portanto, 4 cenários de distribuições de probabilidades das séries amostradas: i) Gumbel; ii) Normal; iii) Pearson 3 com coeficiente de assimetria positivo; e iv) Pearson 3 com coeficiente de assimetria negativo. Em todos os casos, foram consideradas distribuições bivariadas, ou seja, em uma das dimensões tem-se uma única série curta e, em outra, uma única série de referência. A seguir, apresenta-se a metodologia empregada para a amostragem das distribuições bivariadas.

4.3.1 Distribuição Normal Bivariada

Como referencial para a geração das amostras aleatórias, foi utilizada a distribuição normal bivariada, cuja função densidade de probabilidade é dada pela equação

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x-\mu_X}{\sigma_X} \right)^2 - 2\rho \frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \left(\frac{y-\mu_Y}{\sigma_Y} \right)^2 \right] \right\} \quad (4.2)$$

na qual

μ_X e μ_Y representam as médias populacionais de X e Y , respectivamente;

σ_X e σ_Y representam os desvios-padrão populacionais de X e Y , respectivamente; e

ρ representa o coeficiente de correlação populacional entre X e Y .

A utilização da distribuição normal multivariada para simulações de técnicas de regressão possui diversas vantagens, entre as quais estão:

- Facilidade de obtenção de conjuntos de quantis aleatórios pertencentes à distribuição normal multivariada, em virtude de sua difusão no meio científico. Por isso, métodos para a geração de quantis aleatórios da distribuição normal multivariada são amplamente estudados e disponibilizados;

- Pelo Teorema do Limite Central, a soma de n variáveis aleatórias independentes e com distribuição de probabilidades idênticas entre si, com mesma média μ e desvio padrão σ populacionais, tende assintoticamente a uma variável normal padrão após a transformação descrita pela Equação 4.3, quando n é suficientemente grande (ROSS, 2007).

$$Z_n = \frac{X_1 + X_2 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \quad (4.3)$$

Nesta equação,

Z_n é a variável resultado da transformação; e

X_i é a i -ésima variável aleatória com média μ e desvio padrão σ populacionais.

O Teorema do Limite Central encontra dificuldades de aplicação na Engenharia de Recursos Hídricos, envolvendo a indefinição de um valor mínimo de n e a improbabilidade de os valores X_i serem independentes e pertencerem exatamente à mesma distribuição de probabilidades, com os mesmos parâmetros. Naghettini e Pinto (2007) apontam que, se as variáveis pertencerem a distribuições não fortemente assimétricas e obedecerem aos demais requisitos, valores de n próximos a 30 são suficientes para que Z_n convirja para uma variável normal padrão.

Certa flexibilidade é fornecida pelo chamado “Teorema do Limite Central Generalizado”, ao permitir que a hipótese de parâmetros idênticos das distribuições de probabilidade de X_i seja relaxada, ampliando sua aplicabilidade. Dessa forma, é possível que variáveis aleatórias como precipitações acumuladas anuais e vazões médias anuais sejam bem representadas por meio da distribuição normal. O valor mínimo de n , contudo, não é facilmente determinado (NAGHETTINI & PINTO, 2007).

- Na grande maioria dos estudos realizados sobre a extensão de séries hidrológicas, as simulações tiveram como base quantis da distribuição normal multivariada. Dessa forma, possibilita-se a comparação dos resultados desta pesquisa com os resultados já apresentados na literatura científica.

Existem diversos métodos para a obtenção de amostras de quantis pertencentes à distribuição normal multivariada, entre os quais estão os métodos de decomposição de Cholesky e *Eigenvalues* (autovalores). Maiores informações sobre tais métodos fogem ao escopo desta

pesquisa e podem ser obtidos nos trabalhos de Genz (1992, 1993 e 2004) e Genz e Bretz (2009).

Para a geração das amostras pertencentes à distribuição normal multivariada, foi utilizado o pacote “mvtnorm”(MI *et al.*, 2009), elaborado para aplicação na linguagem de programação R. O método de geração dos quantis aleatórios utilizado foi o *Eigenvalues*. O embasamento dessa metodologia é puramente matemático e estatístico e, portanto, ela não será detalhada aqui. Contudo, é importante mencionar que a metodologia utilizada não é capaz de incorporar correlação serial na amostra sintética, uma vez que os dados gerados são considerados aleatórios e independentes.

Além dos valores correspondentes aos cenários de simulação (n_1 , n_2 e ρ), definidos no item 4.1, foram utilizados na geração das amostras aleatórias os valores de $\mu_X = \mu_Y = 0$ e $\sigma_X = \sigma_Y = 1$. Esses valores foram escolhidos por serem os mesmos utilizados nos demais estudos sobre técnicas de extensão de séries hidrológicas, possibilitando a comparação de resultados.

4.3.2 Outras distribuições multivariadas

Existem diversos métodos para a geração de quantis de distribuições multivariadas com distribuições diferentes da normal. Esse é um problema que vem sendo estudado desde o trabalho desenvolvido por Wigner (1932) *apud* Schweizer e Sklar (1986). O tratamento mais comum dado à geração de distribuições multivariadas é feito por meio do uso de cópulas.

A metodologia das cópulas baseia-se no princípio de que uma distribuição multivariada pode ser definida por meio de funções densidade de probabilidade marginais contínuas e por uma função denominada cópula, que descreve a dependência entre as variáveis. Inicialmente, a dependência entre as FDP marginais é traduzida, por meio da função cópula, na dependência entre FDP marginais uniformes. A obtenção das FDP marginais é feita por meio da transformação entre variáveis aleatórias. Neste processo, a correlação entre as variáveis não é preservada, e a amostragem não é exata (DUKIC & MARIĆ, 2013).

A grande vantagem do uso de cópulas na modelagem da dependência entre variáveis aleatórias reside no fato de ser dispensável que as variáveis aleatórias pertençam à mesma família de distribuições univariadas (GENEST & FAVRE, 2007). Além disso, as cópulas têm sido de grande uso na Engenharia de Recursos Hídricos, com aplicações em análises de

frequência regionais e na estimativa de relações vazão, duração e frequência, entre outras. Contudo, não se recomenda que as cópulas sejam utilizadas em estimativas associadas a probabilidades muito baixas, a menos que fortes evidências justifiquem o contrário (RENARD & LANG, 2007).

Outros processos de geração de variáveis aleatórias, inclusive com amostragem exata, são descritos na literatura (BUSTOS *et al.*, 2001; SARABIA & DÉNIZ, 2008; YAHAV & SHMUELI, 2011; DUKIĆ & MARIĆ, 2013). No presente trabalho, será utilizado o método da transformação, descrito por Bustos *et al.* (2001), por sua simplicidade e pela disponibilidade da amostragem da distribuição normal bivariada. Esse método é caracterizado pelas seguintes etapas (BUSTOS *et al.*, 2001):

- Amostragem independente de n amostras, cada uma contendo quantis pertencentes a uma distribuição conveniente;
- Introdução de correlação entre as n amostras geradas;
- Cálculo da probabilidade acumulada de todos os valores de cada uma das n amostras, com base nas FDP marginais acumuladas das distribuições nas quais foram baseadas as amostras independentes;
- Ajuste dos parâmetros das distribuições marginais desejadas por meio dos parâmetros utilizados na amostragem inicial; e
- Cálculo dos quantis das distribuições marginais desejadas utilizando as probabilidades de cada um dos valores gerados.

A transformação que garante a validade desse procedimento baseia-se no cálculo das probabilidades, utilizando a distribuição originalmente adotada. Tais probabilidades pertencem à distribuição uniforme e podem ser transformadas para a distribuição desejada. Após esses procedimentos, é necessário verificar as correlações entre as n amostras, pois a amostragem realizada por meio desse método não é exata (BUSTOS *et al.*, 2001). Para a estimativa dos parâmetros utilizados nas transformações aplicadas nesta pesquisa, foi utilizado o método dos momentos.

A amostragem segundo a distribuição normal bivariada, descrita no item 4.3.1, foi utilizada para suprimir as duas primeiras etapas na amostragem de outras distribuições. Portanto, a média e o desvio padrão populacionais de todas as distribuições consideradas foram os

mesmos das amostras normais ($\mu_X = \mu_Y = 0$ e $\sigma_X = \sigma_Y = 1$). Apesar de os dados normais possuírem coeficiente de assimetria nulo, durante o processo de transformação das amostras para a distribuição Gumbel, adquire-se um coeficiente de assimetria populacional de +1,14, intrínseco a essa distribuição. Na transformação para a distribuição Pearson 3, o método de estimação dos parâmetros permitiu a definição do coeficiente de assimetria populacional. Procedeu-se, portanto, à escolha desse parâmetro.

Foram encontrados, na literatura, diversos estudos nos quais foram desenvolvidas análises de sensibilidade frente à variação do coeficiente de assimetria de variáveis hidrológicas: Chowdury e Stedinger (1991) restringiram os coeficientes de assimetria utilizados em seu estudo para o intervalo $|\gamma| = 1,5$; McCuen (2001) utilizou valores variando entre -1,0 e +1,0; e Griffis *et al.* (2004) utilizaram, no processo de amostragem aleatória, distribuições com coeficiente de assimetria populacional limitados ao intervalo entre -1,4 e +1,4. No presente trabalho, foram adotados os mesmos limites utilizados por Griffis *et al.* (2004).

A título de ilustração, apresentam-se, na Figura 4.2, as FDP marginais utilizadas em cada cenário. Além da clara simetria entre as distribuições Pearson 3 com assimetrias positiva e negativa, nota-se alguma semelhança visual entre os formados das FDP das distribuições com assimetria positiva (Gumbel e Pearson 3+).

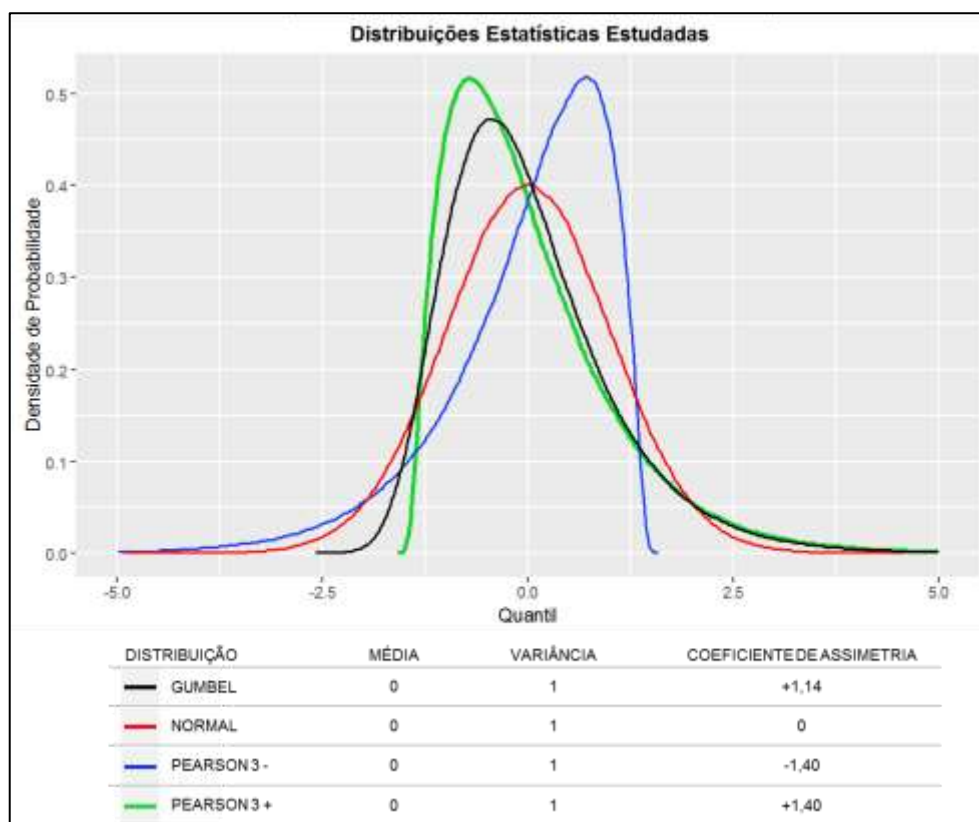


Figura 4.2 – Distribuições marginais utilizadas nas amostragens

4.4 Técnicas de extensão de séries hidrológicas

Entre as técnicas de extensão de séries hidrológicas apresentadas na Revisão da Literatura, serão utilizadas na pesquisa apenas aquelas que consideram uma única série de referência e uma única série curta a ser estendida. Essa premissa foi escolhida pelas seguintes razões: i) simplicidade (caso fosse considerado o caso multivariado, inúmeros cenários de correlação e de extensão das séries curta e de referência deveriam ser analisados, aumentando enormemente o escopo da pesquisa); ii) para o caso univariado, existe uma variedade maior de técnicas de extensão de séries hidrológicas, como apresentado na Revisão da Literatura; e iii) dadas as fragilidades do monitoramento hidrometeorológico brasileiro, como a baixa densidade de estações e as próprias extensões reduzidas dos registros, a aplicação prática envolvendo uma série curta e uma série de referência é mais provável do que cenários específicos com uma ou mais séries curta ou de referência. Com base nesse aspecto, apresentam-se, a seguir, os critérios e as premissas utilizadas para a aplicação das técnicas de extensão.

As técnicas KTRL, KTRL2, MOVE1, MOVE2, MOVE3, MOVE4 e RLOC atendem à premissa fixada acima e foram aplicadas diretamente, conforme descrito pelas formulações

apresentadas na Revisão da Literatura. Ao analisar a pertinência de estudar as técnicas que consideram a autocorrelação amostral, optou-se por se ater ao foco da pesquisa, a estimativa dos quantis máximos populacionais. Como na análise de frequência tradicional adota-se a premissa de observações IID, tais técnicas foram descartadas. A técnica OLS, apesar de possuir uma versão multivariada (regressão linear múltipla), foi agregada ao grupo de técnicas estudadas apenas em seu caso univariado.

Cabe mencionar também que, apesar de a técnica RPN (*Regression Plus Noise*, ou Regressão Mais Ruído) basear-se na técnica OLS e fornecer melhores resultados do que esta no caso populacional, não foi estudada devido à sua desvantagem prática de não fornecer um resultado único em sua aplicação, por ser dependente de um termo aleatório.

As técnicas mencionadas até aqui são de aplicação direta. Já a técnica GLM apresenta, num primeiro momento, a desvantagem de necessitar de definições que podem vir a ser subjetivas para sua aplicação, como a função de ligação a ser considerada, por exemplo. Essas escolhas tiveram que ser automatizadas para a realização das simulações de Monte Carlo, e para tanto, foram utilizados os seguintes critérios:

- Distribuição dos resíduos: adotou-se que os resíduos pertencem a uma distribuição Normal. Para o cenário em que são geradas amostras bivariadas da distribuição Normal, essa premissa é adequada. Porém, ao amostrar os dados sintéticos seguindo outras distribuições (Gumbel e Pearson 3), essa não é a premissa ideal, mesmo sabendo que nenhuma das distribuições aqui estudadas é dada como opção na rotina de ajuste utilizada. Tendo em vista que muitas das outras técnicas estudadas ainda sequer foram estudadas com relação a aspectos estatísticos dos resíduos, acredita-se que esse critério não trará prejuízo sensível para a pesquisa;
- Função de ligação: entre as diversas possibilidades relacionadas às de funções de ligação que os modelos lineares generalizados permitem utilizar, foram escolhidas algumas daquelas que representam uma relação crescente entre x e y e que estão entre as mais utilizadas na prática. Por esse motivo, foram utilizadas as funções de ligação *logit*, *probit*, complemento log-log e logarítmica, obtendo, portanto, quatro modelos de regressão para cada par de amostras. A diversidade das funções de ligação possibilita

que ajustes com formatos distintos sejam utilizados para se aderir melhor aos dados amostrais;

- Ajuste do modelo: foi adotado o método IWLS (*Iterative Weighted Least Squares*), que efetua a resolução numérica das equações geradas pelo método da verossimilhança; e
- Escolha do modelo de regressão mais adequado: existem diversos critérios para a definição do modelo de regressão mais adequado. Nesta pesquisa foi utilizado o Critério de Informação de Akaike (AIC). Para cada par de amostras gerado, o modelo de regressão (baseado em cada uma das funções de ligação consideradas) que apresentasse o menor valor do AIC seria adotado para a extensão da série de dados. Adianta-se que não seria possível utilizar todos os modelos de regressão considerados em todos os pares de amostras gerados, uma vez que o algoritmo utilizado para o ajuste dos modelos não obteve convergência em todos os casos de função de ligação utilizados. Em cada par de amostras gerado, bastou que um dos modelos convergisse para que a técnica GLM retornasse uma série estendida.

O ajuste e a extensão das séries hidrológicas na técnica GLM foi feito por meio da aplicação da rotina “*glm.fit*” da linguagem de programação R.

Em decorrência das funções de ligação utilizadas, tornou-se necessário normalizar os dados gerados para o intervalo (0, 1). Para tanto, foi adotada a normalização Min-Max (JAIN & BHANDARE, 2011). Essa normalização foi adotada pela sua simplicidade, por não alterar o aspecto visual dos dados amostrais, por manter a relação linear entre as séries curta e de referência e pela facilidade de retornar para a escala inicial após a extensão utilizando a técnica GLM. Essa normalização é definida pela expressão

$$N_i = \frac{A_i - \text{mínimo}(A)}{\text{máximo}(A) - \text{mínimo}(A)} (MAX - MÍN) + MÍN \quad (4.4)$$

na qual

N_i é o i -ésimo valor do vetor N de valores normalizados;

A_i é i -ésimo valor do vetor A de valores amostrais segundo a distribuição escolhida (variável dependente ou independente);

$MÍN$ é o valor mínimo do vetor N ; e

$MÁX$ é o valor máximo do vetor N .

Como valores de $MÍN$ e $MÁX$, foram adotados os valores mais baixo e mais alto, respectivamente, da posição de plotagem de Weibull (WEIBULL, 1939), para cada amostra de x e y , calculados por

$$MÍN = 1/(n_1 + 1) \quad (4.5)$$

$$MÁX = n_1/(n_1 + 1) \quad (4.6)$$

Uma desvantagem do uso dessa normalização reside no fato que a escolha dos valores de $MÍN$ e $MÁX$ pode alterar o ajuste da técnica GLM. Todavia, observou-se visualmente que a metodologia utilizada garante que o formato dos dados amostrais seja buscado por meio do critério de escolha da função de ligação. Assim, essa possível vantagem da técnica GLM pôde ser mantida.

Como já mencionado na Revisão da Literatura, transformações dos dados originais podem ser utilizadas em casos específicos, como na análise de vazões baixas. Na literatura, recomenda-se que sejam feitas transformações para normalizar os dados de x e y , tendo em vista a premissa fundamental de algumas técnicas terem sido desenvolvidas considerando dados normais. Contudo, nesta pesquisa, buscou-se avaliar a aplicação direta das técnicas frente a vários cenários de distribuição dos dados. Portanto, nenhuma transformação foi utilizada.

4.5 Definição de critérios para a aplicação das técnicas de extensão de séries hidrológicas sobre dados reais

Para a definição dos critérios de aplicação das técnicas de extensão de séries hidrológicas sobre dados reais, foram considerados apenas os resultados referentes à estimativa do quantil populacional de referência associado ao TR = 100 anos. Dessa forma, foi considerado o seguinte procedimento, baseado na teoria dos estimadores:

- Para cada cenário simulado (cada valor de n_1 , n_2 , ρ e distribuição à qual pertencem os dados), foi verificado o viés das técnicas. As técnicas que geraram séries estendidas cujas estimativas do quantil de referência apresentaram viés superior (em valor absoluto) ao viés da estimativa do quantil de referência com base na série curta foram descartadas;

- Foi verificada a variância da estimativa do quantil de referência com base em séries estendidas por meio de cada técnica remanescente, tendo sido descartadas aquelas que possuísem variância superior à variância da estimativa por meio da série curta; e
- Adicionalmente, as técnicas remanescentes foram submetidas à verificação do erro na estimativa do quantil. As técnicas que originaram séries estendidas cujas estimativas dos quantis apresentaram erro superior ao erro da estimativa baseada na série curta foram descartadas.

As técnicas remanescentes, em cada cenário, foram consideradas como aplicáveis para a obtenção de quantis máximos. O ajuste das distribuições analisadas para a estimativa do quantil associado ao período de retorno de 100 anos foi feito por meio do método dos momentos.

A escolha final da técnica a ser utilizada deve levar em conta aspectos importantes que não foram considerados neste estudo, quais sejam:

- Objetivo da extensão da série hidrológica: algumas técnicas de extensão têm como objetivo obter estimativas melhores dos momentos populacionais em relação à série curta. Tais técnicas, após passar pelos critérios definidos acima, serão indicadas quando o objetivo for obter estimativas mais confiáveis dos quantis máximos extremos. Quando o objetivo estiver relacionado à obtenção de percentis para a série estendida, as técnicas que se baseiam nesses conceitos serão as mais indicadas, também após passar pelos critérios definidos pela metodologia descrita acima; e
- Presença de *outliers*: neste estudo, não foi feito qualquer tipo de tratamento com relação ao comportamento das técnicas na presença de *outliers*. Dessa forma, esse fato deve ser levado em consideração quando da escolha da técnica mais apropriada.

Para a validação dos critérios definidos e análise da extensão de séries hidrológicas em condições nas quais existe relação física entre elas, procedeu-se à aplicação das técnicas de extensão sobre amostras reais.

4.6 Aplicação sobre amostras reais

4.6.1 Critérios para a escolha da região estudada

Para viabilizar a aplicação das técnicas sobre amostras reais, foi necessário definir uma região de interesse. A região estudada deveria possuir as seguintes características:

- número elevado de estações fluviométricas e pluviométricas;
- estações com séries monitoradas extensas; e
- séries históricas consistentes, com número reduzido de falhas de monitoramento.

A consistência, mencionada como um dos requisitos, é importante para garantir que os postos de monitoramento descrevam fielmente o comportamento hidrológico e meteorológico da região estudada. Uma das ferramentas utilizadas para a análise de consistência é o método da Dupla Massa.

O método da Dupla Massa foi desenvolvido pelo *United States Geological Survey* (Serviço Geológico dos Estados Unidos), e é bastante difundido no Brasil. Para a aplicação do método, são selecionados todos os postos de monitoramento em uma região. Os valores, registrados em base mensal, são acumulados e são plotados os valores do posto cuja consistência é analisada (no eixo das ordenadas) contra os valores de um posto de referência (no eixo das abscissas). Outra forma de aplicar o método consiste em considerar os valores médios das séries acumuladas dos demais postos da região com os registros de referência. Se os valores monitorados no posto em análise não apresentarem inconsistências, o gráfico gerado terá o aspecto de uma linha reta (TUCCI, 2009).

4.6.2 Região estudada

Após uma pesquisa na rede de monitoramento hidrometeorológica gerenciada pela Agência Nacional de Águas e disponibilizada por meio do portal Hidroweb (<http://hidroweb.ana.gov.br/>), foi encontrada uma região que atende aos critérios definidos. Trata-se da bacia hidrográfica do rio Piranga com exutório no município de Porto Firme, distante aproximadamente 120 km de Belo Horizonte. As Figuras 4.3 e 4.4 apresentam a localização da região de estudo, das estações hidrometeorológicas pré-selecionadas e a hidrografia da bacia hidrográfica do rio Piranga, a montante da estação fluviométrica Porto Firme (código 56075000).

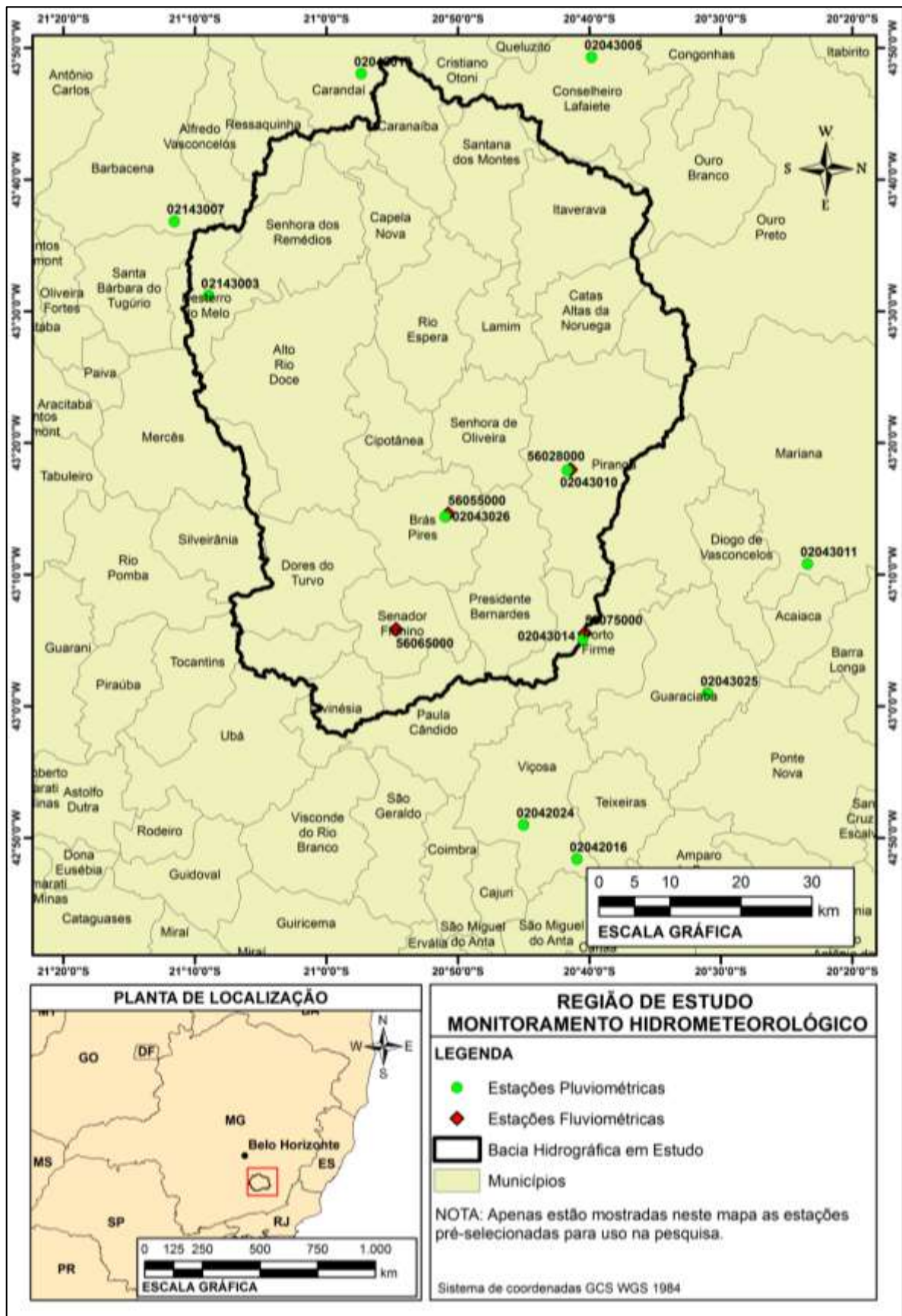


Figura 4.3 – Estações pluviométricas dentro e nas adjacências da bacia hidrográfica em estudo

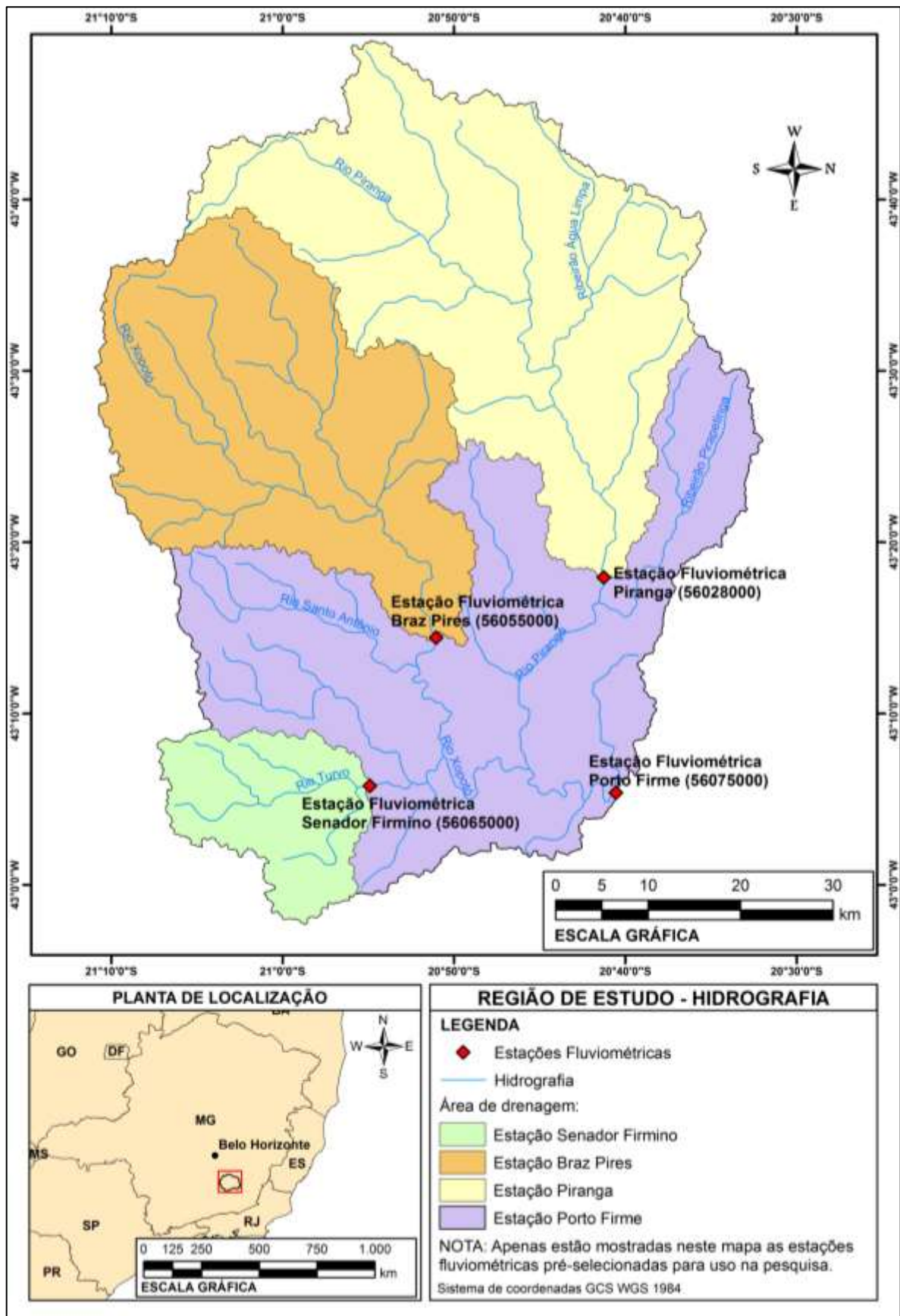


Figura 4.4 - Estações fluviométricas na bacia hidrográfica em estudo

Na região, foram localizadas diversas estações pluviométricas e fluviométricas, tendo sido pré-selecionadas 4 estações fluviométricas e 4 estações pluviométricas dentro da bacia do rio Piranga e 7 estações pluviométricas nos arredores da bacia. Essas estações foram pré-selecionadas pela extensão da série de monitoramento (acima de 50 anos, aproximadamente) e pela disponibilidade de dados no sistema Hidroweb. As características gerais das estações pré-selecionadas são apresentadas nas Tabelas 4.1 e 4.2, e, nas Figuras 4.3 e 4.4, é apresentada a localização dessas estações.

Tabela 4.1 - Estações pluviométricas pré-selecionadas

Código	Nome	Localização		Monitoramento		Extensão (anos)
		Latitude	Longitude	Início	Fim	
Na bacia do rio Piranga em Porto Firme						
02043010	Piranga	20°41'26" S	43°17'58" O	1941	2015	74
02043014	Porto Firme	20°40'13" S	43°05'17" O	1941	2016	75
02043026	Braz Pires	20°50'51" S	43°14'31" O	1967	2016	49
02143003	Desterro do Melo	21°08'57" S	43°31'12" O	1941	2016	75
Nas adjacências da bacia do rio Piranga						
02042016	São Miguel do Anta	20°40'57" S	42°48'24" O	1967	2016	49
02042024	Viçosa	20°45'00" S	42°51'00" O	1920	1998	78
02043005	Conselheiro Lafaiete	20°39'54" S	43°46'40" O	1940	2016	76
02043011	Fazenda Paraíso	20°23'24" S	43°10'49" O	1941	2016	75
02043018	Carandaí	20°57'21" S	43°48'30" O	1941	2015	74
02043025	Usina da Brecha	20°31'00" S	43°01'00" O	1959	2016	57
02143007	Vargem do Engenho	21°11'34" S	43°36'50" O	1941	2015	74

Tabela 4.2 - Estações fluviométricas pré-selecionadas na bacia do rio Piranga

Código	Nome	Localização		Monitoramento		Extensão (anos)	Área de drenagem (km ²)
		Latitude	Longitude	Início	Fim		
56028000	Piranga	20°41'26" S	43°17'58" O	1938	2016	78	1.397
56055000	Braz Pires	20°50'51" S	43°14'31" O	1938	2016	78	1.091
56065000	Senador Firmino	20°54'42" S	43°05'50" O	1941	2016	75	296
56075000	Porto Firme	20°40'13" S	43°05'17" O	1938	2016	78	4.257

Após a coleta das informações das estações, procedeu-se à etapa de verificação de falhas de monitoramento. Para cada posto, no caso da vazão média anual e da precipitação acumulada anual, foram removidos da análise os anos hidrológicos que apresentaram qualquer número de falhas em qualquer mês do ano. No caso da precipitação diária máxima anual e da vazão média diária máxima anual, foram removidos da análise apenas os anos hidrológicos que apresentaram falhas no período chuvoso da região (outubro a março). Os valores máximos diários anuais nos anos que apresentaram falhas no período seco, de abril a setembro, foram mantidos.

As matrizes de correlações para as séries monitoradas são apresentadas, nas Tabelas 4.3 e 4.4, para as estações pluviométricas e fluviométricas dentro da região estudada.

Tabela 4.3 - Matriz de correlação entre as séries de máximos diários

		FLUVIOMÉTRICAS				PLUVIOMÉTRICAS			
		Braz Pires	Piranga	Senador Firmínio	Porto Firme	Braz Pires	Piranga	Desterro do Melo	Porto Firme
FLUVIOMÉTRICAS	Braz Pires	1,00	0,76	0,59	0,78	0,52	0,31	0,22	0,19
	Piranga	0,76	1,00	0,43	0,88	0,35	0,20	0,28	0,19
	Senador Firmínio	0,59	0,43	1,00	0,58	0,54	0,19	0,05	0,07
	Porto Firme	0,78	0,88	0,58	1,00	0,54	0,30	0,16	0,28
PLUVIOMÉTRICAS	Braz Pires	0,52	0,35	0,54	0,54	1,00	0,38	0,10	0,23
	Piranga	0,31	0,20	0,19	0,30	0,38	1,00	0,02	0,14
	Desterro do Melo	0,22	0,28	0,05	0,16	0,10	0,02	1,00	0,09
	Porto Firme	0,19	0,19	0,07	0,28	0,23	0,14	0,09	1,00

Tabela 4.4 - Matriz de correlação entre as séries de precipitação acumulada anual / vazões médias anuais

		FLUVIOMÉTRICAS				PLUVIOMÉTRICAS			
		Braz Pires	Piranga	Senador Firmínio	Porto Firme	Braz Pires	Piranga	Desterro do Melo	Porto Firme
FLUVIOMÉTRICAS	Braz Pires	1,00	0,88	0,81	0,92	0,62	0,66	0,62	0,52
	Piranga	0,88	1,00	0,69	0,89	0,64	0,60	0,58	0,40
	Senador Firmínio	0,81	0,69	1,00	0,87	0,57	0,52	0,37	0,58
	Porto Firme	0,92	0,89	0,87	1,00	0,70	0,67	0,50	0,55
PLUVIOMÉTRICAS	Braz Pires	0,62	0,64	0,57	0,70	1,00	0,86	0,70	0,27
	Piranga	0,66	0,60	0,52	0,67	0,86	1,00	0,60	0,30
	Desterro do Melo	0,62	0,58	0,37	0,50	0,70	0,60	1,00	0,21
	Porto Firme	0,52	0,40	0,58	0,55	0,27	0,30	0,21	1,00

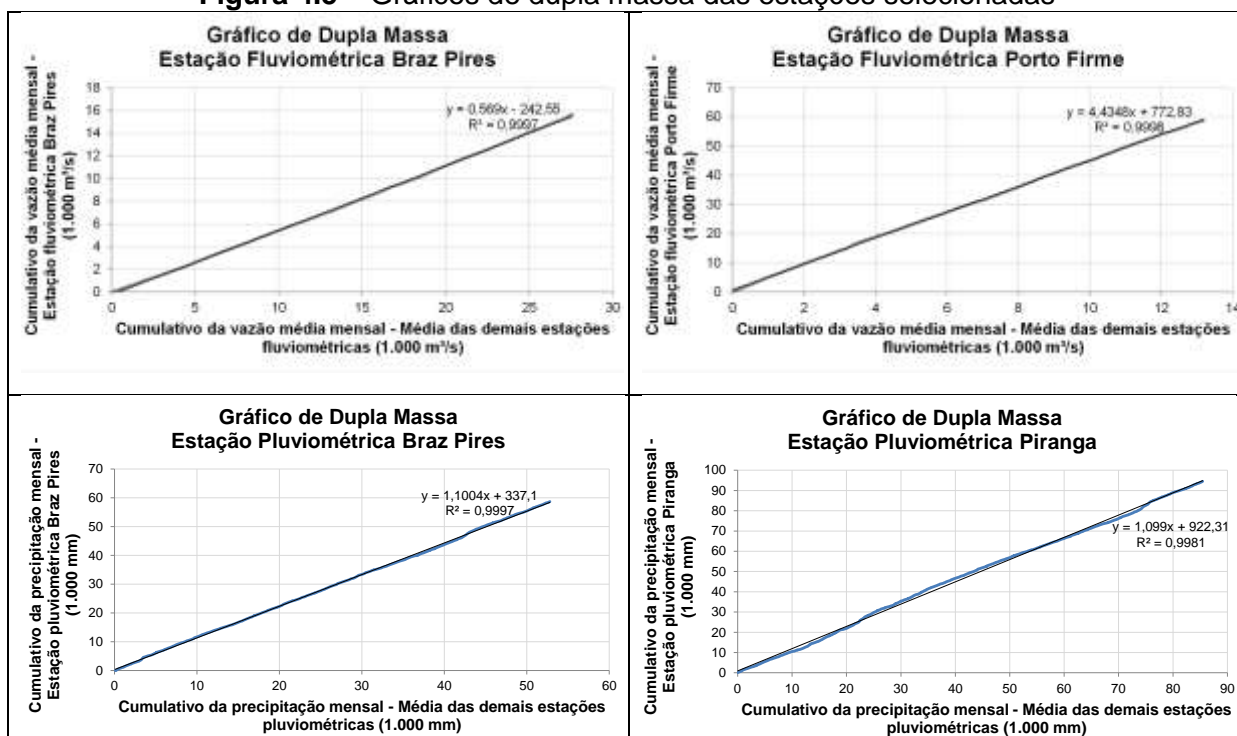
Com base nos níveis de correlação entre as séries monitoradas, foram definidas as séries a serem estudadas em cada um dos cenários definidos no item 4.1.2, a saber:

- Cenário 1 - Extensão de uma série de precipitações tomando como referência uma outra série de precipitações: estações pluviométricas Piranga (série curta) e Braz Pires (série de referência), com nível de correlação de 0,38, para a variável aleatória máxima diária anual, e 0,86, para a variável aleatória acumulada anual; e
- Cenário 2 - Extensão de uma série de vazões tomando como referência uma outra série de vazões: estações fluviométricas Porto Firme (série curta) e Braz Pires (série de referência), com nível de correlação, de 0,78, para a vazão média diária máxima anual, e 0,92, para a vazão média anual.

As séries monitoradas pelas estações listadas acima foram escolhidas por possuir amplitude razoável do nível de correlação (0,38 até 0,92). É importante frisar, contudo, que os valores do coeficiente de correlação apresentados, neste momento, foram calculados para as séries hidrológicas completas, e podem ser diferentes para o período de observações comuns utilizado para ajustar os modelos de extensão.

Com o intuito de verificar a consistência das séries hidrológicas selecionadas para a aplicação das técnicas, foram elaboradas as curvas de dupla massa das estações em base mensal. Os resultados são apresentados na Figura 4.5.

Figura 4.5 – Gráficos de dupla massa das estações selecionadas



Observa-se, na Figura 4.5, que, dentre as quatro estações selecionadas, apenas o gráfico da estação pluviométrica Piranga indica que a relação entre os registros desta estação e os demais registros pluviométricos da região apresentam comportamento distinto. Contudo, não foi observada uma tendência clara nessa divergência de comportamento. Com isso em vista, conclui-se que os registros destas estações estão aptos para a etapa de aplicação das técnicas de extensão de séries hidrológicas sobre dados reais.

4.6.3 Aplicação das técnicas de extensão de séries hidrológicas

Para a aplicação das técnicas sobre as séries hidrológicas selecionadas, foi adotado $n_1 = 10$ e n_2 definido a partir do número de observações restantes da série hidrológica de referência. O nível de correlação entre as séries curta e de referência foi definido pelas séries hidrológicas escolhidas.

De particular importância é a amostragem dos valores pertencentes à série curta. Foram adotados dois cenários para a aplicação das técnicas. O primeiro é aquele associado à série histórica registrada (na sequência cronológica). Foram considerados apenas os n_1 primeiros registros do período de observações comuns das séries históricas para ajustar as técnicas de extensão. Todos os registros restantes da estação de referência, que possuem observações comuns na estação que originou a série curta, foram utilizados para estimar os valores observados na série curta. Este procedimento foi adotado para os dois cenários já definidos.

De forma a não utilizar um único conjunto de valores para representar a série curta, também foram realizadas 5.000 amostragens *bootstrap* a partir de todos os valores registrados nas séries em estudo, em cada um dos dois cenários de aplicação sobre amostras reais. Em cada uma das simulações *bootstrap*, o coeficiente de correlação entre as séries curta e de referência apresentou um valor diferente. As técnicas utilizadas foram aquelas consideradas aplicáveis a cada cenário simulado, a partir do critério de aplicação definido no experimento sobre amostras sintéticas.

A Tabela 4.5 sintetiza os cenários estudados para as séries hidrológicas escolhidas.

Tabela 4.5 - Simulações na aplicação das técnicas de extensão sobre dados reais

	Cenário 1	Cenário 2
Estação de referência	Braz Pires (PLU)	Braz Pires (FLU)
Estação associada à série curta	Piranga (PLU)	Porto Firme (FLU)
Cenários de n_1 e n_2 – Dados diários	$n_1 = 10$ e $n_2 = 33$	$n_1 = 10$ e $n_2 = 45$
Cenários de n_1 e n_2 – Dados anuais	$n_1 = 10$ e $n_2 = 43$	$n_1 = 10$ e $n_2 = 48$
Número de simulações (amostragens <i>bootstrap</i>)	5.000	5.000

4.7 Avaliação dos resultados

A avaliação das técnicas de extensão baseou-se na comparação das estimativas das medidas populacionais por meio da série curta e das séries estendidas por meio das diversas técnicas estudadas. As principais medidas analisadas para cada cenário simulado foram: momentos amostrais (média, variância e coeficiente de assimetria), estimativa dos percentis, estimativa de cada registro individual e estimativa do quantil associado ao TR = 100 anos. Cabe ressaltar que, no cenário de dados seguindo a distribuição Pearson 3 com assimetria negativa, o quantil de referência foi tratado como evento mínimo. Nas demais distribuições, o quantil foi estimado considerando evento máximo. Os indicadores analisados para cada uma das medidas citadas foram:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{s}_i - s_i)^2} \quad (4.7)$$

$$VIÉS = \frac{1}{m} \sum_{i=1}^m (\hat{s}_i - s_i) \quad (4.8)$$

$$VARIÂNCIA = \frac{1}{m} \sum_{i=1}^m (s_i - \bar{s})^2 \quad (4.9)$$

em que

m são as 50.000 simulações no experimento de Monte Carlo; e

\hat{s}_i é a estimativa da medida descritiva populacional s_i .

Para quantificar os avanços com a extensão das séries hidrológicas, os indicadores *RMSE* (*Root Mean Square Error* ou Raiz do Erro Quadrático Médio) e *VARIÂNCIA* das séries estendidas foram apresentados por meio de sua relação com os mesmos indicadores para a série curta. Dessa forma, valores de tais razões inferiores a 1 indicaram avanços no aspecto

analisado (redução no erro na estimativa da variância populacional ou redução variância das estimativas do quantil de referência, por exemplo) em relação à série curta.

Na aplicação sobre os dados reais, diferente das simulações de Monte Carlo, não se dispõe dos valores populacionais para a avaliação das técnicas. Cabe lembrar que, neste caso, parte dos registros estudados foram desprezados, configurando uma série curta, com extensão de n_1 registros. Assim, foram comparados os valores das estimativas dos momentos amostrais, antes ($i = 1, 2, \dots, n_1$) e após a extensão das séries hidrológicas, com os momentos amostrais das séries completas ($i = 1, 2, \dots, n_1 + n_2$).

Apesar de, a rigor, a série completa não possibilitar a estimativa correta das medidas descritivas populacionais das variáveis aleatórias hidrológicas, foi possível, na aplicação sobre dados reais, analisar aspectos como a factibilidade dos quantis estimados a partir das séries estendidas e a influência do nível de correlação no desempenho das técnicas.

5 RESULTADOS E DISCUSSÃO

5.1 Comparação das técnicas de extensão de séries hidrológicas

Neste item, são apresentados os resultados das simulações de Monte Carlo empreendidas com enfoque na obtenção dos momentos amostrais, percentis, registros individuais e quantis das séries estendidas. Será analisada a influência da extensão da série curta, da extensão da série de referência, do nível de correlação linear e da distribuição à qual pertencem os dados nos resultados obtidos.

Cabe mencionar que o comportamento esperado é de o erro, desvio ou variância na estimativa da variável analisada ser menor, em magnitude, para maiores níveis de correlação e para valores maiores de n_1 . Essa tendência foi observada em quase todos os resultados, sendo dado destaque aos casos em que tal comportamento não foi observado.

Apresentam-se, em alguns momentos, os resultados para a série curta e/ou para a série completa. Em outros, é apresentado, como resultado, o indicador da razão entre o $RMSE/VARIÂNCIA$ encontrado para a técnica analisada e o $RMSE/VARIÂNCIA$ encontrado para a série curta. Nesse caso, um valor menor que 1 indica que a técnica analisada teve resultados positivos.

Grande importância também tem o fato de que nem a série curta nem a série completa registrada fornecem estimativas exatas dos momentos populacionais. Não se espera que os resultados indiquem que as técnicas de extensão tenham capacidade de gerar séries estendidas cujas estimativas de medidas descritivas populacionais sejam exatas ou possuam erro absoluto menor do que o erro das estimativas utilizando a série completa. Espera-se, contudo, que as técnicas produzam séries estendidas cujo desempenho seja intermediário entre a série curta e a série completa.

Antes de apresentar os resultados propriamente ditos, alguns aspectos observados durante as simulações cabem ser comentados. Um ponto importante, já ressaltado pelos autores de algumas das técnicas de extensão, diz respeito à capacidade do coeficiente de correlação amostral descrever a relação populacional entre as séries curta e longa, no período de observações comuns. Nas simulações empreendidas, o valor esperado do coeficiente de correlação ficou muito próximo dos valores populacionais, exceto nos cenários em que a

amostra era muito pequena ($n_1 = 10$), em que o erro relativo foi de 4%. Além disso, destaca-se a variabilidade desse indicador. Para $n_1 = 10$, a variância de r foi de 0,07; para $n_1 = 25$, foi de 0,025; e para $n_1 = 50$, foi de 0,012. Esses resultados não foram influenciados pela distribuição de probabilidades das amostras. Pode-se concluir, portanto, que, principalmente nas amostras com extensão reduzida, o coeficiente de correlação linear amostral pode não ser o melhor indicador da relação entre as variáveis estudadas.

Em todas as amostras geradas, de todos os cenários, o ajuste considerando pelo menos uma das funções de ligação utilizadas na técnica GLM obteve convergência. De modo geral, em cenários com $n_1 = 10$, houve equilíbrio entre as funções de ligação utilizadas, enquanto para $n_1 = 25$ e $n_1 = 50$, a função *probit* foi a mais escolhida (50% a 60% das amostras geradas), tendo como base o valor de AIC. A função complemento log-log foi a menos utilizada, sendo aplicada em menos de 10% dos pares de séries sintéticas geradas.

Foi observado que a função de ligação escolhida em cada caso, pelo critério definido, foi aquela que apresentou mais aderência às amostras geradas. Entretanto, em testes desenvolvidos independentemente da simulação de Monte Carlo final, constatou-se que o critério baseado no valor de AIC tendeu a direcionar a escolha da função de ligação para a função identidade, igualando o ajuste da técnica GLM ao ajuste da técnica OLS. Esse resultado também é válido para os resultados apresentados nesse item, com a exceção de que a função de ligação identidade não foi considerada para a aplicação da técnica GLM. Apesar desse efeito ter sido menor no cenário de dados com assimetria positiva, foi capaz de influenciar em todos os resultados correspondentes à técnica GLM.

Como ressaltado no item 4.1.1, foi gerado um volume muito grande de resultados, que foram apresentados e discutidos de forma breve, com enfoque na capacidade das técnicas de gerar séries estendidas com melhores estimativas das medidas descritivas populacionais do que a série curta, ou na própria comparação das técnicas. Nem todos os resultados gerados foram apresentados, mas os resultados que seguem são suficientes para extrair as conclusões mais relevantes sobre a extensão das séries hidrológicas.

5.1.1 Momentos amostrais

5.1.1.1 Estimativa da Média Populacional

De modo geral, a média é o momento populacional estimado com maior acurácia por meio das séries estendidas pelas diversas técnicas aqui analisadas. A Figura 5.1 apresenta o RMSE para a estimativa da média populacional para alguns dos cenários simulados, com dados da série curta e da série de referência pertencentes à distribuição normal.

A técnica MOVE2 destacou-se por ser aquela que retornou a menor redução no erro da estimativa da média com o aumento da correlação populacional. Além disso, em apenas poucos cenários, ela retorna médias com erros menores do que aqueles apresentados pela série curta.

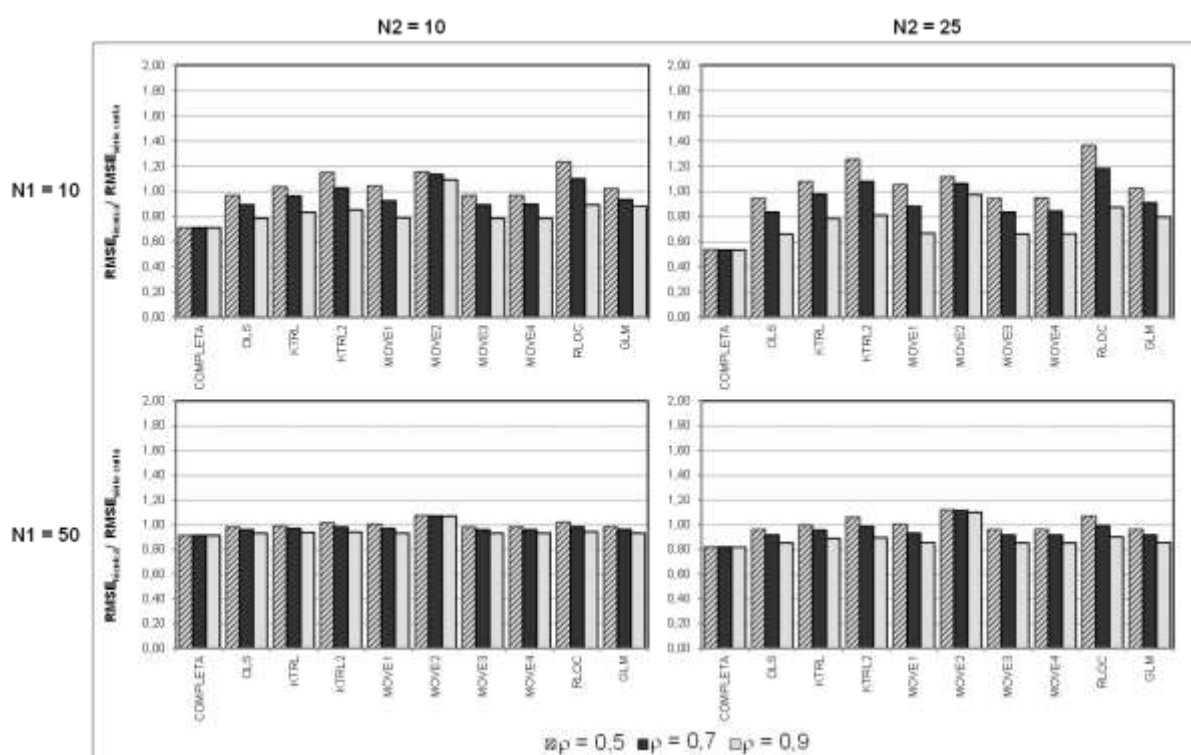


Figura 5.1 – RMSE na estimativa da média populacional – dados pertencentes à distribuição normal

Com o aumento do valor de n_1 , menores ficaram as diferenças entre as técnicas, uma vez que, nesse caso, a série curta possui um peso maior na série estendida. O aumento do valor de n_2 , contudo, tem efeito contrário, e as diferenças entre as técnicas ficam mais evidentes. Nota-se, por exemplo, que, para valores maiores de n_2 , existe uma amplitude maior entre os valores extremos de *RMSE* nos três cenários de correlação populacional simulados.

A técnica OLS, que pode ser considerada como uma referência, dada sua popularidade e difusão no meio científico, apresentou bom desempenho, de forma geral, com $RMSE_{técnica}/RMSE_{série\ curta}$ menor que 1 em todos os cenários simulados. Os destaques negativos foram, além da MOVE2, as técnicas KTRL2 e RLOC, porém com desempenho satisfatório para correlações elevadas.

A Figura 5.2 mostra alguns resultados do RMSE da estimativa da média para dados NOR, GUM, PE3+ e PE3-, considerando o cenário em que $n_1 = 10$ e $n_2 = 10$. A estimativa da média para algumas técnicas é melhor do que para a série curta independente da distribuição dos dados, mesmo com valores de correlação mais baixos ($\rho = 0,5$). O desempenho inferior da técnica MOVE2 também foi observado para dados com outras distribuições além da Normal.

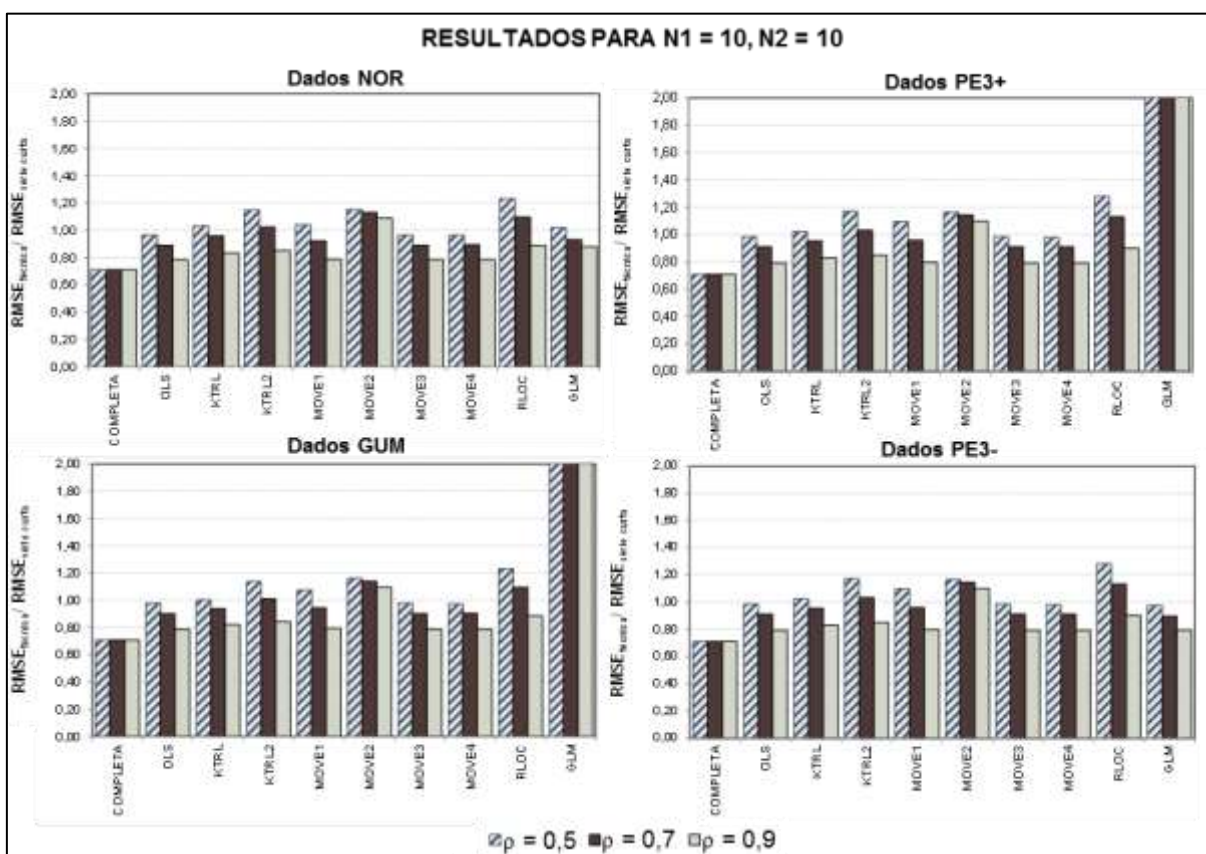


Figura 5.2 – RMSE na estimativa da média populacional – influência da densidade de probabilidade dos dados

Destaca-se, na Figura 5.2, que os dados distribuídos com assimetria positiva apresentaram erros elevados para a técnica GLM, o que provavelmente se deve, além da assimetria já mencionada, à extensão da série, curta, com 10 elementos. Nos cenários em que $n_1 = 25$ ou 50, o erro mostrado na Figura 5.2 foi substituído por um desempenho semelhante ao apresentado para os dados NOR e PE3-.

O viés das séries estendidas por meio da aplicação das técnicas analisadas é apresentado na Figura 5.3, para as quatro distribuições analisadas. Os dados NOR, de forma geral, não apresentaram viés, exceto a técnica GLM, no cenário em que $n_1 = 10$. Apesar de sutil, o viés positivo para esta técnica aumentou em magnitude para um aumento de n_2 . O viés é decorrente da precisão do ajuste do modelo para uma amostra com apenas 10 membros, uma vez que, nos demais cenários de n_1 , n_2 e ρ , todas as técnicas, inclusive a GLM, apresentaram-se não enviesadas.

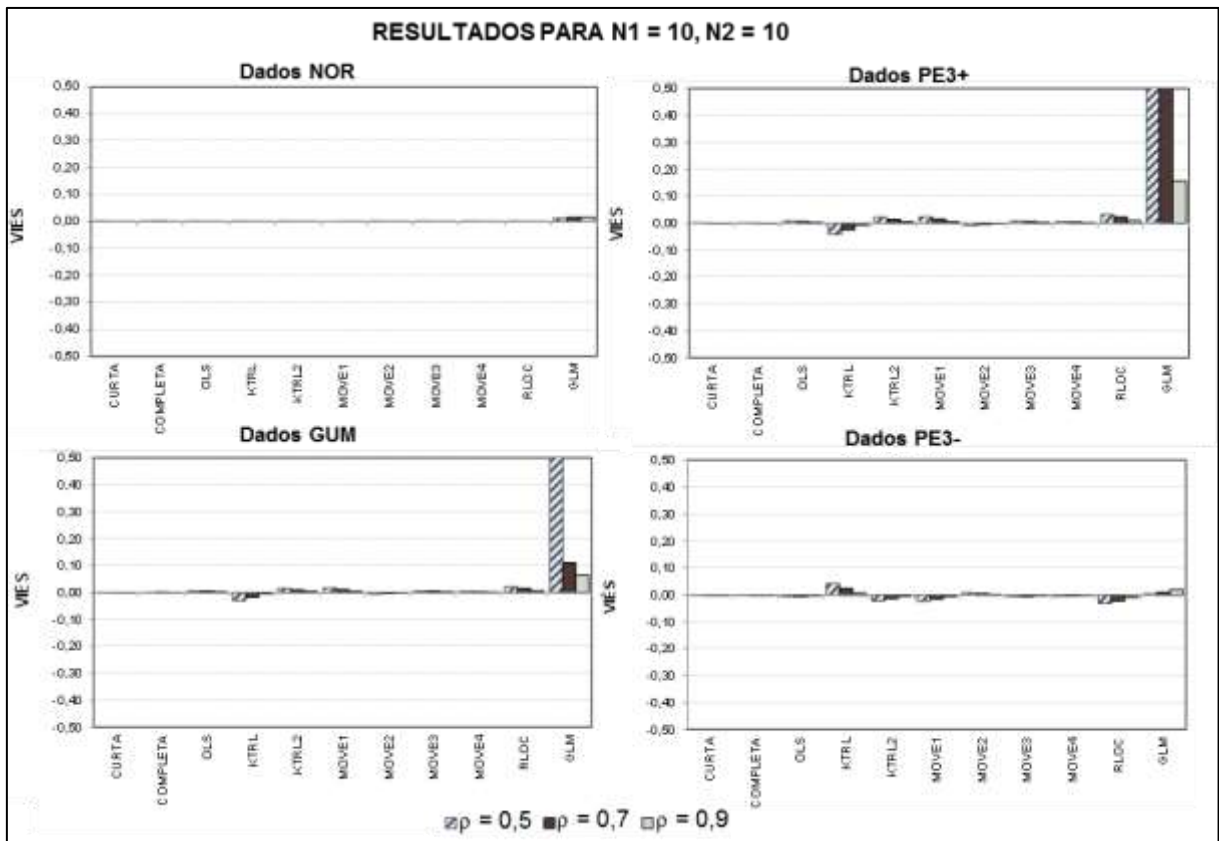


Figura 5.3 – Viés na estimativa da média populacional – influência da densidade de probabilidade dos dados

Para as distribuições com assimetria positiva (GUM e PE3+), as técnicas apresentaram resultados semelhantes, com um viés mais pronunciado da técnica GLM, devido à assimetria das amostras e à pequena extensão da série curta, de apenas 10 elementos. As demais técnicas mostraram-se não enviesadas para níveis altos de correlação, enquanto que, para níveis baixos, apenas OLS, MOVE2, MOVE3 e MOVE4 são técnicas não enviesadas. Com exceção da KTRL, todas as técnicas apresentaram viés no sentido positivo. Com a alteração do sinal do coeficiente de assimetria, os vieses encontrados também mudaram de sinal. As técnicas não enviesadas permanecem as mesmas, com a GLM retornando para valores de viés mais baixos.

A respeito da variância dos estimadores, o comportamento é muito próximo do comportamento do *RMSE*, com maior influência do nível de correlação. Em geral, para $\rho = 0,7$, a eficiência das séries estendidas, independente da técnica utilizada, é maior do que a eficiência do estimador da média tendo como base a série curta.

5.1.1.2 Estimativa da Variância Populacional

A estimativa da variância populacional é um dos pontos mais problemáticos na extensão de séries hidrológicas. A Figura 5.4 apresenta o erro na estimativa da variância para a distribuição Normal. Fica claro que algumas técnicas, no caso populacional, são incapazes de melhorar a estimativa da variância por meio da extensão da série curta.

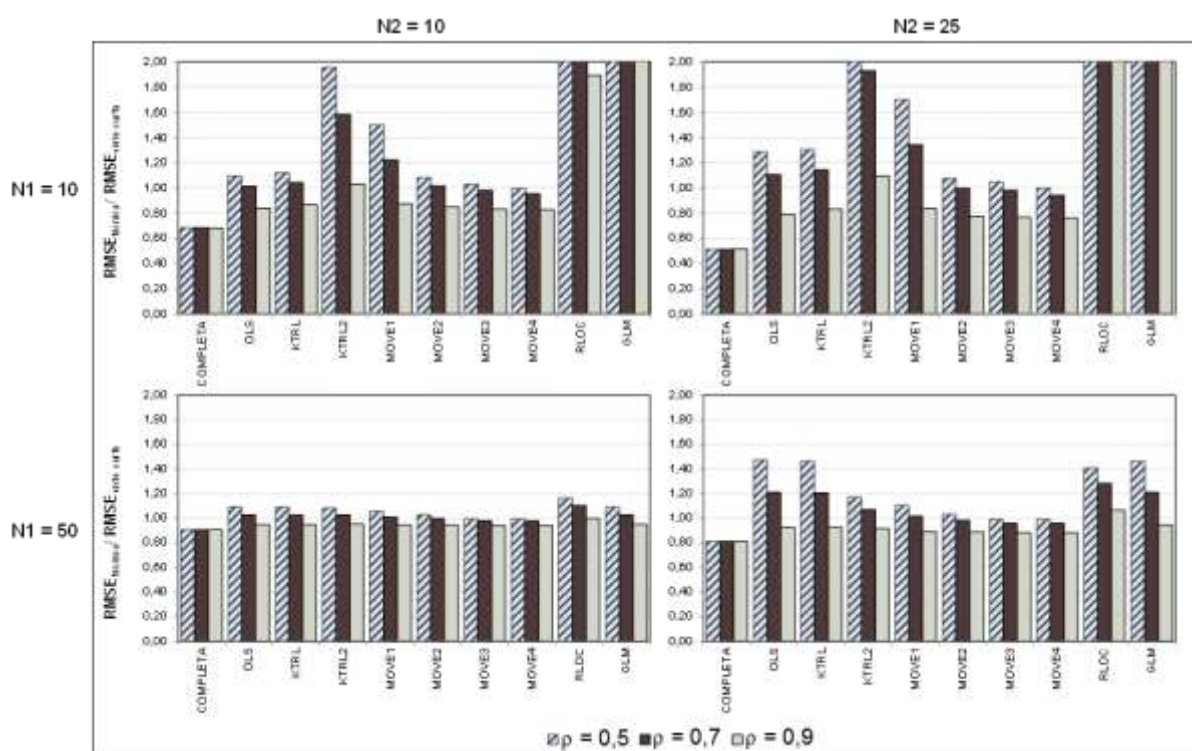


Figura 5.4 – RMSE na estimativa da variância populacional – dados pertencentes à distribuição Normal

Diferente do comportamento para a estimativa da média, as técnicas GLM e RLOC apresentam o maior erro na estimativa da variância. À medida que n_1 aumenta, o erro da estimativa da variância se reduz mais rapidamente para a GLM do que para a RLOC. De fato, dentre as técnicas analisadas, a RLOC é a que utiliza a menor quantidade de parâmetros e, portanto, é de se esperar que ela não seja capaz de reproduzir ou melhorar a estimativa da variância. Nos cenários analisados, a GLM foi capaz de melhorar a estimativa da variância quando n_1 é igual ou superior a 25 e $\rho = 0,9$.

As técnicas KTRL2 e MOVE1 tiveram desempenho ruim para séries curtas com extensão de 10 e 25. Nesses casos, um nível de correlação elevado (0,9) foi mais efetivo para compensar esse problema para a técnica MOVE1. Uma desvantagem da técnica KTRL2 é que, além da correlação entre a série curta e a série de referência, ele depende das estimativas dos percentis nas duas séries serem acuradas para retornar bons resultados. Dessa forma, n_1 torna-se uma variável da maior importância.

Uma das formas de interpretar os resultados apresentados na Figura 5.4 é avaliar para quais técnicas o aumento de n_2 , que representa uma presença mais forte das características das técnicas de extensão, prejudica mais a estimativa da variância populacional. Essa análise desqualifica as técnicas OLS, KTRL e GLM para a estimativa da variância. Nestas, contudo, elevados níveis de correlação reduzem os erros das estimativas. Destacam-se positivamente as técnicas MOVE2, MOVE3 e MOVE4, com os melhores resultados.

A Figura 5.5 apresenta o erro na estimativa da variância para as quatro distribuições analisadas, para o cenário em que $n_1 = 10$ e $n_2 = 25$. De modo geral, observa-se que a distribuição normal é a mais restritiva para as técnicas OLS e KTRL, especialmente para $\rho = 0,5$.

Para as distribuições PE3+ e PE3-, por exemplo, há uma redução do erro da estimativa da variância por meio da técnica KTRL mesmo para $\rho = 0,5$, o que configura esta técnica como uma das melhores para $n_1 = 10$ e dados PE3+ e PE3-. As técnicas OLS, MOVE2, MOVE3, MOVE4 e GLM (PE3- apenas) tem bom desempenho em ambas as distribuições Pearson 3. Dentre essas, as técnicas MOVE promovem um erro maior da estimativa da variância para valores baixos de correlação (0,5) e, de modo geral, para $\rho > 0,7$, é possível dizer que a série estendida leva a uma estimativa mais correta da variância populacional do que a série curta. As técnicas KTRL2, MOVE1, RLOC e GLM (PE3+ apenas) tiveram desempenho ruim, especialmente a técnica RLOC que, nem mesmo para correlações elevadas (0,9), apresentou melhoria na estimativa da variância em relação à série curta. Os resultados para os dados GUM são aproximadamente os mesmos para os dados PE3+.

Não foram observadas grandes diferenças entre o viés da estimativa da variância para as distribuições analisadas. A Figura 5.6 mostra o viés calculado para dados GUM. Cabe

mencionar que a variância populacional considerada foi igual a 1. Portanto, o viés apresentado equivale ao viés percentual na estimativa.

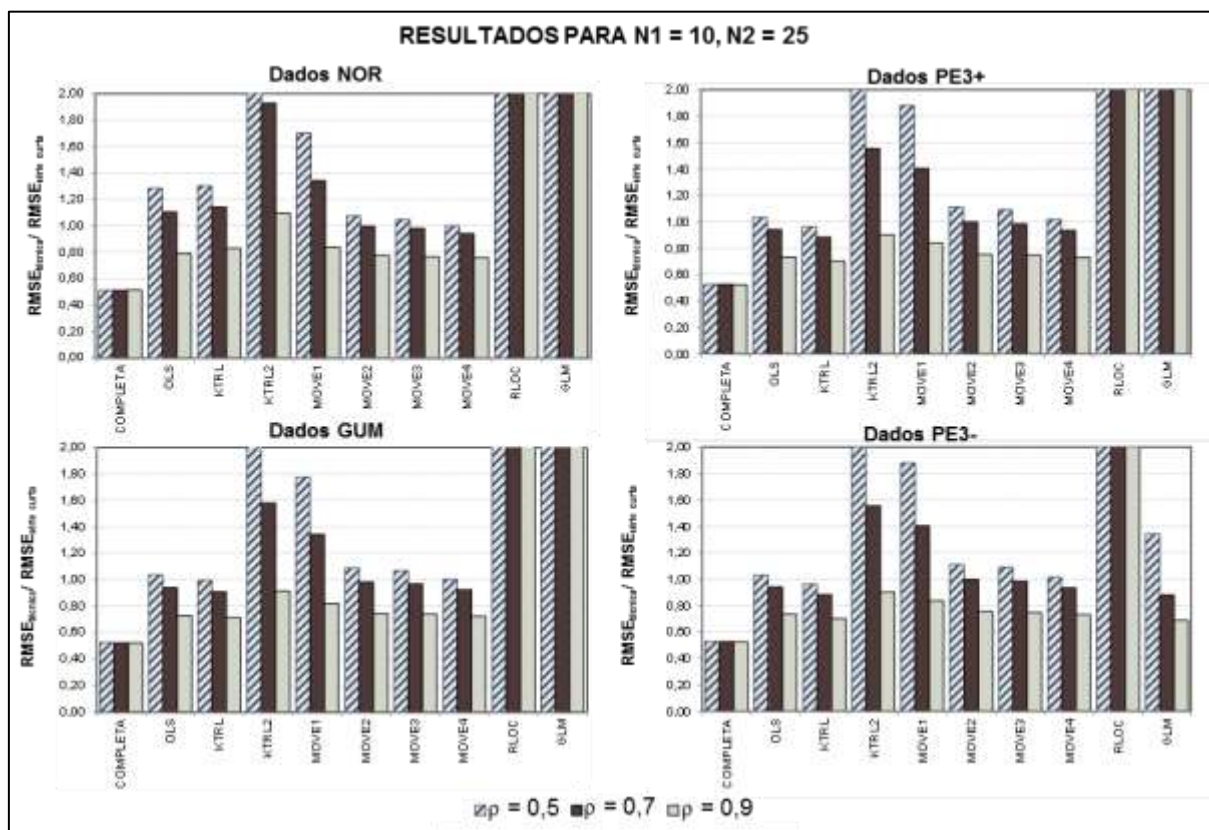
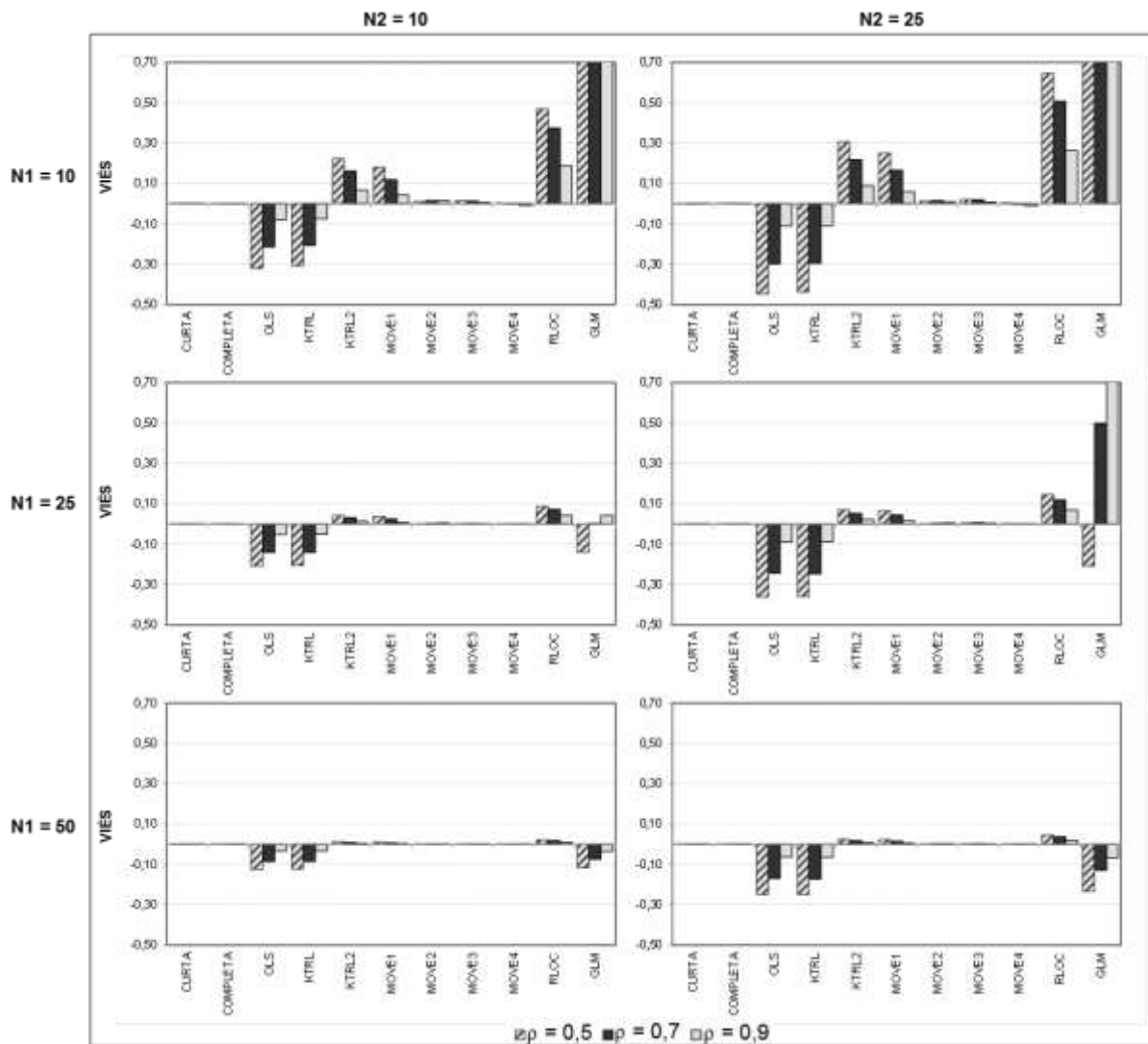


Figura 5.5 – RMSE na estimativa da variância populacional – influência da densidade de probabilidade dos dados

É evidente que as técnicas OLS e KTRL apresentam viés negativo na estimativa da variância. Para a técnica OLS, esse fato já é conhecido, uma vez que a estimativa dos parâmetros de regressão é puramente geométrica. No caso da KTRL, as vantagens decorrentes do uso da mediana nas estimativas não se traduziram na preservação da variabilidade da série curta.

Um outro grupo de técnicas, KTRL2, MOVE1 e RLOC, apresenta viés positivo na estimativa da variância. A técnica KTRL2 apresentou esse problema de forma muito mais significativa para $n_1 = 10$ e para valores maiores de n_2 do que nos demais cenários. É interessante avaliar as diferenças entre os resultados para as técnicas KTRL2 e KTRL. A técnica KTRL2 possui estimadores baseados em 19 pares de percentis, definidos a partir da série curta, e a técnica KTRL é ajustada sem a definição dos percentis, a partir do número exato de pares de dados da série curta. Para $n_1 = 10$, tem-se uma amostra muito pequena para ajustar os modelos, com resultados ruins para as duas técnicas. Contudo, dado um nível de correlação, o viés encontrado para a técnica KTRL2 é inferior, em magnitude, ao viés encontrado para a técnica

KTRL. A diferença aumenta à medida que n_1 também aumenta, com a técnica KTRL2 praticamente perdendo o viés para $n_1 = 50$. É possível concluir, portanto, que os percentis estimados a partir da série curta são mais representativos, do ponto de vista da estimativa da variância, do que a própria série curta.



Como os parâmetros da técnica RLOC são estimados de forma semelhante à técnica KTRL2, porém com simplificações, a RLOC também retorna viés positivo, sempre maior em magnitude do que a KTRL2. A princípio, poderia ser considerada uma surpresa a técnica MOVE1 apresentar o viés positivo para valores baixos de n_1 . Contudo, como apresentado por Hirsch (1982) e reproduzido na presente pesquisa, a MOVE1 foi considerada um avanço por produzir vieses menores em módulo do que a técnica OLS, apesar dela produzir séries com variância enviesada positivamente. É importante mencionar também que Hirsch (1982)

estimou a relação s_Y^2/σ_Y^2 para diversos cenários de n_1 , n_2 e ρ , obtendo os resultados apresentados na Tabela 3.6, que concordam com aqueles apresentados na Figura 5.6, correspondentes a dados GUM.

A técnica GLM apresenta resultados que não seguem um padrão bem definido, com um comportamento consistente, frente aos níveis de correlação, apenas para $n_1 = 50$. Esse comportamento é atribuído a fatores como a diversidade de funções de ligação consideradas, ao critério para escolha do melhor ajuste e à própria capacidade da técnica de modelar dados com maior flexibilidade, especialmente para $n_1 = 10$ e $n_1 = 25$.

A título de informação, a Figura 5.7 apresenta o viés da estimativa da variância para diversas distribuições. A diferença entre os resultados é muito pequena, manifestando-se em maior grau na técnica GLM devido aos fatores já mencionados. As técnicas com menor viés na estimativa da variância foram a MOVE2, MOVE3 e MOVE4, com bom desempenho independente de n_1 e ρ .

As técnicas mais eficientes na estimativa da variância foram OLS e KTRL, apesar de serem técnicas enviesadas. A técnica GLM teve variância próxima à variância da OLS e KTRL para dados com assimetria negativa. Dentre as técnicas não enviesadas (MOVE2, MOVE3 e MOVE4), a eficiência para correlação de 0,5 é igual ou menor à eficiência da série curta, sendo maior para correlação de 0,7 e 0,9. Para correlação alta, de 0,9, todas as técnicas, com exceção da RLOC e GLM (nos cenários em que os dados possuem assimetria positiva), geraram séries estendidas mais eficientes, do ponto de vista da variância, em relação à série curta. A Figura 5.8 sintetiza os resultados apresentados, mostrando a eficiência na estimativa da variância para as quatro distribuições analisadas.

5.1.1.3 Estimativa do Coeficiente de Assimetria Populacional

Assim como a estimativa dos quantis, séries hidrológicas curtas apresentam problemas na estimativa do coeficiente de assimetria populacional. Para dados normais, não foi observado viés para a série curta, completa ou estendida por nenhuma das técnicas, com exceção da técnica GLM. O viés, no sentido de superestimar o valor do coeficiente de assimetria, não foi influenciado pelo coeficiente de correlação populacional, e pôde ser considerado como inexistente para o cenário em que $n_1 = 25$ e $n_2 = 10$ e também para todos os cenários nos quais $n_1 = 50$. Nos demais casos, o viés chegou a ser de +0,20.

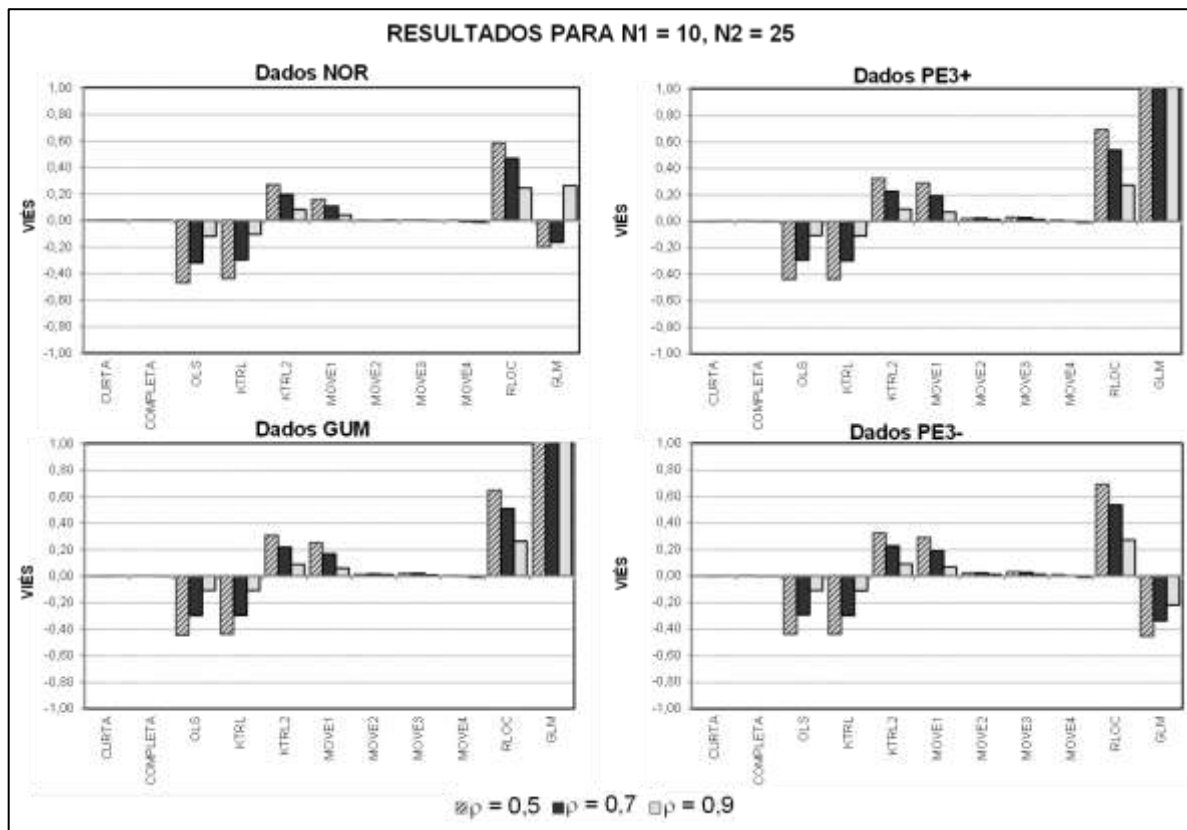


Figura 5.7 – Viés na estimativa da variância populacional – influência da distribuição de probabilidades dos dados

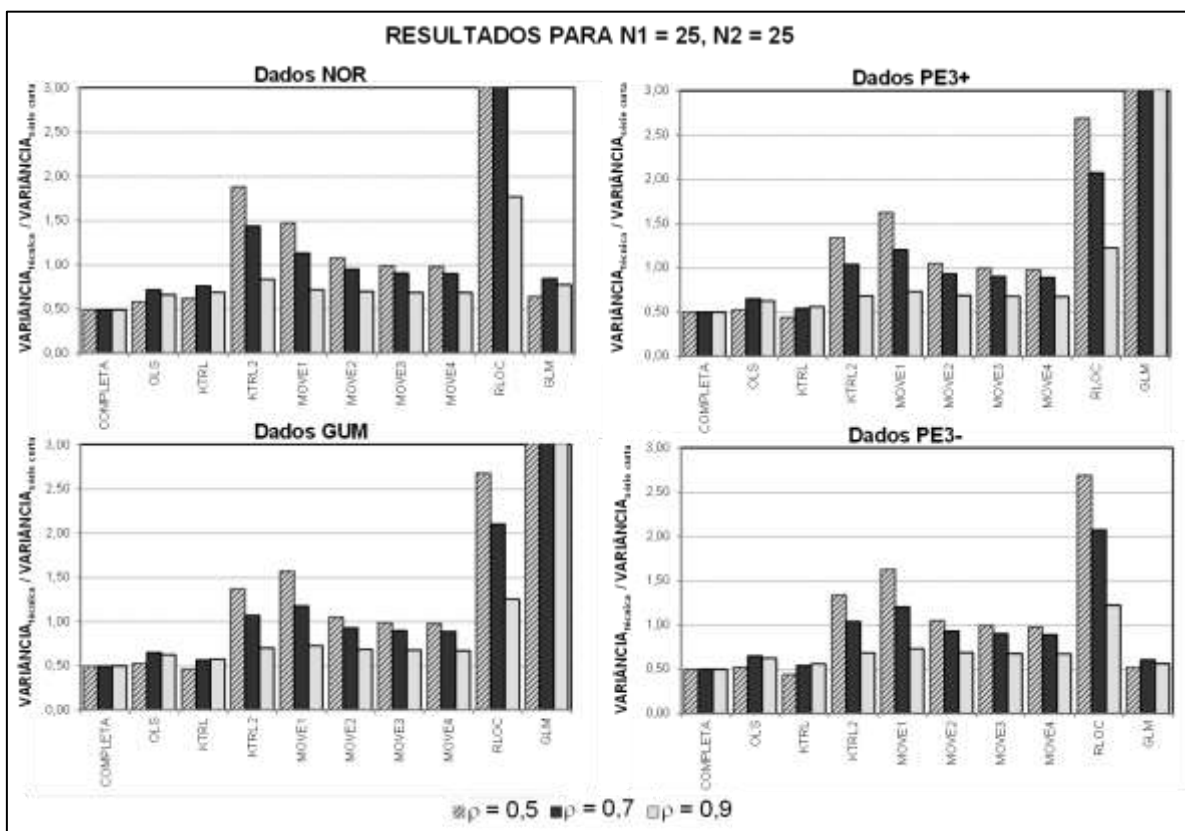


Figura 5.8 – Variância na estimativa da variância populacional

O coeficiente de correlação apresentou pouca influência no erro da estimativa do coeficiente de assimetria no cenário de dados normais. Para as técnicas OLS, KTRL, KTRL2, RLOC e GLM, maiores coeficientes de correlação implicaram menores erros na estimativa. As técnicas MOVE tiveram desempenho independente de ρ . As técnicas MOVE e KTRL2, na verdade, tiveram o melhor desempenho entre as técnicas, com redução do erro da estimativa da assimetria em todos os cenários analisados. Em seguida, destaca-se positivamente a técnica RLOC. A técnica GLM, por outro lado, não apresentou melhoria em qualquer dos cenários. A Figura 5.9 apresenta o erro na estimativa do coeficiente de assimetria populacional para dois cenários extremos (maiores e menores valores de n_1 e n_2).

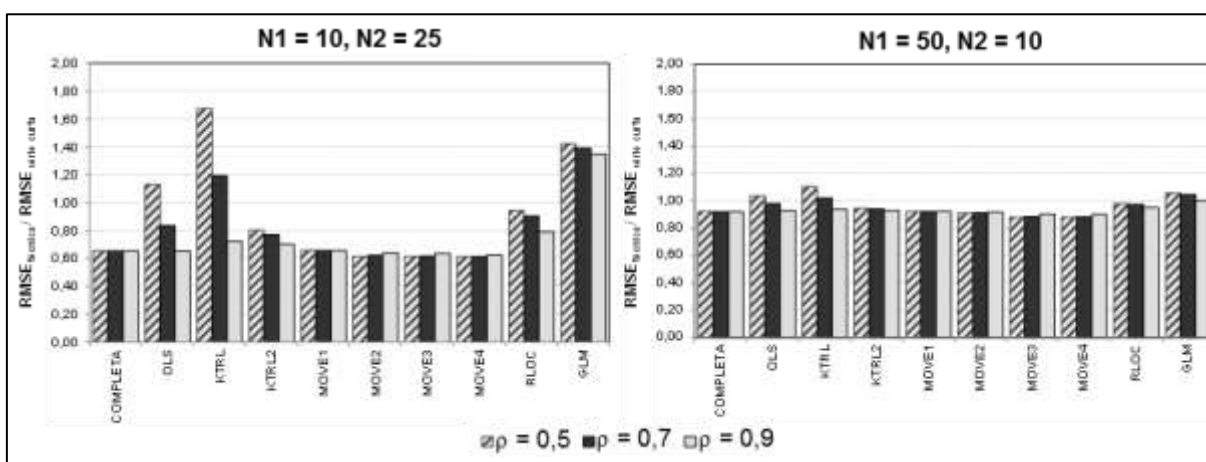


Figura 5.9 – RMSE na estimativa do coeficiente de assimetria populacional – resultados para dados NOR

Atribuem-se como causas para os resultados da Figura 5.9 o fato de que as técnicas OLS, KTRL e GLM podem ser consideradas modelos de regressão. As demais técnicas foram desenvolvidas com o intuito de obter melhorias em relação à estimativa da média e da variância da série estendida e a dedicação à estimativa dessas grandezas acaba por se traduzir em uma melhoria das medidas descritivas de forma geral. Também deve ser mencionado o fato de que o coeficiente de assimetria é um momento para o qual a extensão da série influencia fortemente na qualidade da estimativa. Com relação à variância da estimativa do coeficiente de assimetria, repete-se o aspecto dos resultados do indicador RMSE.

Os dados normais compreendem a condição para a qual foi desenvolvido o gerador das amostras sintéticas utilizado nesta pesquisa, e, portanto, as técnicas tiveram mais facilidade em convergir para o coeficiente de assimetria populacional, sem viés na estimativa utilizando séries estendidas por meio de nenhuma técnica. Para os dados com assimetria positiva (dados GUM e PE3+), todas as técnicas e as próprias séries curta e completa apresentaram viés

negativo, muito semelhantes para os dois cenários de distribuição dos dados, apesar da assimetria populacional variar entre as distribuições (para a distribuição Gumbel, $\gamma = +1,14$, e para a distribuição Pearson 3 foi adotado $\gamma = +1,40$). A Figura 5.10 apresenta os resultados para as distribuições Gumbel, PE3+ e PE3-.

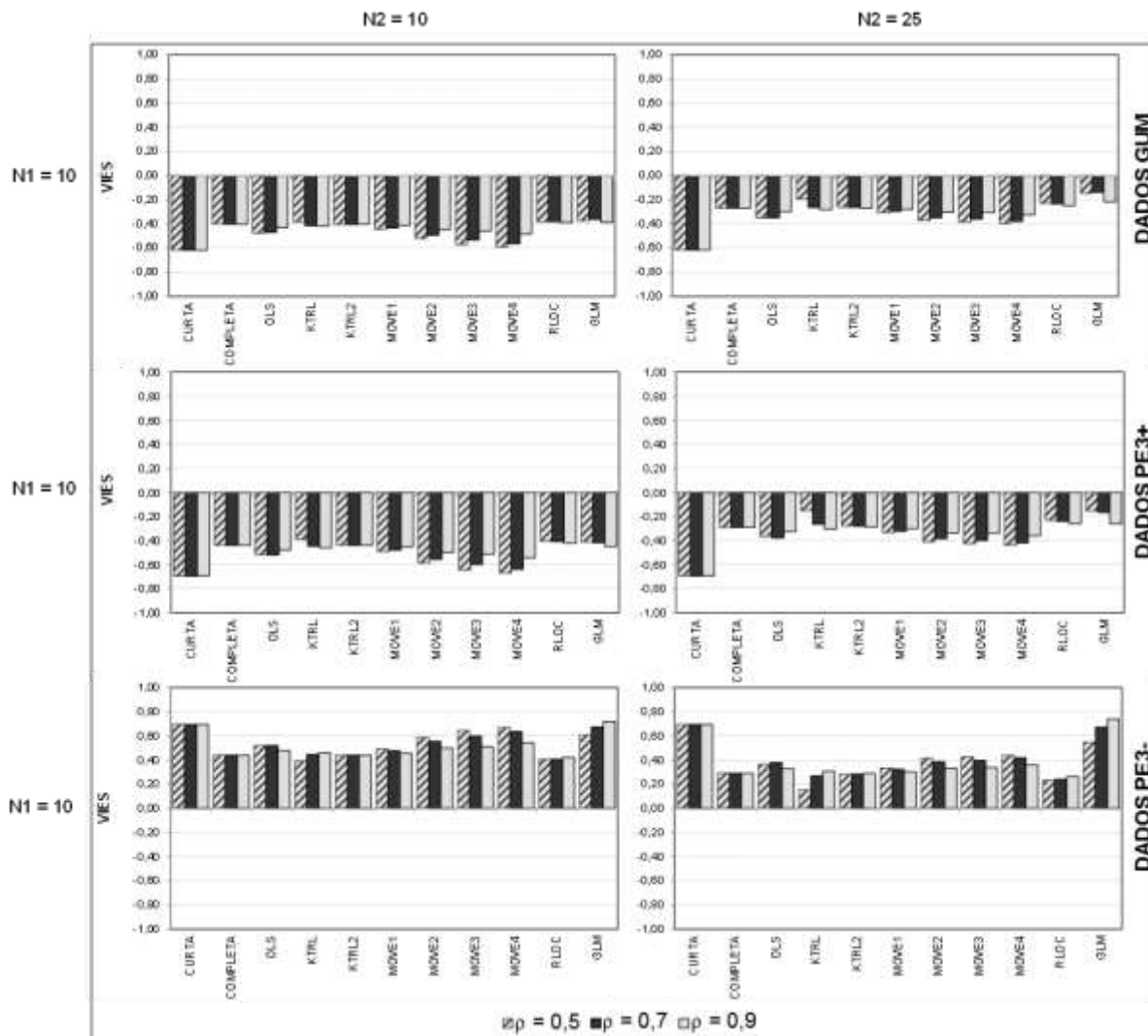


Figura 5.10 – Viés na estimativa do coeficiente de assimetria populacional – resultados para dados GUM, PE3+ e PE3-

Para dados distribuídos conforme a PE3-, foi observado, na Figura 5.10, viés no sentido oposto aos dados PE3+, com mesma magnitude, de modo geral. Infere-se, portanto, que as técnicas geram séries estendidas com coeficiente de assimetria enviesado, tendendo para zero.

No que tange ao erro da estimativa de γ , as conclusões são, de modo geral, iguais às aquelas apresentadas para dados normais. O comportamento das técnicas para dados Gumbel e PE3+ foi o mesmo, tendo-se como melhores técnicas MOVE e KTRL2. Outras técnicas também tiveram bom desempenho (redução no erro da estimativa da assimetria), com exceção da

técnica GLM (em todos os cenários) e KTRL para correlação populacional de 0,5. Para dados pertencentes à distribuição PE3-, os erros foram menores, apesar de as conclusões serem também as mesmas para as demais distribuições.

Apenas para os dados normais houve técnicas não enviesadas para a estimativa do coeficiente de assimetria. Portanto, apresenta-se, na Figura 5.11, apenas a variância das estimativas para dois cenários extremos (maiores e menores valores de n_1 e n_2) da distribuição normal.

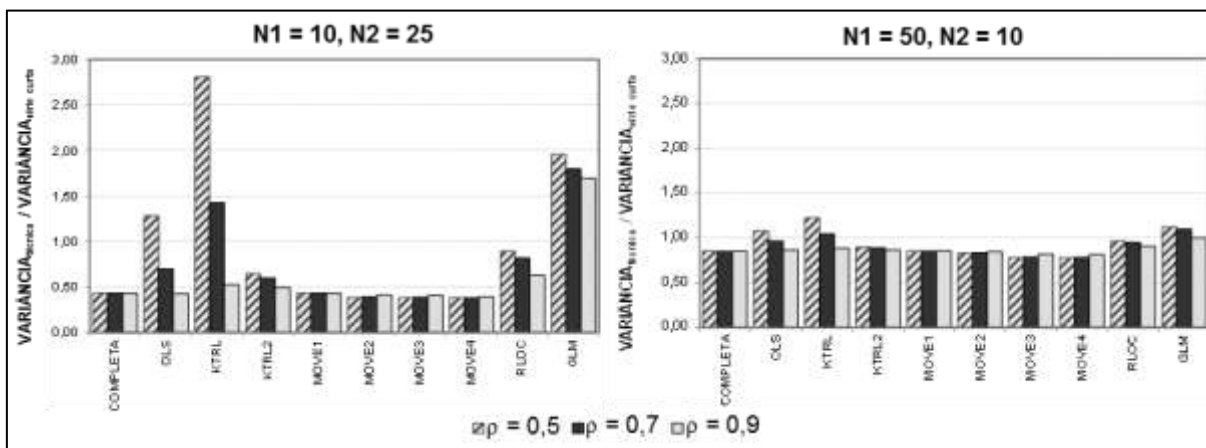


Figura 5.11 – Variância na estimativa do coeficiente de assimetria populacional – resultados para dados NOR

Como mostra a Figura 5.11, em apenas poucos cenários, nas técnicas OLS, KTRL e GLM, as séries estendidas possuem variância maior do que as estimativas da série curta. As técnicas MOVE3 e MOVE4 apresentam o melhor desempenho, muito próximo da variância da série completa, independente do nível de correlação populacional.

5.1.2 Percentis

Foram calculados o erro e o viés na estimativa dos percentis de ordem 5 a 95. Ressalta-se que foram obtidos resultados diferentes para as quatro distribuições analisadas, com principal diferença ocorrendo em função da assimetria dos dados. O comportamento esperado, de maneira geral, consiste em aumentar erros e vieses na estimativa dos percentis extremos (mínimos e máximos).

Para os dados normais, cujos resultados são apresentados na Figura 5.12, foram observados dois comportamentos distintos com relação ao erro da estimativa dos percentis. O primeiro caracteriza-se por um valor aproximadamente constante independente da ordem do percentil analisado. Nesse grupo, enquadram-se as técnicas MOVE e a técnica KTRL2. Destas, as

técnicas MOVE3 e MOVE4 tiveram o melhor desempenho, com redução no erro das estimativas de todos os percentis em todos os cenários de n_1 , n_2 e ρ . Essa constatação deve-se ao fato de essas técnicas considerarem que a combinação da série monitorada, com os valores estimados, fornece os estimadores da média e da variância populacional estimados por Matalas e Jacobs (1964) – técnica MOVE3 – ou aqueles adaptados por Vogel e Stedinger (1985) – técnica MOVE4. Esperava-se que a técnica KTRL2 tivesse desempenho melhor, dada a formulação do seu ajuste, porém apenas propiciou redução do erro das estimativas dos percentis para valores elevados de n_1 (50) e de ρ (0,9). O aumento da extensão da série curta possui implicações óbvias na estimativa dos percentis da série curta, enquanto o nível de correlação elevado propicia que informação relevante da série de referência seja utilizada corretamente na estimativa dos percentis. As técnicas MOVE1 e MOVE2 tiveram desempenho melhor do que a série curta apenas para $\rho = 0,9$. Contudo, a técnica MOVE1 apresentou erro menor na estimativa dos percentis.

O segundo comportamento é caracterizado por erros maiores para percentis extremos (máximo e mínimo). É o caso das técnicas OLS, GLM, KTRL e RLOC. Considerando o caso da técnica OLS, cujo ajuste pode ser considerado geométrico, é de se esperar que houvesse desvios maiores para os percentis extremos, dada a própria função densidade de probabilidade dos dados normais, na qual concentram-se valores maiores nas proximidades da média. Para essas técnicas, esse comportamento foi observado em todos os cenários, porém em menor grau para valores maiores de extensão da série curta e de correlação.

A estimativa dos percentis antes do ajuste da KTRL2 reduziu significativamente os erros para a estimativa dos percentis extremos, em comparação com a técnica KTRL. A técnica RLOC é penalizada por considerar apenas quatro valores em seu ajuste, tendo o pior desempenho dentre todas as técnicas quando são avaliados os percentis. A Figura 5.12 apresenta uma síntese dos resultados obtidos, de forma gráfica, incluindo a estimativa do RMSE e do viés na estimativa de cada percentil, considerando dados normalmente distribuídos.

Cabe mencionar que a técnica GLM, que teve desempenho inferior a técnicas como OLS e RLOC com relação à estimativa dos momentos amostrais, para $n_1 = 10$, teve na estimativa dos percentis desempenho similar a tais técnicas. Maiores erros foram observados apenas para percentis de ordens elevadas (90/95).

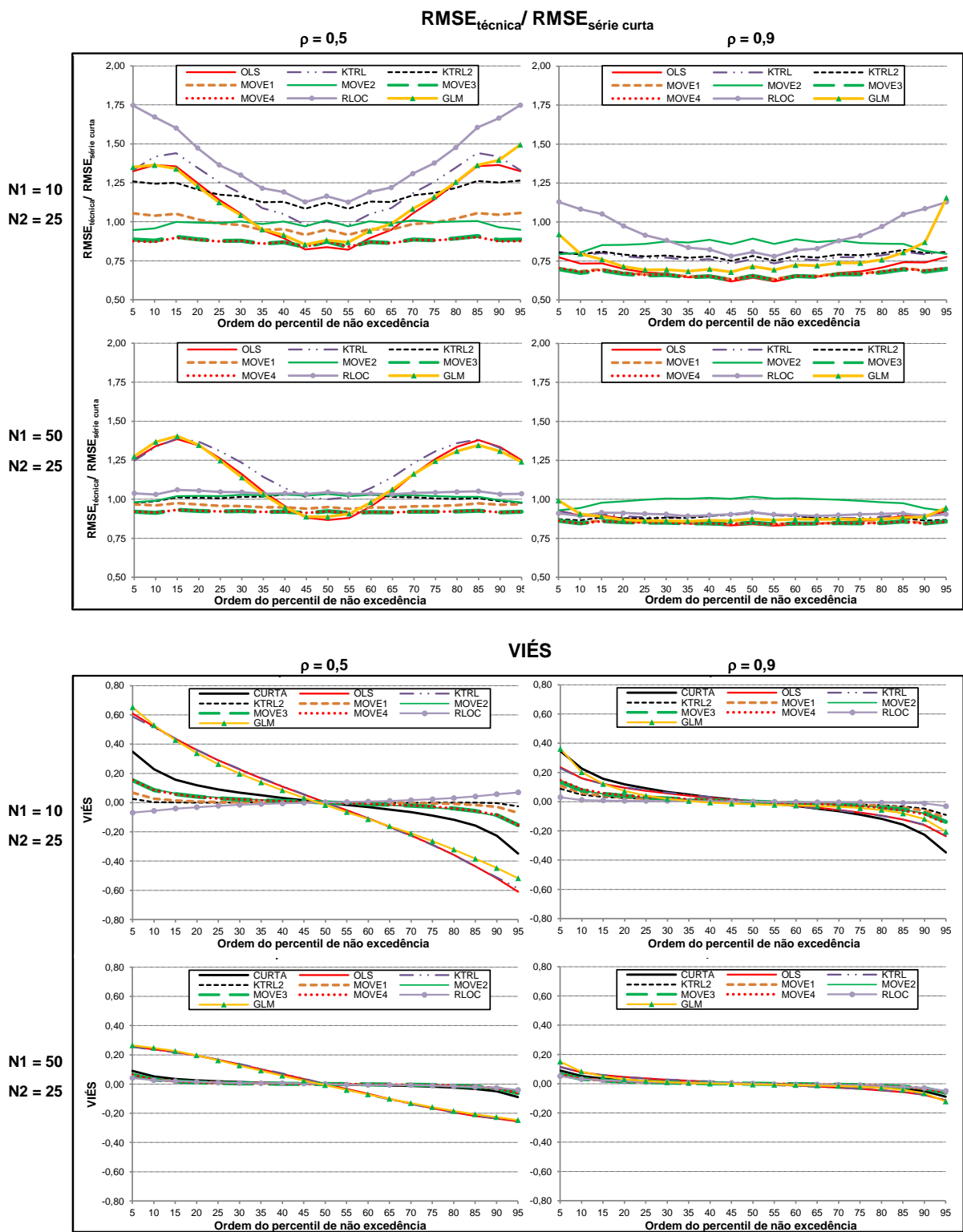


Figura 5.12 – RMSE e viés na estimativa dos percentis – resultados para dados NOR

O viés apresentado pela série curta na estimativa dos percentis é de superestimar os percentis mais baixos e subestimar os percentis mais altos. Essa tendência foi observada também para as técnicas de extensão, com exceção da técnica RLOC. Como já mencionado, o comportamento diferenciado da técnica RLOC é fruto da utilização de um número pequeno de

dados em relação às demais técnicas, especialmente para $n_1 = 10$, quando a estimativa dos percentis é a mais imprecisa entre os cenários analisados.

Como mostra a Figura 5.12, para coeficientes de correlação elevados, todas as técnicas são menos enviesadas do que a série curta ou apresentam viés baixo. Já para valores baixos de correlação, as técnicas GLM, OLS e KTRL são enviesadas, além da já mencionada RLOC. A técnica KTRL2 teve bom desempenho em todos os cenários simulados, com viés próximo de zero.

A Figura 5.13 mostra alguns dos resultados obtidos no cenário de dados GUM. Apesar de, em geral, as conclusões serem as mesmas daquelas para dados normais, os erros geraram gráficos assimétricos, como esperado. As técnicas originalmente conhecidas como de regressão (OLS, GLM e KTRL) foram as mais influenciadas, enquanto as demais técnicas, desenvolvidas no contexto da extensão de séries hidrológicas, com exceção da RLOC, apresentaram uniformes.

Os resultados do viés da estimativa dos percentis indicam que coeficiente de assimetria da distribuição Gumbel é pequeno o suficiente para que não haja diferenças substanciais com relação aos dados normais. Novamente, nota-se um deslocamento no aspecto dos resultados gráficos, com pequeno aumento do valor do percentil para o qual obteve-se viés nulo. Como se trata de distribuição amplamente utilizada na representação de máximos, é importante ressaltar que as técnicas MOVE2, MOVE3 e MOVE4 apresentaram viés no mesmo sentido da série curta, porém com menor magnitude, o que as aponta como melhores técnicas na construção de amostras que, como um todo, permitem boas estimativas de percentis de máximos.

A influência do coeficiente de assimetria no desempenho das técnicas pode ser melhor avaliada por meio dos resultados para os dados PE3+ e PE3- e da comparação com os dados normais.

O comportamento das técnicas para os dados PE3+ e PE3- foi aquele esperado, de simetria entre os resultados gráficos. Os resultados para a distribuição normal podem ser considerados um caso intermediário. A Figura 5.14 mostra os resultados do erro da estimativa dos percentis para as três distribuições e para dois cenários, em que $n_2 = 25$, $\rho = 0,5$ e $n_1 = 10$ ou 50.

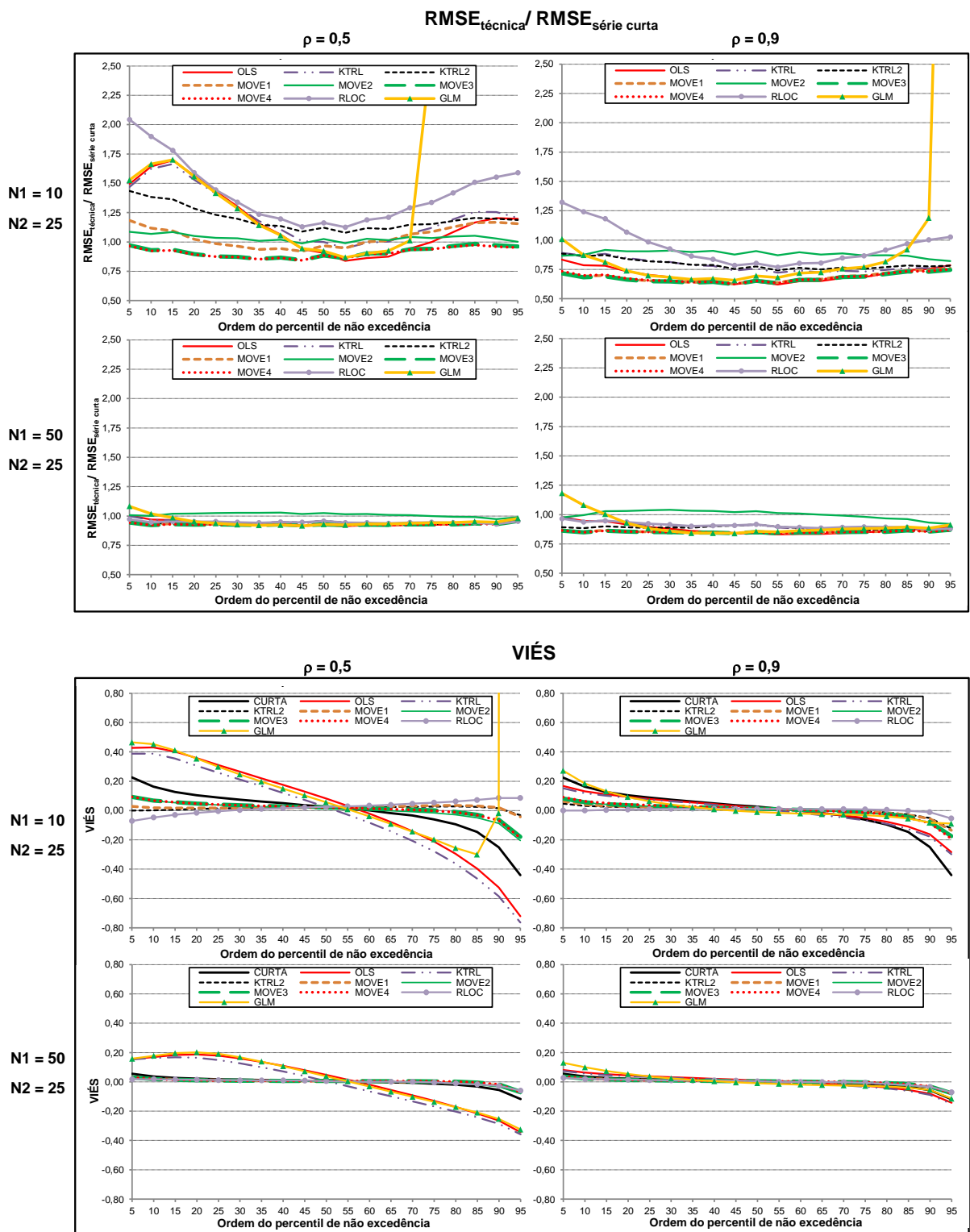


Figura 5.13 – RMSE e viés na estimativa dos percentis – resultados para dados GUM

Para as técnicas OLS e KTRL, ao contrário do que se esperava, os maiores erros na estimativa dos percentis para os dados PE3+ e PE3- não estão nos extremos. A técnica GLM apresentou a mesma tendência, porém, devido a seus problemas de ajuste para dados PE3+, quando $n_1 = 10$, apresentou erros muito superiores às demais técnicas para percentis superiores. O aumento

da extensão da série curta não representou ganho na redução do erro nas estimativas para as técnicas OLS, KTRL e GLM, como se nota para os casos em que $n_1 = 50$.

Para as demais técnicas, incluindo a RLOC, que não teve bom desempenho para dados NOR, o comportamento foi significativamente diferente. Não houve redução do erro nas estimativas dos percentis, mesmo para $n_1 = 50$, com aumento do erro para $n_1 = 10$. Isto vale mesmo para a técnica KTRL2, que foi capaz de, no máximo, representar a função densidade da série curta sem maiores refinamentos.

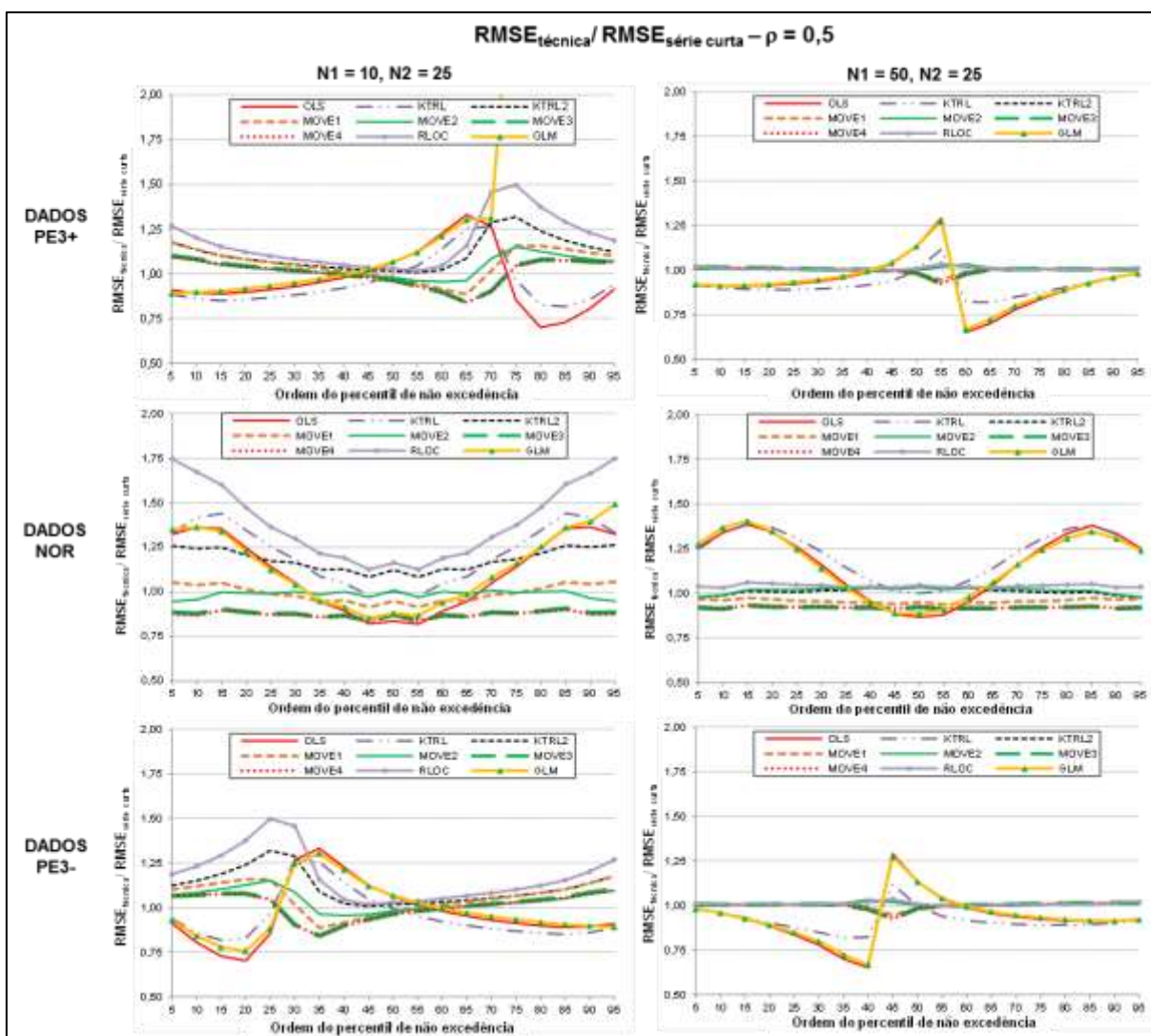


Figura 5.14 – RMSE na estimativa dos percentis – influência da assimetria dos dados

De forma geral, para todas as distribuições, o aumento do nível de correlação aproxima significativamente o desempenho das técnicas.

A Figura 5.15 (atentar para a escala das ordenadas) apresenta o viés das técnicas na estimativa dos percentis para dados PE3+, NOR e PE3-, também para os cenários em que $n_2 = 25$, $\rho = 0,5$ e $n_1 = 10$ e 50. Ao analisar a técnica MOVE1, por exemplo, que apresenta o mesmo ajuste entre a série curta e a série de referência, nota-se que o aspecto visual do viés em função dos percentis é crescente ou decrescente, para as distribuições PE3+ e PE3-, enquanto para a distribuição NOR o aspecto é aproximadamente horizontal. Destaca-se, também, a magnitude do viés, muito mais elevado nas distribuições PE3+ e PE3-. Os resultados para a série curta indicam que as técnicas (com exceção da OLS, GLM e KTRL) apenas extrapolam o comportamento dos dados originais. Com exceção dos dados normais, o aumento de n_1 não elimina o viés das estimativas dos percentis. Apesar de não apresentado, o aumento do nível de correlação não altera significativamente os vieses já apresentados na Figura 5.15, mas aproxima o desempenho das técnicas.

5.1.3 Registros individuais

Para a estimativa dos registros individuais, não foi observada influência da extensão da série curta, da série estendida ou do coeficiente de correlação nos resultados. Essa conclusão é esperada, uma vez que as flutuações ao redor das retas ajustadas das técnicas são aleatórias e independentes dos fatores analisados. A distribuição dos dados originais também não influenciou no erro na estimativa dos registros, porém influenciou no viés das estimativas. Esse resultado é mostrado nas Figuras 5.16 e 5.17 e justifica-se pelo fato de que o coeficiente de assimetria das distribuições é capaz de influenciar nos ajustes das técnicas.

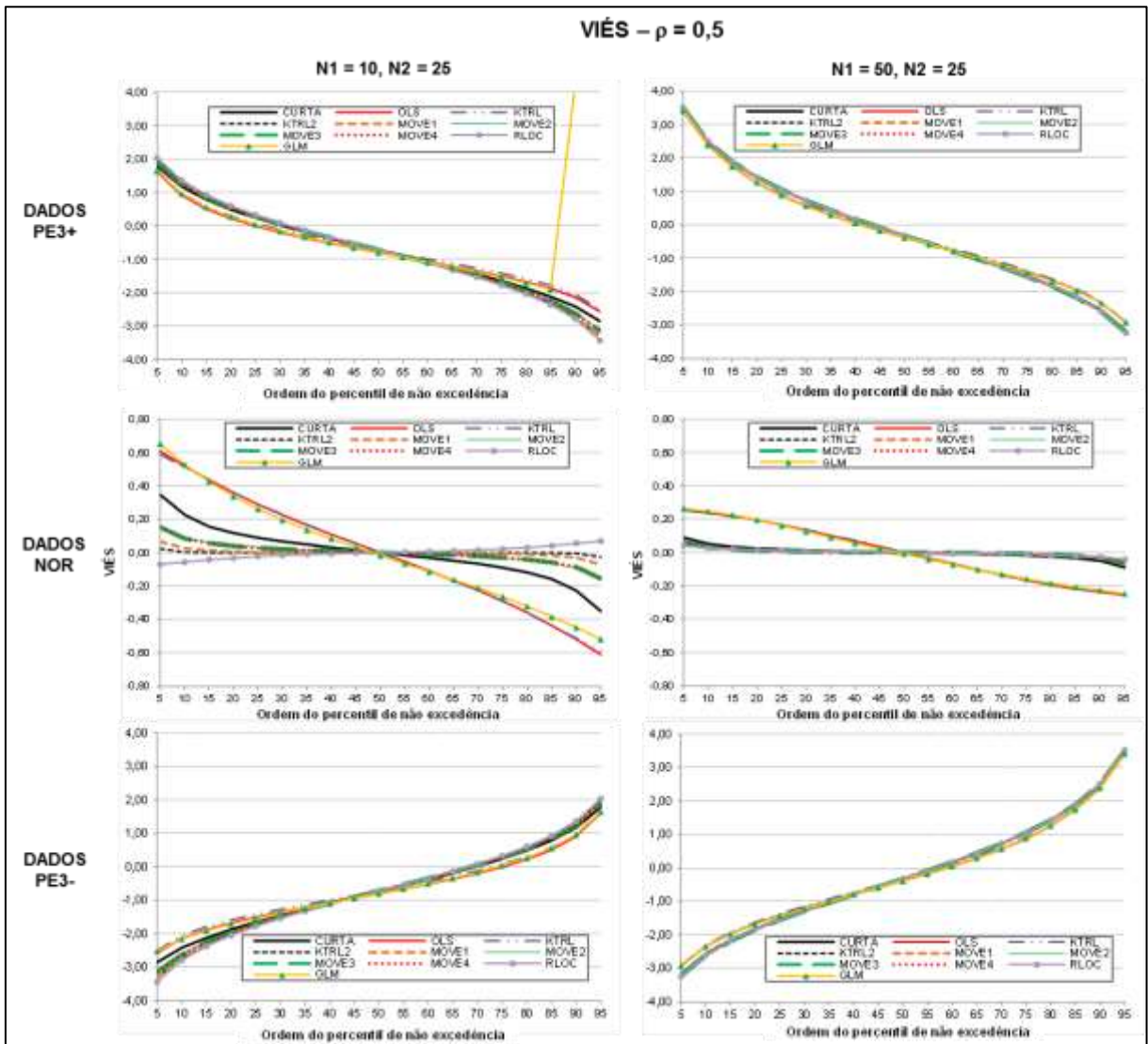


Figura 5.15 – Viés na estimativa dos percentis – influência da assimetria dos dados

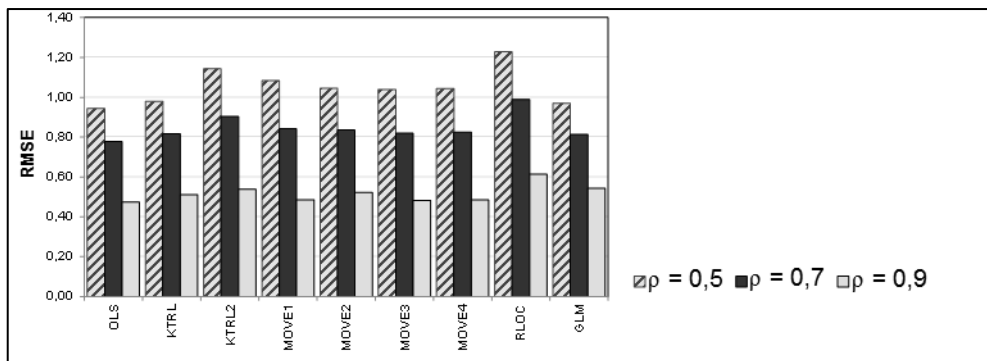


Figura 5.16 – Erro na estimativa dos registros individuais

O erro apresentado na Figura 5.16 é invariável frente às distribuições dos dados e caracteriza todas as técnicas com desempenho relativamente próximo, apesar dos menores valores observados para a OLS, KTRL e GLM. Para os cenários simulados, em que as séries curta e de referência apresentam mesma distribuição, é de se esperar que a técnica OLS tenha melhor

desempenho, pois minimiza os erros na estimativa dos valores da série originalmente curta. Contudo o valor esperado do erro é muito próximo entre as técnicas. É importante ressaltar que a técnica GLM, para dados com distribuição com assimetria positiva e extensão da série curta igual a 10, apresentou erro significativamente superior às demais.

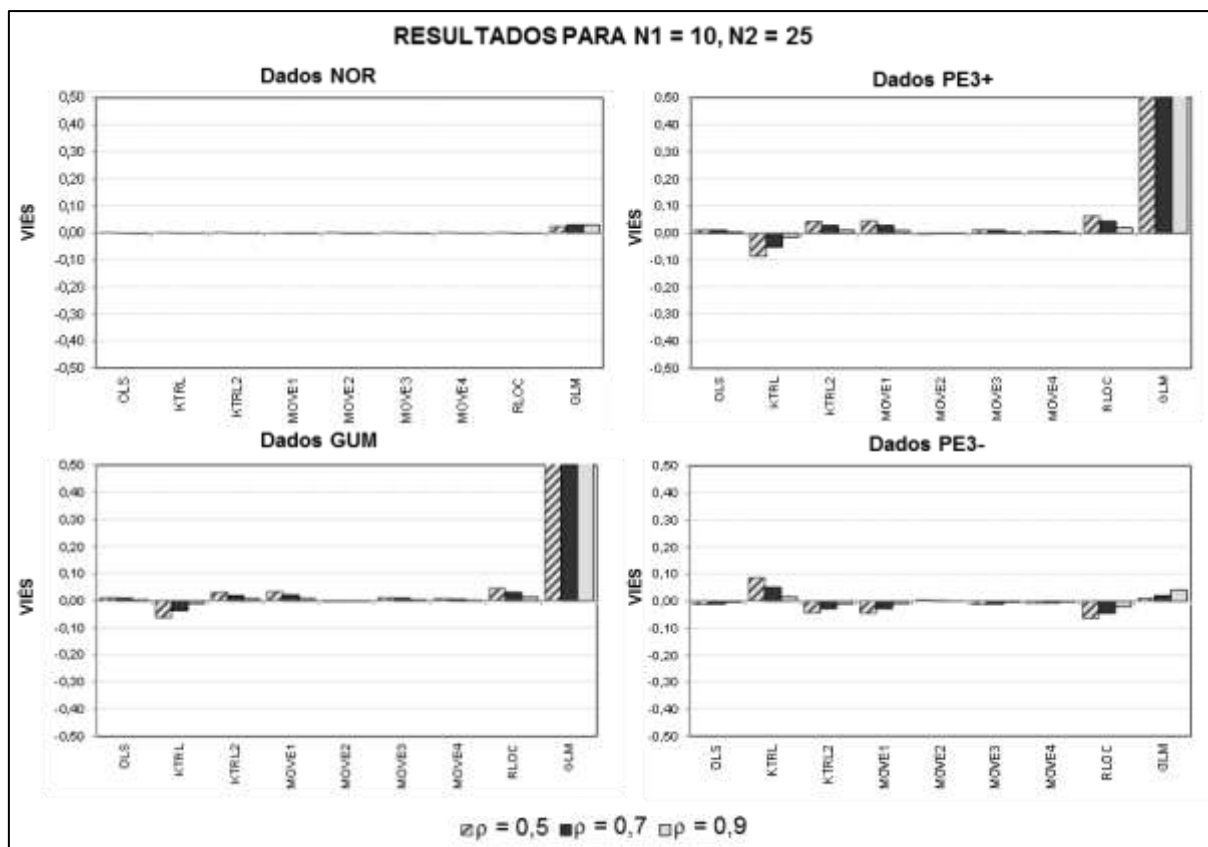


Figura 5.17 – Viés na estimativa dos registros individuais – influência da distribuição dos dados

A Figura 5.17 indica que a assimetria dos dados interfere no viés das técnicas na estimativa dos registros individuais, tendo o poder de alterar a direção da tendência das técnicas. Dados com assimetria positiva inviabilizam o uso da técnica GLM no preenchimento de falhas, tendo em vista o viés apresentado. É importante mencionar que o viés das estimativas dos registros individuais foi o mesmo da estimativa da média, o que era esperado, tendo em vista a formulação matemática do indicador analisado.

A Figura 5.18 apresenta as retas geradas pela aplicação das técnicas de extensão para o caso de uma das amostras sintéticas geradas, com $n_1 = 25$, $n_2 = 25$ e $\rho = 0,7$. As técnicas KTRL2, MOVE1 a MOVE4 e RLOC produziram estimativas mais altas do que a OLS, para valores mais altos de x , e mais baixos, para valores menores de x . A técnica KTRL gerou retas deslocadas, no eixo das ordenadas, em relação à técnica OLS. Apenas para dados PE3- as

estimativas da técnica KTRL foram inferiores às estimativas da técnica OLS. A técnica GLM retornou resultados muito próximos da técnica OLS, com sutis diferenças na estimativa dos valores extremos.

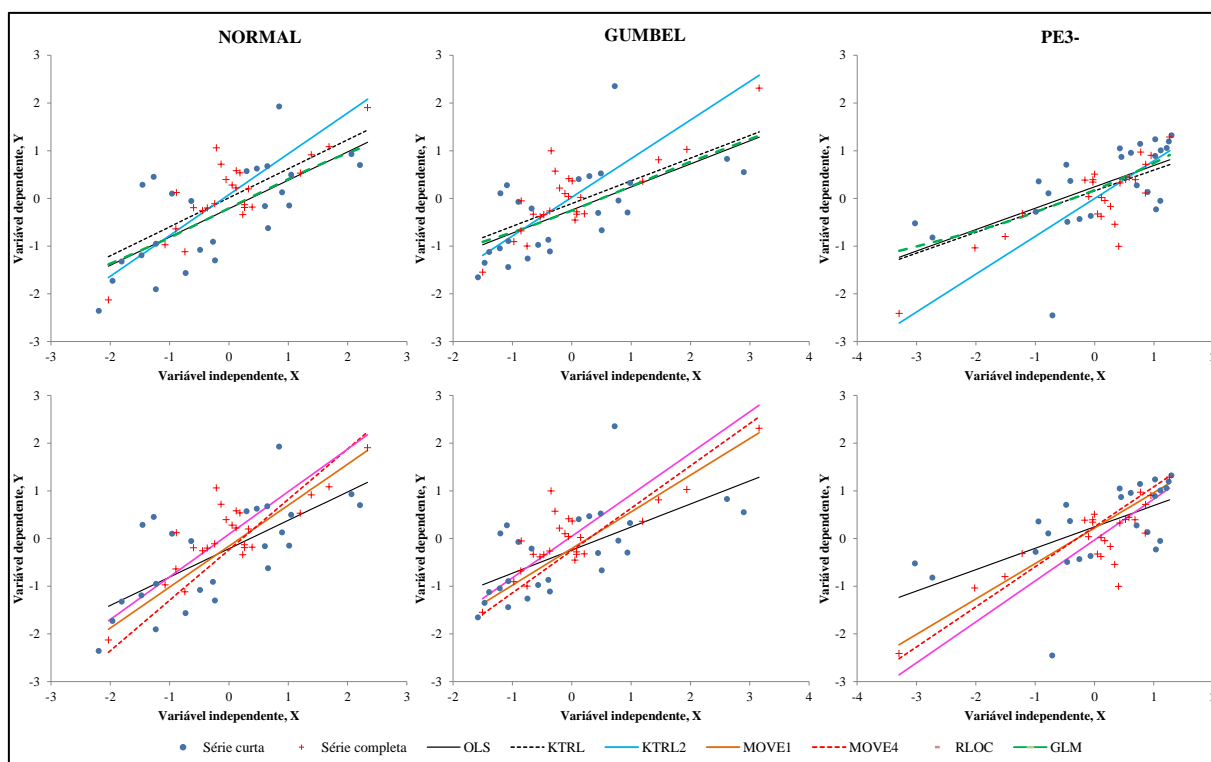


Figura 5.18 – Ajuste das técnicas de extensão para o caso $n_1=25$, $n_2=25$ e $\rho = 0,7$.

Tendo em vista que o indicador *RMSE* (Figura 5.16) retorna resultados para a condição média, conclusões feitas considerando esse indicador podem não ser as mais representativas. Para todos os níveis de correlação, incluindo os mais baixos, os melhores resultados globais foram das técnicas OLS, MOVE2, MOVE3 e MOVE4. Entretanto, a técnica MOVE1 apresenta uma característica interessante, de pressupor erros nos dados de x e y , o que também pode levar a resultados positivos em situações práticas. A escolha da melhor técnica, para a estimativa dos registros individuais, deve ser feita considerando os resultados apresentados na Figura 5.18.

5.1.4 Quantil associado ao TR = 100 anos

De modo geral, para os dados NOR, observa-se que o coeficiente de correlação é muito relevante na extensão da série hidrológica. Para valores baixos de ρ , nenhuma das técnicas de extensão reduziu significativamente o erro das estimativas do quantil de referência. Para valores elevados ($\rho = 0,9$), quase todas as técnicas, com exceção da RLOC e GLM,

promovem séries estendidas cujas estimativas do quantil de referência são mais corretas em relação à série curta. Tais resultados podem ser visualizados na Figura 5.19.

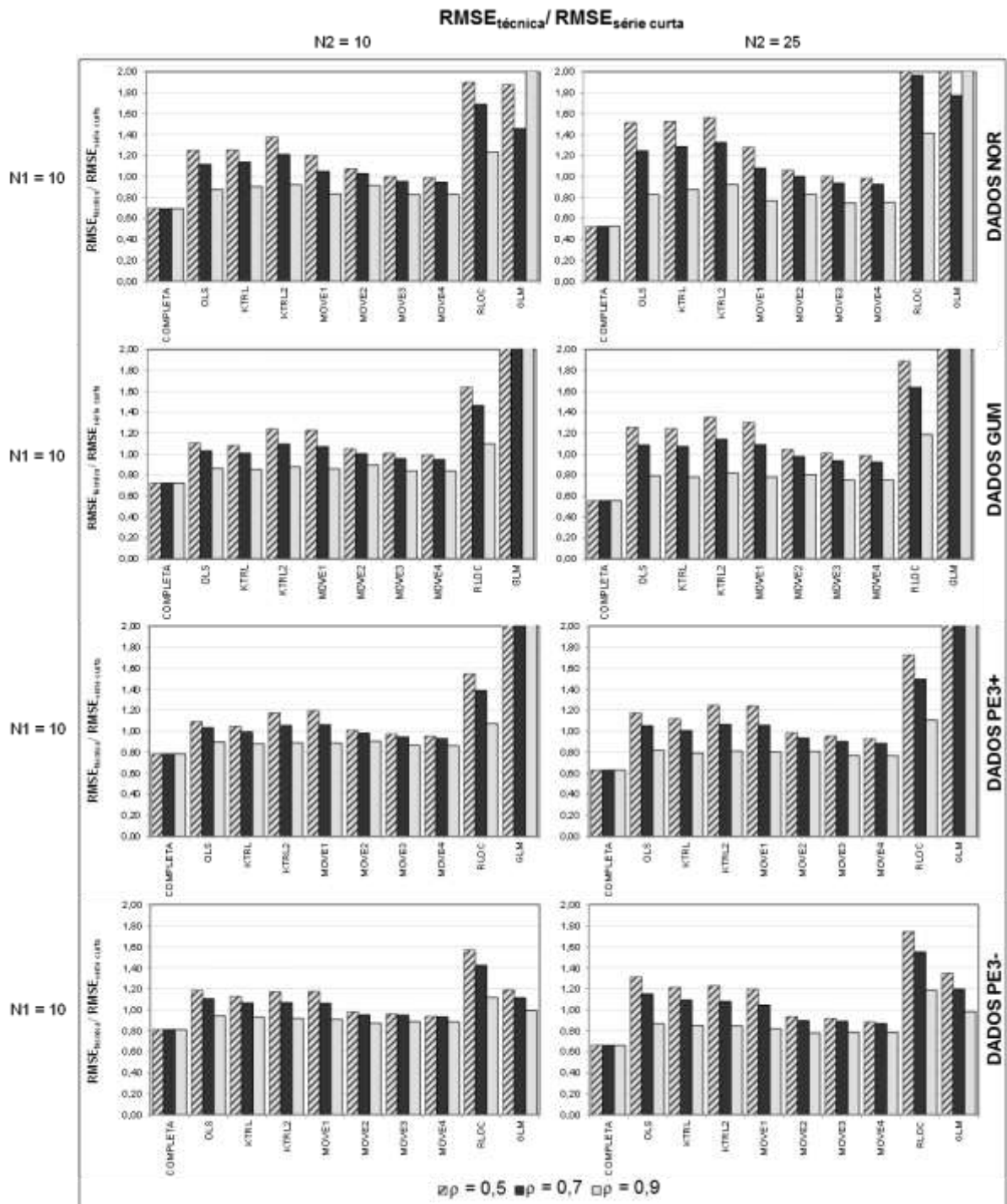


Figura 5.19 – RMSE na estimativa do quantil populacional associado ao TR = 100 anos

Considerando o critério de que as técnicas com bom desempenho, além de melhorarem a estimativa em relação à série curta, melhoram a estimativa para um aumento de n_2 , as técnicas OLS, KTRL, KTRL2, RLOC e GLM, para correlações de 0,5 e 0,7, tiveram

desempenho ruim. As técnicas MOVE foram melhor representadas pela MOVE3 e MOVE4, que melhoraram a estimativa dos quantis para $\rho \geq 0,7$. Cabe mencionar também que, para $n_1 = 50$, é muito mais difícil notar, graficamente, as técnicas com $RMSE_{técnica}/RMSE_{série\ curta} < 1$, dado que a relação n_2/n_1 é menor para maiores valores de n_1 .

Para os dados GUM, inclui-se, no grupo já mencionado, a técnica MOVE1, com erro na estimativa do quantil aumentando para um aumento de n_2 , independente da extensão da série curta, quando $\rho \leq 0,7$. Por outro lado, as demais técnicas MOVE apresentaram bons resultados. A técnica GLM apresentou resultados ruins em todos os cenários estudados. Sobre os demais aspectos, os comentários feitos para a distribuição NOR são válidos para os dados GUM.

De modo geral, os mesmos comentários se aplicam para os dados PE3+ e PE3-. A maior diferença entre esses dois cenários diz respeito à técnica GLM, com melhores resultados para os dados com coeficiente de assimetria negativo.

O viés das estimativas do quantil de referência apresentou diferenças significativas entre os cenários simulados. Primeiramente, como mostrado na Figura 5.20, dados NOR, GUM e PE3+ apresentam viés no sentido de subestimar os quantis ao utilizar o método dos momentos para ajuste das distribuições, enquanto dados PE3- apresentaram viés positivo. Para a série curta, o viés da estimativa pode ser considerado alto para todas as distribuições, exceto NOR.

Para os dados normais, é possível classificar as técnicas em duas categorias. A primeira possui viés no sentido de subestimar os quantis. São as técnicas OLS, KTRL e GLM, com magnitude do viés elevada, e, com viés entre aquele apresentado pela série curta e zero, as técnicas MOVE2, MOVE3 e MOVE4. A segunda inclui as técnicas que superestimam os quantis. São a KTRL2, MOVE1 e RLOC. Pode-se considerar que as técnicas MOVE, de modo geral, são não enviesadas, enquanto a técnica KTRL2 pode ser considerada não enviesada para $n_1 = 50$ ou $\rho = 0,9$. A técnica RLOC possui viés que aumenta à medida que n_2 também aumenta, contudo, para $n_1 = 50$, o viés pode ser considerado suficientemente baixo.

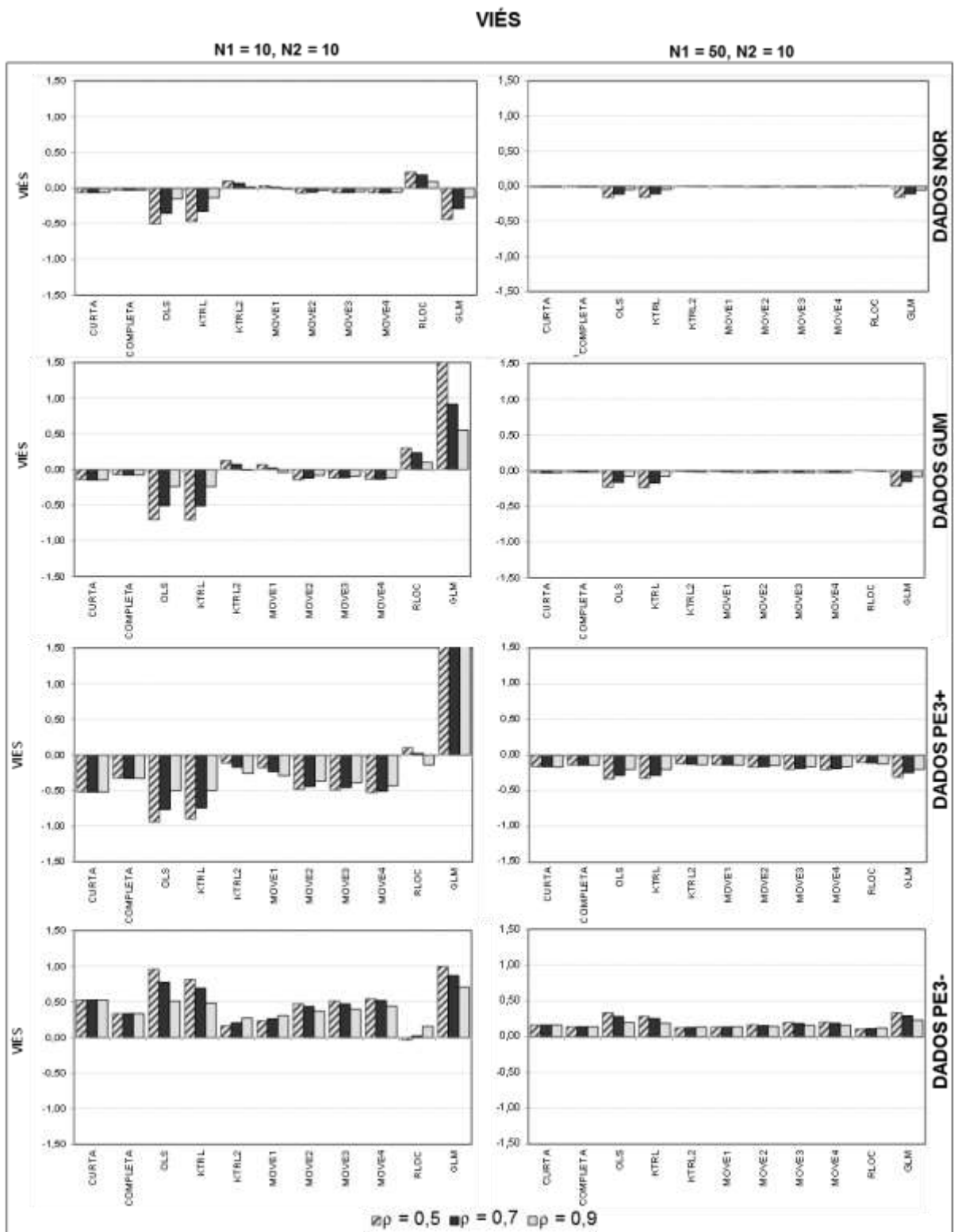


Figura 5.20 – VIÉS na estimativa do quantil populacional associado ao TR = 100 anos

Com relação ao cenário de dados GUM, é possível observar que os vieses calculados foram maiores, em magnitude, do que aqueles calculados para os dados NOR. Como possíveis causas, podem ser citados: i) o fato de o quantil associado ao TR = 100 anos da distribuição Gumbel ser numericamente maior do que o mesmo quantil da distribuição Normal, dados os

valores de média e variância populacional; e ii) os dados NOR gerados foram obtidos a partir de um algoritmo elaborado especialmente para este fim, enquanto os dados GUM foram gerados por meio do método da transformação, e se aderem menos à distribuição Gumbel do que os dados NOR se aderem à distribuição Normal. Contudo, tomando como referência o viés da série curta, os resultados são semelhantes àqueles da distribuição Normal, com as técnicas MOVE não enviesadas e a técnica KTRL2 tornando-se não enviesada à medida que n_1 aumenta. A técnica RLOC pôde ser considerada não enviesada para ρ elevado (0,9) e também quando $n_1 = 50$.

A grande diferença entre o viés na estimativa dos quantis, dos dados NOR e GUM para os dados PE3+ e PE3-, foi o aumento do viés das técnicas que geraram séries estendidas menos enviesadas do que a série curta. Para os dados PE3+, por exemplo, a série curta e a série completa possuem vieses que vão de -0,10 a -0,50, nos cenários simulados, enquanto, nos dados NOR, esses valores vão de 0 a -0,08. A série completa possui viés menor, em magnitude.

Para os dados PE3+, no cenário em que $n_1 = 10$, a técnica RLOC chega a ter um viés no sentido de superestimar o valor do quantil, o que também ocorre com a GLM, pelos problemas de ajuste já mencionados. Nos demais casos, em que $n_1 = 25$ ou 50 , todas as técnicas possuem viés no sentido de subestimar o quantil populacional. O pior desempenho foi observado para as técnicas OLS, KTRL e GLM. A RLOC teve desempenho surpreendente, frente aos resultados para as demais distribuições, com menor viés, em módulo, dentre todas as técnicas.

Algumas técnicas, nas quais incluem-se KTRL2, MOVE1 e RLOC, tiveram viés mais pronunciado para $\rho = 0,9$ do que para $\rho = 0,5$. Essas técnicas são as mesmas que apresentaram, para dados normais, viés no sentido de superestimar o quantil populacional. Como aqui os próprios dados da série curta possuem viés negativo, essas técnicas acabaram por gerar quantis com menor viés em módulo. As demais técnicas MOVE tiveram desempenho mediano, com viés da mesma ordem de grandeza e sinal da série curta.

Para os dados PE3-, podem ser feitos os mesmos comentários dos dados PE3+, porém salientando que o viés é positivo, ou seja, no sentido de superestimar os quantis. As técnicas com menor viés em módulo foram KTRL2, MOVE1 e RLOC.

De modo geral, as conclusões sobre a variância das estimativas dos quantis após a extensão das séries, em relação à variância da série curta, são semelhantes para todas as distribuições, como mostra a Figura 5.21. Para $n_1 = 10$, as técnicas KTRL2, MOVE1, RLOC e GLM possuem variância superior às demais técnicas, com esta característica se reduzindo à medida que n_1 aumenta. Os resultados ruins, para as técnicas mencionadas, são revertidos por um coeficiente de correlação elevado (0,9), cenário no qual quase todas as técnicas (exceção da RLOC e GLM) tiveram variância inferior às demais técnicas na estimativa do quantil. Os menores valores de variância foram obtidos para as técnicas OLS, KTRL, MOVE2, MOVE3 e MOVE4.

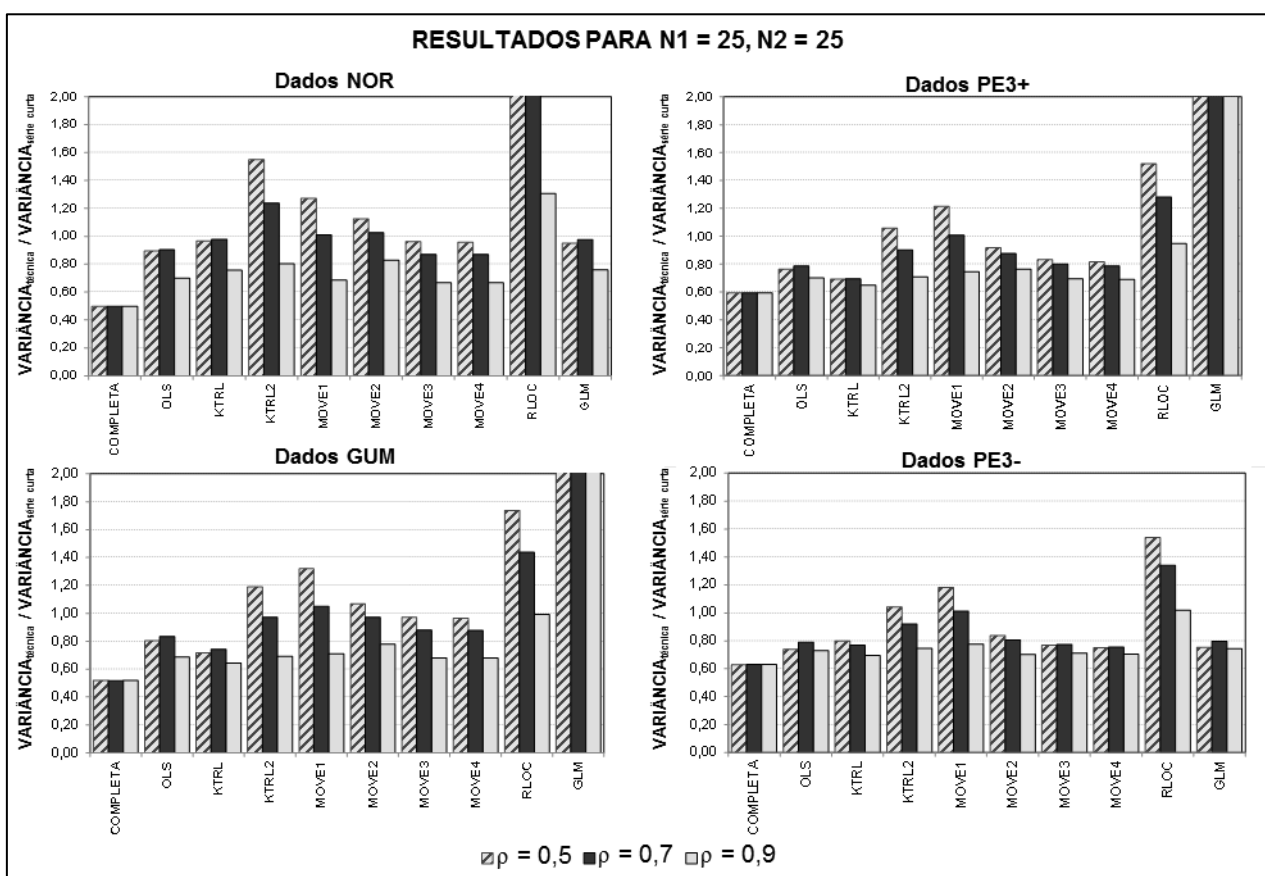


Figura 5.21 – Variância na estimativa do quantil populacional associado ao TR = 100 anos

Os resultados aqui apresentados para a estimativa do quantil de referência foram sintetizados em critérios para a aplicação das técnicas de extensão. Tais critérios, bem como a definição de aspectos práticos a serem considerados quando da extensão das séries hidrológicas, são apresentados a seguir.

5.2 Critérios para a aplicação sobre dados reais

Os critérios aqui apresentados aplicam-se à extensão das séries hidrológicas com objetivo de estimar os quantis de máximos, no caso das distribuições NOR, GUM e PE3+, e de mínimos, no caso da distribuição PE3-. Após a eliminação das técnicas com viés superior ao viés da série curta, foram selecionadas aquelas com menor erro e menor variância na estimativa dos quantis. Os resultados parciais para cada cenário de distribuição de probabilidades dos dados são apresentados nas Figuras 5.22 a 5.25. As técnicas consideradas aplicáveis são apresentadas nas Tabelas 5.1 a 5.4.

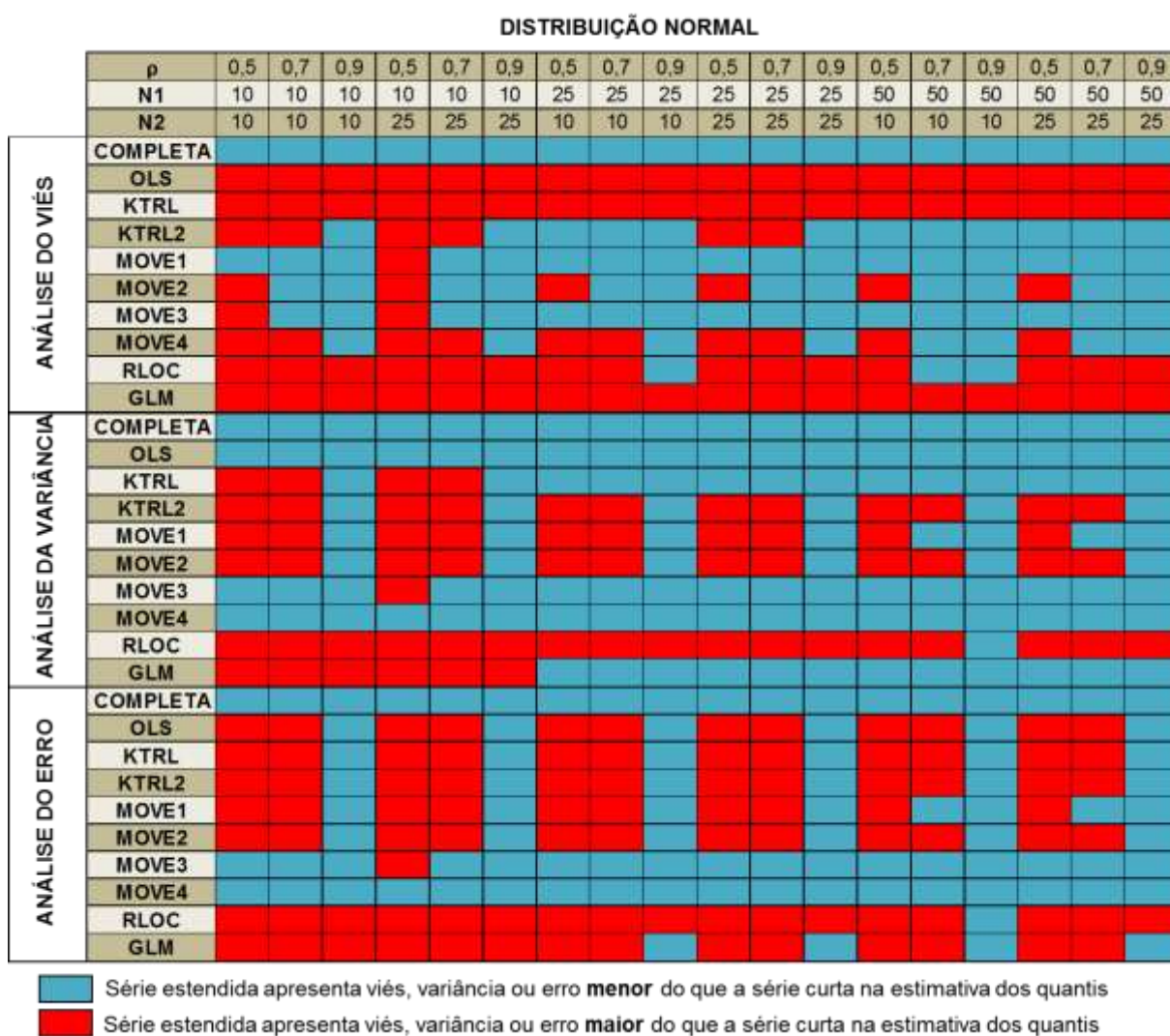


Figura 5.22 – Resultados da análise do viés, da variância e do erro na estimativa de quantis em comparação com a série curta – dados com distribuição Normal

DISTRIBUIÇÃO GUMBEL

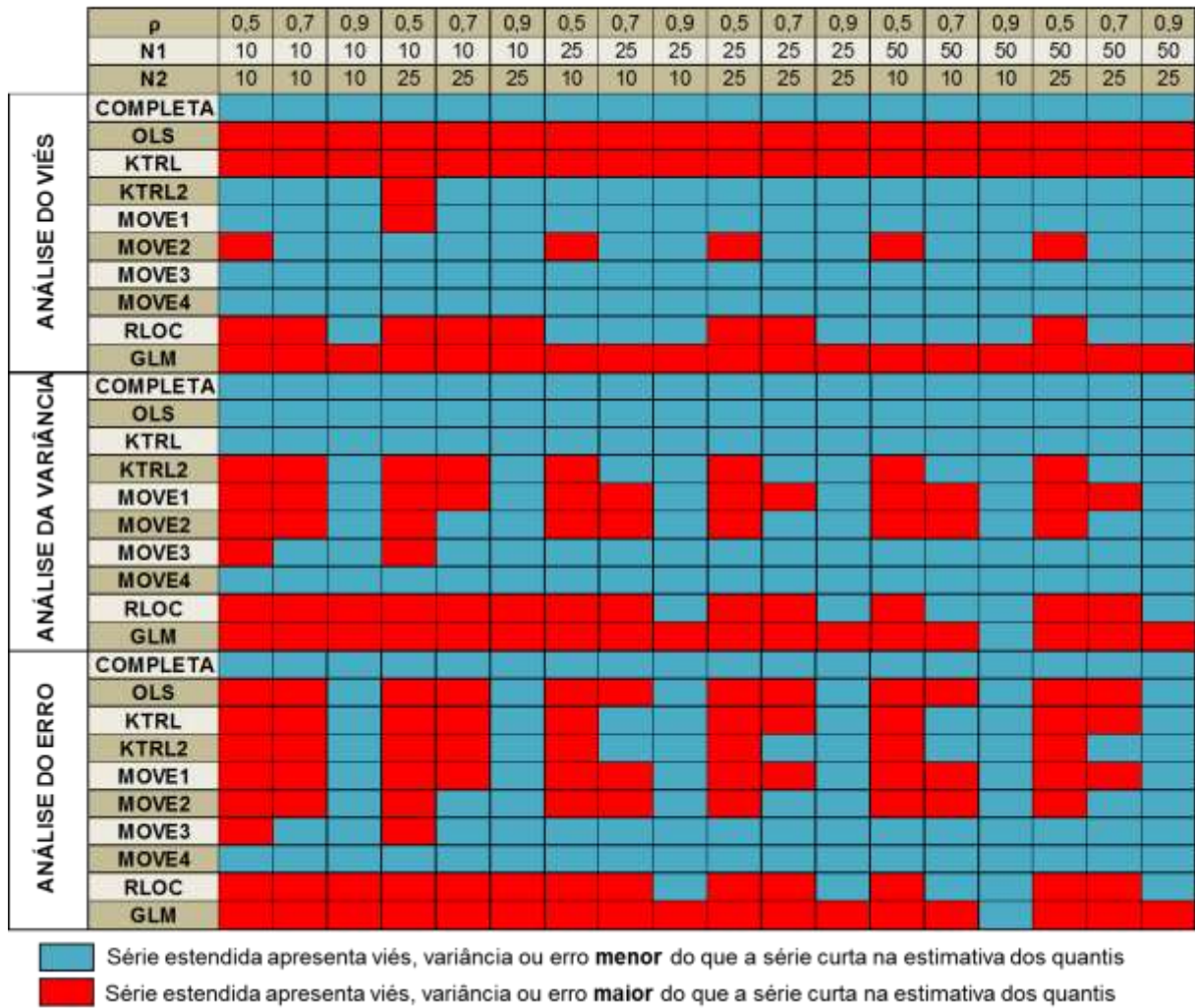


Figura 5.23 – Resultados da análise do viés, da variância e do erro na estimativa de quantis em comparação com a série curta – dados com distribuição Gumbel

DISTRIBUIÇÃO PEARSON 3 COM ASSIMETRIA POSITIVA

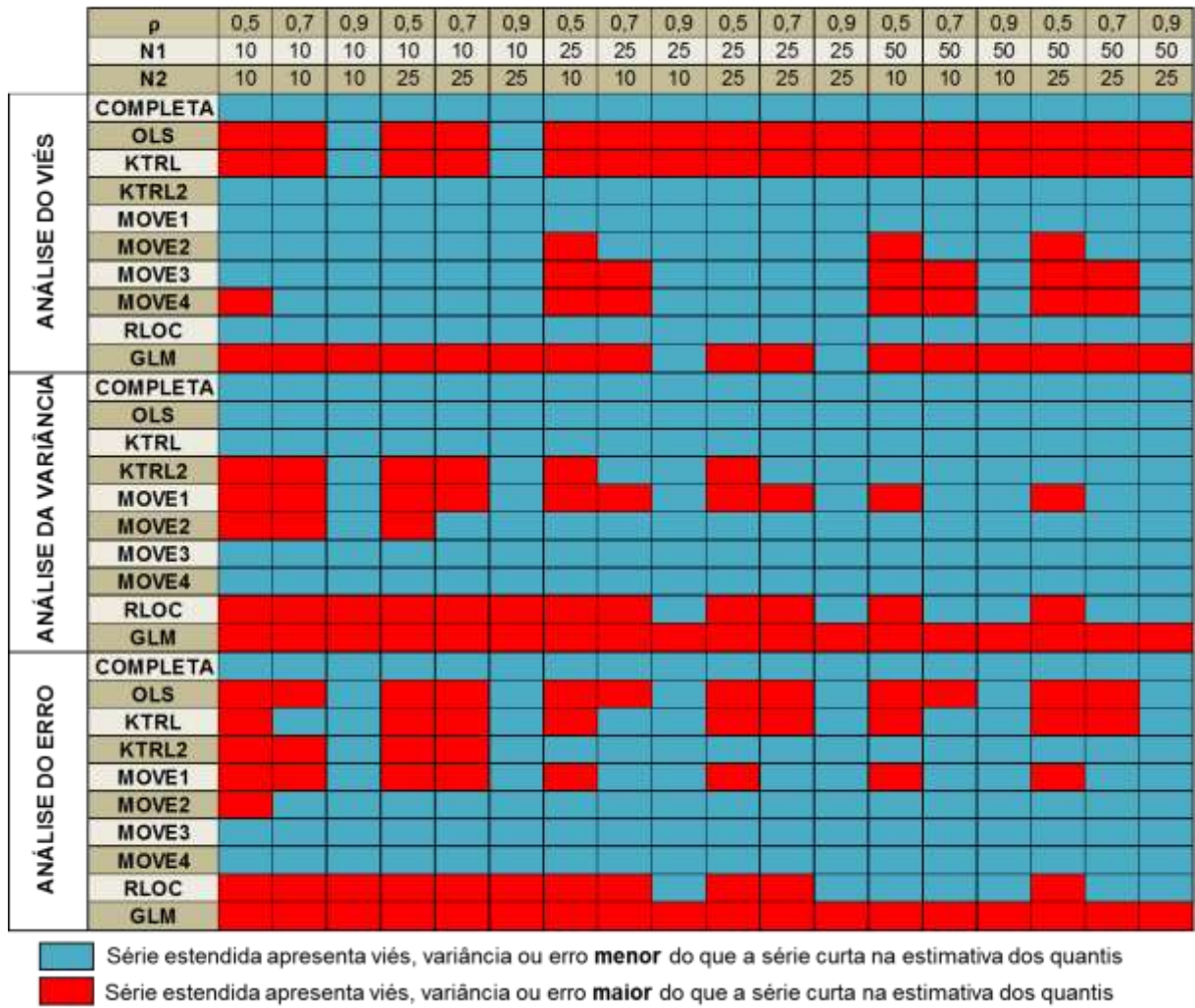


Figura 5.24 – Resultados da análise do viés, da variância e do erro na estimativa de quantis em comparação com a série curta – dados com distribuição Pearson 3, assimetria positiva

DISTRIBUIÇÃO PEARSON 3 COM ASSIMETRIA NEGATIVA

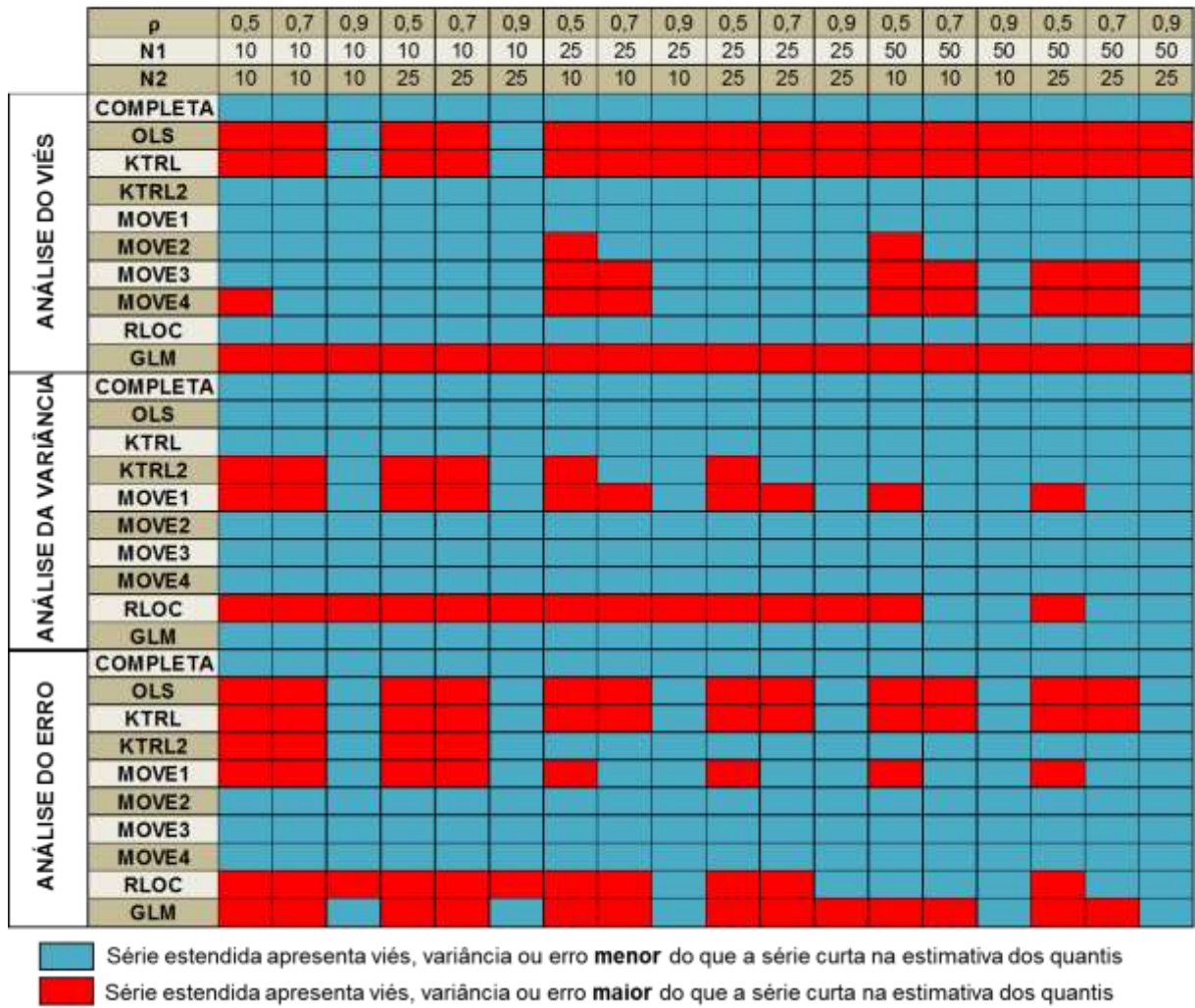


Figura 5.25 – Resultados da análise do viés, da variância e do erro na estimativa de quantis em comparação com a série curta – dados com distribuição Pearson 3, assimetria negativa

Tabela 5.1 – Técnicas aplicáveis na extensão de séries hidrológicas – dados com distribuição Normal

		N2 = 10	N2 = 25
N1 = 10	$\rho = 0,5$	-	-
	$\rho = 0,7$	MOVE3	MOVE3
	$\rho = 0,9$	MOVE4, MOVE3, MOVE2, MOVE1, KTRL2	MOVE4, MOVE3, MOVE2, MOVE1, KTRL2
N1 = 25	$\rho = 0,5$	MOVE3	MOVE3
	$\rho = 0,7$	MOVE3	MOVE3
	$\rho = 0,9$	MOVE4, MOVE3, MOVE2, MOVE1, KTRL2	MOVE4, MOVE3, MOVE2, MOVE1, KTRL2
N1 = 50	$\rho = 0,5$	MOVE3	MOVE3
	$\rho = 0,7$	MOVE4, MOVE3, MOVE1	MOVE4, MOVE3, MOVE1
	$\rho = 0,9$	MOVE4, MOVE3, MOVE2, MOVE1, KTRL2, RLOC	MOVE4, MOVE3, MOVE2, MOVE1, KTRL2

Para dados NOR, destaca-se que, para $n_1 = 10$ e $\rho = 0,5$, apenas a técnicas MOVE1 reduziu o viés na estimativa da série curta, porém mostrou-se menos eficiente e com erro maior na estimativa do quantil. A técnica considerada mais aplicável foi a MOVE3.

Para dados GUM, todos os cenários retornaram pelo menos uma técnica aplicável, sendo MOVE4 a técnica mais robusta. Destaca-se a aplicabilidade de um número maior de técnicas em relação aos dados normais, devido aos seguintes fatores: i) o coeficiente de assimetria dos dados pode diferir significativamente do coeficiente de assimetria da distribuição Gumbel (+1,14), e torna-se mais provável que as séries estendidas melhorem a aderência à distribuição; e ii) os dados normais (série curta) são gerados por meio de algoritmo desenvolvido especialmente para este fim, gerando séries curtas com momentos amostrais mais próximos dos momentos populacionais do que no cenário de dados Gumbel, no qual as amostras foram obtidas por meio do método da transformação.

Tabela 5.2 – Técnicas aplicáveis na extensão de séries hidrológicas – dados com distribuição Gumbel

		N2 = 10	N2 = 25
N1 = 10	$\rho = 0,5$	MOVE4	MOVE4
	$\rho = 0,7$	MOVE4, MOVE3, MOVE2	MOVE4, MOVE3, MOVE2
	$\rho = 0,9$	MOVE4, MOVE3, MOVE2, MOVE1, KTRL2	MOVE4, MOVE3, MOVE2, MOVE1, KTRL2
N1 = 25	$\rho = 0,5$	MOVE4, MOVE3	MOVE4, MOVE3
	$\rho = 0,7$	MOVE4, MOVE3, MOVE2, KTRL2	MOVE4, MOVE3, MOVE2, KTRL2
	$\rho = 0,9$	MOVE4, MOVE3, MOVE2, MOVE1, KTRL2, RLOC	MOVE4, MOVE3, MOVE2, MOVE1, KTRL2, RLOC
N1 = 50	$\rho = 0,5$	MOVE4, MOVE3	MOVE4, MOVE3
	$\rho = 0,7$	MOVE4, MOVE3, MOVE2, KTRL2, RLOC	MOVE4, MOVE3, MOVE2, KTRL2
	$\rho = 0,9$	MOVE4, MOVE3, MOVE2, MOVE1, KTRL2, RLOC	MOVE4, MOVE3, MOVE2, MOVE1, KTRL2, RLOC

As Tabelas 5.3 e 5.4 apresentam os resultados para os dados PE3+ e PE3-. De modo geral, as técnicas MOVE também são aplicáveis em muitos cenários de n_1 , n_2 e ρ . Contudo, foram observadas algumas inconsistências, caracterizadas por: i) técnicas sendo consideradas aplicáveis para $n_2 = 25$, mas não para $n_2 = 10$; e ii) técnicas sendo consideradas aplicáveis para um dado valor de n_1 , mas não para valores maiores. Cabe mencionar que algumas técnicas, como ocorre com a RLOC na Tabela 5.2, podem ser consideradas aplicáveis para $n_2 = 10$ e não para $n_2 = 25$. Caso os valores estimados pela técnica de extensão possuam uma tendência diferente dos demais termos da série curta, podem implicar redução do erro na estimativa dos quantis se utilizados em pequeno número. Caso sejam utilizados para estimar uma quantidade maiores de valores, a tendência pode prevalecer sobre a série curta, e o quantil pode se tornar enviesado.

Os resultados apresentados nas tabelas foram obtidos com base no valor esperado do erro, do viés e da variância das estimativas dos quantis após a aplicação das técnicas, devendo ser resguardadas as diferenças para a extensão das amostrais e para a correlação populacional. A Figura 5.26 sintetiza a capacidade de cada técnica em reduzir o erro na estimativa de quantis após a extensão das séries hidrológicas.

Tabela 5.3 – Técnicas aplicáveis na extensão de séries hidrológicas – dados com distribuição Pearson 3, assimetria positiva

		N2 = 10	N2 = 25
N1 = 10	$\rho = 0,5$	MOVE3	MOVE3, MOVE4
	$\rho = 0,7$	MOVE3, MOVE4	MOVE2, MOVE3, MOVE4
	$\rho = 0,9$	OLS, KTRL, KTRL2, MOVE1, MOVE2, MOVE3, MOVE4	OLS, KTRL, KTRL2, MOVE1, MOVE2, MOVE3, MOVE4
N1 = 25	$\rho = 0,5$	-	MOVE2, MOVE3, MOVE4
	$\rho = 0,7$	KTRL2, MOVE2	KTRL2, MOVE2, MOVE3, MOVE4
	$\rho = 0,9$	KTRL2, MOVE1, MOVE2, MOVE3, MOVE4, RLOC	KTRL2, MOVE1, MOVE2, MOVE3, MOVE4, RLOC
N1 = 50	$\rho = 0,5$	KTRL2	KTRL2
	$\rho = 0,7$	KTRL2, MOVE1, MOVE2, RLOC	KTRL2, MOVE1, MOVE2, RLOC
	$\rho = 0,9$	KTRL2, MOVE1, MOVE2, MOVE3, MOVE4, RLOC	KTRL2, MOVE1, MOVE2, MOVE3, MOVE4, RLOC

Tabela 5.4 – Técnicas aplicáveis na extensão de séries hidrológicas – dados com distribuição Pearson 3, assimetria negativa

		N2 = 10	N2 = 25
N1 = 10	$\rho = 0,5$	MOVE2, MOVE3	MOVE2, MOVE3, MOVE4
	$\rho = 0,7$	MOVE2, MOVE3, MOVE4	MOVE2, MOVE3, MOVE4
	$\rho = 0,9$	OLS, KTRL, KTRL2, MOVE1, MOVE2, MOVE3, MOVE4	OLS, KTRL, KTRL2, MOVE1, MOVE2, MOVE3, MOVE4
N1 = 25	$\rho = 0,5$	-	MOVE2, MOVE3, MOVE4
	$\rho = 0,7$	KTRL2, MOVE2	KTRL2, MOVE2, MOVE3, MOVE4
	$\rho = 0,9$	KTRL2, MOVE1, MOVE2, MOVE3, MOVE4	KTRL2, MOVE1, MOVE2, MOVE3, MOVE4
N1 = 50	$\rho = 0,5$	KTRL2	KTRL2, MOVE2
	$\rho = 0,7$	KTRL2, MOVE1, MOVE2, RLOC	KTRL2, MOVE1, MOVE2, RLOC
	$\rho = 0,9$	KTRL2, MOVE1, MOVE2, MOVE3, MOVE4, RLOC	KTRL2, MOVE1, MOVE2, MOVE3, MOVE4, RLOC

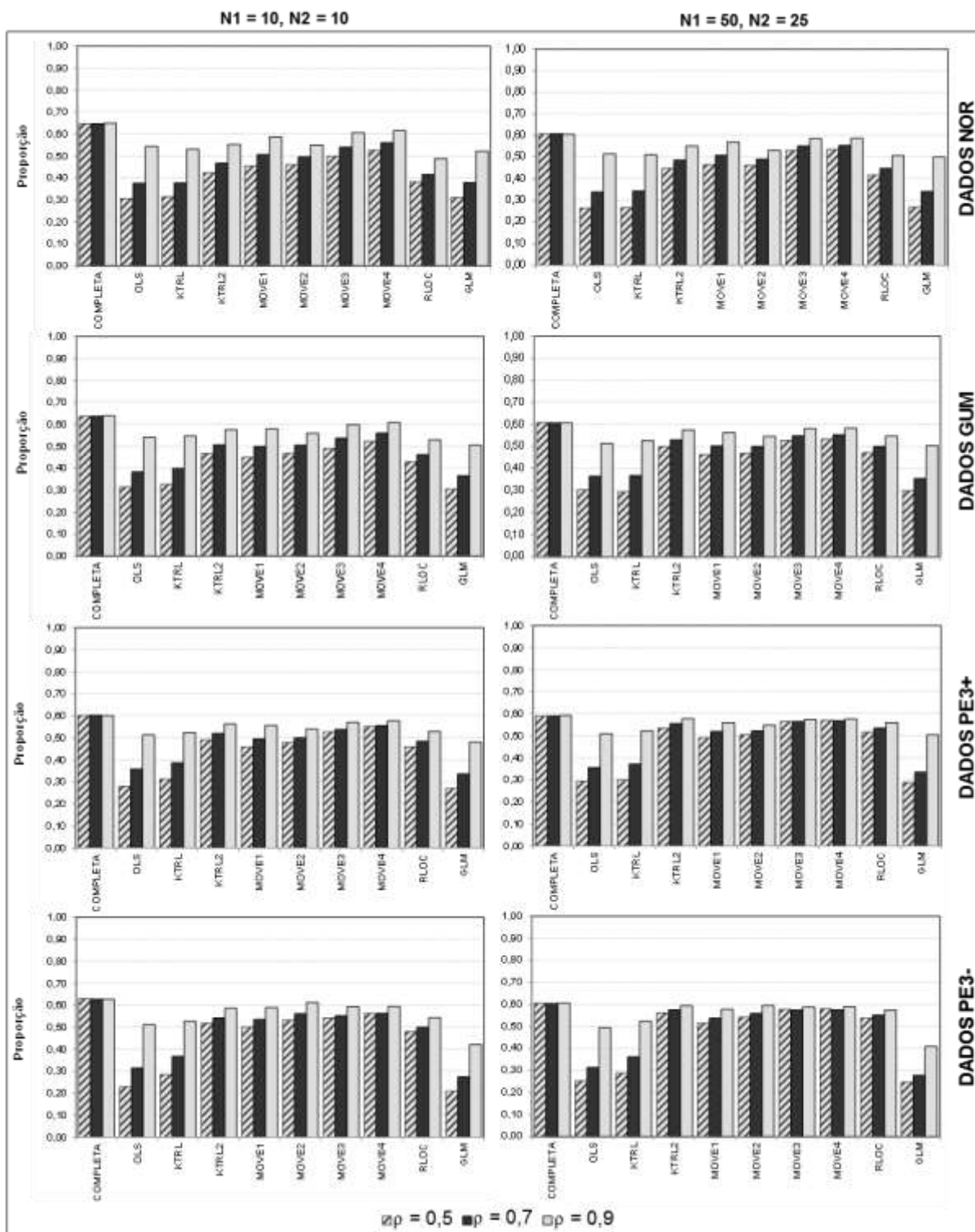


Figura 5.26 – Proporção das amostras na qual houve redução do erro absoluto na estimativa do quantil associado ao TR = 100 anos

Analisando a Figura 5.26, é importante destacar que:

- A série completa (com extensão de $n_1 + n_2$ valores) não representa redução do erro absoluto em todas as amostras analisadas. Porém, a série completa representa redução do erro em cerca de 60% das amostras simuladas, o que pode ser considerado como a

probabilidade de a série completa gerar estimativas de quantis mais próximas do valor populacional do que a série curta; e

- As técnicas KTRL2 e MOVE apresentam redução do erro na estimativa dos quantis mesmo para $\rho = 0,5$, enquanto as técnicas OLS, KTRL e GLM apresentam resultados tão bons quanto as demais técnicas apenas para $\rho = 0,9$.

Cabe ressaltar que, para a elaboração da Figura 5.26, foram considerados todos os casos em que houve redução do erro absoluto na estimativa dos quantis, independente da magnitude da redução.

Um ponto de grande complexidade diz respeito às limitações de aplicação das técnicas. Uma das formas de verificar a aplicabilidade das técnicas de extensão, nesse caso apenas para as técnicas MOVE2, MOVE3 e MOVE4, consiste em comparar a correlação amostral do caso estudado com os valores críticos definidos para a técnica RPN, apresentados na Tabela 3.5. Outra forma de formas de avaliar a aplicabilidade das técnicas, agora de modo mais generalizado, consiste em verificar os valores de correlação amostral para os quais houve redução do erro na estimativa do quantil de referência após a extensão da série. Esses valores são apresentados, resumidamente, na Figura 5.27, sem distinção entre cenários de correlação, extensão da série curta e extensão da série longa. São apresentados apenas os gráficos para os dados NOR e GUM, não tendo sido observada diferença significativa para os dados PE3+ e PE3-.

Observa-se, na Figura 5.27, que as técnicas definidas como aplicáveis em pelo menos um dos cenários requerem coeficiente de correlação inferior em relação às demais técnicas para que haja redução do erro da estimativa do quantil. Apesar de não se traduzir em um critério, os valores de correlação do primeiro quartil podem ser um valor mínimo de referência para a aplicação das técnicas. De maneira geral, o valor de 0,55, correspondente a um R^2 de 0,30, é observado em todas as técnicas. Este valor de referência não reduz significativamente o espectro de aplicação das técnicas, uma vez que, em situações práticas, tal valor do coeficiente de determinação pode, até mesmo, gerar dúvidas sobre a relação entre as variáveis dependente e independente.

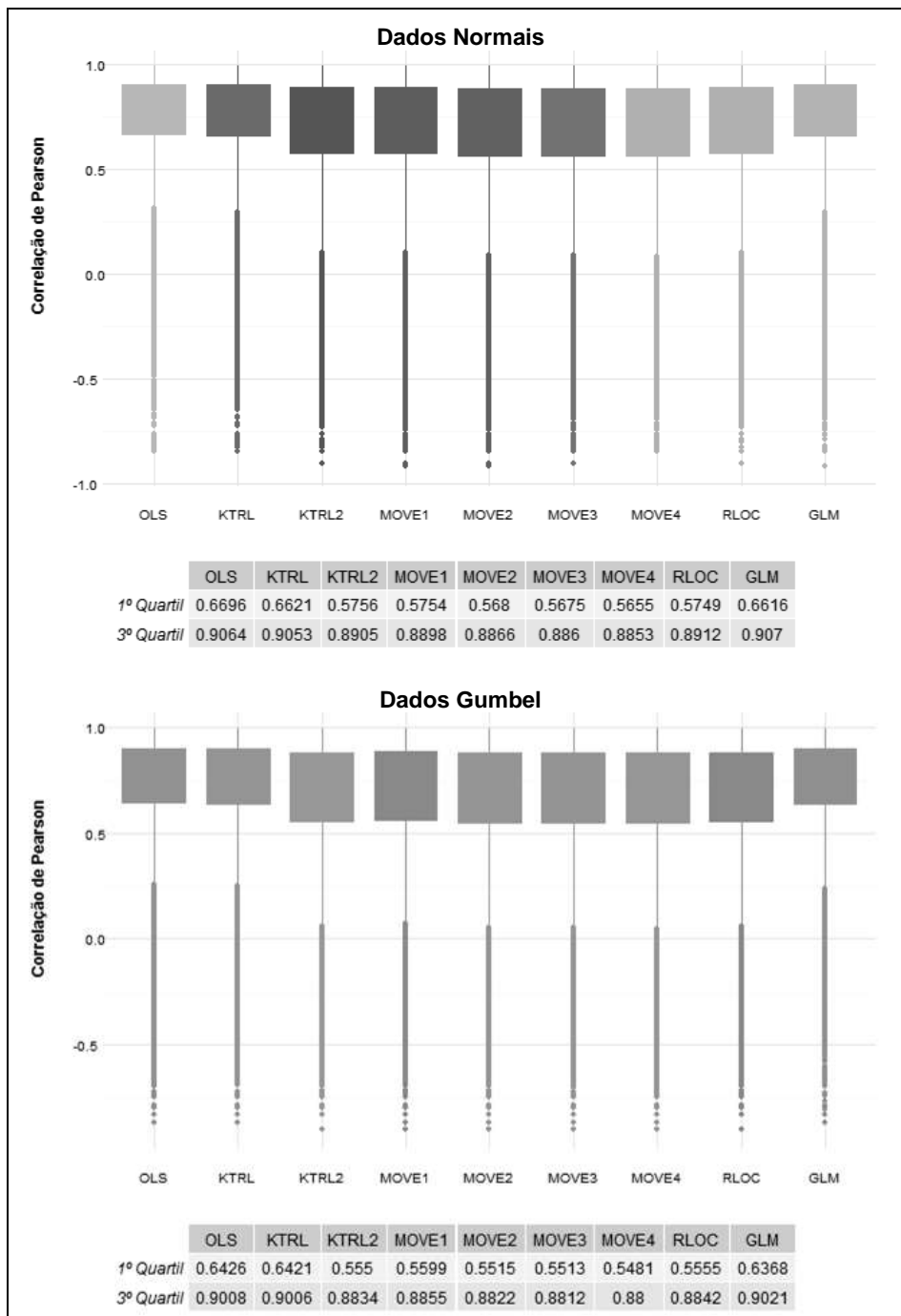


Figura 5.27 – Coeficientes de correlação entre as séries curta e de referência das simulações nas quais houve redução do erro na estimativa dos quantis após a extensão

É importante mencionar, também, que os critérios elaborados foram definidos para séries hidrológicas curta e de referência que são modelados com a mesma distribuição de probabilidades. Essa premissa, adotada no presente estudo, deve seguir como diretriz básica para a aplicação não apenas dos critérios elaborados, mas das técnicas de extensão de séries hidrológicas aqui estudadas.

Com relação às técnicas propriamente ditas, pode-se dizer que as técnicas MOVE1 e MOVE2 têm bom desempenho, de modo geral, na extensão das séries hidrológicas, apesar de se constituírem como técnicas que possuem um embasamento estatístico levemente diferente das técnicas MOVE3 e MOVE4. As duas últimas acabam por ser mais vantajosas em inúmeros casos. Apesar disso, a técnica MOVE1 possui uma característica que pode ser útil em várias situações, de fornecer a mesma relação entre x e y independente de qual é a variável dependente. Esse fato não deve ser ignorado na escolha da técnica de extensão a ser utilizada. As técnicas KTRL2 e RLOC são ajustadas com base em parâmetros muito sensíveis à extensão original da amostra e, portanto, devem ser utilizadas apenas para valores elevados de n_1 . Essa recomendação vale especialmente no caso da técnica RLOC, que é ajustada utilizando uma quantidade de informação significativamente menor do que a técnica KTRL2.

Tratamento especial é dado, na literatura, para a estimativa do coeficiente de assimetria, sendo propostos ajustes, correções e até mesmo novos estimadores da assimetria populacional (BOBÉE & ROBITAILE, 1975; YEVJEVICH & OBEYSEKERA, 1984; BRYS *et al.*, 2004; KIM & WHITE, 2004). Algumas técnicas geraram séries estendidas com maior capacidade do que a série curta de gerar estimativas do coeficiente de assimetria no caso populacional. No entanto, recomenda-se cautela quanto ao uso das técnicas de extensão aqui apresentadas para esse propósito, tendo em vista que, ao basear as estimativas em séries com extensão relativamente curta, o erro na estimativa do coeficiente de assimetria pode ainda ser significativo.

Apesar de os critérios apresentados terem sido definidos para a aplicação das técnicas de extensão com o objetivo de melhorar as estimativas dos quantis máximos, não se pode negar que eles também definem, simultaneamente, casos em que a média, a variância e a assimetria (guardadas as devidas limitações) da série estendida compõem um conjunto de medidas descritivas estimadas com mais acurácia do que utilizando apenas a série curta. Portanto, os critérios elaborados podem ser utilizados em casos nos quais considera-se essa propriedade útil. Por exemplo, pode-se extrapolar os resultados apresentados da distribuição Gumbel para outras distribuições de 2 parâmetros utilizadas para modelar eventos hidrológicos máximos, como a Fréchet e Log-Normal. É importante pontuar que não foi analisado, nesta pesquisa, o comportamento das técnicas frente a amostras contendo *outliers*. Para maiores informações sobre este caso, a pesquisa de Khalil e Adamowski (2014a) pode ser consultada.

Para a validação do critério definido, propõe-se a aplicação das técnicas sobre dados reais, com o intuito, também, de estudar o processo de escolha da melhor técnica entre aquelas definidas como aplicáveis e da melhor série hidrológica de referência..

5.3 Aplicação sobre dados reais

Para a aplicação com dados reais, foram selecionadas, conforme apresentado no item 4.6.3, dois pares de estações na região de interesse. Os registros dessas estações serviram como base para a aplicação das técnicas de duas formas:

- Simulação histórica: aplicação das técnicas em um maior nível de detalhamento, restringindo as observações para a sequência em que aconteceram na série histórica. Foi considerado $n_1 = 10$ e n_2 foi adotado de forma que, após a extensão, a série curta tivesse a mesma extensão da série de referência; e
- Simulação de *bootstrap*: aplicação das técnicas em simulações de *bootstrap*, alterando, aleatoriamente, a sequência em que ocorreram as observações, porém mantendo a correspondência dos pares de registros da série curta e da série de referência observados em períodos de observação comuns. Também foi considerado $n_1 = 10$ e n_2 foi adotado de forma que, após a extensão, a série curta tivesse a mesma extensão da série de referência.

Os resultados foram avaliados tomando como referência as séries históricas completas de cada estação hidrometeorológica.

A Tabela 5.5 apresenta, em síntese, características dos cenários considerados e das estações (incluindo as séries temporais nelas monitoradas) utilizadas para a aplicação das técnicas de extensão de séries hidrológicas.

Tabela 5.5 – Informações sobre os cenários considerados e sobre as séries hidrológicas consideradas na aplicação sobre dados reais

Cenário	1		2	
Simulação	1.1	1.2	2.1	2.2
Variável aleatória analisada	Precipitação diária máxima anual	Precipitação acumulada anual	Vazão média diária máxima anual	Vazão média anual
Estação de referência	Braz Pires (PLU)		Braz Pires (FLU)	
Estação associada à série curta	Piranga (PLU)		Porto Firme (FLU)	
n_1	10			
n_2	33	43	45	48
Coefficiente de correlação – séries históricas completas	0,38	0,86	0,78	0,92
Coefficiente de correlação – período de observações comuns considerado Simulação histórica	0,55	0,87	0,62	0,98
Coefficiente de correlação – período de observações comuns considerado Simulação de <i>bootstrap</i>	Variável	Variável	Variável	Variável
Número de simulações <i>bootstrap</i>	5.000			

5.3.1 Simulação histórica

5.3.1.1 Cenário 1

Cenário 1.1: Precipitação Diária Máxima Anual

Para a aplicação sobre dados reais, neste cenário, a estação pluviométrica Piranga será considerada como provedora da série curta. A estação Piranga possui registros de monitoramento de julho de 1941 a julho de 2015 (75 anos de monitoramento), porém a presença de falhas levou à remoção de alguns registros das análises aqui efetuadas. Como estação de referência, é adotada a série de máximos diários anuais registrada na estação pluviométrica Braz Pires, que possui monitoramento entre março de 1967 e fevereiro de 2016. O período de observações comuns foi o primeiro período na série histórica das duas estações que apresentou 10 observações comuns em cada estação. Cabe mencionar que, no caso da precipitação diária máxima anual, as observações consideradas comuns são aquelas que ocorreram em um mesmo ano hidrológico. Uma síntese dos dados considerados e dos resultados da aplicação das técnicas, para o Cenário 1.1 (precipitação diária máxima anual), é apresentada na Tabela 5.6.

Tabela 5.6 – Resultados da aplicação das técnicas de extensão de séries hidrológicas – Cenário 1.1

APLICAÇÃO PARA MÁXIMOS DIÁRIOS ANUAIS						
Dados de entrada						
Estação pluviométrica provedora da série curta:						Piranga
Estação pluviométrica provedora da série de referência:						Braz Pires
Coeficiente de correlação no período de observações comuns:						0,55
Extensão da série curta (n_1):						10
Extensão da série de referência (n_1+n_2):						43
Extensão da série completa da estação Piranga PLU (removendo anos com falhas):						70
Extensão da série completa da estação Braz Pires PLU (removendo anos com falhas):						44
Medida descritiva	Série curta	Série de referência				
		Curta	Completa			
Média	77,65	61,28	74,00			
Desvio padrão	21,93	18,42	21,74			
Coeficiente de assimetria	0,689	-0,092	1,116			
Coeficiente de variação (CV)	0,2825	0,3007	0,2938			
Quantil Gumbel (TR = 100 anos)	146,45	119,07	142,21			
Quantil GEV (TR = 100 anos)	139,85	100,98	141,90			
Resultados - Série estendida						
Técnica	Média	Desvio padrão	Coeficiente de Assimetria	CV	Quantil Gumbel (TR = 100 anos)	Quantil GEV (TR = 100 anos)
MOVE1	92,80	25,89	1,201	0,2789	173,99	174,89
MOVE2	87,19	22,87	1,192	0,2623	158,91	159,59
MOVE3	86,00	23,00	1,232	0,2675	158,15	159,30
MOVE4	83,94	22,33	1,238	0,2660	154,00	155,17
KTRL2	86,84	21,44	1,103	0,2469	154,10	153,64
RLOC	84,95	18,71	0,943	0,2202	143,64	141,39
OLS	86,00	16,60	0,589	0,1930	138,06	131,78
Série completa	81,33	23,36	1,261	0,2872	154,60	156,13

Para verificar a influência da distribuição de probabilidade adotada nos resultados, foram utilizadas as distribuições Gumbel e GEV na estimativa do quantil associado ao TR = 100 anos. Os resultados para as técnicas consideradas aplicáveis, de acordo com os resultados do item 5.2, são destacados na Tabela 5.6. É possível notar, primeiramente, que o coeficiente de assimetria para as duas séries, no período de observações comuns, é muito diferente daquele para as séries completas, dadas as limitações para cálculo deste momento amostral para no caso em que $n_1 = 10$.

A única técnica considerada aplicável é a MOVE4, que teve excelente desempenho na estimativa do quantil de referência, para as duas distribuições analisadas (Gumbel e GEV).

Cabe destacar que o coeficiente de assimetria foi estimado satisfatoriamente, garantindo bons resultados também para a distribuição GEV. As técnicas MOVE2, MOVE3 e KTRL2, apesar de consideradas não aplicáveis, tiveram também bons resultados. Isso se deve às flutuações aleatórias que os dados reais apresentam, em comparação com os dados considerados para a definição dos critérios de aplicação das técnicas. Por outro lado, a técnica OLS apresentou evidências práticas do viés na variância incorporado durante a extensão, sendo a única técnica que reduziu o valor do quantil estimado em comparação com a série curta.

A título de ilustração, a Figura 5.28 apresenta o ajuste das técnicas MOVE4 e OLS. podendo-se notar que a série estendida por meio da técnica MOVE4 possui amplitude muito mais próxima dos valores registrados na série histórica, em comparação com a técnica OLS.

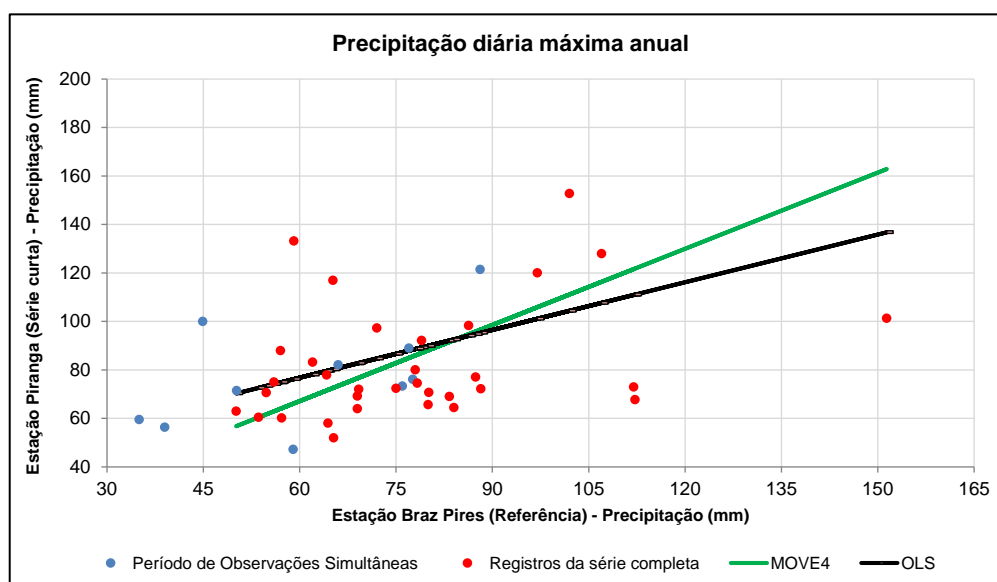


Figura 5.28 – Ajuste das técnicas MOVE4 e OLS – Cenário 1.1

Cenário 1.2: Precipitação Acumulada Anual

Os resultados da aplicação das técnicas de extensão para dados pluviométricos anuais (Cenário 1.2) são apresentados na Tabela 5.7. É possível observar, primeiramente, um nível de correlação mais elevado do que no Cenário 1.1. Esse fato deve-se a uma relação espacial mais clara entre as precipitações acumuladas anuais em comparação com as máximas anuais. É possível verificar, portanto, se essa relação leva a um melhor aproveitamento das técnicas de extensão.

Assim como no Cenário 1.1, são analisados os quantis estimados após o ajuste de duas distribuições, porém, neste caso, foram adotadas Normal e Gumbel. Essa avaliação foi feita

para determinar se a distribuição escolhida para o ajuste dos dados seria capaz de alterar os resultados comparativos entre as técnicas. A distribuição Pearson 3 pode ser mais representativa para as precipitações anuais do que a distribuição Gumbel, tendo em vista a flexibilidade introduzida por um terceiro parâmetro, podendo convergir para a distribuição Normal. Contudo, os ajustes à distribuição Pearson 3, por meio do Método dos Momentos convencional, resultaram em quantis muito diferentes daqueles estimados por meio de outras distribuições e foram, portanto, considerados não aplicáveis. Atribui-se essa dificuldade ao coeficiente de assimetria, que assumiu valores variando de negativos a positivos tanto nas séries curta e completa quanto nas séries estendidas.

Tabela 5.7 – Resultados da aplicação das técnicas de extensão de séries hidrológicas – Cenário 1.2

APLICAÇÃO PARA ACUMULADO ANUAL						
Dados de entrada						
Estação provedora da série curta:						Piranga
Estação provedora da série de referência:						Braz Pires
Coeficiente de correlação no período de observações comuns:						0,87
Extensão da série curta (n_1):						10
Extensão da série de referência (n_1+n_2):						53
Extensão da série completa da estação Piranga PLU (removendo anos com falhas):						70
Extensão da série completa da estação Braz Pires PLU (removendo anos com falhas):						43
Medida descritiva	Série curta	Série de referência				
		Curta	Completa			
Média	1.207	1.142	1.278			
Desvio padrão	369,9	365,5	291,5			
Coeficiente de assimetria	1,379	0,011	-0,447			
Coeficiente de variação (CV)	0,3064	0,3200	0,2281			
Quantil Normal (TR = 100 anos)	2.068	1.993	1.956			
Quantil Gumbel (TR = 100 anos)	2.368	2.289	2.193			
Resultados - Série estendida						
Técnica	Média	Desvio padrão	Coeficiente de Assimetria	CV	Quantil Normal (TR = 100 anos)	Quantil Gumbel (TR = 100 anos)
MOVE1	1.345	295,1	0,038	0,2194	2.031	2.270
MOVE2	1.333	304,3	0,050	0,2282	2.041	2.288
MOVE3	1.328	312,9	0,046	0,2357	2.055	2.309
MOVE4	1.322	324,4	0,034	0,2453	2.077	2.340
KTRL2	1.390	399,9	-0,042	0,2878	2.320	2.644
RLOC	1.456	491,0	-0,039	0,3372	2.598	2.996
OLS	1.328	271,0	0,111	0,2041	1.958	2.178
Série completa	1.395	344,6	-0,121	0,2470	2.197	2.476

As técnicas consideradas aplicáveis, considerando que os dados podem ser modelados pela distribuição Normal, são MOVE (todas) e KTRL2. A técnica que mais se aproximou do coeficiente de assimetria da série completa foi a KTRL2, provavelmente devido a sua capacidade de extrapolar a função densidade de probabilidade das séries curta e de referência para a série estendida. Contudo, foi a técnica que mais errou na estimativa do coeficiente de variação.

As técnicas consideradas não aplicáveis tiveram desempenho inferior às demais, especialmente com relação à estimativa do desvio padrão da série completa. A técnica OLS, como já determinado na literatura e confirmado pelos resultados apresentados no item 5.1.2.2, possui um viés na estimativa da variância, o que se traduziu nos resultados da Tabela 5.7.

As técnicas MOVE tiveram desempenho muito próximo entre si, sendo difícil a tarefa de escolher uma delas supondo que não se conhecem as medidas descritivas da série completa. Duas abordagens podem ser adotadas: i) escolher a técnica com maior probabilidade de reduzir o erro da estimativa do quantil populacional, que é a MOVE4, com 64,9%, como determinado na Simulação de Monte Carlo; ou ii) analisar os quantis nas estações adjacentes e adotar a técnica que mais se aproxime deles. A Figura 5.29 apresenta as medidas descritivas das séries hidrológicas completas registradas nas estações pluviométricas próximas. Adotando como limites regionais inferidos para o quantil de referência da estação Piranga de 1.956 mm e 2.170 mm para a distribuição Normal, todas as técnicas MOVE retornam resultados consistentes, e qualquer uma delas poderia ser utilizada.

Os resultados apresentados na Tabela 5.7 revelam ainda que os momentos amostrais estimados não provocaram diferenças significativas entre as técnicas quando da estimativa dos quantis pelas duas distribuições (Normal e Gumbel), guardadas as diferenças inerentes à utilização de uma ou outra distribuição de probabilidade.

A Figura 5.30 apresenta os percentis e registros individuais estimados pelas técnicas MOVE4, KTRL2 e OLS, em comparação com aqueles correspondentes à série curta e à série completa. Notam-se comportamentos opostos nas técnicas KTRL2 e OLS, com relação à variabilidade dos dados, e a técnica MOVE4 apresenta valores intermediários, ratificando os resultados da Tabela 5.7.

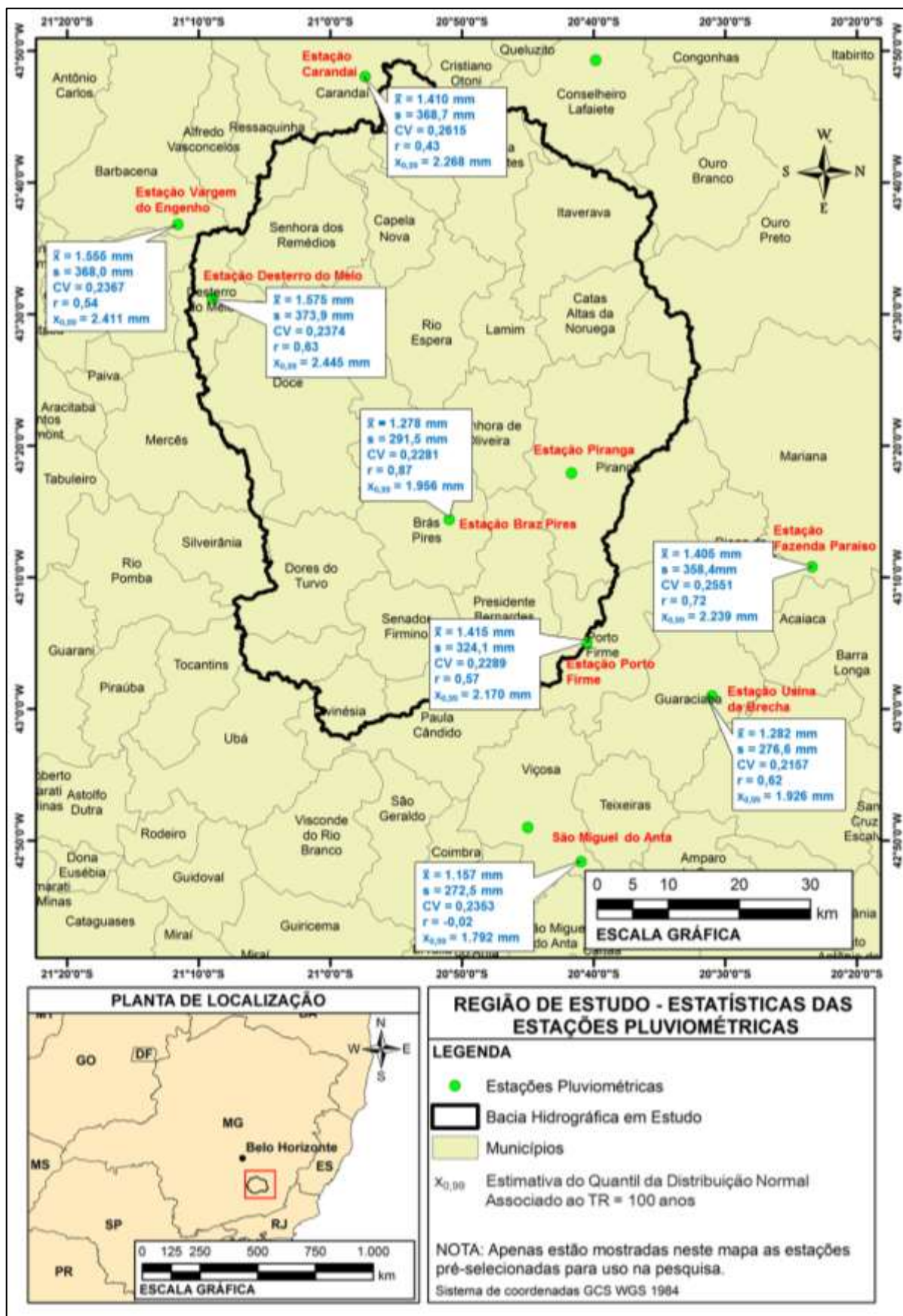


Figura 5.29 – Cenário 1.2 – Medidas descritivas das séries históricas completas - estações próximas à estação pluviométrica Piranga

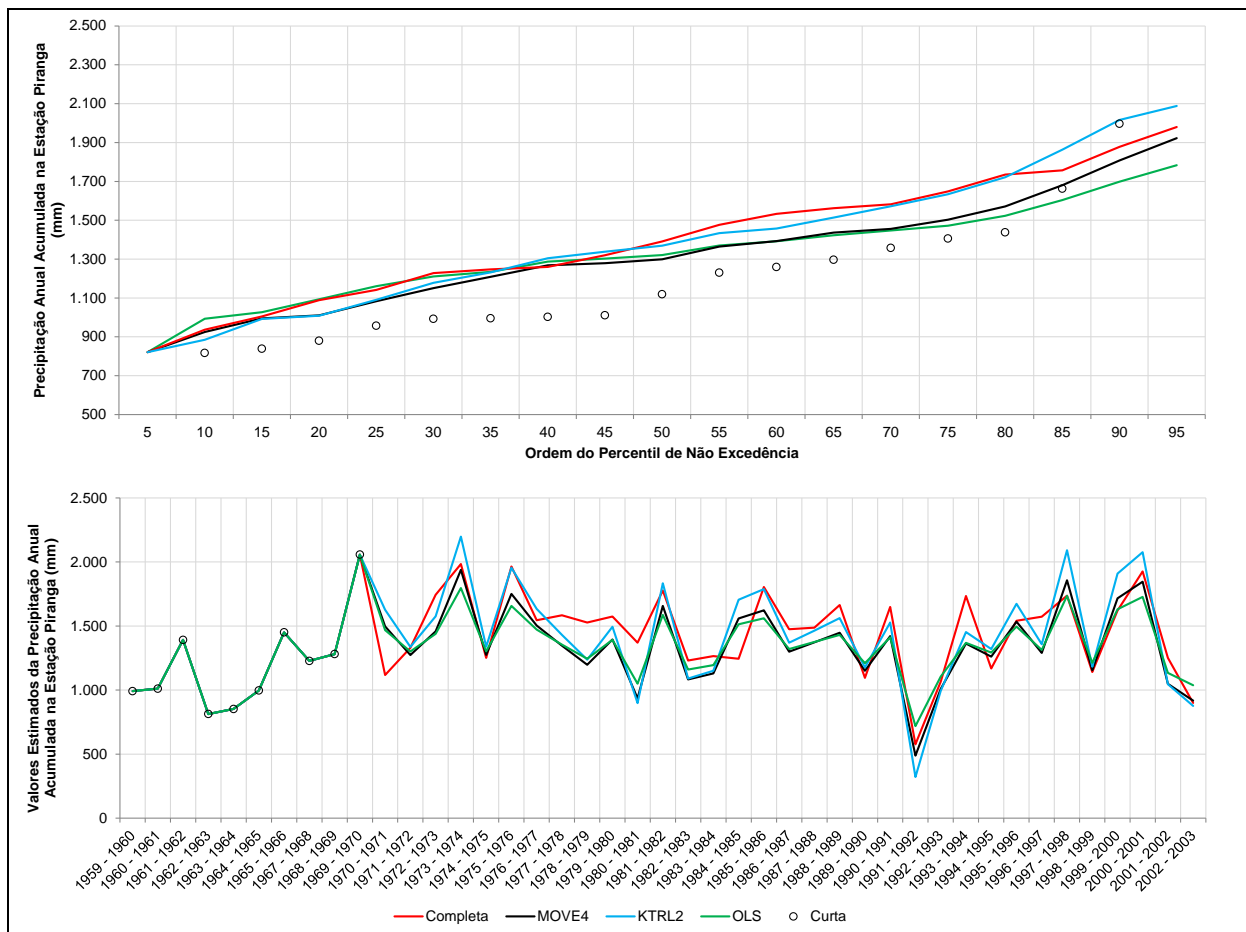


Figura 5.30 – Cenário 1.2 – Estimativa dos percentis e registros individuais da estação pluviométrica Piranga

5.3.1.2 Cenário 2

Cenário 2.1: Vazão Média Diária Máxima Anual

Os estudos de extensão de séries hidrológicas aqui apresentados, para dados fluviométricos, diferem significativamente daqueles para dados pluviométricos, tendo em vista a relação física muito mais evidente entre vazões monitoradas, a montante e a jusante em um curso de água, em comparação com as precipitações monitoradas, mais sujeitas à variabilidade espacial.

No Cenário 2, será considerada a estação pluviométrica Porto Firme, no exutório da bacia hidrográfica em estudo, como provedora da série curta. A estação Porto Firme possui registros de monitoramento de junho de 1938 a dezembro de 2015 (77 anos de monitoramento), porém a presença de falhas levou à remoção de alguns registros das análises aqui efetuadas. Como estação de referência, é adotada a série de vazões médias diárias máximas anuais registrada na estação Braz Pires, que possui monitoramento entre agosto de 1938 e janeiro de 2016. O

período de observações comuns foi o primeiro período na série histórica das duas estações que apresentou 10 observações comuns em cada uma delas. Assim como mencionado no Cenário 1, é importante ressaltar que, no caso da vazão média diária máxima anual, as observações consideradas comuns são aquelas que ocorreram em um mesmo ano hidrológico, e não necessariamente no mesmo dia. Uma síntese dos dados considerados e dos resultados da aplicação das técnicas, para o Cenário 2.1 (vazão média diária máxima anual), é apresentada na Tabela 5.8.

Tabela 5.8 – Resultados da aplicação das técnicas de extensão de séries hidrológicas – Cenário 2.1

APLICAÇÃO PARA VAZÃO MÉDIA DIÁRIA MÁXIMA ANUAL						
Dados de entrada						
Estação provedora da série curta:					Porto Firme	
Estação provedora da série de referência:					Braz Pires	
Coeficiente de correlação no período de observações comuns:					0,62	
Extensão da série curta (n_1):					10	
Extensão da série de referência (n_1+n_2):					55	
Extensão da série completa da estação Porto Firme FLU (removendo anos com falhas):					71	
Extensão da série completa da estação Braz Pires FLU (removendo anos com falhas):					68	
Medida descritiva	Série curta	Série de referência				
		Curta		Completa		
Média	356,3	99,9		109,4		
Desvio padrão	86,9	25,1		33,7		
Coeficiente de assimetria	0,543	-0,404		0,268		
Coeficiente de variação (CV)	0,2439	0,2516		0,3081		
Quantil Gumbel (TR = 100 anos)	628,9	178,7		215,2		
Quantil GEV (TR = 100 anos)	592,7	147,1		193,1		
Resultados - Série estendida						
Técnica	Média	Desvio padrão	Coeficiente de Assimetria	CV	Quantil Gumbel (TR = 100 anos)	Quantil GEV (TR = 100 anos)
MOVE1	389,4	116,6	0,313	0,2994	755,1	683,6
MOVE2	<i>378,1</i>	<i>100,0</i>	<i>0,283</i>	<i>0,2644</i>	<i>691,6</i>	627,8
MOVE3	<i>376,9</i>	<i>98,8</i>	<i>0,281</i>	<i>0,2620</i>	<i>686,7</i>	623,4
MOVE4	<i>373,1</i>	<i>92,5</i>	<i>0,272</i>	<i>0,2479</i>	<i>663,3</i>	603,1
KTRL2	359,3	100,1	0,256	0,2785	673,3	606,8
RLOC	357,0	90,8	0,258	0,2543	641,7	581,7
OLS	376,9	77,0	0,241	0,2042	618,3	566,3
Série completa	363,7	171,3	2,131	0,4710	901,0	965,1

Observa-se, na Tabela 5.8, um comportamento distinto do coeficiente de variação na série curta e na série completa. As técnicas definidas como aplicáveis, neste caso, são MOVE2, MOVE3 e MOVE4. Não há grande variação entre as estimativas do quantil de referência,

uma vez que estas aumentaram para as três técnicas consideradas aplicáveis. As diferenças entre os valores estimados e o quantil da série completa devem-se a um aumento da variabilidade dos registros após o período de observações comuns utilizado nos ajustes das técnicas.

A Figura 5.31 apresenta um mapa contendo medidas descritivas das estações fluviométricas a montante da estação Porto Firme. Observa-se um comportamento distinto entre as estações no que diz respeito ao coeficiente de variação. Os valores apresentados são correspondentes ao período completo de observações em cada posto. A estação Piranga apresenta coeficiente de variação muito mais alto que as demais estações, ainda apresentando um coeficiente de correlação alto (0,74). Ao utilizar a estação Piranga como referência, tem-se o resultado apresentado na Tabela 5.9.

A Tabela 5.9 mostra que, ao utilizar a estação Piranga como referência, os resultados são melhores do que ao utilizar a estação Braz Pires como referência. Esse resultado convida-nos à discussão em relação à escolha da série de referência. A série com maior correlação com a estação Porto Firme pertence ao posto Senador Firmino, enquanto a maior área de drenagem, entre as três estações, pertence ao posto Piranga. Além da relação física, observa-se, no posto Piranga, certa variabilidade do coeficiente de variação no decorrer do tempo, de 0,41 a 0,65. Fica claro que o coeficiente de correlação entre a série curta e a série de referência não deve ser a única variável analisada quando da escolha da estação de referência, para propósitos de extensão de séries hidrológicas.

Apesar da alteração da estação de referência, as técnicas consideradas aplicáveis (MOVE2, MOVE3 e MOVE4) tiveram desempenho semelhante, com erro significativo na estimativa do quantil de referência da série completa. A técnica MOVE1 apresentou o melhor desempenho, como mostrado nas Tabelas 5.8 e 5.9, e o fato de representar a mesma relação entre as duas estações, independente de qual é considerada a estação de referência, pode ser uma vantagem nos casos em que a relação física entre as estações é mais pronunciada. Um exemplo é o caso em que duas estações fluviométricas em um mesmo curso de água são estudadas para propósitos de extensão da série hidrológica de uma delas. A estimativa da técnica KTRL2 melhorou significativamente quando a estação Piranga foi considerada como referência, por representar melhor a variabilidade da série histórica da estação Porto Firme.

As estimativas da média e coeficiente de assimetria foram muito semelhantes entre as técnicas, prevalecendo aquelas com melhor estimativa do desvio padrão. Como consequência, o desempenho comparativo das técnicas foi igual independente do cenário de distribuição ajustada aos dados.

Tabela 5.9 – Resultados da aplicação das técnicas de extensão de séries hidrológicas – Cenário 2.1 (estação Piranga como referência)

APLICAÇÃO PARA VAZÃO MÉDIA DIÁRIA MÁXIMA ANUAL						
Dados de entrada						
Estação provedora da série curta:						Porto Firme
Estação provedora da série de referência:						Piranga
Coeficiente de correlação no período de observações simultâneas:						0,74
Extensão da série curta (n_1):						10
Extensão da série de referência (n_1+n_2):						63
Extensão da série completa da estação Porto Firme FLU (removendo anos com falhas):						71
Extensão da série completa da estação Piranga FLU (removendo anos com falhas):						66
Medida descritiva	Série curta	Série de referência				
		Curta		Completa		
Média	356,3	198,6		209,8		
Desvio padrão	86,9	82,3		137,2		
Coeficiente de assimetria	0,543	0,674		1,726		
Coeficiente de variação (CV)	0,2439	0,4145		0,6543		
Quantil Gumbel (TR = 100 anos)	628,9	456,9		640,3		
Quantil GEV (TR = 100 anos)	592,7	431,2		676,1		
Resultados - Série estendida						
Técnica	Média	Desvio padrão	Coeficiente de Assimetria	CV	Quantil Gumbel (TR = 100 anos)	Quantil GEV (TR = 100 anos)
MOVE1	363,4	141,4	1,848	0,3890	806,8	849,2
MOVE2	<i>361,7</i>	<i>120,8</i>	<i>1,793</i>	<i>0,3340</i>	<i>740,6</i>	774,8
MOVE3	<i>361,5</i>	<i>119,5</i>	<i>1,789</i>	<i>0,3305</i>	<i>736,4</i>	770,1
MOVE4	<i>361,0</i>	<i>108,3</i>	<i>1,747</i>	<i>0,2999</i>	<i>700,6</i>	729,6
KTRL2	366,0	140,0	1,851	0,3825	805,2	847,3
RLOC	361,6	119,8	1,790	0,3315	737,5	771,3
OLS	361,5	107,1	1,744	0,2963	697,6	726,2
Série completa	363,7	171,3	2,131	0,4710	901,0	965,1

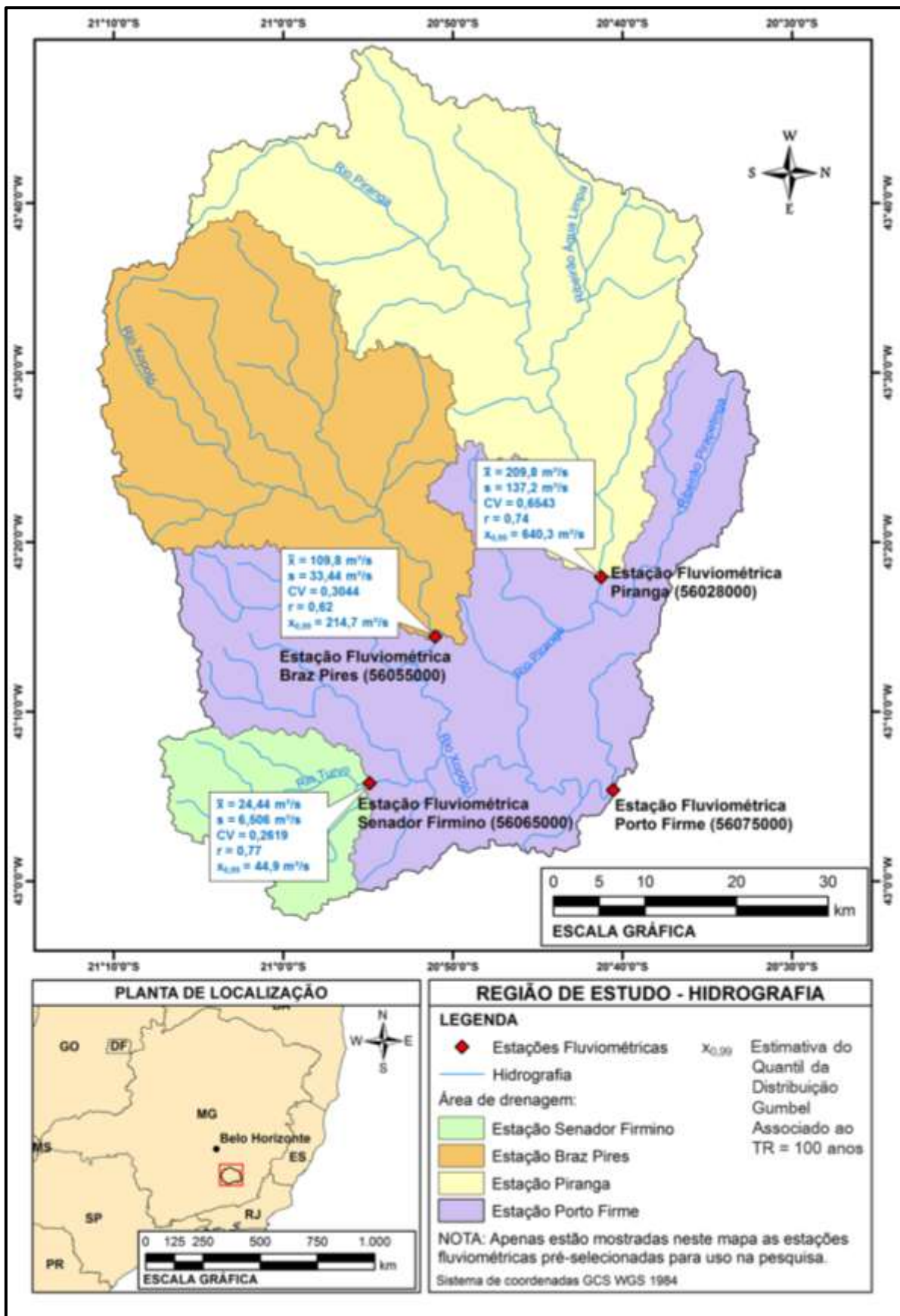


Figura 5.31 – Cenário 1.2 – Medidas descritivas das séries históricas completas - estações à montante da estação fluviométrica Porto Firme

Cenário 2.2: Vazão Média Anual

O Cenário 2.2 apresenta o comportamento das técnicas para o caso em que o coeficiente de correlação entre as séries curta e de referência é próximo de 1. Esse cenário é favorecido, como já mencionado, pela relação física evidente entre as séries históricas em análise. A Tabela 5.10 apresenta os resultados da aplicação das técnicas.

Tabela 5.10 – Resultados da aplicação das técnicas de extensão de séries hidrológicas – Cenário 2.2

APLICAÇÃO PARA VAZÃO MÉDIA ANUAL						
Dados de entrada						
Estação provedora da série curta:					Porto Firme	
Estação provedora da série de referência:					Braz Pires	
Coeficiente de correlação no período de observações comuns:					0,98	
Extensão da série curta (n_1):					10	
Extensão da série de referência (n_1+n_2):					58	
Extensão da série completa da estação Porto Firme FLU (removendo anos com falhas):					67	
Extensão da série completa da estação Braz Pires FLU (removendo anos com falhas):					62	
Medida descritiva	Série curta	Série de referência				
		Curta	Completa			
Média	89,9	21,3	19,7			
Desvio padrão	16,20	3,80	5,01			
Coeficiente de assimetria	1,848	1,546	0,538			
CV	0,1801	0,1790	0,2546			
Quantil Normal (TR = 100 anos)	127,6	30,1	31,4			
Quantil Gumbel (TR = 100 anos)	140,8	33,2	35,4			
Resultados - Série estendida						
Técnica	Média	Desvio padrão	Coeficiente de Assimetria	CV	Quantil Normal (TR = 100 anos)	Quantil Gumbel (TR = 100 anos)
MOVE1	83,3	21,35	0,556	0,2563	133,0	150,3
MOVE2	83,4	21,19	0,559	0,2540	132,7	149,9
MOVE3	83,4	21,18	0,560	0,2538	132,7	149,9
MOVE4	83,5	21,05	0,562	0,2522	132,4	149,5
KTRL2	85,7	18,90	0,630	0,2207	129,7	145,0
RLOC	85,7	18,90	0,630	0,2207	129,7	145,0
OLS	83,4	20,99	0,562	0,2515	132,3	149,3
Série completa	75,1	19,5	0,593	0,2603	120,5	136,3

Como indicado na Tabela 5.10, o desempenho de todas as técnicas, mesmo as consideradas não aplicáveis, é muito próximo, uma vez que as técnicas convergem para resultados próximos quando r se aproxima tanto de 1. Esse resultado mostrou ser invariável frente à distribuição ajustada aos dados. Assim como mostrado na Tabela 5.10, em alguns casos

simulados ao longo de toda a pesquisa, as técnicas KTRL2 e RLOC apresentaram ajuste idêntico. Esse resultado é facilmente explicado pelo fato de que, quando as amostras da série curta e da série de referência estão alinhadas para a estimativa dos percentis, a mediana das relações entre os percentis (utilizada para ajuste do modelo KTRL2) tende a ser igual à declividade obtida pela razão das AIQ das séries curta e de referência.

A Tabela 5.10 mostra que os quantis estimados por meio das séries estendidas se distanciaram mais do quantil da série completa do que em relação à série curta. A principal causa desse comportamento foi a reprodução das características da série de referência nas séries estendidas, com uma redução menor da média do que aquela observada na série completa. É possível concluir que é extremamente relevante analisar as medidas descritivas da série de referência, no período de observações comuns e em todo o período registrado, antes da extensão.

A Figura 5.32 apresenta os percentis e registros individuais estimados pelas técnicas MOVE4, KTRL2 e OLS. Em ambos os gráficos, as técnicas MOVE4 e OLS apresentam-se muito próximas, dado o elevado nível de correlação entre as séries curta e de referência. A técnica KTRL2 não converge para os resultados da técnica OLS mesmo para $r \approx 1$. Em geral, as técnicas superestimaram os registros individuais e percentis.

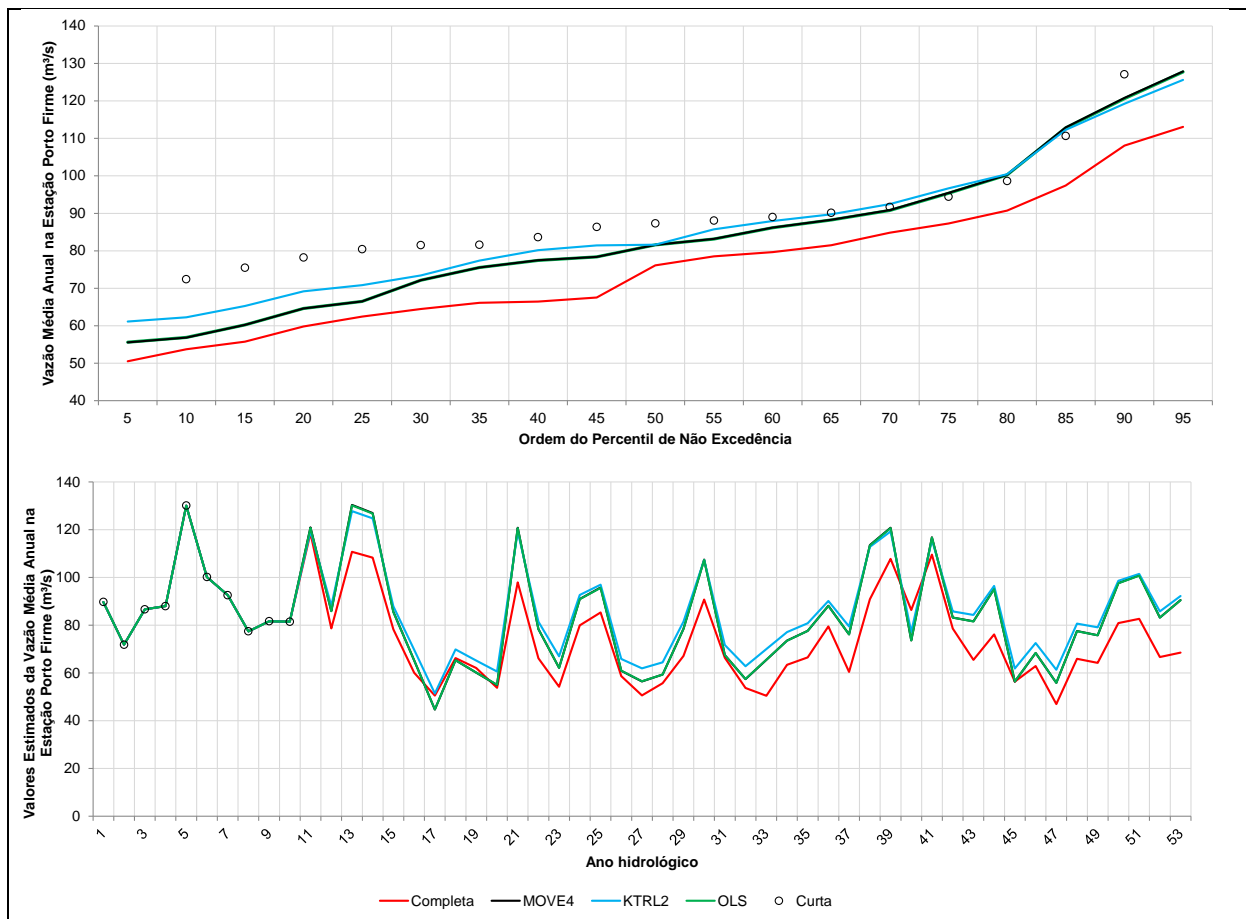


Figura 5.32 – Cenário 2.2 – Estimativa dos percentis e registros individuais da estação fluviométrica Porto Firme

5.3.2 Simulações *bootstrap*

Os resultados das simulações *bootstrap* são apresentados na Figura 5.33. Em cada amostragem da simulação, foram obtidos novos coeficientes de correlação entre a série curta e a série de referência. Para cada técnica, a extensão só foi realizada nos casos em que foi considerada aplicável pelo critério definido no item 5.2.

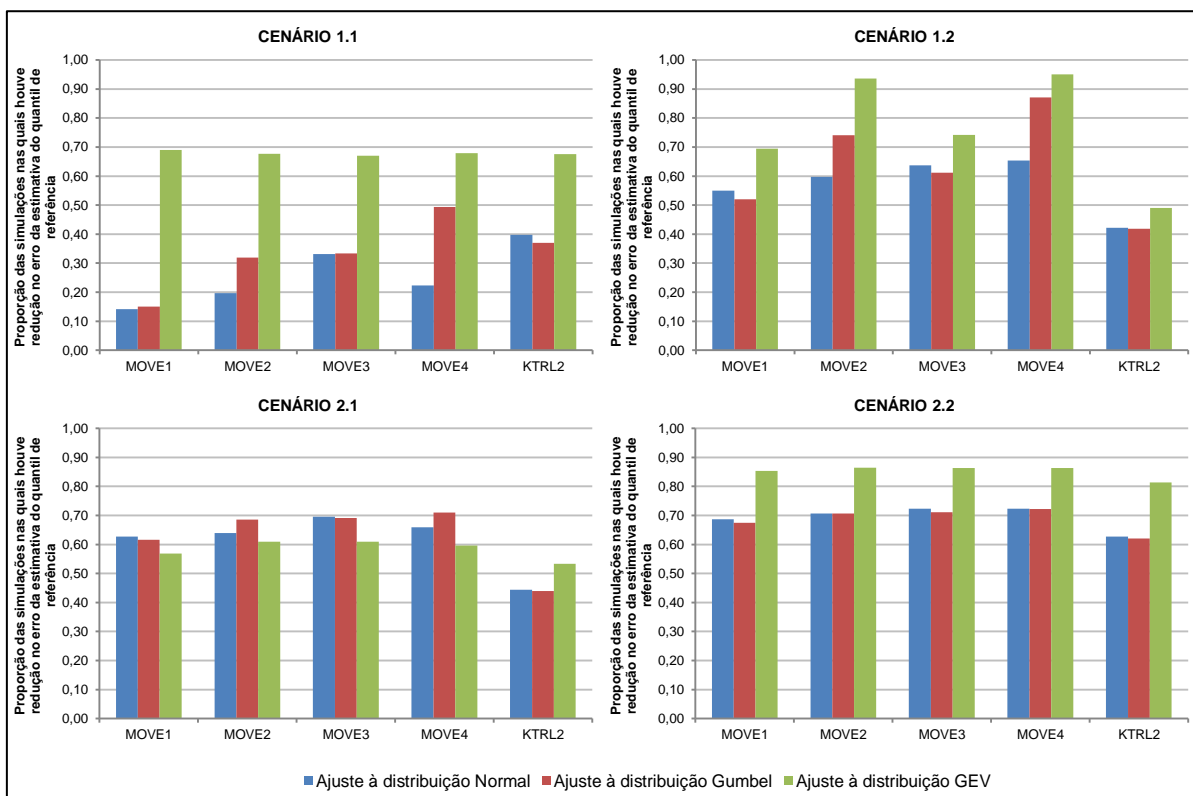


Figura 5.33 – Resultados das simulações de *bootstrap*

Ao analisar a Figura 5.33, observa-se, primeiramente, que, no Cenário 1.1, os resultados são significativamente piores do que nos demais cenários, para as distribuições Normal e Gumbel. Atribui-se esse resultado aos seguintes fatores: i) o coeficiente de correlação mais baixo do que nos demais cenários (0,38 para toda a série histórica). Apesar de se tratar de uma simulação de *bootstrap*, com coeficiente de correlação variável, em um número maior de simulações houve valores baixos, em comparação com os demais cenários; e ii) o comportamento do coeficiente de assimetria nas simulações de *bootstrap*, que favoreceu os resultados para a distribuição GEV. Como foram analisadas precipitações diárias máximas anuais, a relação física entre a série curta e a série de referência não é tão clara. Apesar disso, ao estimar o quantil de referência por meio da distribuição GEV, houve redução significativa do erro em relação à estimativa a partir da série curta. Isso se deve ao aumento da confiabilidade do coeficiente de assimetria, afinal, no Cenário 1.1, $n_2 = 33$.

Os resultados da Figura 5.33 não caracterizaram bem as técnicas, devido aos desempenhos ora próximos, ora distantes entre si. Contudo, nos Cenários 1.2 e 2.1, é possível notar que a técnica KTRL2 possui desempenho inferior às demais técnicas, o que se deve ao valor de n_1 adotado, de apenas 10.

Com exceção do Cenário 1.1, as técnicas tiveram bom desempenho, com redução do erro da estimativa do quantil, em relação à série curta, em pelo menos 60% das amostras.

Por fim, é possível concluir que o modelo utilizado para a caracterização das técnicas, cujos resultados são apresentados no item 5.1, não é capaz de representar, em sua totalidade, a realidade dos dados hidrológicos. Além da limitação óbvia de não representar eventual relação física entre a série curta e a série de referência, os dados reais apresentam características inesperadas, como mudanças na variabilidade dos dados ao longo da série histórica (ver item 5.3.1.2), ou mesmo não estacionariedade dos dados, erros sistemáticos, entre outros.

6 CONCLUSÕES E RECOMENDAÇÕES

Na presente dissertação, apresentou-se um estudo com o objetivo de caracterizar algumas das técnicas de extensão de séries hidrológicas e apresentar critérios para a aplicação no sentido de estimar quantis máximos a partir de séries hidrológicas reduzidas. A partir dos resultados apresentados, é possível concluir que:

- As técnicas desenvolvidas dentro do contexto da extensão de séries hidrológicas possuem desempenho superior às técnicas de regressão (inclusive os Modelos Lineares Generalizados) na estimativa dos momentos populacionais. Esse fato deve-se, principalmente, à variabilidade da série estendida imposta pelas técnicas de extensão. Portanto, não é recomendada a estimativa de quantis de séries estendidas por meio das técnicas OLS, KTRL e GLM;
- Para a estimativa da média populacional, de modo geral, as técnicas mostraram-se não enviesadas para dados normais e levemente enviesadas no caso das demais distribuições. Dados com assimetria positiva levaram a estimativas superiores da média, e o contrário também é verdadeiro. Apenas a técnica KTRL apresentou viés no sentido oposto ao coeficiente de assimetria. A técnica MOVE2 não gerou séries estendidas com média próxima da média populacional, enquanto as técnicas OLS, MOVE3 e MOVE4 levaram aos melhores resultados;
- A estimativa da variância populacional proporcionou a maior diferenciação do desempenho das técnicas de extensão. As técnicas OLS, KTRL, RLOC e GLM mostraram que não podem ser aplicadas quando o objetivo da extensão for a obtenção de uma estimativa melhor da variância populacional. As técnicas KTRL2 e MOVE1 apresentaram desempenho satisfatório em casos particulares, sendo melhor o desempenho das técnicas MOVE2, MOVE3 e MOVE4;
- Com relação à estimativa do coeficiente de assimetria, conclui-se que, para dados NOR, todas as técnicas, com a exceção dos Modelos Lineares Generalizados, geraram séries estendidas sem viés na estimativa da assimetria populacional. Para dados com distribuição assimétrica, todas as técnicas retornaram viés significativo, porém inferior ao viés da série curta, com sinal oposto à assimetria da distribuição.
- No que tange ao erro da estimativa, apenas as técnicas MOVE e KTRL2 se destacaram, com valor esperado do coeficiente de assimetria próximo daquele para a

série completa, independente do cenário analisado. Pontua-se aqui que, nessas técnicas, a influência do coeficiente de correlação no erro da estimativa da assimetria populacional foi bastante reduzida;

- Na estimativa dos registros individuais, as técnicas desenvolvidas no contexto da extensão das séries hidrológicas (MOVE1 a MOVE4, KTRL2 e RLOC) apresentaram estimativas menores do que a técnica OLS, para valores baixos de x , e maiores, para valores altos de x . As estimativas da técnica KTRL apresentam-se transladadas em relação à técnica OLS, enquanto a técnica GLM retorna estimativas muito próximas da última, com sutis diferenças nos extremos. Dentre as técnicas aqui estudadas, a MOVE1 destaca-se pela sua característica de pressupor erros de observação nas amostras de x e y , o que também pode levar a resultados positivos em situações práticas;
- A estimativa dos percentis foi fortemente influenciada pelo coeficiente de assimetria dos dados. Para as distribuições de 2 parâmetros analisadas (Normal e Gumbel), o viés das estimativas flutuou ao redor de zero. Já para as distribuições de 3 parâmetros, que envolvem a estimativa da assimetria populacional no ajuste, o viés não ficou centralizado em zero;
- A estimativa dos percentis para as distribuições de dois parâmetros ficou caracterizada por vieses e erros elevados para as técnicas de regressão (OLS, KTRL e GLM). A técnica com menor viés na estimativa dos percentis foi a KTRL2, apesar de outras técnicas também apresentarem viés inferior à série curta. É possível concluir que a extensão das séries hidrológicas utilizando as técnicas KTRL2, MOVE e RLOC é capaz de melhorar a estimativa dos percentis populacionais em relação à série curta. No entanto, essa conclusão não deve ser generalizada, permanecendo como recomendação avaliar, caso a caso, o desempenho das técnicas;
- Em síntese, a distribuição de probabilidades dos dados interfere significativamente apenas na estimativa dos percentis e do coeficiente de assimetria (e, por conseguinte, na estimativa dos quantis). Para a estimativa dos quantis para as distribuições de três parâmetros, utilizando o método dos momentos, é necessário a estimativa da assimetria amostral. Tendo em vista que o coeficiente de assimetria das séries estendidas é menos enviesado do que a mesma medida descritiva da série curta, um

número maior de técnicas pode ser considerado aplicável para a extensão de séries hidrológicas reduzidas que podem ser modeladas por distribuições de três parâmetros;

- O critério elaborado para a extensão de séries hidrológicas, com o objetivo de melhorar a estimativa dos quantis populacionais, reúne técnicas que geram séries com desempenho melhor do que a série curta com relação à média e variância (dados NOR e GUM) e também com relação ao coeficiente de assimetria (dados PE3+ e PE3-). Esse mesmo critério também pode ser utilizado outras situações nas quais a estimativa dos momentos populacionais é necessária;
- Apesar de as técnicas abordarem apenas aspectos estatísticos, sua aplicação, bem como a escolha da estação de referência, não deve se basear apenas no coeficiente de correlação. A relação física entre as séries curta e de referência é fundamental para obter bons resultados na extensão. Portanto, pode-se esperar que os melhores resultados sejam obtidos na aplicação utilizando séries fluviométricas em uma mesma bacia hidrográfica. A aplicação para séries pluviométricas aproveita-se do princípio de que o comportamento estatístico regional é homogêneo;
- O critério elaborado, apesar de definido para algumas distribuições específicas, pode ser aplicado para quaisquer distribuições que representem o mesmo fenômeno. O critério elaborado para dados GUM, por exemplo, é aplicável para variáveis hidrológicas máximas; e
- A aplicação do critério elaborado não elimina a necessidade de analisar as medidas descritivas de outros postos de monitoramento próximos à hidrometeorológica que originou a série curta. Na verdade, esse procedimento pode ser considerado fundamental para detectar erros na escolha da estação de referência ou para escolher a técnica de extensão com melhor desempenho, entre aquelas definidas como aplicáveis. Recomenda-se, também, que durante o processo de escolha da estação de referência, seja levada em consideração a evolução temporal das medidas descritivas mais relevantes na análise de frequência (média, variância e coeficiente de assimetria), que pode influenciar no ajuste da técnica de extensão adotada ou na estimativa dos registros não observados da série curta.

De modo geral, é possível concluir que as técnicas de extensão de séries hidrológicas mostraram-se como uma ferramenta útil na extensão das séries hidrológicas reduzidas, com o objetivo de estimar quantis máximos de precipitação e vazão. É importante mencionar que

parte significativa da pesquisa, incluindo a caracterização das técnicas e a definição dos critérios para sua aplicação, foi desenvolvida sobre amostras geradas sinteticamente. Essa ressalva deve ser conhecida durante a aplicação das técnicas. Aspectos como a não estacionariedade dos dados, previsibilidade das séries hidrológicas em estudo e a homogeneidade entre as séries hidrológicas analisadas devem ser investigados em uma etapa anterior à própria extensão das séries hidrológicas reduzidas.

Apesar do enfoque da pesquisa na estimativa de quantis máximos a partir de séries hidrológicas reduzidas, muitos dos resultados aqui apresentados podem ser utilizados na extensão das séries hidrológicas completas. Neste caso, recomenda-se a leitura da Revisão da Literatura da presente dissertação, bem como das principais referências sobre o assunto.

É importante pontuar que, nesta pesquisa, não foi analisado o desempenho das técnicas frente à presença de *outliers*, bem como em séries hidrológicas que possuem autocorrelação. Além disso, muitas outras técnicas que podem ser utilizadas para a extensão de séries hidrológicas não foram estudadas aqui, tendo em vista o enfoque dado às técnicas desenvolvidas no contexto da Engenharia de Recursos Hídricos (com exceção dos modelos de regressão OLS, KTRL e GLM).

Por fim, sugere-se para trabalhos futuros:

- Analisar o desempenho das técnicas de extensão aqui estudadas sob cenários de amostras contendo *outliers* e duas ou mais séries curtas e de referência;
- Avaliar se uso das técnicas de extensão de séries hidrológicas aqui estudadas, combinado com outros métodos de estimação dos parâmetros das distribuições (Máxima Verossimilhança e Momentos-L, por exemplo), pode ser mais vantajoso do que aplicando o Método dos Momentos;
- Avaliar o desempenho de outras técnicas que não foram analisadas aqui, como Redes Neurais Artificiais e Método dos Mínimos Quadrados Generalizado;
- Avaliar o desempenho de modelos chuva-vazão para a extensão de séries hidrológicas;
- Avaliar o desempenho de modelos estocásticos para a extensão cronológica das séries hidrológicas completas, em comparação com as técnicas de extensão aqui apresentadas; e

- Realizar a aplicação das técnicas de extensão aqui apresentadas sobre um número maior de séries hidrológicas reais, por meio de rotinas automatizadas para a seleção das estações hidrometeorológicas em estudo.

REFERÊNCIAS

- ADLOUNI, S. E.; BOBÉE; B.; OUARDA, T. B. M. J. On the tails of extreme events distributions in hydrology. *J Hydrol*, v. 355, n. 1, p. 16-33, 2008.
- AGÊNCIA NACIONAL DE ÁGUAS (ANA). Cadernos de Capacitação em Recursos Hídricos. In: _____. Outorga de Direito de Uso de Recursos Hídricos. Brasília: 2011. v. 6.
- AGÊNCIA NACIONAL DE ÁGUAS. Evolução da Rede Hidrometeorológica Nacional. Brasília: Agência Nacional de Águas, 2007. 15 p.
- AGÊNCIA NACIONAL DE ÁGUAS. Inventário das Estações Fluviométricas. 2. ed. Brasília: Agência Nacional de Águas, 2009a. 195 p.
- AGÊNCIA NACIONAL DE ÁGUAS. Inventário das Estações Pluviométricas. 2. ed. Brasília: Agência Nacional de Águas, 2009b. 331 p.
- AGÊNCIA NACIONAL DE ÁGUAS. Orientações para consistência de dados pluviométricos: Versão julho de 2012. Brasília: Agência Nacional de Águas, 2012. Disponível em: <<http://arquivos.ana.gov.br/infohidrologicas/cadastro/OrientacoesParaConsistenciaDadosFluviometricos-VersaoJul12.pdf>>. Acesso em: 04 Abr. 2016.
- ALBEK, E. Estimation of point and diffuse contaminant loads to streams by non-parametric regression analysis of monitoring data. *Water Air Soil Poll*, v. 147, n. 1, p. 229-243, 2003.
- AMBROSINO, C.; CHANDLER, R. E.; TODD, M. Southern African monthly rainfall variability: An analysis based on generalized linear models. *J Climate*, v. 24, p. 4600-4617, 2011.
- BENSON, M. A. Evolution of methods for evaluating the occurrence of floods. *US Geological Survey Water-Supply Paper*, v. 1580, p. A1-A30, 1960.
- BENSON, M. A. Uniform flood-frequency estimating methods for federal agencies. *Water Resour Res*, v. 4, n 5, 1968.
- BOBÉE, B.; ROBITAILLE, R. Correction of bias in the estimation of the coefficient of skewness. *Water Resour Res*, v. 11, n. 6, p. 851-854, 1975.
- BOWERS, M. C.; TUNG, W. W.; GAO, J. B. On the distributions of seasonal river flows: Lognormal or power law? *Water Resour Res*, v. 48, n. 5, p. 1-12, 2012.
- BRYN, G.; HUBERT, M.; STRUYF, A. A robust measure of skewness. *J Comput Graph Stat*, v. 13, n. 4, p. 996-1017, 2004.
- BUSTOS, O. H.; FLESIA, A. G.; FRERY, A. C. Generalized Method for Sampling Spatially Correlated Heterogeneous Speckled Imagery. *EURASIP J Adv Signal Process.*, v. 2001, n. 2, p. 158–200, Jun. 2001.
- CAVANAUGH, N. R.; GERSHUNOV, A.; PANORSKA, A. K.; KOZUBOWSKI, T. J. The probability distribution of intense precipitation. *Geophys Res Lett*, v. 42, n. 5, p. 1-8, 2015.
- CHOWDURY, J. U.; STEDINGER, J. R. Confidence interval for design floods with estimated skew coefficient. *J Hydraul Eng*, v. 117, n. 7, p. 811-831, 1991.
- CONOVER, W. L. Practical nonparametric statistics. 2. ed. New York: John Wiley and Sons, 1980. 493 p.

- COSTA, K. T. Avaliação de Distribuições de Probabilidades das Vazões Médias Diárias Máximas Anuais do Brasil. Dissertação (Mestrado em Saneamento, Meio Ambiente e Recursos Hídricos) – Escola de Engenharia. Universidade Federal de Minas Gerais. Belo Horizonte, 2014, 195 p.
- DECKER, K. M. The Monte Carlo method in science and engineering: theory and application. *Comput Method Appl M*, v. 89, n. 1, p. 463-483, 1991.
- DEPINÉ, H.; CASTRO, N. M. R.; PINHEIRO, A.; PEDROLLO, O. Preenchimento de falhas de dados horários de precipitação utilizando Redes Neurais Artificiais. *Revista Brasileira de Recursos Hídricos*, v. 19, n. 1, p. 51-63, 2014.
- DÉRY, S.; MLYNOWSKI, T. J.; HERNÁNDEZ-HENRÍQUEZ, M. A.; STRANEO, F. Interannual variability and interdecadal trends in Hudson Bay streamflow. *J Marine Syst*, v. 88, n. 3, p. 341-351, 2011.
- DUAN, W.; TAKARA, K.; HE, B.; LUO, P.; NOVER, D.; YAMASHIKI, Y. Spatial and temporal trends in estimates of nutrient and suspended sediment loads in the Ishikari River, Japan, 1985 to 2010. *Sci Total Environ*, v. 461-462, p. 499-508, 2013.
- DUKIĆ, V. M.; MARIĆ, N. On minimum correlation in construction of multivariate distributions. *Phys. Rev. E.*, v. 87, n. 1, Fev. 2013.
- ELETOBRÁS. Diretrizes para estudos e projetos de pequenas centrais hidrelétricas. Eletrobrás, 2000. 458 p.
- ENG, K.; KIANG, J. E.; CHEN, Y. Y.; CARLISLE, D. M.; GRANATO, G. E. Causes of systematic over- or underestimation of low streamflows by use of index-streamgauge approaches in the United States. *Hydrol Processes*, v. 25, n. 14, p. 2211-2220, 2011.
- FIERING, M. B. Use of correlation to improve estimates of the mean and variance. *US Geol Surv Prof Paper*, v. 434-C, p. C1-C9, 1963.
- FISHER, R. A.; TIPPETT, L. H. C. Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Math Proc Camb Phil Soc*, v. 24, n. 2, p. 180-190, 1928.
- FOX, J. Applied Regression Analysis and Generalized Linear Models. 3. ed. Sage, 2015. 688 p.
- GENEST, C.; FAVRE, A. C. Everything you always wanted to know about copula modeling but were afraid to ask. *J Hydrol Eng*, v. 12, n. 4, p. 347-368, 2007.
- GENZ, A. Comparison of Methods for the Computation of Multivariate Normal Probabilities. *Computing Science and Statistics*, v. 25, p. 400-405, 1993.
- GENZ, A. Numerical Computation of Multivariate Normal Probabilities. *J. Comp. Graph. Stat.*, v. 1, n. 2, p. 141-149, Jun. 1992.
- GENZ, A. Numerical Computation of Rectangular Bivariate and Trivariate Normal and t Probabilities. *Stat Comput*, v. 14, n. 3, p. 251-260, Ago. 2004.
- GENZ, A.; BRETZ, F. Computation of Multivariate Normal and t Probabilities. New York: Springer, 2009. New York, 2009. 124 p.
- GOTTSCHALK, L.; YU, K.; LEBLOIS, E.; XIONG, L. Statistics of low flow: theoretical derivation of the distribution of minimum streamflow series. *J Hydrol*, v. 481, p. 204-219, 2013.

- GRANATO, G.E. Kendall-Theil Robust Line (KTRLine—version 1.0): A visual basic program for calculating and graphing robust nonparametric estimates of linear-regression coefficients between two continuous variables. Techniques and Methods of the U.S. Geological Survey, Book 4, Chapter A7. [Richmond]: U.S. Geological Survey, 2006. 31 p.
- GREENWOOD, J. A.; LANDWEHR, J. M.; MATALAS, N. C.; WALLIS, J. R. Probability weighted moments: definition and relation to parameters of several distributions expressible in inverse form. *Water Resour Res*, v. 15, n 5, p. 1049-1054, 1979.
- GRIFFIS, V. W.; STEDINGER, J. R. Log-Pearson type 3 distribution and its application in flood frequency analysis. III: Sample skew and weighted skew estimators. *J Hydrol Eng*, v. 14, n. 2, p. 121-130, 2009.
- GRIFFIS, V. W.; STEDINGER, J. R.; COHN, T. A. Log Pearson type 3 quantile estimators with regional skew information and low outlier adjustments. *Water Resour Res*, v. 40, n. 7, p. 1-17, 2004.
- GRYGIER, J. C.; STEDINGER, J. R.; YIN, H. B. A generalized Maintenance of Variance Extension procedure for extending correlated series. *Water Resour Res*, v. 25, n. 23, p. 345-349, 1989.
- HELSEL, D. R.; HIRSCH, R. M. Statistical methods in water resources. Techniques of Water Resources Investigations of the United States Geological Survey, Book 4, Chapter A3. U.S. Geological Survey. [S.l.]: Set. 2002. Disponível em: <http://water.usgs.gov/pubs/twri/twri4a3/>. Acesso em: 16 Jan 2016.
- HERNÁNDEZ-HENRIQUEZ, M. A.; MLYNOWSKI, T. J.; DÉRY, S. J. Reconstructing the natural streamflow of a regulated river: a case study of La Grande Rivière, Québec, Canada. *Can Water Resour J*, v. 35, n. 3. p.301-316, 2010.
- HIRSCH, R. M. A Comparison of Four Streamflow Record Extension Techniques. *Water Resour Res*, v. 18, n. 4, p. 1081-1088, Ago. 1982.
- HIRSCH, R. M. An evaluation of some record reconstruction techniques. *Water Resour Res*, v. 15, n. 6, p.1781-1790, 1979.
- HIRSCH, R.M.; ALEXANDER, R. B.; SMITH, R. A. Selection of methods for the detection and estimation of trends in water quality. *Water Resour Res*, v. 27, n. 5, p. 803-813, 1991.
- HOSKING, J. R. M.; WALLIS, J. R. Parameter and quantile estimation for the Generalized Pareto distribution. *Technometrics*, v. 29, n. 3, p. 339-349, 1987.
- JAIN, Y. K.; BHANDARE, S. K. Min-Max normalization based data perturbation method for privacy protection. *Int J Comput Comm Tec*, v. 2, n. 8, p. 45-50, 2011.
- JAMES, F. Monte Carlo theory and practice. *Rep Prog Phys*, v. 43, n. 9, p. 1145-1189, 1980.
- JENKINSON, A. F. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q J Roy Meteor Soc*, v. 81, n. 348, p. 158-171, 1955.
- JIA, Y.; CULVER, T. Bootstrapped artificial neural networks for synthetic flow generation with a small data sample. *J Hydrol*, v. 331, n. 3, p. 580-590, 2006.
- KATZ, R. W.; PARLANGE, M. B.; NAVEAU, P. Statistics of extremes in hydrology. *Adv Water Res*, v. 25, n. 8, p. 1287-1304, 2002.
- KHALIL, B.; ADAMOWSKI, J. Record extension for short-gauged water quality parameters using a newly proposed robust version of the line of organic correlation technique. *Hydrol Earth Syst Sci Discuss*, v. 16, n. 7, p. 4667–4702, 2012.

- KHALIL, B.; AWADALLAH, A. G.; ADAMOWSKI, J. A novel record-extension technique for water quality variables based on L-moments. *Water Air Soil Pollut*, v. 227, n. 6, p. 1-20, 2016.
- KHALIL, B.; OUARDA, T. B. M. J.; ST-HILAIRE, A. Comparison of Record-Extension Techniques for Water Quality Variables. *Water Resour Manage*, v. 26, n. 14, p. 4259–4280, 2012.
- KHALIL, B.; ADAMOWSKI, J. Evaluation of the Performance of Eight Record-Extension Techniques Under Different Levels of Association, Presence of Outliers and Different Sizes of Concurrent Records: A Monte Carlo Study. *Water Resour Manage*, v. 28, n. 14, p. 5139–5155, 2014a.
- KHALIL, B.; ADAMOWSKI, J. Comparison of OLS, ANN, KTRL, KTRL2, RLOC, and MOVE as record-extension techniques for water quality variables. *Water Air Soil Pollut*, v. 225, n. 6, Jun. 2014b.
- KIM, T. H.; WHITE, H. On more robust estimation of skewness and kurtosis. *Financ Res Lett*, v. 1, n. 1, p. 56-73, 2004.
- LORD, F. M. Estimation of parameters from incomplete data. *J Am Stat Assoc*, v. 50, n. 271, p. 870-876, 1955.
- MALAMUD, B. D.; TURCOTTE, D. L. The applicability of power-law frequency statistics to floods. *J Hydrol*, v. 332, n. 1, p. 168-180, 2006.
- MATALAS, N. C.; JACOBS, B. A correlation procedure for augmenting hydrologic data. *US Geol Surv Prof Paper*, v. 434, p. E1–E7, 1964.
- MATTHAI, A. Estimation of parameters from incomplete data with application to design of sample surveys. *Indian J Stat*, v. 11, n. 2, p. 145-152, 1951.
- MCCUEN, R. H. Generalized flood skew: map versus watershed skew. *J Hydrol Eng*, v. 6, n. 4, 2001.
- MI, X.; MIWA, T.; HOTHORN, T. mvtnorm: New Numerical Algorithm for Multivariate Normal Probabilites. *The R Journal*, v. 2009, n. 1, p. 37-39, Mai. 2009.
- MUN, J. Modeling Risk – Applying Monte Carlo Simulation, Real Options Analysis, Forecasting, and Optimization Techniques. New Jersey: Wiley Finance Series, 2006. 623 p.
- NAGHETTINI, M.; PINTO, E. J. A. Hidrologia estatística. Belo Horizonte: CPRM, 2007. 552p.
- NELDER, J. A.; WEDDERBURN, R. W. M. Generalized Linear Models. *J Roy Statist Soc Ser A*, v. 135, n. 3, p. 370-384, 1972.
- OLIVEIRA, L. F. C.; FIOREZE, A. P.; MEDEIROS, A. M. M.; SILVA, M. A. A. Comparação de metodologias de preenchimento de falhas de séries históricas de precipitação pluvial anual. *Rev Bras Eng Agríc Ambient*, v. 14, n. 11, p. 1186-1192, 2010.
- OLSSON, O.; GASSMANN, M.; WEGERICH, K; BAUER, M. Identification of the effective water availability from streamflows in the Zerafshan river basin, Central Asia. *J Hydrol*, v. 390, n. 3, p. 190-197, 2010.
- PAPALEXIOU, S. M.; KOUTSOYIANNIS, D. Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resour Res*, v. 49, n. 1, 187–201, 2013.

- PAPALEXIOU, S. M.; KOUTSOYIANNIS, D.; MAKROPOULOS, C. How extreme is extreme? An assessment of daily rainfall distribution tails. *Hydrol Earth Syst Sci*, v. 17, n. 2, p. 851-862, 2013.
- REINFELDS, I.; BROOKS, A. J.; HAEUSLER, T.; WILLIAMS, S. Temporal patterns and effects of surface-water diversions on daily flows and aquatic habitats: Bega-Bemboka River, New South Wales, Australia. *Geogr Res*, v. 44, n. 4, p. 401-417, 2006.
- RENARD, B.; LANG, M. Use of a gaussian copula for multivariate extreme value analysis: some case studies in hydrology. *Adv Water Resour*, v. 30, n. 4, p. 897-912, 2007.
- ROSBJERG, D.; MADNESEN, H.; RASMUSSEN, P. F. Prediction in partial duration series with Generalized Pareto-distributed exceedances. *Water Resour Res*, v. 28, n. 11, p. 3001-3010, 1992.
- ROSS, S. Introduction to Probability Models. 9. ed. San Diego: Elsevier, 2007. 801 p.
- RUBINSTEIN, R. Y. Simulation and the Monte Carlo Method. 2. ed. Wiley-Interscience, 1981. 298 p.
- SARABIA, J. M.; DÉNIZ, E. G. Construction of multivariate distributions: a review of some recent results. *SORT*, v. 32, n. 1, p. 3-36, 2008.
- SCHWEIZER, B.; SKLAR, A. Probability distributions with given margins: note on a paper by Finch and Groblicki. *Found Phys*, v. 16, n. 10, 1986.
- SEN, P. K. Estimates of the Regression Coefficient Based on Kendall's Tau. *J Amer Statist Assoc*, v. 63, n. 324, p. 1379-1389, Dez. 1968.
- SERINALDI, F.; KILSBY, C. G. Rainfall extremes: toward reconciliation after the battle of distributions. *Water Resour Res*, v. 50, n. 1, p. 336-352, 2014.
- SMAKHTIN, V. U. Low flow hydrology: a review. *J Hydrol*, v. 240, n. 3, p. 147,186, 2001.
- STAHL, K.; HISDAL, H.; HANNAFORD, J.; TALLASKEN, L. M.; VAN LANEN, H. A. J.; SAUQUET, E.; DEMUTH, S.; FENDEKOVA, M.; JÓDAR, J. Streamflow trends in Europe: evidence from a dataset of near-natural catchments. *Hydrol Earth Syst Sc*, v. 14, n. 12, p. 2367-2382, 2010.
- STENDINGER, J. R.; TASKER, G. D. Regional hydrologic analysis 1: Ordinary, weighted, and generalized least squares compared. *Water Resour Res*, v. 21, n. 9, p. 1421-1432, 1985.
- STIGLER, S. M. The epic story of maximum likelihood. *Stat Sc*, v. 22, n. 4, p. 598-620, 2007.
- TARAWNEH, Z. S. Augmenting short hydrological records to improve water resources studies. In: INTERNATIONAL WATER TECHNOLOGY CONFERENCE, 15, 2011, Alexandria: 2011. 13 p.
- TASKER, G. D.; STEDINGER, J. R. Regional skew with weighted LS regression. *J Water Resour Plan Manag*, v. 112, n. 2, p. 225-237, 1986.
- THEIL H. A rank-invariant method of linear and polynomial regression analysis, 1, 2, and 3. *Ned AkadWentsch Proc*, v. 53, p. 386-392, 521-525, e 1397-1412, 1950.
- TUCCI, C. E. Hidrologia – Ciência e Aplicação. 4. ed. Porto Alegre: Editora UFRGS, 2009, 943 p.
- UNITED STATES WATER RESOURCES COUNCIL. Guidelines for Determining Flood Flow Frequency: Bulletin 17B. Washington (DC): US WRC, 1981, 194 p.

- VAN DER HEIJDEN, F.; DUIN, R. P. W.; DE RIDDER, D.; TAX, D. M. J. Classification, parameter estimation and state estimation: An engineering approach using Matlab. Chichester: John Wiley & Sons, 2004. 423 p.
- VAN GELDER, P. H. A. J. M. Statistical Estimation Methods in Hydrological Engineering, Irkutsk: Publishing House of the Institute of Geography SB RAS, 2004, 41p.
- VOGEL, R. M.; STEDINGER, J. R. Minimum variance streamflow record augmentation procedures. *Water Resour Res*, v. 21, n. 5, p. 715–723, 1985.
- WEIBULL, W. A statistical theory of the strength of materials. *Ing Vet Ak Handl*, n. 151, p. 1-45, 1939.
- WILKS, S. S. Moments and distributions of estimates of population parameters from fragmentary samples. *Ann Math Statist*, v. 3, n. 3, p. 163-195, 1932.
- WOOD, S. N. Generalized Additive Models: An Introduction with R. Boca Raton: Chapman & Hall/CRC, 2006. 392 p.
- YAHAV, I.; SHMUELI, G. On generating multivariate Poisson data in management science applications. *Appl Stochastic Models Bus*, v. 28, n. 1, p. 91-102, 2011.
- YALCIN, G. C.; RABASSA, P.; BECK, C. Extreme event statistics of daily rainfall: Dynamical systems approach. *J Phys A: Math Theor*, v. 49, n. 15, p. 1-18, 2016.
- YANG, C.; CHANDLER, R. E.; ISHAM, V. S.; WHEATER, H. S. Spatial-temporal rainfall simulation using generalized linear models. *Water Resour Res*, v. 41, n. 11, p. 1-13, 2005.
- YEVJEVICH, V.; OBEYSEKERA, J. T. B. Estimation of skewness of hydrologic variables. *Water Resour Res*, v. 20, n. 7, p. 935-943, 1984.
- YUE, S.; HASHINO, M. Probability distribution of annual, seasonal and monthly precipitation in Japan. *Hydrol Sc*, v. 52, n. 5, 2007.

APÊNDICE A

***SCRIPTS* PARA AJUSTE DAS TÉCNICAS DE
EXTENSÃO DE SÉRIES HIDROLÓGICAS NA
LINGUAGEM DE PROGRAMAÇÃO R**

Os *scripts* apresentados a seguir foram utilizados no decorrer da pesquisa e consideram as seguintes informações:

- A série curta e a série de referência formam uma matriz com $n_1 + n_2$ linhas e 2 colunas. Esta matriz é apresentada como o objeto *data*;
- Na coluna 1 constam os dados da série de referência e na coluna 2 constam os dados da série curta;
- Na coluna 2 (série curta) são utilizados apenas os $n_1 = N_y$ primeiros termos; e
- Os resultados são apresentados em duas variáveis: i) *a*, o intercepto da equação da reta da técnica de extensão; e ii) *b*, o coeficiente angular da equação da reta da técnica de extensão.

KTRL2

```
perc = matrix(1:38, nrow = 19, ncol = 2)
perc[,1] = as.numeric(quantile(data[1:Ny,1], c(0.05, 0.10, 0.15, 0.20,
      0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60,
      0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95)))
perc[,2] = as.numeric(quantile(data[1:Ny,2], c(0.05, 0.10, 0.15, 0.20,
      0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60,
      0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95)))

v_med = NULL
for (i in 1:19)
{
  for (j in 2:19)
  {
    m = length(v_med)
    v_med[m+1] = ifelse(i<j,(perc[j,2]-perc[i,2])/(perc[j,1]-perc[i,1]),NA)
  }
}
bad = is.na(v_med)
v_med = v_med[!bad]
b = median(v_med)

a = median(data[1:Ny,2])-b*median(data[1:Ny,1])

parameters = c(a, b)
```

MOVE1

```
n1 = Ny
n2 = length(data[,1]) - n1
Nx = length(data[,1])
bols = cor(x = data[1:Ny,1], y = data[1:Ny,2])*sd(data[1:Ny,2])/sd(data[1:Ny,1])
sy1 = sd(data[1:Ny,2])
sx1 = sd(data[1:Ny,1])
sx2 = sd(data[(Ny+1):Nx,1])
r = bols*sx1/sy1
my1 = mean(data[1:Ny,2])
mx1 = mean(data[1:Ny,1])

b = sy1/sx1
a = my1 - b*mx1

parameters = c(a, b)
```

MOVE2

```
n1 = Ny
n2 = length(data[,1]) - n1
Nx = length(data[,1])
bols = cor(x = data[1:Ny,1], y = data[1:Ny,2])*sd(data[1:Ny,2])/sd(data[1:Ny,1])
sy1 = sd(data[1:Ny,2])
sx1 = sd(data[1:Ny,1])
sx2 = sd(data[(Ny+1):Nx,1])
r = bols*sx1/sy1

a = sqrt(n2*(n1-4)*(n1-1)/((n2-1)*(n1-3)*(n1-2)))
my1 = mean(data[1:Ny,2])
mx2 = mean(data[(Ny+1):Nx,1])
mx1 = mean(data[1:Ny,1])
mi = my1 + ((n2)/(n1+n2))*bols*(mx2 - mx1)
sd2 = (1/(n1+n2-1))*((n1-1)*sy1^2 + (n2-1)*(bols^2)*sx2^2 +
(n2-1)*(a^2)*(1-r^2)*sy1^2 + (n1*n2)/(n1+n2)*(bols^2)*(mx2-mx1)^2)

a = mi
b = sqrt(sd2)/sd(data[,1])
a = a - b*mean(data[,1])

parameters = c(a, b)
```

MOVE3

```
n1 = Ny
n2 = length(data[,1]) - n1
Nx = length(data[,1])
bols = cor(x = data[1:Ny,1], y = data[1:Ny,2])*sd(data[1:Ny,2])/sd(data[1:Ny,1])
sy1 = sd(data[1:Ny,2])
sx1 = sd(data[1:Ny,1])
sx2 = sd(data[(Ny+1):Nx,1])
r = bols*sx1/sy1

a = sqrt(n2*(n1-4)*(n1-1)/((n2-1)*(n1-3)*(n1-2)))

my1 = mean(data[1:Ny,2])
mx2 = mean(data[(Ny+1):Nx,1])
mx1 = mean(data[1:Ny,1])
mi = my1 + ((n2)/(n1+n2))*bols*(mx2 - mx1)

sd2 = (1/(n1+n2-1))*((n1-1)*sy1^2 + (n2-1)*(bols^2)*sx2^2 +
(n2-1)*(a^2)*(1-r^2)*sy1^2 + (n1*n2)/(n1+n2)*(bols^2)*(mx2-mx1)^2)

a = ((n1 + n2)*mi - n1*my1)/n2
b = ((n1+n2-1)*sd2-(n1-1)*sy1^2-n1*(my1-mi)^2-n2*(a-mi)^2)/((n2-1)*sx2^2)
b = sqrt(b)
a = a - b*mx2

parameters = c(a, b)
```

MOVE4

```
n1 = Ny
n2 = length(data[,1]) - n1
Nx = length(data[,1])
bols = cor(x = data[1:Ny,1], y = data[1:Ny,2])*sd(data[1:Ny,2])/
      sd(data[1:Ny,1])
sy1 = sd(data[1:Ny,2])
sx1 = sd(data[1:Ny,1])
sx2 = sd(data[(Ny+1):Nx,1])
r = bols*sx1/sy1

a = sqrt(n2*(n1-4)*(n1-1)/((n2-1)*(n1-3)*(n1-2)))
t1 = ((n1-3)*(r^2))/((n1-4)*(r^2)+1)
t2 = (n1-4)*(r^2)/((n1-8.5)*(r^2)+4.5)

my1 = mean(data[1:Ny,2])
mx2 = mean(data[(Ny+1):Nx,1])
mx1 = mean(data[1:Ny,1])
miy = my1 + ((n2)/(n1+n2))*bols*(mx2 - mx1)
mi = my1 + t1*(n2/(n1+n2))*bols*(mx2 - mx1)

sd2 = (1/(n1+n2-1))*((n1-1)*sy1^2 + (n2-1)*(bols^2)*sx2^2 +
      (n2-1)*(a^2)*(1-r^2)*sy1^2 + (n1*n2)/(n1+n2)*(bols^2)*(mx2-mx1)^2)
sd2 = (1-t2)*(sy1^2)+t2*sd2

a = ((n1 + n2)*mi - n1*my1)/n2
b = ((n1+n2-1)*sd2-(n1-1)*sy1^2-n1*(my1-mi)^2-n2*(a-mi)^2)/((n2-1)*sx2^2)
b = sqrt(b)

a = a - b*mx2

parameters = c(a, b)
```

RLOC

```
perc = matrix(1:4, nrow = 2, ncol = 2)
perc[,1] = as.numeric(quantile(data[1:Ny,1], c(0.25,0.75)))
perc[,2] = as.numeric(quantile(data[1:Ny,2], c(0.25,0.75)))

b = (perc[2,2] - perc[1,2])/(perc[2,1] - perc[1,1])

a = median(data[1:Ny,2])-b*median(data[1:Ny,1])

parameters = c(a, b)
```
