

**UNIVERSIDADE FEDERAL DO PARANÁ  
PPGERHA - PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE  
RECURSOS HÍDRICOS E AMBIENTAL**

**DANIEL HENRIQUE MARCO DETZEL**

**MODELAGEM DE SÉRIES HIDROLÓGICAS: UMA ABORDAGEM DE  
MÚLTIPLAS ESCALAS TEMPORAIS**

**CURITIBA**

**2015**

**DANIEL HENRIQUE MARCO DETZEL**

**MODELAGEM DE SÉRIES HIDROLÓGICAS: UMA ABORDAGEM DE  
MÚLTIPLAS ESCALAS TEMPORAIS**

Tese apresentada ao Programa de Pós Graduação em Engenharia de Recursos Hídricos e Ambiental da Universidade Federal do Paraná como requisito parcial para obtenção do título de Doutor em Engenharia de Recursos Hídricos.

Linha de Pesquisa: Hidrologia estocástica e estudos energéticos.

Orientadora: Prof.<sup>a</sup> Dr.<sup>a</sup> Miriam Rita Moro  
Mine

**CURITIBA**

**2015**

---

D484m

Detzel, Daniel Henrique Marco

Modelagem de séries hidrológicas : uma abordagem de múltiplas escalas temporais/ Daniel Henrique Marco Detzel. – Curitiba, 2015.

218 f. : il. color. ; 30 cm.

Tese - Universidade Federal do Paraná, Setor de Tecnologia, Programa de Pós-graduação em Engenharia de Recursos Hídricos e Ambiental, 2015.

Orientador: Miriam Rita Moro Mine .

Bibliografia: p. 185-195.

1. Análise de séries temporais. 2 Análise estocástica. 3. Hidrologia - Modelos matemáticos. I. Universidade Federal do Paraná. II.Mine, Miriam Rita Moro. III. Título.

CDD: 551.48011

---



**TERMO DE APROVAÇÃO**  
**DANIEL HENRIQUE MARCO DETZEL**

**“Modelagem de Séries Hidrológicas: Uma Abordagem de Múltiplas Escalas Temporais”**

Tese aprovada como requisito parcial à obtenção do grau de Doutor, pelo Programa de Pós-Graduação em Engenharia de Recursos Hídricos e Ambiental do Setor de Tecnologia da Universidade Federal do Paraná, pela comissão formada pelos professores:

PRESIDENTE:

**Miriam Rita Moro Mine**  
Universidade Federal do Paraná  
Orientadora

MEMBROS:

**Mauro da Cunha Naghettini**  
Universidade Federal de Minas Gerais

**Reinaldo Castro Souza**  
Pontifícia Universidade Católica do Rio de Janeiro

**Eloy Kaviski**  
Universidade Federal do Paraná

**Marcelo Rodrigues Bessa**  
Universidade Federal do Paraná

**Curitiba, 02 de outubro de 2015**

À Susan, minha inesgotável fonte de  
motivação e inspiração.

## AGRADECIMENTOS

Muito se ouve que um trabalho do porte de um doutorado é feito de forma solitária, ou quase, produto de esforço individual do autor e com participação de seu orientador. A grande realidade é que a realização de um doutorado só é possível com as diversas formas de apoio oferecidas por pessoas queridas presentes na vida do autor. Não raro, essas pessoas não percebem o importante papel que elas desempenharam e, por esse motivo, explico meus mais sinceros agradecimentos:

À minha esposa Susan Farago Pastega Detzel, cujo apoio no decorrer desses quase quatro anos pode ser entendido, pura e simplesmente, como uma grande prova de amor. Mesmo longe da esfera acadêmica, esse suporte foi determinante para a execução e conclusão do trabalho. Ficarei feliz se um dia conseguir, ao menos em parte, retribuí-la.

Ao meu filho Bernardo Pastega Detzel, que nasceu (e cresceu um pouquinho...) durante a elaboração da tese. Você foi uma enorme fonte de inspiração, muito embora sei que demorará alguns anos para conseguir entender o que isso significa.

Aos meus pais, Cesar Divonsir Detzel e Catarina Elisabete Marco Detzel, pelo estímulo e amor dedicados. Os diversos almoços e jantas oferecidos tiveram um papel muito maior do que uma simples reunião de família, servindo como momentos de descontração muito bem-vindos. Também à minha irmã Fernanda Marco Detzel, meu cunhado Adilson Redivo e meu sobrinho Danilo Detzel Redivo, sempre presentes de alguma forma, mesmo que distantes geograficamente.

Aos meus sogros, Douglas Ruy Pastega e Rosemari Farago Pastega, pelo apoio a mim e companhia à Susan e ao Bernardo nos momentos em que me ausentei para trabalhar no doutorado. Da mesma forma, agradeço ao meu cunhado e futuro doutor Diego Farago Pastega pela amizade e pelas conversas, nem sempre técnicas, mas sempre de alto nível, e cujo ápice acontecia nas noites de ano novo.

À minha orientadora, Miriam Rita Moro Mine, pela confiança no meu trabalho. A senhora foi a grande responsável por eu ter chegado onde estou agora. Desde a época do mestrado, sempre norteou meus desenvolvimentos e espero continuar dessa mesma forma. Seu modo de trabalho e orientação me serviram como valiosas lições que carregarei por toda minha vida.

Aos meus amigos Ana Paula Oening, Débora Cintia Marcilio e Lucio de Medeiros, por serem muito mais do que colegas de trabalho. Tenho certeza que formamos um grande time.

Ao PPGERHA, em nome dos coordenadores Tobias Bernward Bleninger e Cristóvão Scapulatempo Vicente Fernandes, por ter aberto as portas do curso para que eu pudesse realizar esse trabalho mesmo sem dedicação exclusiva. Além disso, agradeço por terem me convidado para ser membro do colegiado na forma de representante discente, atividade que tive imenso prazer em exercer e com a qual aprendi muito. Nesse sentido, devo agradecer também aos alunos de doutorado que confiaram em mim para ser sua voz nas decisões de colegiado.

Ao professor Cristóvão Scapulatempo Vicente Fernandes, agora não mais como vice coordenador do PPGERHA, mas como pessoa cuja honestidade e paixão pelo trabalho são únicas e extremamente inspiradoras. Fico feliz em dizer que consegui me espelhar em sua dedicação, mesmo trabalhando em uma área de concentração distinta.

Aos professores Marcelo Rodrigues Bessa e Eloy Kaviski, por aceitar fazer parte da minha comissão de acompanhamento, sempre com ótimas sugestões. Incluo também os professores Mauro da Cunha Naghettini e Reinaldo Castro Souza, por aceitar fazer parte de minhas bancas de avaliação e contribuir enormemente com o trabalho. Novo agradecimento deve ser feito ao professor Marcelo Rodrigues Bessa, desta vez não em caráter técnico, mas pela amizade no decorrer desses anos.

Aos bons amigos que fiz no PPGERHA, os quais resisto em citar os nomes por receio de cometer alguma injustiça. Ainda assim, não posso deixar de dedicar especial agradecimento a Michael Männich, Heloise Knapik, Robinson Ploszai, Luis Miguel Samussone Tomas Buchir e Livia Yu Iwamura.

Aos Institutos Lactec, em nome de Fábio Alessandro Guerra, Alexandre Rasi Aoki e Eduardo Kazumi Yamakawa pelo estímulo e por possibilitar que esse trabalho fosse feito.

Ao secretário do PPGERHA, Celmar de Mattos, pelo auxílio com as questões administrativas do doutorado.

Por fim, mas não menos importante, eu agradeço aos Senseis Roney Alberti e Julio Cesar Bassan, além de todos os amigos que tenho na academia Sol Karatê. Por reviravoltas da vida eu tive que interromper meus quase 16 anos de treino, porém tenho absoluta certeza de que a disciplina que tive para completar meu doutorado teve origem dentro do dojo.

*1. Esforçar-se para o aprimoramento do caráter*

*1. Desenvolver o espírito de luta*

*1. Seguir o caminho da verdade*

*1. Respeitar as regras de cortesia*

*1. Reprimir a coragem insensata*

*Dojo kun*



## RESUMO

Ao longo dos anos, as observações dos diversos fenômenos ocorridos na natureza produziram as séries temporais que servem de base para a elaboração de modelos. No caso específico das séries de vazão, análises evidenciaram peculiaridades associadas a comportamentos não estacionários, tradicionalmente justificados como sendo consequência de alterações no uso do solo de bacias hidrográficas e de variabilidades ou mudanças climáticas. Para contornar essa situação e permitir a aplicação de modelos estocásticos tradicionais, técnicas matemáticas como a representação da alteração do regime hidrológico através de tendências lineares passaram a ser adotadas. Por outro lado, estudos recentes questionam esse artifício principalmente por não ser possível obter uma função determinística que explique adequadamente as tendências observadas. Assim sendo, esse novo enfoque considera que as flutuações observadas nas séries são parte de seu comportamento natural esperado, fruto da interação de eventos que ocorrem em diversas escalas temporais. A classe de modelos estocásticos que trabalha com essas premissas é chamada de modelo de múltiplas escalas temporais (SSS), tema central da presente tese. Até então, o potencial do modelo SSS havia sido apenas explorado em bases quase teóricas através de séries hidrológicas únicas e de excepcional comprimento, o que motivou sua aplicação em séries comuns à hidrologia operacional. Dessa maneira, a hipótese a ser testada é a de que os modelos de SSS são viáveis para utilização na hidrologia operacional, oferecendo representatividade estatística superior quando comparados a formulações tradicionais. Os modelos de múltiplas escalas temporais foram ajustados a seis séries de vazões médias da bacia do rio Iguaçu nas escalas anual e mensal, sendo confrontados com formulações Markovianas [MAR(1)] e de Box & Jenkins (CARIMA), respectivamente, por meio de geração de séries sintéticas. A inovação do trabalho está no algoritmo de médias móveis simétricas dessazonalizadas (SMAD), solução proposta para a geração de séries sintéticas sazonais utilizando os modelos de múltiplas escalas temporais. Previamente à aplicação dos modelos estocásticos, contudo, a tese aborda questões relacionadas com a análise de tendências em séries hidrológicas. Como análise preliminar, as principais técnicas pesquisadas foram aplicadas às séries hidrológicas em questão, mostrando resultados conflitantes que ficam à mercê da subjetividade do hidrólogo. Para as séries sintéticas, os resultados obtidos permitiram confirmar a hipótese levantada principalmente para as séries mensais, nas quais o modelo SSS melhorou a representatividade dos desvios padrão mensais, a estrutura de persistência e a matriz de correlações espaciais.

Palavras-chave: modelo de múltiplas escalas temporais, modelos estocásticos, séries sintéticas, estacionariedade estatística

## ABSTRACT

Over the years, the observation of diverse natural phenomena produced the time series used as input for models. Regarding streamflow series, analyses have shown peculiarities associated with nonstationarity behavior, traditionally explained as a consequence of watersheds land use modifications and climate variability or change. In order to bypass this situation and allow the use of traditional stochastic models, researchers have adopted mathematical tools such as the representation of hydrological regime disturbances by means of linear trends. On the other hand, recent studies have questioned this solution due to the impossibility in obtaining a deterministic function capable to explain these trends. One of these new approaches considers the observed fluctuations as part of a natural phenomenon expected behavior, consequence of events occurring in multiple time scales. The class of stochastic models that considers these assumptions is called simple scaling stochastic processes (SSS), the main subject of this thesis. So far, literature explored SSS potential only in quasi-theoretical basis, using a unique and extraordinary long hydrological time series, which motivated its application in operational hydrology time series. Thus, the hypothesis to be tested is that SSS model is feasible to operational hydrology use, offering better statistical performance when compared to traditional formulations. SSS models were adjusted do six streamflow gauges in Iguazu river basin, southern Brazil. The generated synthetic series were confronted with Markovian [MAR(1)] and Box & Jenkins (CARIMA) models, for annual and monthly scales respectively. This work novelty is the deseasonalized symmetrical moving average algorithm (SMAD), proposed solution to seasonal synthetic series generation using SSS model. Prior to stochastic models results, the thesis considers issues related to hydrological series trend analyses. As a preliminary investigation, the main researched methods were applied to the six considered gauges. The results were contradictory, being dependent on the hydrologist subjectivity. For the synthetic series, the obtained results allowed to confirm the proposed hypothesis mainly for monthly scale, in which SSS model improved the representation of standard deviations, persistence structure and spatial correlation matrix.

Keywords: simple scaling stochastic model, stochastic models, synthetic series, statistical stationarity

## LISTA DE FIGURAS

Figura 1 – Representação de cinco realizações de um processo estocástico com destaque a uma amostra da série temporal .....	38
Figura 2 – Índice de oscilação multidecadal do oceano Atlântico (AMO). As linhas tracejadas delimitam os ciclos observados (Fonte: NOAA - <i>National Oceanic and Atmospheric Administration</i> ).....	42
Figura 3 – Série heterocedástica submetida à transformação Box-Cox .....	45
Figura 4 – Série de níveis mínimos anuais do rio Nilo em Rhoda, com destaque ao período entre 680 a 780 [Fonte: adaptado de Koutsoyiannis (2006), com dados coletados em <a href="http://lib.stat.cmu.edu/S/beran">http://lib.stat.cmu.edu/S/beran</a> ].....	73
Figura 5 – Climacograma das séries de níveis mínimos anuais do rio Nilo e de números pseudoaleatórios com as mesmas estatísticas (Fonte: adaptado de Koutsoyiannis, 2006).....	74
Figura 6 – Método gráfico para remoção do comportamento não estacionário utilizado em Detzel <i>et al.</i> (2011). As linhas tracejadas ajustadas sobre a série original evidenciam a quebra de tendências.....	90
Figura 7 – Procedimento iterativo de Box & Jenkins (Fonte: adaptado de Box, Jenkins e Reinsel, 2008).....	95
Figura 8 – Experimentos numéricos desenvolvidos .....	105
Figura 9 – Procedimento para geração de novos valores de série sintética, após a geração do primeiro elemento.....	108
Figura 10 – Incertezas nas estatísticas de curto termo em função do número de séries sintéticas geradas para União da Vitória, considerando escala mensal .....	113
Figura 11 – Bacias hidrográficas do Estado do Paraná (Fonte: SEMA, 2010) .....	117
Figura 12 – Bacia do rio Iguaçu e estações fluviométricas selecionadas.....	120
Figura 13 – Médias e desvios padrão mensais para as estações fluviométricas consideradas.....	125
Figura 14 – Nilômetro da ilha de Rhoda, no Egito: (a) localização geográfica indicada pela estrela (Fonte: <a href="http://maps.google.com">maps.google.com</a> ) (b) “ <i>The nilometer on the Isle of Rhoda, Cairo</i> ”, litogravura de David Roberts (série <i>Egypt &amp; Nubia</i> , 1845-1849) que ilustra o funcionamento da estrutura	

(Fonte: albion-prints.com) (c) interior do nilômetro de Rhoda nos dias atuais (Fonte: cairobserver.com) .....	129
Figura 15 – FAC das seis séries estudadas .....	132
Figura 16 – Regressões lineares ajustadas.....	135
Figura 17 – Resultados do teste de Mann-Kendall sequencial.....	138
Figura 18 – Resultados da transformada wavelet contínua. Tons mais escuros indicam sinais mais fortes e vice-versa .....	140
Figura 19 – Séries com tendências lineares removidas. Em Rio Negro o procedimento resultou em vazão negativa em 2006.....	142
Figura 20 – Valores mínimos, médios, máximos e desvios-padrão do expoente de Hurst obtidos para diversas subamostras da série do Nilômetro de Rhoda .....	146
Figura 21 – Parâmetros $\theta_j$ para os níveis mínimos anuais do rio Nilo, usando $q = 100$ .....	148
Figura 22 – FAC das séries sintéticas de níveis mínimos anuais do rio Nilo, geradas pelo modelo SSS .....	150
Figura 23 – Climacogramas de União da Vitória, nas escalas anual (direita) e mensal (esquerda) .....	151
Figura 24 – Parâmetros $\theta_j$ para séries da bacia do rio Iguaçu, nas escalas anual (direita) e mensal (esquerda) .....	153
Figura 25 – FAC e FACP para União da Vitória .....	155
Figura 26 – Histogramas para as estatísticas de curto termo em União da Vitória, para séries sintéticas anuais .....	160
Figura 27 – FAC das séries sintéticas anuais .....	162
Figura 28 – Histogramas para as estatísticas de longo termo em União da Vitória, para séries sintéticas anuais .....	166
Figura 29 – Mapas de correlações espaciais para as séries anuais .....	167
Figura 30 – Histogramas para as estatísticas de longo termo em União da Vitória, para séries sintéticas mensais .....	171
Figura 31 – Estatísticas mensais das séries sintéticas .....	172

Figura 32 – FAC das séries sintéticas mensais.....	173
Figura 33 – Histogramas das estatísticas de longo termo em União da Vitória, para séries sintéticas mensais .....	177
Figura 34 – Mapas de correlações espaciais para as séries mensais .....	178
Figura 35 – FAC e FACP para São Bento.....	197
Figura 36 – FAC e FACP para Porto Amazonas.....	197
Figura 37 – FAC e FACP para Jangada do Sul .....	198
Figura 38 – FAC e FACP para Rio Negro .....	198
Figura 39 – FAC e FACP para Fazenda Maracanã .....	198
Figura 40 – Histogramas para as estatísticas de curto termo em São Bento, para séries sintéticas anuais.....	199
Figura 41 – Histogramas para as estatísticas de curto termo em Porto Amazonas, para séries sintéticas anuais .....	200
Figura 42 – Histogramas para as estatísticas de curto termo em Jangada do Sul, para séries sintéticas anuais .....	201
Figura 43 – Histogramas para as estatísticas de curto termo em Rio Negro, para séries sintéticas anuais .....	202
Figura 44 – Histogramas para as estatísticas de curto termo em Fazenda Maracanã, para séries sintéticas anuais .....	203
Figura 45 – Histogramas para as estatísticas de longo termo em São Bento, para séries sintéticas anuais.....	204
Figura 46 – Histogramas para as estatísticas de longo termo em Porto Amazonas, para séries sintéticas anuais .....	205
Figura 47 – Histogramas para as estatísticas de longo termo em Jangada do Sul, para séries sintéticas anuais .....	206
Figura 48 – Histogramas para as estatísticas de longo termo em Rio Negro, para séries sintéticas anuais.....	207
Figura 49 – Histogramas para as estatísticas de longo termo em Fazenda Maracanã, para séries sintéticas anuais .....	208

Figura 50 – Histogramas para as estatísticas de curto termo em São Bento, para séries sintéticas mensais .....	209
Figura 51 – Histogramas para as estatísticas de curto termo em Porto Amazonas, para séries sintéticas mensais .....	210
Figura 52 – Histogramas para as estatísticas de curto termo em Jangada do Sul, para séries sintéticas mensais .....	211
Figura 53 – Histogramas para as estatísticas de curto termo em Rio Negro, para séries sintéticas mensais .....	212
Figura 54 – Histogramas para as estatísticas de curto termo em Fazenda Maracanã, para séries sintéticas mensais .....	213
Figura 55 – Histogramas para as estatísticas de longo termo em São Bento, para séries sintéticas mensais .....	214
Figura 56 – Histogramas para as estatísticas de longo termo em Porto Amazonas, para séries sintéticas mensais .....	215
Figura 57 – Histogramas para as estatísticas de longo termo em Jangada do Sul, para séries sintéticas mensais .....	216
Figura 58 – Histogramas para as estatísticas de longo termo em Rio Negro, para séries sintéticas mensais .....	217
Figura 59 – Histogramas para as estatísticas de longo termo em Fazenda Maracanã, para séries sintéticas mensais .....	218

## LISTA DE TABELAS

Tabela 1 – Classificação dos modelos estocásticos.....	36
Tabela 2 – Distinção entre processos e séries temporais, nos contextos de modelos e fenômenos naturais .....	37
Tabela 3 – Comportamentos teóricos das FAC e FACP .....	96
Tabela 4 – Estações fluviométricas selecionadas .....	123
Tabela 5 – Taxa de falhas encontradas nos dados mensais coletados .....	124
Tabela 6 – Estatísticas descritivas mensais das séries coletadas .....	126
Tabela 7 – Estatísticas descritivas anuais das séries coletadas.....	127
Tabela 8 – Estatísticas descritivas dos níveis mínimos anuais do rio Nilo, medidas no nilômetro da Ilha de Rhoda.....	129
Tabela 9 – Divisão das séries históricas para execução dos testes estatísticos .....	133
Tabela 10 – Resultados dos testes estatísticos em termos de p-valores .....	136
Tabela 11 – Resultados do teste de Pettitt .....	137
Tabela 12 – Coeficientes do estimador não paramétrico de Sen.....	141
Tabela 13 – Valores de $\theta-q = \theta 2q + 1$ para diversas quantidades q .....	148
Tabela 14 – Estatísticas de curto termo das séries sintéticas de níveis mínimos anuais do rio Nilo, geradas pelo modelo SSS .....	149
Tabela 15 – Expoentes de Hurst e desvios padrão para séries da bacia do rio Iguaçu.....	151
Tabela 16 – Quantidades q de parâmetros $\theta_q$ a serem usadas no modelo SSS.....	152
Tabela 17 – Matrizes-parâmetros para modelo MAR(1) .....	154
Tabela 18 – Matrizes-parâmetros para modelo MAR(1)* .....	154
Tabela 19 – Resultados do critério BIC para as variantes do modelo CARIMA.....	155
Tabela 20 – Parâmetro do modelo CARIMA(1,0,0) e p-valores da validação teórica.....	156
Tabela 21 – Médias das estatísticas de curto termo das séries sintéticas anuais.....	158
Tabela 22 – Incertezas das estatísticas de curto termo das séries sintéticas anuais.....	159

Tabela 23 – Médias das estatísticas de longo termo das séries sintéticas anuais .....	163
Tabela 24 – Incertezas das estatísticas de longo termo das séries sintéticas anuais.....	165
Tabela 25 – Médias das estatísticas de curto termo das séries sintéticas mensais .....	169
Tabela 26 – Incertezas das estatísticas de curto termo das séries sintéticas mensais.....	169
Tabela 27 – Médias das estatísticas de longo termo das séries sintéticas mensais .....	175
Tabela 28 – Incertezas das estatísticas de longo termo das séries sintéticas mensais .....	176



## LISTA DE ABREVIATURAS E SIGLAS

AMO	Índice de Oscilação Multidecadal do Oceano Atlântico
ANA	Agência Nacional de Águas.
AR	Modelo autorregressivo de ordem $p$ .
ARMA	Modelo autorregressivo de médias móveis de ordens $p$ , e $q$ .
ARIMA	Modelo autorregressivo integrado e de médias móveis de ordens $p$ , $d$ , e $q$ .
BIC	Critério de informação bayesiano (do inglês <i>Bayesian Information Criterion</i> ).
CARIMA	Modelo autorregressivo integrado e de médias móveis contemporâneo, de ordens $p$ e $q$ .
FAC	Função de autocorrelação da série.
FACP	Função de autocorrelação parcial da série.
FDP	Função densidade de probabilidades.
FDE	Função de densidade espectral.
FGN	Modelo Ruído Gaussiano Fractal (do inglês <i>fractional Gaussian noise</i> ).
FM	Posto fluviométrico de Fazenda Maracanã.
HK	Processo de Hurst-Kolmogorov.
JS	Posto fluviométrico de Jangada do Sul.
LN3	Distribuição log-normal a três parâmetros.
LSSD	Método de Mínimos Quadrados sobre o Desvio Padrão (do inglês <i>Least Squares Based on Standard Deviation</i> ).
LSV	Método de Mínimos Quadrados sobre a Variância (do inglês <i>Least Squares Based on Variance</i> ).
MA	Modelo de médias móveis de ordem $q$ .
MAR(1)	Modelo Markoviano de primeira ordem.
MAR(1)*	Modelo Markoviano de primeira ordem, aplicado a séries sem tendência.
MPA	Massa de ar Polar Atlântica.
MTC	Massa de ar Tropical Continental.
MTA	Massa de ar Tropical Atlântica.
RN	Posto fluviométrico de Rio Negro.
SARIMA	Modelo autorregressivo integrado de médias móveis sazonal de ordens $p$ , $d$ , $q$ , $P$ , $D$ e $Q$ .
SB	Posto fluviométrico de São Bento.
SIN	Sistema Interligado Nacional.
SMA	Algoritmo de Médias Móveis Simétricas (do inglês <i>Symmetric Moving Average</i> ).

- SMAD Algoritmo de Médias Móveis Simétricas Dessazonalizadas (do inglês *Deseasonalized Symmetric Moving Average*).
- SSS Processo estocástico escalonado simples (do inglês *Simple Scaling Stochastic Process*), ou modelo de múltiplas escalas temporais.
- PA Posto fluviométrico de Porto Amazonas.
- PAR Modelo periódico autorregressivo de ordem  $p$  e período  $\kappa$ .
- UV Posto fluviométrico de União da Vitória.

## LISTA DE SÍMBOLOS

$a_t$	Série de resíduos do ajuste de um modelo, ou série de números pseudoaleatórios usados na geração de séries sintéticas com distribuição normal padrão.
$a'_t$	Série de números pseudoaleatórios usados na geração de séries sintéticas com distribuição log-normal a três parâmetros.
$a_{ij}, b_i$	Funções auxiliares para estimador do expoente de Hurst via método LSSD ( $i = j = 1, 2$ ).
$A_\kappa, B_\kappa, C_\kappa$	Coefficientes de Fourier para o harmônico $\kappa$ .
$\mathbf{A}, \mathbf{B}$	Matrizes-parâmetro utilizadas na geração multivariada de séries sintéticas.
$B$	Operador defasagem da formulação ARIMA.
$b, c$	Parâmetros do método não paramétrico de Sen para remoção de tendências.
$c_v$	Coefficiente de variação.
$d$	Grau ou ordem de diferenciação (I) não sazonal.
$D$	Grau ou ordem de diferenciação (I) do modelo sazonal.
$D_k$	Função de amortecimento, ou janela de defasagem, da função de densidade espectral.
$e$	Base dos logaritmos naturais ( $e = 2,718281 \dots$ ).
$e^2$	Erro quadrático de um ajuste.
$f_Z(z)$	Função densidade de probabilidades de $Z$ . Interpreta-se como a probabilidade da variável aleatória $Z$ assumir o valor $z$ .
$H$	Expoente de Hurst.
$H_0$	Hipótese nula.
$H_A$	Hipótese alternativa.
$k$	Defasagem, ou <i>lag</i> , da série temporal.
$l$	Número de locais modelados.
$L$	Função de verossimilhança.
$\ln$	Logaritmo natural de base $e$ .
$m$	Número de harmônicos a utilizar em uma análise espectral, ou número de termos significativos da função de autocorrelação amostra.
$M_n$	Soma parcial máxima.
$m_n$	Soma parcial mínima.
$\mathbf{M}_0, \mathbf{M}_1$	Matrizes auxiliares para a estimação dos parâmetros do modelo MAR(1).
$MK$	Estatística do teste de Mann-Kendall.
$n$	Tamanho da amostra ou comprimento da série histórica.

$n'$	Tamanho efetivo da amostra, ou tamanho das séries usadas para compor um climacograma.
$N$	Número de realizações de um processo estocástico.
$p$	Ordem do modelo autorregressivo (AR) não sazonal.
$P$	Ordem do modelo autorregressivo (AR) sazonal.
$PT$	Estatística do teste de Pettitt.
$q$	Ordem do modelo de médias móveis (MA) não sazonal.
$Q$	Ordem do modelo de médias móveis (MA) sazonal.
$r$	Coeficiente de correlação.
$R_t$	Amplitude.
$R_t^*$	Amplitude ajustada.
$R_t^{**}$	Amplitude re-escalada.
$\mathbb{R}$	Conjunto dos números reais.
$S(f)$	Função de densidade espectral.
$S_t$	Soma parcial.
$SSQ$	Função soma dos quadrados dos resíduos.
$s$	Parâmetro de dilatação da transformada <i>wavelet</i> .
$t$	Variável pertencente ao espaço dos instantes ( $t \in T$ ), nesta tese representando o intervalo de tempo. Representa também a estatística do teste t-Student.
$T$	Espaço dos instantes, nesta tese representando o espaço do tempo.
$u, v$	Localidades modeladas.
$w_t$	Série temporal diferenciada.
$W$	Estatística do teste de Wilcoxon.
$W_\tau(s)$	Transformada <i>wavelet</i> contínua.
$X, x$	Variável aleatória genérica e o valor assumido por ela, respectivamente.
$Y_t^{(\eta)}$	Processo estocástico agregado discreto.
$y_t^{(\eta)}, y_t$	Uma série temporal agregada.
$y_{t,u}$	Uma série temporal agregada para a localidade $u$ .
$y_t'$	Uma série temporal agregada dessazonalizada.
$Z$	Processo estocástico.
$Z_t$	Processo estocástico discreto.
$Z(t)$	Processo estocástico contínuo.
$z_t$	Uma realização de um processo estocástico discreto, ou uma série temporal.
$z_t'$	Série temporal transformada via Box-Cox ou distribuição log-normal a três parâmetros.

$z_t^*$	Série temporal com tendência linear removida (estacionárias na média).
$z$	Variável normal padrão.
$\mathbb{Z}$	Conjunto dos números inteiros.
$\alpha$	Nível de significância para testes de hipóteses.
$\beta$	Parâmetro que traduz a tolerância para o erro na estimativa do termo $\theta_0$ do modelo SMA.
$\gamma$	Covariância do processo estocástico, ou matriz de variância-covariância das suas realizações, caso seja exibido em negrito.
$\gamma_{uv}$	Elementos da matriz de variância-covariância entre as localidades $u$ e $v$ .
$\hat{\gamma}_k$	Autocovariância da série temporal.
$\delta$	Variável auxiliar para estimação dos parâmetros da distribuição log-normal a três parâmetros.
$\Delta$	Variável auxiliar para determinação das estatísticas de longo termo referentes aos déficits.
$\varepsilon_{t,u}$	Séries de resíduos independentes no tempo, mas correlacionados espacialmente para um mesmo instante $t$ .
$\epsilon$	Constante.
$\zeta$	Parâmetro de deslocamento da distribuição log-normal a três parâmetros.
$\eta$	Inteiro positivo multiplicador da escala temporal básica, representada por $t$ .
$\theta_q$	Parâmetro do modelo não sazonal de médias móveis de ordem $q$ [MA( $q$ )].
$\Theta_Q$	Parâmetro do modelo sazonal de médias móveis de ordem $Q$ [(S)MA( $Q$ )].
$\kappa$	Harmônico, número de ciclos entre $t = 0$ e $t = n$ , ou ainda período de um processo periódico.
$\lambda$	Parâmetro da transformação Box-Cox.
$\Lambda$	Variável auxiliar para determinação das estatísticas de longo termo referentes às corridas.
$\mu$	Média populacional, ou média conjunta do processo estocástico.
$\hat{\mu}_z$	Média amostral da série temporal.
$\xi$	Assimetria populacional.
$\hat{\xi}_z$	Coeficiente de assimetria da série temporal.
$\rho$	Autocorrelação do modelo ou do processo estocástico.
$\hat{\rho}_{z,k}$	Autocorrelação da série temporal $z$ , de ordem $k$ .
$\sigma^2$	Variância populacional, ou variância conjunta do processo estocástico.
$\hat{\sigma}_z^2$	Variância amostral da série temporal.
$\hat{\sigma}_z$	Desvio padrão amostral da série temporal.
$\Sigma$	Operador somatório.
$\tau$	Parâmetro de translação temporal de um processo estocástico ou de uma série.

$v$	Número de graus de liberdade.
$\phi_p$	Parâmetro do modelo não sazonal autorregressivo de ordem p [AR(p)].
$\Phi_P$	Parâmetro do modelo sazonal autorregressivo de ordem P [(S)AR(P)].
$\varphi$	Autocorrelação parcial do modelo ou do processo estocástico.
$\hat{\varphi}_{kk}$	Autocorrelação parcial da série temporal.
$\psi(t)$	<i>Wavelet</i> -mãe.
$\omega$	Evento pertencente ao espaço de probabilidades ( $\omega \in \Omega$ ).
$\Omega$	Espaço de probabilidades.
$\nabla$	Operador diferença.
$E[ \ ]$	Operador valor esperado.
$VAR[ \ ]$	Operador variância.
$STD[ \ ]$	Operador desvio padrão.
$COV[ \ ]$	Operador covariância.
$sgn[ \ ]$	Sinal da operação entre colchetes.



1.4.2.2	Remoção do comportamento não estacionário das séries históricas	89
1.4.3	<i>Modelos Box &amp; Jenkins</i>	91
1.4.3.1	Modelos estocásticos lineares estacionários	91
1.4.3.2	Modelos estocásticos lineares não estacionários	93
1.4.3.3	Modelos estocásticos lineares sazonais	94
1.4.3.4	O procedimento iterativo de Box & Jenkins	95
1.4.3.5	Extensão para o caso multivariado – modelo CARIMA	99
1.4.4	<i>Estado-da-arte da modelagem de séries de vazões médias</i>	101
1.5	Resumo do capítulo .....	102
<b>2</b>	<b>MÉTODO DE PESQUISA.....</b>	<b>105</b>
2.1	Estrutura do estudo .....	105
2.2	Sistemática para implementação dos algoritmos.....	106
2.2.1	<i>Modelo SSS</i>	107
2.2.2	<i>Modelo MAR(1)</i>	108
2.2.3	<i>Modelo CARIMA</i>	110
2.3	Métricas para análises comparativas .....	112
2.4	Resumo do capítulo .....	115
<b>3</b>	<b>ÁREA DE ESTUDO .....</b>	<b>117</b>
3.1	Características gerais .....	118
3.2	Clima.....	119
3.3	Estações fluviométricas selecionadas .....	122
3.4	Níveis mínimos anuais do rio Nilo .....	127
3.5	Resumo do capítulo .....	130
<b>4</b>	<b>ANÁLISE PRELIMINAR: ESTACIONARIEDADE DAS SÉRIES.....</b>	<b>131</b>
4.1	Persistência das séries.....	131
4.2	Premissas para execução das análises.....	133
4.3	Resultados.....	134
4.4	Remoção do comportamento não estacionário .....	141
4.5	Resumo do capítulo .....	143
<b>5</b>	<b>RESULTADOS E DISCUSSÕES.....</b>	<b>145</b>
5.1	Detalhes do ajuste dos modelos estocásticos.....	145



5.1.1	<i>Modelo SSS</i>	145
5.1.1.1	Estimação do expoente de Hurst	145
5.1.1.2	Estrutura dos coeficientes do algoritmo SMA	147
5.1.1.3	Ajuste do modelo SSS à série de níveis mínimos anuais do rio Nilo	149
5.1.1.4	Parâmetros do modelo SSS para a bacia do rio Iguaçu	150
5.1.2	<i>Modelo MAR(1)</i>	153
5.1.3	<i>Modelos CARIMA</i>	154
5.2	Análise comparativa SSS <i>versus</i> MAR(1) para séries anuais .....	157
5.3	Análise comparativa SSS <i>versus</i> CARIMA para séries mensais .....	168
5.4	Resumo do capítulo .....	179
<b>CONCLUSÃO.....</b>		<b>181</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS .....</b>		<b>185</b>
<b>APÊNDICES .....</b>		<b>197</b>
A.1.	Funções de autocorrelação (FAC) e de autocorrelação parcial .....	197
A.2.	Histogramas das estatísticas de curto termo para séries anuais .....	199
A.3.	Histogramas das estatísticas de longo termo para séries anuais .....	204
A.4.	Histogramas das estatísticas de curto termo para séries mensais .....	209
A.5.	Histogramas das estatísticas de longo termo para séries mensais .....	214



## INTRODUÇÃO

A necessidade de tomar decisões frente a problemas do cotidiano da humanidade sempre foi o principal estímulo para a ciência na tarefa de desenvolver ferramentas de auxílio a tais escolhas. O primeiro passo na busca de respostas é, sem dúvida, o bom entendimento da questão. Em se tratando de situações envolvendo variáveis e fenômenos naturais, a modelagem matemática emergiu como linguagem universal para comunicação entre cientistas, ajudando-os a compreender e representar os complexos processos existentes.

Ao longo dos anos, as observações dos diversos fenômenos ocorridos na natureza produziram as séries temporais que servem de base para a elaboração de modelos. Em hidrologia, as variáveis mais comuns são chuvas e níveis de rios, estes últimos transformados em vazões através de curvas-chave. No Brasil, as mais longas séries de chuvas têm início antes dos anos 1900 (a estação de Curitiba iniciou os registros em 1884, Pinto *et al.*, 1976, p. 9), enquanto que as maiores séries de vazões tiveram seus primeiros registros a partir de 1930. Dessa maneira, até meados da década de 1970 as amostras possuíam tamanhos que variavam em torno de 40 a 50 anos, para os quais era possível assumir uma condição de equilíbrio estatístico chamada de estacionariedade. Em outras palavras, considerava-se que o comportamento das séries históricas observadas seguiria padrão semelhante no futuro.

Com o passar dos anos as séries hidrológicas foram se tornando cada vez maiores, evidenciando peculiaridades associadas a comportamentos não estacionários. Em muitos lugares do Brasil, principalmente na região Sul, essas alterações em séries de vazão foram registradas a partir do final da década de 1970, justificada através de intensas mudanças no uso dos solos das bacias hidrográficas. Nessa mesma época se intensificaram estudos sobre variabilidade e mudança climáticas, a primeira impulsionadas por anomalias como oscilações decadais e multidecadais na temperatura superficial dos oceanos, e a segunda pelo aumento na emissão dos gases de efeito estufa na atmosfera, que alguns pesquisadores também apontam como indutoras de distúrbios nos regimes hidrológicos. Porém, independentemente da causa, ficou evidente que a aplicação das séries observadas diretamente nos tradicionais modelos estatísticos estacionários desenvolvidos até então não renderia resultados confiáveis.

Para contornar a situação, algumas técnicas matemáticas passaram a ser utilizadas. A principal delas é a representação da alteração do regime hidrológico através de tendências lineares. Fisicamente, isso significa dizer que as vazões de um rio sofreram um aumento (ou redução) linear no decorrer dos anos. A detecção de tal tendência, por si só, resultou em um

grande número de trabalhos voltados ao desenvolvimento de técnicas específicas. No entanto, estudos recentes questionam esse artifício principalmente por não ser possível obter uma função determinística que explique adequadamente as tendências observadas. Ademais, análises sobre séries de longa duração permitiram observar que as flutuações existentes seriam bem modeladas através de uma abordagem estocástica.

Esse novo enfoque se traduz em uma linha de pensamento ainda pouco explorada na literatura hidrológica e, por esse motivo, é o tema central da presente tese. Ao contrário das técnicas tradicionais, considera-se que as flutuações observadas nas séries são parte de seu comportamento natural esperado, fruto da interação de eventos que ocorrem em diversas escalas temporais. Dessa maneira, utilizam-se os conceitos de processos estocásticos agregados em uma classe de modelos que abordam múltiplas escalas temporais, cuja premissa é, justamente, representar a citada interação entre eventos.

Até hoje a aplicação dos modelos de múltiplas escalas temporais estava limitada a séries hidrológicas particularmente longas, como a sequência de níveis mínimos anuais do rio Nilo medida na ilha de Rhoda, em Cairo, no Egito, que totaliza 663 anos (Koutsoyiannis, 2002). Evidentemente, séries hidrológicas com tal comprimento são extremamente raras (senão únicas), fazendo com que a aplicação desses modelos em séries de comprimento condizente com as trabalhadas na hidrologia operacional seja uma investigação de grande potencial. Assim sendo, a **hipótese** a ser testada nesta tese é a de que os modelos de múltiplas escalas temporais são viáveis para utilização na hidrologia operacional, oferecendo representatividade estatística superior quando comparados a formulações tradicionais.

Como abordagem adotada para validar a hipótese supramencionada, optou-se pela utilização de modelos computacionais e, em particular, o método de Monte Carlo. Dentre outras situações, os modelos computacionais cumprem importante papel quando se trabalha com problemas (i) de difícil solução teórica, (ii) que exigem experimentos laboratoriais potencialmente perigosos, ou ainda (iii) que são onerosos do ponto de vista de tempo e recursos financeiros (U.S. DOE, 2000, p. 1). A complexidade de fenômenos naturais é, por si só, uma justificativa válida para o emprego de modelos computacionais, encaixando-se na situação (i) mencionada. Como desdobramento dessa complexidade, tem-se que a incerteza associada ao comportamento das variáveis naturais pode ser devidamente considerada através de múltiplas simulações estocásticas, base do método de Monte Carlo (Srivastav e Simonovic, 2014).

Dentro do contexto da hidrologia estocástica, além dos argumentos expostos no parágrafo anterior, as simulações para geração de séries sintéticas são úteis principalmente para o melhor conhecimento da informação histórica existente. Considera-se que a série histórica é

apenas uma realização do processo estocástico inerente ao fenômeno natural que lhe deu origem; uma vez ajustado um modelo que representa esse processo, podem ser geradas múltiplas séries sintéticas diferentes da histórica, mas com igual probabilidade de ocorrência. Na presente tese, os modelos de múltiplas escalas temporais são ajustados a séries de vazões médias nas escalas anual e mensal, sendo confrontados com formulações Markovianas e de Box & Jenkins, respectivamente. Destaca-se que a **inovação** do trabalho está no algoritmo de médias móveis simétricas dessazonalizadas (SMAD), solução proposta para a geração de séries sintéticas sazonais utilizando os modelos de múltiplas escalas temporais.

O **objetivo principal** da tese é investigar a aplicabilidade dos modelos de múltiplas escalas temporais em séries hidrológicas da região Sul do Brasil. A pesquisa é norteada pelos seguintes **objetivos específicos**:

1. Elaboração de revisão bibliográfica referente a tópicos de não estacionariedade aplicados a séries hidrológicas.
2. Conceituação teórica e matemática, quando aplicável, dos termos estatísticos envolvidos.
3. Distinção conceitual entre a abordagem de múltiplas escalas temporais e os modelos Markoviano e de Box & Jenkins.
4. Estudo da influência do eventual comportamento não estacionário das séries hidrológicas no ajuste dos modelos Markoviano e de Box & Jenkins, incluindo a aplicação de técnicas para a detecção e correção desse comportamento.
5. Aplicação prática comparativa das técnicas trabalhadas.

A tese está estruturada da seguinte maneira: o capítulo 1 apresenta a fundamentação teórica, contendo conceitos estatísticos de interesse, técnicas para detecção de tendências em séries hidrológicas e modelos de geração de séries sintéticas, incluindo o algoritmo SMAD proposto como inovação da tese (em específico, na subseção 1.4.1.1.5, página 83); o capítulo 2 relata o método de pesquisa e explica como foi feita a junção dos conceitos teóricos para aplicação na área de estudo considerada; o capítulo 3 caracteriza esta área, incluindo as estações fluviométricas utilizadas e também uma descrição da série de níveis mínimos anuais do rio Nilo, utilizada como depuração do modelo de múltiplas escalas temporais; o capítulo 4 exhibe resultados preliminares relacionados com a análise da estacionariedade estatística das séries; o capítulo 5 mostra os resultados principais da tese, que incluem análises comparativas das séries sintéticas geradas pelos modelos considerados. O documento é finalizado com as conclusões, lista de Referências Bibliográficas utilizadas e apêndices gerados.

# 1 FUNDAMENTAÇÃO TEÓRICA

Em hidrologia o uso de modelos matemáticos tem início na metade do século XIX com o método racional proposto por Mulvaney (1850), que relaciona as vazões máximas em uma bacia hidrográfica com a intensidade da chuva de duração igual ao seu tempo de concentração. Cem anos mais tarde a modelagem hidrológica foi impulsionada pelo uso dos computadores, que permitiram o aparecimento de modelos determinísticos semiconceituais (Tucci, 2005). Na época, pesquisadores visavam reproduzir as fases do ciclo hidrológico, considerando processos físicos e equações empíricas. Novo direcionamento foi dado na década de 1970 com a publicação do trabalho de Box & Jenkins (atualmente em sua quarta edição, Box, Jenkins e Reinsel, 2008), a partir do qual se expandiu a abordagem da análise de séries temporais aos modelos hidrológicos. No caso univariado, um processo físico é simulado utilizando unicamente as observações históricas do próprio fenômeno no tempo. Hoje em dia, formulações determinísticas e modelos de séries temporais permanecem como as duas grandes vertentes da modelagem hidrológica. Este capítulo, em concordância com as propostas da tese, foca-se na classe de modelos de séries temporais (conhecidos na literatura internacional como *data driven models*, em referência a formulações cuja única fonte de informação é a própria série. Ver Todini, 2007).

Uma série temporal é um encadeamento de observações de uma variável no tempo. Em raras ocasiões, o espaçamento entre elementos de uma série pode ser desigual, como, por exemplo, uma série hidrológica cujos registros dependam de campanhas de medição em campo que não sejam realizadas periodicamente. Entretanto, na grande maioria das vezes consideram-se os elementos de uma série igualmente espaçados no tempo. Como característica inerente das séries temporais naturais, tem-se que as observações adjacentes guardam certo grau de dependência entre si. Por esse motivo, em sua análise são comumente aplicadas técnicas que permitem estudar essa relação de dependência (Box, Jenkins e Reinsel, 2008, p. 1).

Um dos conceitos de maior importância na análise de séries temporais é o de modelos estatisticamente estacionários. Diz-se que um modelo é estacionário se seus momentos estatísticos não se alteram com o tempo, ou ainda, o processo possui propriedades estatísticas que não são função do tempo (Hipel e McLeod, 1994, p. 67). Portanto, um modelo estacionário traz consigo uma condição de equilíbrio estatístico. Séries de fenômenos naturais, no entanto, não apresentam esse equilíbrio, sendo marcadas por sazonalidades e outros processos que as

tornam impróprias para aplicação em formulações estacionárias. Dessa maneira, a modelagem de tais séries precisa contemplar minúcias matemáticas que permitam contornar essa condição.

A estacionariedade estatística carrega consigo outros termos relacionados, tais como processos estocásticos, homocedasticidade, ergodicidade, e outros cujas definições fundamentais são formalmente apresentadas. Revisar-se-ão também os principais trabalhos na área da hidrologia estocástica que discutem, direta ou indiretamente, a questão da não estacionariedade de séries hidrológicas.

## 1.1 Definições e conceitos fundamentais

### 1.1.1 Modelos determinísticos e estocásticos

A ocorrência de um fenômeno natural está atrelada a um ou mais fatores influentes. Um modelo determinístico é aquele em que é possível mapear tais fatores de forma a escrever uma relação matemática que traduza exatamente suas influências sobre a variável de interesse (Box, Jenkins e Reinsel, 2008, p. 7). Alternativamente, Hipel e McLeod (1994, p. 22) argumentam que modelos determinísticos possuem a capacidade de reproduzir todos os estados de um sistema.

Evidentemente, fenômenos completamente determinísticos são raros, pois dificilmente é possível relevar todos os seus fatores influentes. Celka *et al.* (2006, p. 90) vão além ao afirmar que mesmo considerando as leis clássicas da física, que atestam que um objeto qualquer tem uma posição e uma velocidade em um período específico de tempo, o conceito de determinismo se tornou incerto desde o surgimento da mecânica quântica. Em hidrologia, Naghettini e Pinto (2007, p. 3-4) exemplificam que um pulso de precipitação conhecido, uniforme e homogêneo sobre uma superfície de geometria simples, impermeável e completamente conhecida causa uma resposta hidrológica determinística.

Ainda que sejam situações irreais na natureza, alguns fenômenos podem ser representados por modelos determinísticos. A utilização da análise de Fourier para caracterização dos harmônicos de uma série, por exemplo, é um método clássico com boa aceitação na literatura (Fleming e Weber, 2012). Desta forma, utilizando uma abordagem menos restritiva, pode-se caracterizar modelos hidrológicos determinísticos como aqueles nos quais uma mesma entrada gera sempre uma mesma saída, independentemente de quantas vezes o modelo é executado (Tucci, 2005, p. 26).

Quando os fatores que causam um determinado fenômeno não têm seus valores exatos conhecidos, eles podem ser escritos em termos probabilísticos. Nesses casos, as equações

matemáticas que descrevem o sistema formam um modelo estocástico. Segundo Hipel e McLeod (1994, p. 23), a grande maioria dos fenômenos naturais segue (ou aparenta seguir) premissas probabilísticas, o que justifica seu uso amplo em hidrologia. Os autores exibem uma classificação específica para modelos estocásticos, diferenciando as considerações referentes ao tempo e aos estados assumidos. Tal categorização é mostrada na Tabela 1.

Os quatro tipos de modelos estocásticos listados encontram aplicação em hidrologia. Para citar alguns exemplos, Wilks (1998) utiliza cadeias de Markov para modelar a ocorrência de precipitações diárias em múltiplos postos meteorológicos; Abi-Zeid, Parent e Bobée (2004) aplicam processos pontuais sobre eventos de estiagens de rios canadenses; Breinholt *et al.* (2012) avaliam a incerteza de modelos chuva-vazão através de equações diferenciais estocásticas. Modelos de séries temporais, em particular, recebem atenção especial no decorrer deste capítulo, pois entram na classe de principal interesse desta tese. Dessa maneira, os estudos revisados nas próximas seções se referem a aplicações de modelos de séries temporais.

**Tabela 1 – Classificação dos modelos estocásticos**

		Estados	
		Discretos	Contínuos
Tempo	Discreto	Cadeias de Markov	Modelos de séries temporais
	Contínuo	Processos pontuais	Equações diferenciais estocásticas

Fonte: Hipel e McLeod (1994, p. 23)

Por fim, há que se apontar que em hidrologia não são incomuns modelos compostos conjuntamente por porções determinísticas e estocásticas em suas equações, muito embora essa mistura de enfoques seja criticada por alguns autores (e.g. Koutsoyiannis, 2006). É uma abordagem considerada, por exemplo, quando as variáveis de interesse possuem algum tipo de periodicidade, como sazonalidade. Séries de vazões médias de um rio podem ser modeladas por arranjos desse tipo, em geral escritos como uma adição entre os componentes periódicos (determinísticos) e estocásticos (Adamowski, 1971; Yevjevich e Obeysekera, 1985).

### 1.1.2 Processos estocásticos

Os processos estocásticos descrevem sistemas cuja evolução ocorre de acordo com leis probabilísticas. Matematicamente, um processo estocástico  $Z$  é uma família de funções definidas por (Papoulis, 1965, p. 279; Morettin e Toloi, 2004, p. 21):



$$Z = \{Z(t, \omega), t \in T; \omega \in \Omega\} \tag{1.1}$$

onde  $t$  é uma variável pertencente ao espaço de instantes  $T$  e  $\omega$  é um evento do espaço de probabilidades  $\Omega$ . Portanto, um processo estocástico pode ser visto como uma função de duas variáveis, na qual para cada evento  $\omega$  é atribuída uma variável  $t$ . Se  $T \equiv \mathbb{Z} = \{1, \dots, t\}$ , diz-se que o processo é discreto, sendo representado por  $Z_{t,\omega}$ . Por outro lado, se  $T \in \mathbb{R}$ , o processo é contínuo e passa a ser representado por  $Z(t, \omega)$ .

Tipicamente omite-se a dependência do processo estocástico dos eventos  $\omega$ , reduzindo-se as notações para  $Z_t$  ou  $Z(t)$  nos casos discretos ou contínuos respectivamente. Ademais, na grande maioria das aplicações de processos estocásticos à modelagem de fenômenos naturais, assume-se o espaço dos instantes  $T$  como sendo finito ( $T \equiv \mathbb{Z}$ ) e tendo o tempo como unidade. Nessas configurações, o processo estocástico definido na equação (1.1) representa quatro situações, dependendo do comportamento de  $t$  e  $\omega$ :

1. Uma família de funções temporais ( $t$  e  $\omega$  variáveis)
2. Uma função temporal única ( $t$  variável e  $\omega$  fixo)
3. Uma variável aleatória ( $t$  fixo e  $\omega$  variável)
4. Um número ( $t$  e  $\omega$  fixos)

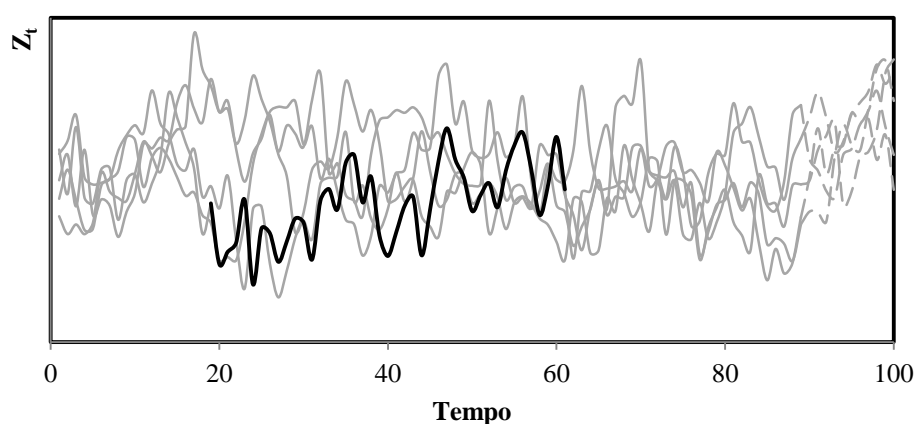
O caso dois é de especial interesse por ser a definição de uma série temporal  $z_t$ , cuja ideia inicial foi dada na parte introdutória deste capítulo. Nesse contexto, é comum referir-se à série temporal como uma realização infinita do processo estocástico que governa o fenômeno natural estudado. Devido à importância dessa explicação no contexto da presente tese, cabe aqui aprofundá-la visando melhor entendimento do leitor quanto às diferenças entre processos **naturais** e processos **estocásticos**, bem como entre séries **observadas** e **realizações** de processos estocásticos. Koutsoyiannis (2006) exhibe um esquema interessante que auxilia esse entendimento, cuja adaptação é mostrada na Tabela 2.

**Tabela 2 – Distinção entre processos e séries temporais, nos contextos de modelos e fenômenos naturais**

	<b>Processos (variáveis)</b>	<b>Séries temporais (valores)</b>
<b>Fenômeno natural</b>	Processo natural	Séries observadas (únicas e finitas)
<b>Modelo</b>	Processo estocástico	Realizações (infinitas)

Fonte: adaptado de Koutsoyiannis (2006)

Assim, processos naturais (e.g. chuvas em determinado posto pluviométrico ou vazões de um rio em uma seção específica de seu curso) não devem ser confundidos com processos estocásticos, matematicamente definidos no início dessa seção. No caso das séries temporais, observações de fenômenos naturais são consideradas como sendo uma amostra (portanto, finitas) de uma realização de um processo estocástico. A Figura 1 mostra uma representação gráfica desses conceitos e como séries observadas se relacionam com os processos estocásticos teóricos.



**Figura 1 – Representação de cinco realizações de um processo estocástico com destaque a uma amostra da série temporal**

O gráfico da Figura 1 exibe apenas cinco séries de um processo com  $N$  realizações, limitadas às suas primeiras 100 ocorrências. A linha destacada simula uma série temporal finita, sendo ela, por hipótese, a única informação real disponível para aplicação em modelos estocásticos. Retomando-se o caso das séries de chuvas e rios mencionados no parágrafo anterior, essa linha representa o histórico medido de cada variável hidrológica. É importante ressaltar que as inferências estatísticas de todo o processo são feitas com base nessa única amostra de realização. Por esse motivo são adotadas as condições restritivas de estacionariedade e ergodicidade do processo estocástico, conceitos a serem explicados em detalhes nas seções 1.1.3 e 1.1.5, respectivamente.

Um processo estocástico é estatisticamente determinado quando se conhece a distribuição conjunta das suas  $N$  funções densidade de probabilidade (FDP). Tal função é representada pela equação (1.2):

$$f_{Z_1, Z_2, \dots, Z_N}(z_1, \dots, z_N; t_1, \dots, t_N) = P\{Z(t_1) \leq z_1, \dots, Z(t_N) \leq z_N\} \quad (1.2)$$

Em outras palavras, a determinação de um processo estocástico  $Z$  envolve o conhecimento da distribuição de probabilidades conjunta de todas as suas realizações  $z_t$  ( $t =$

$1, \dots, \infty$ ) . Como parâmetros importantes associados à FDP conjunta (1.2), têm-se a média  $\mu_t$  e a variância  $\sigma_{z_t}^2$  conjuntas<sup>1</sup> do processo, dadas pelas equações (1.3) e (1.4), respectivamente:

$$\mu_t = E[z_t] = \int_{-\infty}^{\infty} z f_z(z; t) dz, \quad t \in T \quad (1.3)$$

$$\sigma_{z_t}^2 = E[(z_t - \mu_t)^2] = VAR[z_t] = \int_{-\infty}^{\infty} (z - \mu)^2 f_z(z; t) dz, \quad t \in T \quad (1.4)$$

onde  $E[ \ ]$  e  $VAR[ \ ]$  são os operadores valor esperado e variância, respectivamente, e  $f_{z_i}(z_i; t_i)$ , ( $i = 1, \dots, N$ ) é a FDP definida na equação (1.2). Os parâmetros apresentados se referem aos momentos conjuntos de primeira e segunda ordens do processo estocástico. Defina-se também a função de covariância  $\gamma(t_1, t_2)$  do processo como sendo:

$$\gamma(t_1, t_2) = COV[z_{t_1}, z_{t_2}] = E[z_{t_1} z_{t_2}] - E[z_{t_1}]E[z_{t_2}], \quad t_1, t_2 \in T \quad (1.5)$$

onde  $COV[ \ ]$  é o operador covariância. Na realidade, a variância definida pela equação (1.4) é um caso particular da equação (1.5) quando  $t_1 = t_2 = t$ . Utilizando novamente o exemplo genérico da Figura 1, as equações (1.3) e (1.4) definem a média e a variância conjunta das cinco séries. A aplicação da equação (1.5), por sua vez, resulta na matriz de covariâncias, de tamanho 5x5, na qual a diagonal principal representa as variâncias de cada uma das séries.

Todavia, em nenhuma situação real é possível ter o conhecimento de todas as realizações de um processo estocástico, ainda menos de sua distribuição de probabilidades conjunta e de seus momentos conjuntos. Como explicado anteriormente, isso decorre do fato da única informação disponível ser a série histórica observada de uma variável hidrológica. Por esse motivo, trabalha-se com a distribuição de probabilidades marginal dessa realização, bem como de seus momentos estatísticos amostrais. Os estimadores de média  $\hat{\mu}$ , variância  $\hat{\sigma}_z^2$  e covariância  $\hat{\gamma}_k$  são obtidos através das equações (1.6) a (1.8):

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n z_t \quad (1.6)$$

$$\hat{\sigma}_z^2 = \frac{1}{n} \sum_{t=1}^n (z_t - \hat{\mu})^2 \quad (1.7)$$

---

<sup>1</sup> Na literatura internacional, é comum encontrar o termo *ensemble* para designar o conjunto de todas as realizações de um processo estocástico, bem como seus parâmetros (e.g. Hamilton, 2004).

$$\hat{\gamma}_k = COV[z_t, z_{t+k}] = E[z_t z_{t+k}] - E[z_t]E[z_{t+k}] \quad (1.8)$$

onde  $n$  é o número de elementos da série temporal e  $k$  é a defasagem (ou *lag*) entre observações da série em intervalos de tempo. Nota-se que a equação (1.8) expressa a relação entre as observações da própria série temporal, separadas pelo *lag*  $k$ . Por esse motivo, essa relação ganha o nome de autocovariância e é uma generalização da equação (1.5) para os casos de  $k > 0$ .

Outra relação de grande importância na análise de uma série temporal é a autocorrelação amostral  $\hat{\rho}_k$ , expressa pela razão entre a covariância  $\hat{\gamma}_k$  e a variância da amostra  $\hat{\sigma}_z^2$ , para um *lag*  $k$ :

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\sigma}_z^2} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0} \quad (1.9)$$

O gráfico formado pelo coeficiente  $\hat{\rho}_k$  como função do *lag*  $k$  é chamado de correlograma, ou função de autocorrelação amostral (FAC) da série. A importância desse gráfico está em sua capacidade de exibir a relação existente entre as observações de uma série, permitindo a identificação de sazonalidades e outros padrões de comportamento.

### 1.1.3 Estacionariedade estatística

O conceito de estacionariedade está intrinsecamente relacionado com um estado de equilíbrio de um processo estocástico, de tal forma que suas propriedades estatísticas independem do tempo (Hipel e McLeod, 1994, p. 67; Morettin e Toloi, 2004, p. 25; Box, Jenkins e Reinsel, 2008, p. 7). Assim, as características de  $Z(t + \tau)$  são as mesmas de  $Z(t)$ , para qualquer  $\tau \in T$ . Formalmente são considerados dois tipos de estacionariedade: (i) estrita, ou forte e (ii) ampla, ou fraca. No caso (i), translações temporais não afetam as FDPs, ou seja:

$$f_Z(z_1, \dots, z_N; t_1 + \tau, \dots, t_N + \tau) = f_Z(z_1, \dots, z_N; t_1, \dots, t_N) \quad (1.10)$$

Nessas condições, tem-se que (Morettin e Toloi, 2004, p. 26; Papoulis, 1965, p. 301):

- a)  $E[z_t] = \mu_t = \mu = \text{cte.}$
- b)  $VAR(z_t) = \sigma_{z_t}^2 = \sigma_z^2.$
- c)  $\gamma(t_1, t_2)$  é função de  $|t_1 - t_2|$ , ou seja, depende somente das próprias diferenças entre as séries, não importando a referência temporal de cada uma.

Na estacionariedade do tipo (ii), as condições recaem somente sobre os itens (a), (b) e (c), desconsiderando-se a restrição sobre a FDP do processo. Na prática, um processo é fracamente estacionário quando sua média e variância independem do tempo. Por esse motivo, é comum categorizar esse caso como estacionariedade de segunda ordem, em referência aos dois momentos estatísticos citados. O único caso no qual um processo é considerado tanto fracamente quanto estritamente estacionário é quando ele possui distribuição normal (multivariada), pois os parâmetros que definem essa distribuição são, justamente, a média e a variância.

Além dos dois tipos de estacionariedade mencionados, Papoulis (1965, p. 302-303) fornece definições para casos específicos de estacionariedade estrita, elencados a seguir:

- Um processo é estacionário de ordem limitada  $\kappa$  se sua FDP [condição (1.10)] for verdadeira para  $n \leq \kappa$  ( $n = 1, \dots, N$ ). Nesse caso, se ela for verdadeira para  $n = \kappa$ , então também será verdadeira para  $n < \kappa$ , pois a FDP de ordem  $\kappa$  define todas as FDP de ordens inferiores.
- Um processo é assintoticamente estacionário se o limite:

$$\lim_{\tau \rightarrow \infty} f_Z(z_1, \dots, z_N; t_1 + \tau, \dots, t_N + \tau)$$

existe e independe de  $\tau$ .

- Um processo é estacionário em um intervalo se a FDP da condição (1.10) é verdadeira somente em um intervalo de tempo.
- Um processo é periodicamente estacionário (ou cicloestacionário) com período  $\kappa$  se a FDP da condição (1.10) for verdadeira para  $\tau = N\kappa$ .
- Um processo  $y_t$  possui incrementos estacionários se:

$$y_t = z_{t+\tau} - z_t$$

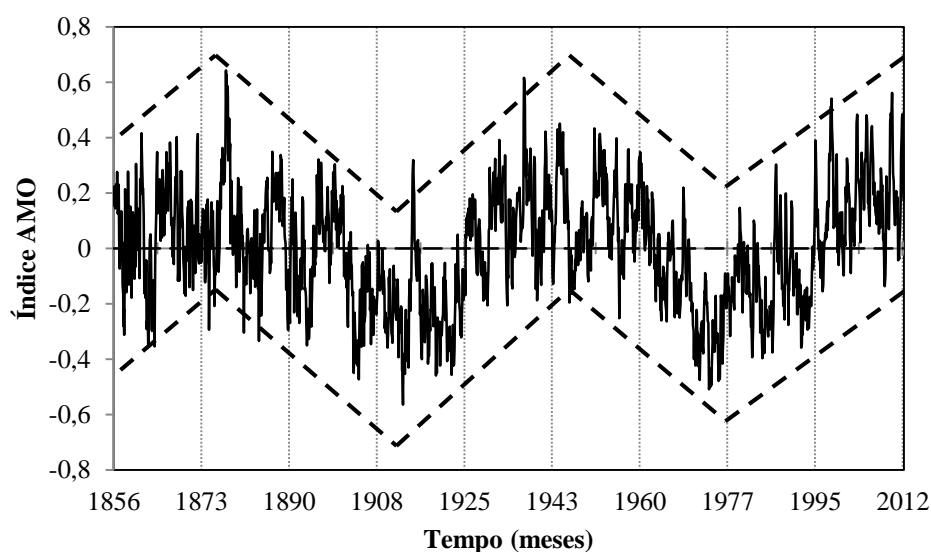
é estacionário para todo  $\tau$ .

Deve-se ressaltar que os conceitos explicados até aqui são definidos para o processo estocástico como um todo. No caso de séries temporais, pode-se encontrar definições equivalentes em referência aos seus momentos estatísticos, ou seja, diz-se que uma série possui *comportamento* estacionário se suas estatísticas não variam com o tempo. O termo em destaque recebeu essa ênfase porque a denominação “estacionariedade” é restrita a modelos e processos estocásticos (Klimes, 1974; Hipel e McLeod, 1994, p. 67; Box, Jenkins e Reinsel, 2008, p. 7; Koutsoyiannis, 2011). Dessa maneira, considera-se conceitualmente incoerente a classificação de uma série como estacionária ou não estacionária. Ainda, Klimes (1974) afirma que essa classificação é mais uma questão filosófica do que matemática, justamente por estar o termo

“estacionariedade” atrelado a modelos ou processos estocásticos. Ainda assim, talvez por mera simplicidade, é extremamente comum na literatura o uso desse termo para séries temporais. No decorrer da tese, a distinção será feita somente em situações necessárias, nas quais poderá haver algum tipo de confusão entre os conceitos.

Makridakis, Wheelwright e Hyndmon (1998) apontam três tipos de comportamentos não estacionários comuns a séries do cotidiano: (i) sazonalidade, evidente em diversos fenômenos naturais e motivada principalmente por variações climáticas entre estações do ano; (ii) ciclos, quando uma série exibe momentos de ascensão e decréscimo sem respeitar um período fixo e (iii) tendências, associadas a crescimentos ou decréscimos de longo termo em uma série.

A principal diferença entre a sazonalidade e o ciclo é que a sazonalidade possui duração aproximadamente constante e ocorre quase periodicamente em uma série. Por sua vez, o ciclo não possui duração fixa e sua magnitude é mais pronunciada, sendo frequentemente associado a fenômenos de grande escala. Portanto, comportamentos cíclicos só são percebidos com clareza em séries que possuam longos históricos. A título de ilustração, a Figura 2 mostra a série histórica do chamado índice de oscilação multidecadal do oceano Atlântico (AMO). Trata-se de um indicador climático relacionado à temperatura da superfície do oceano Atlântico, em sua porção norte. Nesse gráfico estão plotados índices históricos mensais com mais de 150 anos de registros (1856 a 2012), juntamente com linhas tracejadas que delimitam os ciclos da série. Dentro de cada porção ascendente ou descendente do ciclo é possível observar o comportamento sazonal do índice.



**Figura 2 – Índice de oscilação multidecadal do oceano Atlântico (AMO). As linhas tracejadas delimitam os ciclos observados (Fonte: NOAA - *National Oceanic and Atmospheric Administration*)**

Contudo, grande parte das análises de estacionariedade de séries hidrológicas tem como meta a identificação de tendências, comumente associadas a respostas às atividades antrópicas nas bacias hidrográficas (Milly *et al.*, 2008; Naik e Jay, 2011). Ainda, segundo alguns autores (Kahya e Kalayci, 2004), as tendências podem ser consequência de mudanças climáticas, embora este seja um tema de grande controvérsia (Salas *et al.*, 2012).

Somada à discussão sobre a causa das tendências, há também a questão da forma com a qual ela é considerada. Na grande maioria dos estudos, as tendências são tratadas como uma função determinística (e muitas vezes linear), tanto para sua detecção quanto para sua modelagem (Delgado, Apel e Merz, 2010; Gautam, Acharya e Tuladhar, 2010; Yue *et al.*, 2002). Entretanto, Koutsoyiannis (2006) alerta que há uma contradição entre a aplicação de técnicas estocásticas nas séries e a atribuição de uma função determinística para a tendência, inclusive questionando a possibilidade de se estabelecer tal função.

À parte das discussões supramencionadas, é fato que a maioria das séries de fenômenos naturais exibe uma combinação dos padrões de sazonalidade, ciclo e tendência, oferecendo desafios à sua modelagem. Ainda assim, esses comportamentos não estacionários podem ser removidos das séries usando ferramentas relativamente simples (ex.: diferenciação da série, Box, Jenkins e Reinsel, 2008, p. 96, com comentários na seção 1.4.3.2).

De um modo geral, observa-se na literatura um uso indiscriminado dos conceitos de comportamento estacionário e homogeneidade de uma série hidrológica, no sentido de que ambos são aplicados às mesmas situações. Há de se apontar, entretanto, que esses conceitos apresentam distinções. Naghettini e Pinto (2007, p. 265) explicam que uma série hidrológica homogênea é aquela formada por observações extraídas de uma mesma população. Por outro lado, uma série não homogênea pode ser formada por observações extraídas de populações distintas (i) no tempo, o que recai na definição de uma série com comportamento não estacionário tal como exposto anteriormente ou (ii) nos mecanismos de causa. O caso (ii) pode ser exemplificado através de uma série de precipitações médias anuais em um local que sofre a influência de fenômenos como El Niño. Os valores médios de precipitação podem ser significativamente distintos entre esses anos, podendo eventualmente ser considerados amostras de populações diferentes. Outro exemplo ocorre em bacias hidrográficas que possuem contribuições de água de degelo. Nesse caso, os picos de vazões máximas mensais são diferentes quando provocados por chuvas extremas ou quando são resultantes de degelo. Dessa maneira, entende-se que o conceito de homogeneidade é mais abrangente do que o de comportamento estacionário.

Outras causas de heterogeneidade na hidrologia podem ser devidas a terremotos, erupções vulcânicas, grandes deslizamentos de terra, entre outros (Salas *et al.*, 1985, p. 44). Especificamente, a homogeneidade de uma série de vazões pode ser afetada pelo barramento de um rio, principalmente quando o propósito envolve regularização de vazões e controle de cheias. A realocação de estações de medições de variáveis hidrológicas é outra causa potencial para a não homogeneidade (Potter, 1979).

É válido mencionar, entretanto, que na literatura clássica de análise de séries temporais o conceito de homogeneidade é associado a séries com comportamento não estacionário (Hipel e McLeod, 1994, p. 76; Morettin e Toloí, 2004, p. 4; Souza e Camargo, 2004, p. 25; Box, Jenkins e Reinsel, 2008, p. 95). Assim, uma série com comportamento não estacionário e homogênea é aquela cujas características não estacionárias podem ser eliminadas através de diferenciações na série (ver seção 1.4.3.2). Todavia, essa interpretação não será adotada no presente trabalho.

#### 1.1.4 Homoscedasticidade e heteroscedasticidade

Retomando-se a situação geral de um processo estocástico com realizações independentes, a homoscedasticidade é observada quando as variâncias das diversas séries não diferem entre si. Caso isso não ocorra e as variâncias das realizações sejam (estatisticamente) distintas, tem-se um processo heteroscedástico.

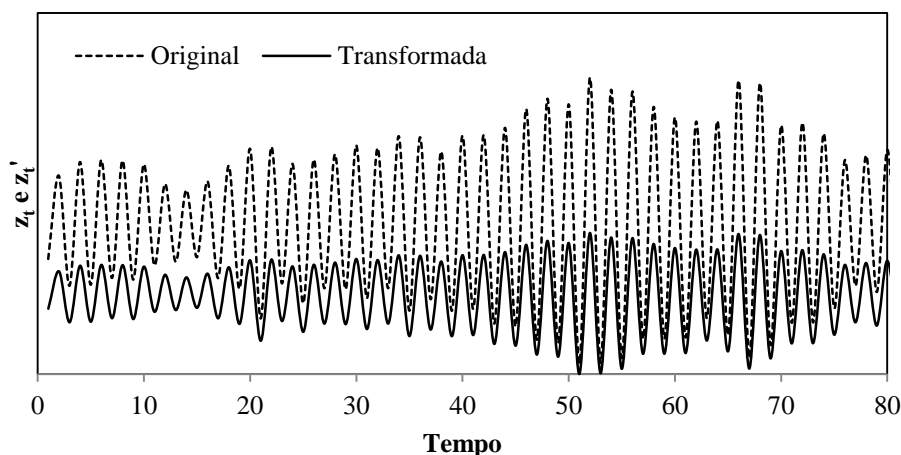
Para uma série temporal específica, a homoscedasticidade é observada quando sua variância se mantém constante ao longo do tempo. Da mesma forma que as condições de não estacionariedade, a heteroscedasticidade é indesejável para modelagem, pois caracteriza um desequilíbrio estatístico da série em questão dificultando o ajuste de modelos. Entretanto, existem artifícios que podem estabilizar a variância de uma série heteroscedástica e a tornar menos volátil. Um dos mais utilizados é a transformação numérica de Box-Cox (Box e Cox, 1964), dada pela equação (1.11):

$$z'_t = \begin{cases} \frac{(z_t + \epsilon)^\lambda - 1}{\lambda} & ; \lambda \neq 0 \\ \ln(z_t + \epsilon) & ; \lambda = 0 \end{cases} \quad (1.11)$$

onde  $z_t$  é a série original,  $z'_t$  é a série transformada Box-Cox,  $\epsilon$  é uma constante e  $\lambda$  é o parâmetro da transformação. Essa transformação não permite a existência de números negativos na série; caso ocorram tais valores, faz-se uso da constante  $\epsilon$  de forma a obter somente valores positivos. Em geral, a estimação do parâmetro  $\lambda$  segue o método da máxima verossimilhança,



recomendado pelos próprios autores em seu trabalho original. Como característica interessante da transformação Box-Cox, tem-se que ela conduz a uma mudança da forma da distribuição de probabilidades da amostra em análise, sendo utilizada por muitos autores para aproximar séries com distribuições diversas da distribuição normal (Wilks, 2006, p. 45).



**Figura 3 – Série heterocedástica submetida à transformação Box-Cox**

A Figura 3 exibe uma série heterocedástica que foi submetida à transformação Box-Cox na tentativa de estabilizar sua variância. Percebe-se que a transformação aplicada promove um abatimento na dispersão da série original, resultando em uma série transformada que se aproxima mais de uma condição de homocedasticidade.

Dependendo do nível de não estacionariedade da série com a qual se está trabalhando, pode ser necessário o uso conjunto de diferenciações e transformações na intenção de obter uma série própria para a modelagem.

#### 1.1.5 Ergodicidade

A teoria ergódica tradicional tem ramificações em diversas áreas do conhecimento, envolvendo principalmente o entendimento de sistemas dinâmicos. Nesse contexto, procura-se estudar as trajetórias de partículas em um sistema qualquer depois de transcorrido um prolongado período de tempo. De um modo geral, o sistema é dito ergódico quando as trajetórias das partículas têm a mesma probabilidade de passar pelos diversos estados que compõem o sistema durante o período de tempo considerado (Choe, 2005). A importância de se ter um sistema ergódico, portanto, reside no fato de que as características do sistema inteiro podem ser inferidas com base na trajetória de apenas uma partícula.

Traçando-se um paralelo entre as trajetórias de partículas em um sistema dinâmico com as realizações de um processo estocástico, tem-se que um processo é ergódico quando suas

estatísticas podem ser inferidas a partir de apenas uma série temporal. Dessa maneira, admitir a condição de ergodicidade do processo estocástico é de vital importância, visto que a única informação disponível é a própria série histórica.

Um processo ergódico de primeira ordem, dito ergódico na média, é aquele em que a relação (1.12) é verdadeira (Papoulis, 1965, p. 328; Hamilton, 2004, p. 46-47):

$$\lim_{n \rightarrow \infty} \left( \frac{1}{n} \sum_{t=1}^n z_t \right) = E[Z_t] = \mu \quad (1.12)$$

onde  $z_t$  é a série temporal de tamanho  $n$ ,  $E[Z_t]$  é o valor esperado do processo estocástico  $Z_t$  e  $\mu$  é sua média populacional. Por sua vez, um processo ergódico de segunda ordem, ou ergódico na autocovariância, atende à relação (1.13):

$$\lim_{n \rightarrow \infty} \left[ \frac{1}{n} \sum_{t=1}^{n-k} (z_t - \mu_z)(z_{t+k} - \mu_z) \right] = COV[Z_t] = \gamma \quad (1.13)$$

onde  $k$  é a defasagem temporal (ou *lag*) da série,  $COV[Z_t]$  representa o operador covariância do processo estocástico  $Z_t$  e  $\gamma$  é sua autocovariância populacional. É evidente que as relações (1.12) e (1.13) não possuem aplicação prática, pois a princípio não se conhecem os parâmetros  $\mu$  e  $\gamma$  do processo estocástico. Entretanto, para se ter um processo ergódico de segunda ordem é suficiente que as autocovariâncias amostrais  $\hat{\gamma}_k$  respeitem a condição (Wei, 2006, p. 20):

$$\sum_{k=0}^{\infty} |\hat{\gamma}_k| < \infty \quad (1.14)$$

Hamilton (2004, p. 47) comenta que em muitas situações as condições de estacionariedade e ergodicidade de um processo estocástico resultam nas mesmas restrições impostas às suas realizações. Ademais, Fill (2011) argumenta que em se tratando de séries hidrológicas, quando se considera que o processo é estacionário, é usual assumi-lo ergódico.

É importante apontar que caso a série tenha comportamento não estacionário, a condição (1.14) pode não se verificar, ou seja, as autocovariâncias sucessivas das defasagens  $k$  não formam uma série convergente. Dessa forma, uma maneira de respeitar a ergodicidade seria remover esse comportamento através de diferenciações sucessivas na série (ver seção 1.4.3.2).

## 1.2 A estacionariedade estatística na hidrologia

A questão do comportamento (não) estacionário de séries hidrológicas é tema de discussões na comunidade científica há mais de quatro décadas (e.g. Klemes, 1974). Em grande parte, isso se deve ao crescimento do uso de modelos estocásticos lineares do tipo Box & Jenkins a partir da segunda metade da década de 1970, formulações estas que exigem que as séries analisadas tenham características estacionárias. O ajuste de modelos estocásticos a séries hidrológicas, seja visando geração de cenários ou a realização de previsões, é prejudicado enormemente quando os momentos estatísticos das séries variam com o tempo.

Duas são as causas frequentemente apontadas como indutoras de distúrbios nas séries hidrológicas (Clarke, 2007; Milly *et al.*, 2008): (i) ações antrópicas na bacia hidrográfica e (ii) mudanças climáticas<sup>2</sup>. No primeiro caso, destacam-se as alterações no uso do solo de bacias hidrográficas provocadas principalmente pela substituição da cobertura vegetal original por campos agropecuários e impermeabilização devido à urbanização. Fisicamente, quando se altera a vegetação de uma bacia hidrográfica, muda-se o balanço entre interceptação vegetal, evapotranspiração, infiltração no solo e escoamento superficial. Na grande maioria dos casos, o tempo entre a chuva e a chegada da água no rio é reduzido, afetando picos de cheias (Tucci e Braga, 2003). Alternativamente, podem ser observadas diminuições nas vazões afluentes de rios devido ao aumento da demanda hídrica.

O segundo aspecto está relacionado com a emissão de gases de efeito estufa na atmosfera, possivelmente repercutindo em um aumento gradativo da temperatura do planeta. No entanto, muito embora essa hipótese de aquecimento da superfície da Terra tenha boa aceitação entre os cientistas (Matalas, 1997), seu efeito nas séries hidrológicas é menos evidente (Salas *et al.*, 2012). Ainda assim, autores como Kahya e Kalayci (2004) argumentam que a velocidade de mudanças na geomorfologia da bacia hidrográfica é lenta e, portanto, tendências observadas nas séries podem ser referidas a mudanças no clima.

Adicionalmente às duas causas mencionadas, Salas *et al.* (2012) apontam a ocorrência de eventos naturais de grande magnitude (e.g. erupções vulcânicas e incêndios florestais) e os fenômenos atmosféricos-oceânicos de baixa frequência (e.g. El Niño e La Niña) como fatores que afetam o equilíbrio estatístico de séries hidrológicas.

---

<sup>2</sup> Na literatura (Tucci e Braga, 2003, p. 54) é comum associar o termo “mudanças climáticas” com alterações antrópicas, enquanto que o termo “variabilidade climática” com flutuações devido aos fenômenos naturais.

Independentemente das causas, alguns autores publicaram trabalhos com fortes opiniões acerca da violação do comportamento estacionário de séries hidrológicas. A intensidade de seus argumentos pode ser avaliada pelos próprios termos utilizados por esses autores. Milly *et al.*, (2008) intitulam seu trabalho com a frase “a estacionariedade está morta” [tradução livre de *Stationarity is dead (...)*]; Poveda e Álvarez (2012) falam em *colapso* da hipótese de estacionariedade [*El colapso de la hipótesis de estacionariedad (...)*]; Poveda (2012) clama o fim do *diletantismo* sobre mudanças climáticas e interferências antrópicas nas bacias, convencido de que não há dúvidas que essas são questões verdadeiras. O ponto comum desses estudos é a discussão sobre o que fazer para contornar essa situação que, segundo os autores, é irreversível.

Além da verificação dos momentos estatísticos das séries históricas para fins de adequação aos modelos estocásticos, um grande número de estudos foi desenvolvido com o propósito único de analisar de que forma a não estacionariedade se manifesta em séries hidrológicas. Destes, a maioria utiliza técnicas diversas para detecção de tendências lineares nos registros, possíveis indicadores de alteração no regime hidrológico. A escolha quase sempre recai em análises sobre séries de vazão, por ser considerada como unidade integradora das variáveis climáticas nas bacias hidrográficas (McCabe Jr. e Wolock, 1997; Genta, Perez-Iribarren e Mechoso, 1998). Nas próximas seções serão detalhados os métodos mais comuns de detecção de tendências em variáveis hidrológicas, acompanhados de comentários acerca da aplicabilidade de cada um.

### 1.2.1 Detecção de tendências em séries hidrológicas

Uma análise sobre os diversos estudos para detecção de tendências em séries hidrológicas permite apontar os testes de hipóteses (ou testes de significância) como métodos mais utilizados. Tratam-se de ferramentas estatísticas empregadas para confrontar suposições acerca dos dados de interesse. A hipótese a ser testada recebe a denominação de hipótese nula ( $H_0$ ) e está sempre acompanhada de uma hipótese alternativa ( $H_A$ ). Se as hipóteses forem complementares,  $H_A$  é adotada como verdadeira no caso de rejeição de  $H_0$ . Um exemplo básico da estruturação de tais hipóteses no contexto da detecção de tendências em séries hidrológicas seria:

- $H_0$ : a série não apresenta tendências no tempo.
- $H_A$ : a série apresenta tendência de crescimento (ou decaimento) no tempo.

Uma vez determinadas as hipóteses, a estatística do teste é calculada com base na amostra e comparada a valores obtidos de sua distribuição de probabilidades. O chamado nível de significância ( $\alpha$ ) do teste estabelece a região na distribuição de probabilidades que determina a rejeição (ou não) da hipótese nula. Em outras palavras, o nível de significância define um valor crítico para a estatística do teste, acima do qual há rejeição da hipótese nula. Como valores típicos para  $\alpha$ , trabalha-se com 1%, 5% ou 10% (Wilks, 2006, p. 133). Ressalta-se que os testes podem ser feitos de forma bilateral, na qual o que importa é a presença ou não da tendência independentemente de seu sentido (e.g. crescimento ou decaimento), ou unilateral, no qual pode ser testado um sentido específico de tendência, bastando que se fixe a região de rejeição em uma das caudas da distribuição de probabilidades.

Os diversos testes de hipóteses existentes podem ser agrupados em duas categorias: (i) testes paramétricos e (ii) não paramétricos. Na primeira categoria assume-se conhecer a distribuição de probabilidades teórica representativa dos dados em análise, bem como do teste empregado. Por outro lado, testes não paramétricos não requerem essa informação, sendo formulados de tal sorte que a distribuição de probabilidades dos dados não é relevante (Wilks, 2006, p. 131). É evidente que, em se tratando de variáveis hidrológicas, a distribuição de probabilidades em geral não é conhecida, sendo necessário admitir apriori uma distribuição aproximada para os dados. Dessa forma, diversos pesquisadores optam pelos testes não paramétricos mesmo sabendo que eles são menos poderosos (em termos de probabilidade de rejeitar  $H_0$  quando ela realmente for falsa) do que os testes paramétricos (Chebana, Ouarda, e Duong, 2013).

Dentre as premissas para utilização dos testes de significância (paramétricos e não paramétricos), tem-se que a elaboração das hipóteses deve sempre ser feita previamente à análise da série, de forma que o investigador não seja influenciado pelo comportamento da variável (von Storch e Navarra, 1999, p. 13; Koutsoyiannis, 2006; Naghettini e Pinto, 2007, p. 245). Além disso, os elementos das amostras devem ser independentes entre si, fato que não ocorre na maioria das séries hidrológicas, principalmente em dados de vazões médias de durações subanuais. Em geral, a presença de autocorrelação entre os registros pode resultar em períodos longos de valores acima ou abaixo da média da série (fenômeno conhecido como “Efeito Joseph”, Mandelbrot e Wallis, 1968), levando os testes a indicar a presença de tendências quando, na realidade, elas não existem (Fleming e Weber, 2012). Essas duas premissas fazem com que a aplicação dos testes de hipóteses sobre variáveis hidrológicas seja criticada por alguns pesquisadores (e.g. Koutsoyiannis, 2006), porém, ainda assim é muito

evidente na literatura a preferência pelo uso de tais técnicas para analisar o comportamento estatístico das séries hidrológicas.

Outra forma de se estudar possíveis tendências em séries é através de análise espectral. Essa abordagem tem por base a decomposição da série em harmônicos, utilizando-os como indicadores de flutuações e variações de uma série. Aplicam-se, para tanto, as transformadas de Fourier e *wavelet* após as quais é possível analisar as contribuições que processos com velocidades ou escalas diferentes fazem à série temporal (Wilks, 2006, p. 383).

Em meio à importância dos testes de hipóteses e da análise espectral no estudo de séries hidrológicas, os próximos subitens se destinam a detalhar as principais técnicas e aplicações desses métodos no contexto da estacionariedade estatística.

#### 1.2.1.1 Regressão linear

Como mencionado anteriormente, tendências lineares são consideradas como as mais simples formas de manifestação da não estacionariedade em séries hidrológicas. Por esse motivo, o ajuste de uma reta de regressão linear aos dados e análise do coeficiente angular resultante é o método mais simples e direto. A regressão linear estima a relação entre uma variável dependente  $y$  (tipicamente a série hidrológica) e uma variável independente  $x$  (em geral os índices de tempo) de acordo com a equação (1.15):

$$y = c_1x + c_2 + a_t \quad (1.15)$$

onde  $c_1$  e  $c_2$  são os coeficientes do modelo e  $a_t$  representa os resíduos do ajuste. Em geral, as estimações dos coeficientes são feitas através do método dos mínimos quadrados (ver Naghettini e Pinto, 2007, p. 363). Nesta equação, o coeficiente angular  $c_1$  possui grande importância, pois revela a presença de uma tendência na série caso seu valor seja positivo (tendência de crescimento) ou negativo (tendência de redução). Assumindo que a série de resíduos possui distribuição normal e é homocedástica, a significância do coeficiente  $c_1$  pode ser testada sob a hipótese nula de ser estatisticamente igual a zero, caracterizando uma série sem tendência. A inferência é baseada no teste t-Student, cuja estatística é dada pela equação (1.16) (Naghettini e Pinto, 2007, p. 372):

$$t = \frac{c_1}{\hat{\sigma}_{c_1}} \quad (1.16)$$

O desvio padrão  $\hat{\sigma}_{c_1}$  do coeficiente  $c_1$  é determinado calculando-se a equação (1.17):

$$\hat{\sigma}_{c_1} = \sqrt{\frac{\hat{\sigma}_{a_t}^2}{\sum_{i=1}^n (x_i - \hat{\mu}_x)^2}} \quad (1.17)$$

onde  $\hat{\mu}_x$  é a média amostral da série  $x_i$ , de comprimento  $n$ . Por sua vez, a variância  $\hat{\sigma}_{a_t}^2$  da série de resíduos do ajuste é definida pela equação (1.18):

$$\hat{\sigma}_{a_t}^2 = \frac{\sum_{i=1}^n a_i^2}{n - 2} \quad (1.18)$$

A hipótese nula é rejeitada caso  $t > t_{1-\alpha/2, v}$ , sendo  $v$  o número de graus de liberdade ( $v = n - 2$ ),  $\alpha$  o nível de significância e  $t_{1-\alpha/2, v}$  o valor limite retirado de tabelas da distribuição t-Student. Esse teste é bilateral do tipo paramétrico e requer que a série tenha distribuição normal.

Bormann, Pinter, e Elfert (2011) utilizaram a regressão linear como um dos métodos para detectar tendências em 78 rios espalhados pela Alemanha. O foco do estudo foi a análise de frequência de cheias e, portanto, as séries utilizadas foram de vazões e níveis máximos dos rios, ambos em escala anual. Ainda em território alemão, Pinter *et al.* (2006) submeteram séries de vazões médias do rio Reno e de precipitação em sua bacia hidrográfica a análises via regressão. Na região comum aos dois trabalhos, tendências positivas nas vazões foram detectadas, entretanto não acompanhadas por tendências nas precipitações.

Regressões foram também empregadas por Sharma *et al.* (2000) na bacia do rio Kosi, na cordilheira do Himalaia. O trabalho desses autores foi mais abrangente, na medida em que envolveu séries de temperatura, precipitação e vazão média em escalas mensais e anuais. Para as séries mensais, os pesquisadores as dividiram em períodos sazonais previamente à aplicação dos testes. Os resultados mostraram tendências positivas para temperaturas e chuvas e tendências negativas para vazões, o que levou os autores a suspeitar de alterações no regime de neves na região.

Séries de vazões médias com comprimentos superiores a 100 anos, observadas em rios suecos, foram submetidas a regressões no estudo de Lindström e Bergström (2004). Dados de diversos postos foram resumidos a uma série média, sobre a qual a análise se desenvolveu. O interessante nesse trabalho foi que no período completo disponível não se detectou tendência significativa, porém, ao dividir a série em subperíodos, as tendências apareceram. Assim, os autores apontaram a escolha dos períodos como determinante para os resultados.

## 1.2.1.2 Teste t-Student para duas amostras

Quando se deseja analisar diferenças entre as médias de duas amostras, o teste t-Student é uma das inferências paramétricas mais conhecidas na literatura. Este teste assume que as amostras provêm de populações com distribuição normal e possui algumas variantes, dependendo do conhecimento ou desconhecimento e da igualdade ou desigualdade das variâncias populacionais. Para o caso mais genérico no qual se assume as variâncias desiguais e desconhecidas, a estatística do teste é dada pela equação (1.19) (Naghettini e Pinto, 2007, p. 257):

$$t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}}} \quad (1.19)$$

onde  $\hat{\mu}$  é a média amostral,  $\hat{\sigma}^2$  a variância amostral e  $n$  o tamanho da amostra. Os índices subscritos indicam as duas amostras consideradas. Para o teste bilateral, a hipótese nula de igualdade entre as médias amostrais é rejeitada se  $t > t_{1-\alpha/2,v}$ , sendo  $v$  o número de graus de liberdade ( $v = n - 2$ ),  $\alpha$  o nível de significância e  $t_{1-\alpha/2,v}$  retirado de valores tabelados da distribuição t-Student. Para amostras pequenas, o número de graus de liberdade pode ser aproximado por  $v = \min(n_1, n_2) - 1$  (Wilks, 2006, p. 141). Alternativamente, ele pode ser calculado de acordo com a expressão (1.20) (Naghettini e Pinto, 2007, p. 257):

$$v = \frac{\left(\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}\right)^2}{\frac{\hat{\sigma}_1^2/n_1}{n_1 - 1} + \frac{\hat{\sigma}_2^2/n_2}{n_2 - 1}} \quad (1.20)$$

Na análise da condição de estacionariedade de séries hidrológicas, o procedimento tradicional é dividir o histórico em duas subamostras e aplicar o teste. Se a hipótese nula for rejeitada, uma tendência de aumento ( $\hat{\mu}_1 < \hat{\mu}_2$ ) ou redução ( $\hat{\mu}_1 > \hat{\mu}_2$ ) nas médias das séries é detectada (considerando que a subamostra dois representa o período mais recente). Este esquema foi empregado, por exemplo, nos trabalhos de Müller, Krüger e Kaviski (1998), Guetter e Prates (2002), Batista *et al.* (2009), Detzel *et al.*, (2011) e Fill (2011). Diferentemente destes, Liang, Li, e Liu (2011) aplicaram o teste t-Student sucessivas vezes em janelas móveis de cinco e 10 anos, na intenção de identificar mudanças abruptas nas séries de precipitação no nordeste da China.



Com relação à premissa de normalidade das séries a serem testadas, é muito comum a aplicação de transformações numéricas do tipo Box-Cox [equação (1.11)]. Especificamente para séries de vazões médias, uma simples transformação logarítmica garante boa aproximação (Kelman, 1987, p. 370). Por outro lado, a premissa de independência merece maiores cuidados, principalmente devido à forte correlação serial presente em dados hidrológicos. Na realidade, a aplicação do teste t-Student para duas amostras requer dois tipos de independência: (i) das amostras entre si e (ii) entre os elementos internos de cada amostra. Trabalhando com escalas anuais, Fill (2011) solucionou o primeiro tipo eliminando da análise o ano limítrofe utilizado para separação das amostras. Para o segundo tipo, Wilks (2006, p. 144) recomenda a utilização do conceito de tamanho efetivo da amostra  $n'$ , calculado de acordo com a equação (1.21):

$$n' = n \frac{1 - r}{1 + r} \quad (1.21)$$

onde  $r$  é o coeficiente de autocorrelação de primeira ordem da amostra. Sua utilização como único parâmetro para a determinação do tamanho efetivo da amostra sugere que a persistência entre seus elementos seja de primeira ordem, isto é, uma observação  $z_t$  depende somente de sua observação  $z_{t-1}$  imediatamente anterior. Nestes casos, a série pode ser representada por um modelo autorregressivo de primeira ordem [AR(1)]. Uma vez calculados  $n'_1$  e  $n'_2$ , o teste pode ser executado substituindo-se estes valores no lugar de  $n_1$  e  $n_2$  nas equações (1.19) e (1.20).

Uma alternativa para lidar com séries correlacionadas é fornecida por von Storch e Navarra (1999, p. 17). Trata-se de uma transformação numérica também baseada na equação do modelo AR(1). Sendo  $z_t$  a série original, obtém-se a série transformada  $x_t$  aplicando-se a expressão (1.22):

$$x_t = z_t - r \cdot z_{t-1} \quad (1.22)$$

Ao processo de passar a série de  $z_t$  para  $x_t$ , dá-se o nome de “branqueamento” da série (em inglês *pre-whitening*), em referência à eliminação da influência da persistência nos dados. Dessa maneira, o teste t-Student é realizado sobre a série “branca”  $x_t$ . Vale ressaltar que essa técnica só é válida para séries com comportamento similar a um processo AR(1); se este não for o caso, von Storch e Navarra (1999, p. 18) recomendam a execução de análises baseadas no método de Monte Carlo.

### 1.2.1.3 Teste de Wilcoxon

Também chamado de teste de Mann-Whitney, essa inferência é não paramétrica e depende da divisão da amostra principal em duas subamostras. A hipótese nula busca verificar se as duas subamostras, em princípio independentes, fazem parte de uma mesma população. Caso o teste resulte em rejeição dessa hipótese, considera-se que a amostra em análise possui comportamento não estacionário.

Após a divisão em subamostras, formam-se dois conjuntos totalizando  $n_1$  e  $n_2$  elementos em cada uma. Na sequência, os valores são ordenados em conjunto, atribuindo-se índices  $(1, 2, \dots, n_1 + n_2)$ . A estatística do teste  $W$  é obtida através da soma dos índices de cada amostra,  $z_1$  e  $z_2$ , o que resultar no menor valor. Para amostras consideradas grandes ( $n_1 > 10$  ou  $n_2 > 10$ ) a distribuição amostral de  $W$  aproxima-se de uma normal, com variável padrão determinada pela equação (1.23) (Naghetini e Pinto, 2007, p. 265-266):

$$z = \frac{W - n_1(n_1 + n_2 + 1)/2}{\sqrt{n_1 \cdot n_2(n_1 + n_2 + 1)/12}} \quad (1.23)$$

No teste bilateral, a hipótese nula é rejeitada se  $|z| > z_{1-\alpha/2}$ , onde  $\alpha$  é o nível de significância e  $z_{1-\alpha/2}$  é retirado de tabelas da distribuição normal.

Por ser um teste não paramétrico, a distribuição de probabilidades da série em análise não é importante, entretanto a independência entre as subamostras sim. As mesmas considerações feitas na seção anterior sobre essa questão são também válidas para esse teste, sendo a utilização do conceito de tamanho efetivo da amostra recomendada por Cluis e Laberge (2001).

O teste de Wilcoxon encontra aplicações em séries de vazões de alguns rios importantes em todo o mundo. Batista *et al.* (2009) e Fill (2011) o empregaram no rio Iguaçu, na região Sul do Brasil, no qual a hipótese de estacionariedade foi rejeitada. Detzel *et al.* (2011) usaram esse teste em conjunto com outros quatro para analisar as condições estatísticas das aflúncias às principais usinas hidrelétricas brasileiras. Das 146 séries, 75 apresentaram comportamento não estacionário, inclusive em usinas operando no rio Iguaçu, corroborando os estudos mencionados. Genta *et al.* (1998) avaliaram séries dos rios Paraná, Paraguai, Uruguai e Negro, todos na América do Sul, chegando à conclusão de um aumento na mediana das vazões médias após a década de 1970. Cluis e Laberge (2001) aplicaram o teste em diversos rios da Ásia, detectando tendências negativas nos rios localizados no sudeste do continente. Thomas (2007)

usou o teste de Wilcoxon no rio Colorado, nos Estados Unidos, visando estudar as relações entre as vazões deste rio e variáveis atmosféricas em meio à variabilidade climática.

Por fim, séries de precipitações anuais sobre o Japão foram estudadas no trabalho de Xu, Takeuchi e Ishidaira (2003). Embora tendências de longo termo não tenham sido detectadas, algumas quebras em valores médios nas séries foram encontradas através do teste de Wilcoxon.

#### 1.2.1.4 Teste de Pettitt

Da mesma forma que o teste de Wilcoxon explicado na seção anterior, a inferência de Pettitt (Pettitt, 1979) é não paramétrica e tem como objetivo verificar se duas subamostras pertencem a uma mesma população. No entanto, diferentemente do primeiro, esse teste não necessariamente precisa da prévia divisão da amostra principal por parte do analista (embora isso possa ser feito). Ao invés disso, o possível limite que divide a amostra em duas subamostras estatisticamente diferentes é o resultado do teste. Por esse motivo, o teste de Pettitt é utilizado quando se deseja identificar pontos de mudanças abruptas nas séries históricas (e.g. Kundzewicz e Robson, 2004; Villarini *et al.*, 2011; Rougé, Ge e Cai, 2013).

Sendo  $\tau$  ( $1 \leq \tau \leq n$ ) um ponto arbitrário de divisão da amostra  $z_t$  ( $t = 1, \dots, n$ ), o teste de Pettitt se inicia com o cálculo das equações (1.24) e (1.25):

$$PT(\tau) = \sum_{i=1}^{\tau} \sum_{j=\tau+1}^n \text{sgn}[z_j - z_i] \quad (1.24)$$

$$\text{sgn}[z_j - z_i] = \begin{cases} 1, & \text{se } (z_j - z_i) > 0 \\ 0, & \text{se } (z_j - z_i) = 0 \\ -1, & \text{se } (z_j - z_i) < 0 \end{cases} \quad (1.25)$$

onde  $PT(\tau)$  é um vetor que guarda as totalizações em cada subdivisão da amostra e  $\text{sgn}[ ]$  é o sinal resultante da operação entre colchetes. A estatística final do teste é dada pela equação (1.26):

$$PT = \max|PT(\tau)|, \quad 1 \leq \tau \leq n \quad (1.26)$$

Evidentemente, o índice temporal associado a  $PT$  representa a data da possível subdivisão da amostra principal. Ademais, por ser empregado o módulo do vetor  $PT(\tau)$  na equação, o teste é bilateral. O nível de significância da estatística de Pettitt é aproximado pela equação (1.27) (Zhang *et al.*, 2009):

$$\alpha \sim \exp\left(\frac{-6PT^2}{n^3 + n^2}\right) \quad (1.27)$$

Assim, a hipótese nula de subamostras pertencentes a uma mesma população é rejeitada caso  $\alpha$  seja menor do que o nível de significância estabelecido pelo analista.

Zhang *et al.* (2009) aplicaram o teste de Pettitt para detectar descontinuidades nas séries de níveis máximos e mínimos anuais em 18 postos hidrométricos em grandes rios chineses. Embora quebras nas séries de máximos quase não tenham sido identificadas, mudanças abruptas ocorreram para níveis mínimos em mais de 80% dos postos estudados. De acordo com os autores, essas alterações foram consequência de atividades antrópicas na bacia ocorridas nas décadas de 1980 e 1990.

A identificação de quebras em séries de vazões anuais máximas na região central dos Estados Unidos é um dos objetivos de Villarini *et al.* (2011). Usado o teste de Pettitt, os autores concluíram que das 196 estações consideradas, 40% apresentaram quebra no valor médio, sendo 63% delas com crescimento nas médias após a mudança. Outras 36% apresentaram quebras na variância e 21% em ambos os momentos estatísticos. Ainda, em 25% dos casos as quebras se deram em um período menor que 10 anos, podendo indicar atividades antrópicas na bacia.

#### 1.2.1.5 Teste do Coeficiente de Correlação de Spearman

O teste de Spearman (em inglês conhecido como *Spearman's rho* em referência à letra grega comumente utilizada para expressar o coeficiente de correlação populacional) é uma técnica consistente na detecção de tendências em séries temporais, não estando sujeito à influência de valores extremos na série (Müller, Krüger e Kaviski, 1998). De concepção não paramétrica, o teste é executado sobre a série inteira, sem separação em subamostras.

O procedimento se inicia com a atribuição de índices à série original; a seguir a série é ordenada e novos índices são atribuídos. A diferença  $d_i$  entre o índice que um elemento adquire na série original e o que este mesmo elemento assume na série ordenada determina a variável chave do teste. O coeficiente de correlação de Spearman é calculado através da expressão (1.28) (Naghetini e Pinto, 2007, p. 267):

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \quad (1.28)$$

onde  $n$  é o tamanho da amostra. A estatística do teste, por sua vez, é dada pela expressão (1.29):

$$\rho_s = \frac{r_s}{\sqrt{\text{VAR}[r_s]}} \quad (1.29)$$

onde  $\text{VAR}[\ ]$  representa a variância da estatística do teste. Para amostras de tamanho suficiente (nesse caso,  $n > 10$ ),  $\rho_s$  segue uma distribuição normal padrão. Portanto, para um teste bilateral, a hipótese nula de série sem tendência é rejeitada quando  $|\rho_s| > z_{1-\alpha/2}$ , sendo  $\alpha$  o nível de significância desejado. Da mesma forma que nos testes apresentados anteriormente, deve-se tomar cuidado na aplicação dessa inferência em séries que apresentam estrutura de autocorrelação.

Nos estudos para detecção de tendências em séries, alguns autores optam por diferenciar mudanças abruptas (ou quebras) nas séries<sup>3</sup> de tendências monotônicas graduais. O trabalho de Villarini *et al.* (2011) é um exemplo, nos quais os autores usaram o teste de Pettitt para quebras (ver comentários na seção anterior) e o teste de Spearman para a identificação de tendências. Apesar de quebras encontradas nos valores médios de um grande número de séries, praticamente não foram detectadas alterações graduais de tendência. Outro estudo que faz essa distinção, aplicando o teste de Spearman para as tendências progressivas, é o anteriormente citado (seção 1.2.1.3) artigo de Cluis e Laberge (2001).

Fleming e Weber (2012) utilizam o teste de Spearman em seu estudo sobre séries anuais de aflúncias a reservatórios canadenses. Tecem comentários sobre a influência da autocorrelação em séries que serão submetidas a testes estatísticos, porém alegam que a fraca correlação das séries anuais (inferior a 0,20) dispensa qualquer tratamento específico. As análises mostraram aumento generalizado das vazões no período de inverno (dezembro e janeiro) e outros aumentos esporádicos nos volumes anuais médios. As explicações são baseadas em estudos anteriores, relacionando-se com tendências observadas nas temperaturas e precipitações. O teste de Spearman foi também aplicado por Gautam *et al.* (2010) na região do Nepal, no qual foram detectadas tendências de crescimento no regime de chuvas e vazões. Demais aplicações desse teste podem ser encontradas em McCabe Jr. e Wolock (1997), Müller, Krüger e Kaviski (1998), Kahya e Kalayci (2004) e Detzel *et al.* (2011).

#### 1.2.1.6 Teste de Mann-Kendall

O teste de Mann-Kendall (Mann, 1945; Kendall, 1975) é a inferência mais difundida na literatura para a detecção de tendências em séries naturais, sendo o teste recomendado pelo

---

<sup>3</sup> Para quebras nas séries, é comum o uso de testes que confrontam estatísticas entre subamostras, tais como t-Student, Wilcoxon e Pettitt descritos nas seções 1.2.1.2, 1.2.1.3 e 1.2.1.4, respectivamente.

World Meteorological Organization para este fim (Liang *et al.*, 2011). Trata-se de uma técnica não paramétrica que emprega a amostra completa em sua formulação. Além da independência entre os elementos de uma amostra, o teste se baseia no pressuposto de que a probabilidade de ocorrência desses elementos é a mesma.

A base do teste está na comparação da grandeza entre os elementos da amostra  $z_t$  ( $t = 1, \dots, n$ ), contabilizando os sinais resultantes das subtrações entre eles. Em outras palavras, fixando-se um ano  $i$ , verificam-se quantos elementos posteriores ( $j, j + 1, j + 2, \dots$ ) são maiores ou menores do que ele. A equação (1.30) define matematicamente o procedimento (Yue *et al.*, 2002):

$$MK = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \operatorname{sgn}[z_j - z_i] \quad (1.30)$$

onde  $\operatorname{sgn}[\ ]$  é o sinal resultante da operação entre colchetes ao qual se associa  $-1, 0$  ou  $1$ , de forma idêntica à definida pela equação (1.25). Para amostras com mais de oito elementos, a estatística  $MK$  possui distribuição aproximadamente normal com  $E[MK] = 0$  e variância calculada pela equação (1.31):

$$\operatorname{VAR}[MK] = \frac{n(n-1)(2n+5)}{18} \quad (1.31)$$

Na realidade, a equação (1.31) possui um complemento para o caso de elementos “empatados”, ou seja, quando  $\operatorname{sgn}[z_j - z_i] = 0$  na equação (1.25). Entretanto, dada a reduzida probabilidade de elementos iguais em uma série temporal de fenômenos naturais, optou-se por omitir essa parcela do equacionamento sem perda de generalidade (o leitor é direcionado a Yue *et al.* [2002] para a formulação completa). A variável normal padrão é, por fim, calculada de acordo com a expressão (1.32):

$$z = \begin{cases} \frac{MK - 1}{\sqrt{\operatorname{VAR}[MK]}}, & \text{se } MK > 0 \\ 0, & \text{se } MK = 0 \\ \frac{MK + 1}{\sqrt{\operatorname{VAR}[MK]}}, & \text{se } MK < 0 \end{cases} \quad (1.32)$$

Para um teste bilateral, a hipótese nula de série sem tendência é rejeitada quando  $|z| > z_{1-\alpha/2}$ , sendo  $\alpha$  o nível de significância adotado.

A ampla utilização dessa inferência inspirou estudos específicos a respeito do poder do teste (Yue, Pilon e Cavadias, 2002) e da influência que a autocorrelação nos dados exerce em sua capacidade de detectar tendências (Yue *et al.*, 2002). Ambos utilizaram simulações de Monte Carlo para as investigações; no primeiro artigo, chegou-se à conclusão de que a força do teste aumenta com a intensidade da tendência, tamanho da amostra disponível e nível de significância utilizado. No entanto, se a série exibir elevada variabilidade a capacidade de detecção de tendências é afetada. Cabe apontar que Yue, Pilon e Cavadias (2002) fizeram os mesmos ensaios para o teste de Spearman, o qual se mostrou tão eficiente quanto o teste de Mann-Kendall para a identificação de tendências.

Com relação à persistência nos dados, concluiu-se que sua presença resulta em um aumento na probabilidade de se detectar uma tendência. Yue *et al.* (2002) também notaram que a aplicação da técnica de “branqueamento” da série [equação (1.22)] para remoção da persistência pode retirar uma porção da tendência existente, levando o teste a subestimar sua significância. Sugeriram, portanto, um procedimento que envolve a estimação e remoção da tendência da série antes de seu “branqueamento”. Depois disso, a tendência estimada é incorporada novamente à série e esta é submetida ao teste.

A comprovação da eficiência do teste de Mann-Kendall para detecção de tendências lineares em séries hidrológicas é dada pela quantidade de trabalhos encontrados na literatura (Fraedrich *et al.*, 1997; Müller, Krüger e Kaviski, 1998; Lins e Slack, 1999; Xu, Takeuchi e Ishidaira, 2003; Cunderlik e Burn, 2004; Lindström e Bergström, 2004; Marengo e Alves, 2005; Wang *et al.*, 2008; Assani *et al.*, 2010; Bormann, Pinter e Elfert, 2011; Detzel *et al.*, 2011; Kottegoda, Natale, e Raiteri, 2011; Liang, Li e Liu, 2011; Villarini *et al.*, 2011; Li *et al.*, 2012; Rougé, Ge e Cai, 2013). Estes estudos fazem uso dessa inferência na análise de tendências de séries em localidades específicas distribuídas pelo mundo.

O teste de Mann-Kendall oferece também a possibilidade de detecção não somente de uma tendência gradual, mas do momento da série em que se inicia essa tendência. Isso é possível através da divisão da amostra estudada em  $n - 1$  subamostras, sendo a primeira composta pelos elementos  $(z_1, z_2)$ , a segunda pelos elementos  $(z_1, z_2, z_3)$  e assim por diante. A estatística do teste é calculada sequencialmente, de forma que se obtenha um vetor  $MK1$ , em contraste à estatística única  $MK$  obtida em sua versão tradicional. Invertendo-se a ordem da amostra principal, o procedimento é repetido para as subamostras  $(z_n, z_{n-1})$ ,  $(z_n, z_{n-1}, z_{n-2})$ , etc., obtendo-se um segundo vetor de estatísticas  $MK2$ . Traçando-se um gráfico dos vetores  $MK1$  e  $MK2$  em relação ao tempo, têm-se duas curvas que se cruzarão em algum momento da série. Se a intersecção for significativa a um nível  $\alpha$ , tem-se caracterizado um ponto crítico da

série, geralmente atribuído a uma quebra na tendência. Esse procedimento foi utilizado nos estudos de Gerstengarbe e Werner (1999), Zhang *et al.*, (2006) e Liang *et al.* (2011), nos quais maiores detalhes do teste de Mann-Kendall sequencial podem ser consultados.

### 1.2.1.7 Análises no domínio da frequência

Nas seções anteriores foram mostradas diversas técnicas comumente empregadas quando se trabalha com séries cujos elementos estão representados em função do tempo. Quando se transfere o foco para o domínio da frequência, o tempo passa a não ser importante, pois os dados são representados em função das contribuições que fazem à série como um todo.

Em sua forma tradicional, as flutuações e variações de uma série  $z_t$  são traduzidas através de uma combinação de senos e co-senos. Uma série com  $n$  elementos pode ser representada através da soma de  $n/2$  funções harmônicas, como mostra a equação (1.33) (Wilks, 2006, p. 378):

$$z_t = \hat{\mu}_z + \sum_{\kappa=1}^{n/2} \left\{ A_{\kappa} \cos \left[ \frac{2\pi\kappa t}{n} \right] + B_{\kappa} \sin \left[ \frac{2\pi\kappa t}{n} \right] \right\} \quad (1.33)$$

onde  $\hat{\mu}_z$  é a média amostral da série,  $\kappa$  é o ciclo (ou harmônico) entre  $t = 0$  e  $t = n$  e  $A_{\kappa}$  e  $B_{\kappa}$  são os coeficientes de Fourier, estimados de acordo com as equações (1.34):

$$A_{\kappa} = \frac{2}{n} \sum_{t=1}^n z_t \cos \left[ \frac{2\pi\kappa t}{n} \right] \quad (1.34)$$

$$B_{\kappa} = \frac{2}{n} \sum_{t=1}^n z_t \sin \left[ \frac{2\pi\kappa t}{n} \right] \quad (1.34a)$$

É importante lembrar que a boa representatividade de uma série não necessariamente implica na utilização dos  $n/2$  harmônicos. Não há uma regra definitiva sobre a quantidade  $m$  de harmônicos a serem empregados, sendo que a escolha depende da característica da série em análise. Ainda assim, a literatura oferece algumas alternativas: Wilks (2006, p. 379), em concordância com o princípio da parcimônia, atesta que o uso apenas dos primeiros harmônicos podem ser suficientes; Chu e Katz (1989) citam uma regra empírica na qual, para séries com mais de 200 elementos, um número apropriado de harmônicos seria  $\kappa < n/6$ ; Kite (1989) utiliza um teste de significância sobre os diversos harmônicos para a escolha.

Uma característica interessante da abordagem de harmônicos é que a propriedade de ortogonalidade das funções seno e co-seno permite o cálculo dos coeficientes  $A_{\kappa}$  e  $B_{\kappa}$  de forma



independente para qualquer harmônico. Isso possibilita representar a proporção da variância explicada por um harmônico simplesmente aplicando-se a equação (1.35):

$$R_k^2 = \frac{(n/2)C_k}{(n-1)\hat{\sigma}_z^2} \quad (1.35)$$

onde  $C_k = \sqrt{A_k^2 + B_k^2}$  e  $\hat{\sigma}_z^2$  é a variância amostral da série. Sabendo que cada processo responde por uma porção da variância da série, podem-se analisar as contribuições que processos com velocidades ou escalas diferentes trazem à série temporal.

Outra ferramenta da análise espectral muito utilizada é a função de densidade espectral (FDE), que resulta da transformada de Fourier aplicada à função de autocorrelação (FAC) da série. Matematicamente, a FDE é estimada através da equação (1.36) (Salas *et al.*, 1985, p. 51):

$$S(f) = 2 \left( 1 + 2 \sum_{k=1}^m D_k \hat{\rho}_k \cos 2\pi f k \right) \quad (1.36)$$

onde  $f$  é a frequência ( $f = 1/2m$ ) para  $m$  harmônicos,  $\hat{\rho}_k$  é a autocorrelação da série e  $D_k$  é a função de amortecimento, ou janela de defasagem, empregada para assegurar que a estimação é não tendenciosa. Existem muitas funções  $D_k$  possíveis de se empregar; Andreo *et al.* (2006) recomendam a utilização da janela de Tukey:

$$D_k = \frac{1 + \cos(\pi k / m)}{2} \quad (1.37)$$

O gráfico formado pela relação densidade espectral  $S(f)$  versus frequência  $f$  é usualmente chamado de espectro da série.

O citado trabalho de Andreo *et al.* (2006) é indicado para entender, em termos práticos, as diferenças entre as análises via FAC (domínio do tempo) e FDE (domínio da frequência). Os autores aplicam essas análises a longas séries de precipitação, temperatura e vazão médias, na intenção de caracterizar a variabilidade de cada uma. De grande importância no contexto da busca por tendências nas séries é a constatação de que a FDE permite identificá-las prontamente através de altas densidades próximas à frequência nula (origem do espectro da série). A superioridade da FDE foi confirmada inserindo-se uma tendência artificial decrescente nas séries de chuvas e analisando o correlograma e o espectro resultantes. Na comparação entre os gráficos antes e depois da tendência, apenas o espectro mostrou diferenças.

A influência de uma tendência em séries temporais sobre o espectro é também apresentada no estudo de Venema *et al.* (1997). O trabalho foi desenvolvido sobre os volumes afluentes do rio Senegal, no continente africano, identificando-se uma forte tendência decrescente após a década de 1960. Tal tendência foi removida da série, de modo que tornou possível o ajuste de um modelo estocástico linear para o estudo de diversos cenários visando o planejamento dos recursos hídricos da região.

Kite (1989) emprega a análise espectral e, em particular, a FDE para averiguar indícios de mudanças climáticas em séries de níveis médios de lagos e vazões médias de rios canadenses. Como premissa, o autor usou um modelo aditivo composto por uma parcela periódica, outra de tendência linear e uma terceira de componentes aleatórios para representar as séries hidrológicas. A análise se baseou em nove passos, partindo da seleção dos dados e passando pela remoção sucessiva dos elementos que compõem o modelo aditivo. A cada alteração na série uma análise via FDE foi empregada para identificar a porção explicada dos elementos na variância da série. Como conclusão, não foi possível identificar indícios estatisticamente significantes de mudanças climáticas nas séries estudadas.

A análise harmônica clássica, no formato apresentado até este momento, tem como premissa a estacionariedade das séries hidrológicas. Ao representar uma série pela combinação de senos e co-senos, assume-se que esta configuração é válida para todo o histórico, o que não é verdade se a série possui comportamento não estacionário. Mesmo com a possibilidade de identificar tendências através da FDE, esta informação é meramente qualitativa caso não seja analisada em conjunto a outras técnicas. Por exemplo, no estudo de Venema *et al.* (1997) citado anteriormente, só foi possível identificar que a tendência decrescente na série de precipitações se manifestou após a década de 1960 através do traçado de seu pluviograma paralelamente ao espectro.

Uma alternativa seria aplicar a análise harmônica utilizando janelas móveis sobre a série. Entretanto, esse método é contestado por ser dependente do tamanho da janela e pelo fato de que componentes de determinadas frequências (e.g. baixas frequências, representando eventos hidrológicos extremos) eventualmente não se encontram dentro de todas as janelas, gerando imprecisões no método (Torrence e Compo, 1998; Santos, Freire e Torrence, 2013).

A solução adotada por grande número de pesquisadores é a utilização da transformada *wavelet*<sup>4</sup>, que consiste na convolução da série temporal  $z_t$  com uma função  $\psi(t)$ , chamada de *wavelet*-mãe. Tal função nada mais é do que uma onda de duração finita e frequência específica,

---

<sup>4</sup> A tradução do termo *wavelets* a língua portuguesa é *ondaletas*. Entretanto, o termo original em inglês continuará a ser utilizado no restante do trabalho.

que é dilatada e transladada através da série permitindo sua caracterização simultânea no domínio tempo-frequência. A transformada *wavelet* contínua é representada pela equação (1.38) (Torrence e Compo, 1998):

$$W_{\tau}(s) = \sum_{\tau'=0}^{n-1} z_{\tau'} \psi^* \left[ \frac{(\tau' - \tau)\delta t}{s} \right] \quad (1.38)$$

onde  $\tau$  é o parâmetro de translação no tempo,  $s$  é o parâmetro de dilatação,  $\delta t$  é um intervalo constante de tempo que separa os elementos da série  $z_{\tau'}$ , e o asterisco representa o conjugado complexo da *wavelet*-mãe. Dentre as diversas *wavelets*-mães existentes, a grande maioria dos estudos aplicados a séries hidrológicas faz uso das funções de Morlet (Pasquini e Depetris, 2007), Haar (Saco e Kumar, 2000) e Paul (Rossi *et al.*, 2009).

De maneira semelhante ao correlograma e ao espectro da série, a forma típica de análise da transformada *wavelet* contínua é também através de gráficos. Por relevar três grandezas simultaneamente (tempo, frequência e força do sinal), em geral utilizam-se diagramas com as frequências nas ordenadas, o tempo nas abscissas e escalas de cores para representar a intensidade do sinal. Alguns autores (e.g. Smith, Turcotte e Isacks, 1998; Adamowski, 2008) se referem a esse gráfico como escalograma.

Uma das aplicações pioneiras de *wavelets* a séries hidrológicas ocorreu no trabalho de Smith, Turcotte e Isacks (1998). Entretanto, ao invés da busca por tendências, os autores utilizaram a técnica para a caracterização de periodicidades e variabilidades em séries de vazões médias diárias. Os trabalhos de Anctil e Coulibaly (2004) e Liang, Li e Liu (2011) também foram dedicados ao estudo da variabilidade de séries hidrológicas, o primeiro sobre vazões médias anuais na região de Quebec, Canadá, e o segundo sobre precipitações mensais e anuais no nordeste da China.

Andreo *et al.* (2006), além das análises via FAC e FDE mencionadas anteriormente, aplicaram a transformada *wavelet* a séries de precipitação mensal com comprimentos superiores a 100 anos, evidenciando o diferencial dessa técnica na caracterização da variabilidade das séries. Através dos gráficos obtidos, os autores conseguiram perceber uma alteração entre componentes periódicos plurianuais antes e depois da década de 1950: nos anos anteriores, tais componentes tinham uma frequência de dois a três anos e passaram a ter frequências de quatro a seis anos após essa data.

Embora a análise via transformada *wavelet* sobre uma determinada série torne evidente a alteração de padrões de variabilidade, o estudo de tendências requer interpretações mais

elaboradas. Isso porque nem sempre uma mudança nas componentes periódicas das séries pode ser ligada diretamente a tendências de aumento ou redução de suas magnitudes. Nas séries de precipitação estudadas por Andreo *et al.* (2006), por exemplo, apesar de ser identificada a alteração de periodicidades a partir da década de 1950, não fica claro se isso significou aumento, redução ou nenhuma diferença significativa nos volumes precipitados. Poucos são os estudos que fazem essa distinção, podendo-se citar os trabalhos de Pasquini e Depetris (2007) e Rossi *et al.* (2009), nos quais os autores atrelam as tendências a mudanças na intensidade do sinal no gráfico da transformada. Em outras palavras, as cores utilizadas para a representação da intensidade tornam-se mais vivas para tendências positivas ou mais opacas para tendências negativas.

Existe ainda outra forma de se trabalhar com as *wavelets*, aplicando-as na decomposição das séries em diferentes bandas de frequência. Partal e Kuçuk (2006) decompõem séries de chuvas anuais em cinco modais, referentes aos componentes periódicos com frequências de dois, quatro, oito, dezesseis e trinta e dois anos. Através da análise combinada entre esses diversos componentes e utilizando o teste de Mann-Kendall como métrica, os autores foram capazes de apontar o modal com frequência de 16 anos como o responsável pela tendência existente na série.

Craigmile, Guttorp e Percival (2004), por sua vez, apresentam maneiras de decompor a série na intenção de estimar as tendências existentes. Apesar de ter forte cunho teórico, esse trabalho é interessante por permitir a estimativa de tendências polinomiais de diversas ordens, diferenciando-se da tradicional tendência linear considerada na grande maioria dos estudos. Os autores apresentam um estudo de caso para uma série de temperatura da superfície do oceano, decompondo-a em seis modais e ajustando uma tendência polinomial de ordem quatro.

### **1.3 Uma visão alternativa sobre a questão do comportamento estacionário**

Em meio a um comportamento não estacionário de uma série hidrológica, a grande maioria dos pesquisadores procura alternativas para remover essa característica da série e, conseqüentemente, torná-la apropriada para a aplicação em modelos. Essa prática de trabalhar no ajuste dos dados para submissão aos modelos é criticada por alguns autores, como Koutsoyiannis (2006), que defende a adaptação dos modelos para receber os dados em sua forma original.

A questão levantada por Koutsoyiannis (2006) vem de inconsistências conceituais que se iniciam na própria atribuição do termo “(não) estacionariedade” a séries temporais. Como

discutido inicialmente na seção 1.1.3, este autor enfatiza que o termo é aplicável somente a modelos ou a processos estocásticos, de forma que seu uso a uma série temporal não tem sentido. Uma exceção ocorreria no caso da série temporal ser uma realização de um processo estocástico, estacionário ou não estacionário, mas com mecanismo gerador conhecido. Claramente esse não é o caso para nenhuma série temporal de fenômenos físicos.

Klemes (1974) também defende esse ponto de vista, afirmando que as discussões acerca da classificação de séries temporais em estacionárias ou não estacionárias são “um exercício de futilidade”. Complementa criticando o argumento de que quanto mais longa a série, maior a probabilidade de ela apresentar comportamento não estacionário, alegando que em uma escala ainda maior de tempo esse comportamento poderia representar variações amostrais de uma série estacionária. Em suma, não é possível tirar tais conclusões acerca do equilíbrio estatístico de uma série, pois ela é a única informação disponível para as análises.

Outra inconsistência apontada principalmente nos trabalhos de Koutsoyiannis (2003a, 2003b, 2006, 2011, 2013) é a caracterização da tendência como um componente determinístico da série, expresso através de uma função do tipo  $z_t = f(t)$ , onde  $f(t)$  é, quase sempre, uma função linear do tempo  $t$ . Como visto na seção 1.2.1 e seus subitens, esse é um procedimento usual adotado na literatura; a própria formulação das hipóteses avaliadas em alguns testes estatísticos (e.g. Mann-Kendall) assume tendências lineares de crescimento ou redução da variável em questão. Porém, Koutsoyiannis critica essa abordagem, primeiro por sua falta de significado físico e segundo pelo paradoxo criado ao modelar estocasticamente uma variável aleatória composta por um elemento determinístico. Para apoiar seu argumento, o autor plota séries naturais de grande duração (dentre elas, níveis mínimos anuais do rio Nilo com 663 observações, série apresentada em maiores detalhes na seção 3.4, na página 127) juntamente com médias móveis de cinco e 25 anos de duração. Ao analisar o gráfico, fica evidente que o comportamento das médias móveis possui certo grau de aleatoriedade, ora representando tendências ascendentes e ora descendentes (uma reprodução dessa análise é exibida na seção 1.4.1.1). Dado este panorama, Koutsoyiannis refuta o emprego de uma tendência determinística, preferindo associar o comportamento a flutuações de larga escala.

Em trabalho recente, porém aplicado à análise de frequência de cheias, Serinaldi e Kilsby (2015) defendem o mesmo argumento ao comentar que o uso de modelos não estacionários só tem justificativa quando a evolução do processo pode ser descrita em termos determinísticos. Entretanto, os autores alertam que essa solução é teoricamente inconsistente (devido às mesmas razões levantadas por Klemes, 1974) e bastante complexa, por entenderem

que as séries históricas observadas são produtos da interação de múltiplos mecanismos e variáveis. Ademais, eles relembram que a condição de ergodicidade é violada quando se trabalha com processos não estacionários, prejudicando toda e qualquer inferência feita a partir da série.

Retomando a questão da variabilidade dos valores médios de uma série, é possível explicá-los através da aglomeração de valores elevados ou baixos durante certo período de tempo, consequência da persistência observada em séries hidrológicas. Mandelbrot e Wallis (1968) nomearam esse fenômeno como “Efeito Joseph” e trouxeram à tona sua importância frente aos objetivos dos modelos hidrológicos, principalmente os destinados ao dimensionamento de reservatórios. Juntamente com outro fenômeno intitulado “Efeito Noah”, que se refere a eventos extremos incomuns na série, os autores elaboraram uma classe de modelos (Mandelbrot e Wallis, 1969a; 1969b; 1969c) visando sua reprodução, alegando que os clássicos modelos Markovianos não conseguem contabilizá-los e acabam por subestimar as flutuações hidrológicas. A base teórica da modelagem proposta por esses autores está no fenômeno de Hurst, tema da próxima seção.

### 1.3.1 O fenômeno de Hurst

O fenômeno de Hurst surgiu a partir dos trabalhos empíricos realizados pelo engenheiro inglês Harold Edwin Hurst na década de 1950, documentados em Hurst (1951). Esse engenheiro estava interessado no estudo de níveis e vazões do rio Nilo, com a intenção de dimensionar a capacidade do reservatório da Represa Alta de Assuã, no Egito. Para tanto, ele empregou o método de Rippl, baseando suas considerações em amplitudes acumuladas.

Previamente à apresentação do fenômeno de Hurst é necessário introduzir conceitos relacionados a somas parciais e amplitudes. Seja uma série temporal  $z_t$ , da qual se definem as chamadas soma parcial  $S_t$ , soma parcial máxima  $M_n$ , soma parcial mínima  $m_n$  e a amplitude  $R_n$  (Gomide, 1978):

$$S_t = z_1 + z_2 + \dots + z_t, \quad t = 1, 2, \dots, n \quad (1.39)$$

$$M_n = \max(0, S_1, S_2, \dots, S_n) \quad (1.39a)$$

$$m_n = \min(0, S_1, S_2, \dots, S_n) \quad (1.39b)$$

$$R_n = M_n - m_n \quad (1.39c)$$

A soma parcial definida na equação (1.39) pode ser ajustada de acordo com a média amostral da série  $\hat{\mu}_z$ , resultando nas equações (1.40):

$$S_t^* = S_t - t\hat{\mu}_z, \quad t = 1, 2, \dots, n \quad (1.40)$$

$$M_n^* = \max(0, S_1^*, S_2^*, \dots, S_n^*) \quad (1.40a)$$

$$m_n^* = \min(0, S_1^*, S_2^*, \dots, S_n^*) \quad (1.40b)$$

$$R_n^* = M_n^* - m_n^* \quad (1.40c)$$

onde os asteriscos indicam que as variáveis estão ajustadas. A razão entre a amplitude ajustada e o desvio padrão amostral da série  $\hat{\sigma}_z$  define a chamada amplitude re-escalada (do Inglês *rescaled range*), conforme mostra a equação (1.41):

$$R_n^{**} = \frac{R_n^*}{\hat{\sigma}_z} \quad (1.41)$$

A amplitude re-escalada é proporcional a uma quantidade  $n^H$ , onde  $H$  é uma constante que varia entre 0,5 e 1. Em teoria, para uma amostra de tamanho grande (porém finito) com elementos independentes e identicamente distribuídos, o valor de  $H$  converge assintoticamente para 0,5 (Mandelbrot e Wallis, 1968; Klemes, 1974). Entretanto, as observações de Hurst, inicialmente sobre os níveis mínimos do rio Nilo e posteriormente em outros fenômenos naturais, apontaram para um valor de  $H \sim 0,7$ . Posteriormente, Mandelbrot e Wallis (1969b) encontraram um valor de expoente ainda maior para o rio Nilo, chegando a  $H = 0,85$ . Essa diferença entre o valor teórico e os obtidos empiricamente para  $H$  ficou conhecida como fenômeno de Hurst. Da mesma forma, atribuiu-se à constante  $H$  o nome do engenheiro, passando a ser chamada de expoente de Hurst.

Desde sua descoberta, o fenômeno do Hurst atraiu a atenção de diversos hidrólogos interessados em suas possíveis causas. Mandelbrot e Wallis (1968) trabalharam com a hipótese de que as variáveis hidrológicas possuem um elevado grau de interdependência, aproximando seu comportamento de um processo com longa memória. Utilizando um exemplo hipotético da vazão média anual medida em um rio, a longa memória pode ser explicada fisicamente

considerando que essa vazão recebeu influência de vazões de décadas, ou até mesmo séculos atrás. Da mesma forma, a vazão observada em uma data atual exercerá influência sobre o regime do rio durante muito tempo no futuro.

Klemes (1974), Potter (1976) e Boes e Salas (1978), por sua vez, alegam que as flutuações do valor médio das séries (“não estacionariedade” na média) podem explicar o fenômeno de Hurst. Klemes (1974) confirmou essa possibilidade através de alguns experimentos com números aleatórios normalmente distribuídos, sobre os quais se impôs uma alternância nas tendências centrais. Boes e Salas (1978) também demonstram que tais flutuações podem ser explicadas utilizando a estrutura de um processo autorregressivo com médias móveis de primeira ordem (ARMA[1,1]).

Gomide (1978) argumenta que o fenômeno de Hurst é um efeito transiente observado nas séries, causado possivelmente por perturbações nas medições das variáveis. Em concordância com essa explicação, Potter (1979) mostra que a não homogeneidade encontrada em séries hidrológicas pode ser indutora do fenômeno de Hurst.

As explicações mencionadas nos parágrafos anteriores aparentemente amenizaram a busca pela causa do fenômeno de Hurst, pois dificilmente se podem encontrar trabalhos específicos após o início da década de 1980. A exceção é o trabalho de Poveda e Mesa (1991), no qual os autores alegam que o fenômeno é inexistente, sendo produto de um problema historicamente mal formulado. Para apoiar essa hipótese, trabalha-se com o chamado “teorema funcional do limite central sobre a distribuição assintótica da amplitude re-escalada em processos de memória curta”. Em outras palavras, ao invés de buscar preservar o fenômeno de Hurst através de modelos estocásticos específicos, os autores trabalham com processos de memória curta [ARMA( $p,q$ )] para preservar a amplitude re-escalada. Os experimentos elaborados sobre séries geofísicas resultaram em expoentes  $H \sim 0,5$  para todas elas, o que levou os autores a concluir que o fenômeno de Hurst não existe.

Koutsoyiannis (2002) levanta novamente a questão e propõe uma explicação baseada no efeito combinado de flutuações aleatórias em séries climáticas que ocorrem em diversas escalas temporais. Desse modo, séries de níveis e vazões nos rios são consequência de tal variabilidade. Segundo o autor, esse comportamento não só resulta no fenômeno de Hurst como causa a interdependência que levou Mandelbrot e Wallis (1968) a considerar a presença de longa memória entre as observações. É interessante observar que a combinação das flutuações climáticas em múltiplas escalas temporais foi, de certo modo, notada nos trabalhos de Mandelbrot e Wallis (1968) e Potter (1976):



*“Whereas precipitation fluctuations of wave-length near one day or one year may participate in several physical mechanisms, those of wavelength near one lifetime<sup>5</sup> are likely to participate in one mechanism only”*

(Mandelbrot e Wallis, 1968)

*“(...) the nonlinearity of the atmosphere-ocean-earth system may result in a number of distinct regimes such that the general circulation tends to persist in a given regime until by chance it acquires just such a state as to allow it to pass into another regime.”*

(Potter, 1976)

Em nenhum dos trabalhos, todavia, esse comportamento foi associado diretamente ao fenômeno de Hurst.

Depois de todos esses estudos acerca das descobertas de Hurst, diversos pesquisadores (com exceção de Poveda e Mesa, 1991) concordaram que o fenômeno de Hurst não é característica de uma série específica (como as vazões do rio Nilo), mas vigora na grande maioria das séries de variáveis naturais (Klemes, 1974; Koutsoyiannis, 2002). Sob esse ponto de vista parece bastante lógico que o caminho a ser seguido em termos de modelagem passa pela escolha de formulações que sejam capazes de reproduzir o expoente de Hurst, em detrimento da utilização de técnicas para a remoção de eventuais tendências nas séries históricas. Essa é a hipótese de trabalho desta tese.

#### **1.4 Modelos estocásticos aplicados a séries de vazões médias**

Formulações estocásticas utilizadas na modelagem de séries hidrológicas de vazões médias são o tema desta seção. De acordo com os objetivos da tese, maior detalhamento é oferecido a três modelos específicos que, dentre outros propósitos, podem ser utilizados para a geração de séries sintéticas de vazão. A primeira abordagem a ser apresentada é o modelo de múltiplas escalas temporais, tema principal do trabalho e que se enquadra em uma classe de formulações inspiradas no fenômeno de Hurst. Depois de mostradas as premissas dessa técnica, descreve-se o algoritmo computacional utilizado para a geração dos cenários nas escalas anual e mensal, sendo que a última constitui a proposta de inovação da tese. Da mesma forma, explica-se o algoritmo utilizado para a estimação do expoente de Hurst, parâmetro de primordial importância para a abordagem aqui selecionada.

---

<sup>5</sup> Período de tempo com duração entre 50 e 100 anos, conforme definição dada pelos autores.

Na sequência são descritos os modelos alternativos, que servem de base de comparação para a formulação de múltiplas escalas temporais. Trabalha-se com abordagens tradicionais Markovianas de primeira ordem e de Box & Jenkins para as escalas anual e mensal, respectivamente. Em todas as opções são consideradas as formulações multivariadas, a fim de gerar séries sintéticas em diversos locais simultaneamente. Para o modelo Markoviano, em particular, é apresentada também uma técnica difundida na literatura para remover o comportamento não estacionário das séries históricas, a fim de adequá-las às exigências da formulação.

A última parte da seção contempla um sucinto levantamento do estado-da-arte da modelagem de séries de vazões médias, no qual são destacados os principais trabalhos publicados sobre o assunto nos últimos cinco anos.

#### 1.4.1 Modelos inspirados no fenômeno de Hurst

A possível relação entre o fenômeno de Hurst e os efeitos Joseph e Noah (ver página 66) norteou Mandelbrot e Wallis (1968) na elaboração de um modelo conhecido como Ruídos Gaussianos Fractais (do inglês *Fractional Gaussian Noise* – FGN), cuja formulação é diferente dos tradicionais modelos Markovianos amplamente utilizados em hidrologia. Como mencionado anteriormente, os autores argumentam que estes modelos Markovianos não são capazes de reproduzir nem o fenômeno de Hurst, nem tampouco os dois efeitos que caracterizam comportamentos importantes observados em séries hidrológicas. Formulações Markovianas priorizam as variações de alta frequência (escalas inferiores à anual), ou seja, as que ocorrem rotineiramente. A proposta dos autores foi elaborar um modelo que invertesse essa abordagem, priorizando flutuações de baixa frequência (escala de 50 a 100 anos), como os efeitos Joseph e Noah.

Para tanto, Mandelbrot e Wallis recorreram ao conceito de longa memória, que se verifica na modelagem FGN através do lento decaimento de sua função de autocorrelação. Koutsoyiannis (2002) exemplifica que, para a série do rio Nilo ( $H = 0,85$ ), a função de autocorrelação retorna valores da ordem de 0,15 para uma defasagem de 100 anos. Apesar de pequena, considera-se que essa autocorrelação contribui para todo o processo. A título de comparação, modelos de memória curta (Markovianos) apresentam autocorrelação praticamente nula para defasagens menores que 10 anos. Uma completa descrição do modelo FGN pode ser consultada em Mandelbrot e Wallis (1969a, 1969b, 1969c).

Contudo, Klemes (1974) criticou a ideia de que séries efetivamente apresentem longa memória, alegando a falta de um significado físico. Segundo este autor, na tentativa de reproduzir o expoente de Hurst, Mandelbrot e Wallis chegaram a uma sofisticação matemática que dificilmente poderia tornar o modelo operacional. Dessa maneira, para se chegar à formulação FGN foram adotadas premissas que distanciaram o modelo da realidade física observada em séries hidrológicas.

Em concordância com as observações de Klemes, Koutsoyiannis (2002) apresentou outro modelo capaz de reproduzir o expoente de Hurst. Porém, é conveniente notar que entre os estudos de Mandelbrot e Wallis (1968) e Koutsoyiannis (2002) há uma diferença superior a 30 anos. Nesse intervalo entrou em destaque toda a questão das tendências em séries hidrológicas, conforme extensivamente discutido na seção 1.2.1. Isso levou Koutsoyiannis (2002) a ressaltar que sua abordagem considera essas flutuações como comportamento esperado de séries naturais, conectando-as com o efeito Joseph debatido por Mandelbrot e Wallis (1968). No entanto, o autor refuta a presença de longa memória nas séries, preferindo atrelar seu comportamento peculiar à influência de flutuações climáticas que ocorrem em múltiplas escalas temporais. Por esse motivo ele denominou o processo que representa o comportamento das séries naturais de Processo Estocástico Escalonado Simples (do inglês *Simple Scaling Stochastic Process* – SSS).

Ainda de acordo com Koutsoyiannis, a importância na reprodução do expoente de Hurst está no fato de que este funciona como uma evidência da influência dos diversos níveis de flutuações observadas em séries naturais. Detalhamento matemático e estatístico da abordagem SSS e do modelo de múltiplas escalas temporais pode ser encontrado em Koutsoyiannis (2002, 2003a, 2003b) e também será tema da seção 1.4.1.1. Por fim, em Koutsoyiannis (2006) é oferecida uma interessante comparação conceitual entre o enfoque tradicional que trabalha com o comportamento não estacionário das séries e as premissas do processo SSS.

De todas as características dos modelos FGN e SSS expostas, talvez a mais interessante seja a de que ambas são formulações estacionárias que se propõem a modelar séries com comportamento não estacionário. É nesse ponto que se observa a maior diferença entre essas abordagens e as práticas tradicionais em hidrologia estocástica. Apesar de discordar no quesito longa memória, os modelos FGN e SSS são importantes por priorizar efeitos hidrológicos relevantes. Considerando, por exemplo, o dimensionamento de um reservatório, a geração de cenários sintéticos de vazões médias que preservem a estrutura de persistência e os fenômenos de baixa frequência pode ser de grande valia. No caso de um reservatório existente, essa possibilidade oferece subsídios para incrementar o planejamento de sua operação. A

investigação dessas questões faz parte do escopo desta tese e, por esse motivo, um maior detalhamento dos modelos será oferecido nos capítulos subsequentes. A prioridade será dada à abordagem SSS, por ter claro significado físico e também por envolver uma linha de raciocínio ainda pouco explorada na literatura específica.

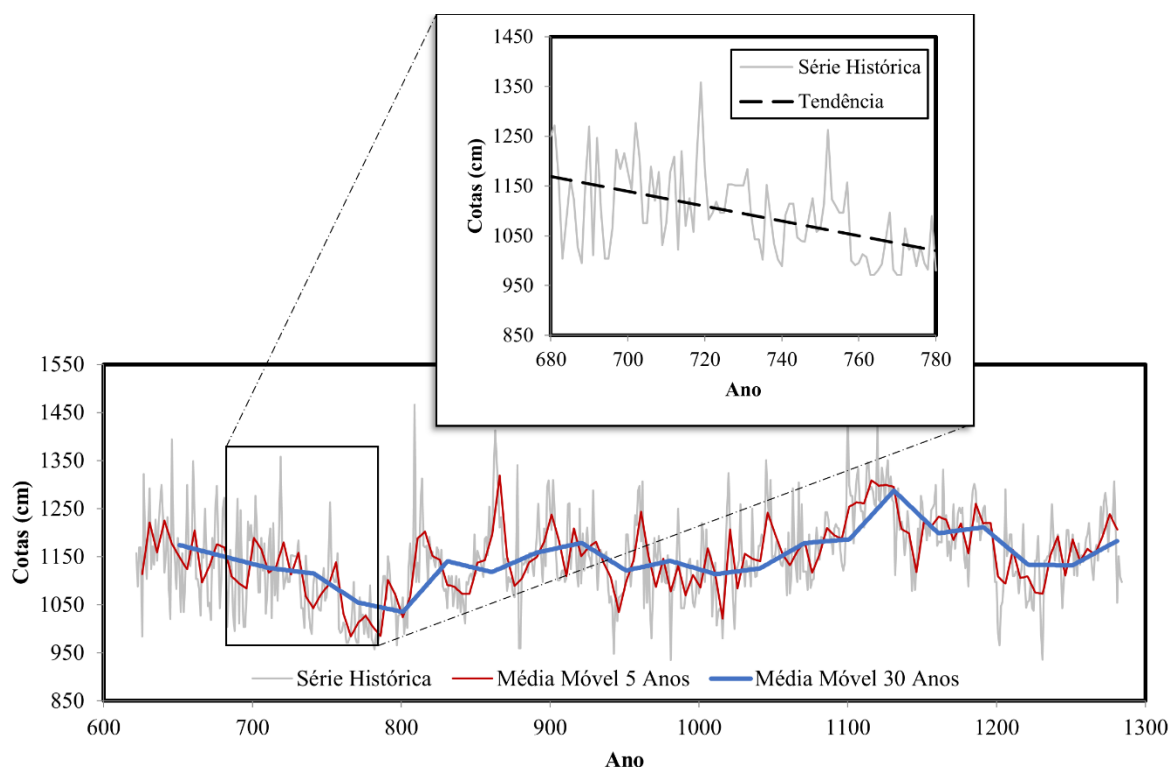
#### 1.4.1.1 Modelo de múltiplas escalas temporais

A ideia inicial do modelo de múltiplas escalas temporais foi introduzida por Demetris Koutsoyiannis há pouco mais de uma década (Koutsoyiannis, 2002, 2003a e 2003b), muito embora um tratamento matemático geral sobre essa abordagem tenha sido publicada antes disso em Koutsoyiannis (2000). Como o próprio nome sugere, a filosofia por trás dessa abordagem se baseia na percepção das escalas de tempo envolvidas nos fenômenos naturais e sua relação com o comportamento observado nas séries hidrológicas. A fim de evidenciar essa questão o autor trabalhou com a anteriormente mencionada série anual de níveis mínimos do rio Nilo medida no nilômetro<sup>6</sup> da ilha de Rhoda, em Cairo. Tal série é a mesma utilizada por Hurst (1951) em seus estudos sobre a amplitude re-escalada (ver subitem 1.3.1, página 66), também apresentada em maiores detalhes na seção 3.4, página 127. A Figura 4 foi reproduzida do trabalho de Koutsoyiannis (2006) e é uma forma útil para o entendimento da filosofia de modelagem.

O gráfico em destaque na parte superior da Figura 4 reúne 100 anos de observações, sobre as quais o ajuste linear sugere uma tendência negativa na série. O gráfico na parte inferior mostra todos os 663 anos (de 622 a 1284) de registros, destacando a posição dos 100 anos exibidos no gráfico superior. Fica claro que a tendência, antes decrescente, passa a variar em um padrão aleatório no restante da série. Segundo Koutsoyiannis (2006), a manifestação dessas flutuações, representadas pelas escalas anuais e de médias móveis de cinco e 30 anos, evidencia a presença de mecanismos que atuam em diversas escalas temporais.

---

<sup>6</sup> “Nilômetro” é o nome dado a uma estrutura específica para a medição dos níveis de água do rio Nilo, no Egito. Diversos nilômetros foram construídos desde épocas faraônicas, três deles existentes até o dia de hoje (embora inoperantes). Dentre eles o nilômetro de Rhoda é o mais famoso, estrutura que forneceu a série histórica de excepcionais 663 anos utilizada pelos autores citados nesta seção. Os registros estão disponíveis em <http://lib.stat.cmu.edu/S/beran>. Maiores detalhes sobre os nilômetros e as estatísticas descritivas da série do Nilômetro de Rhoda são fornecidos na seção 3.4, página 103.



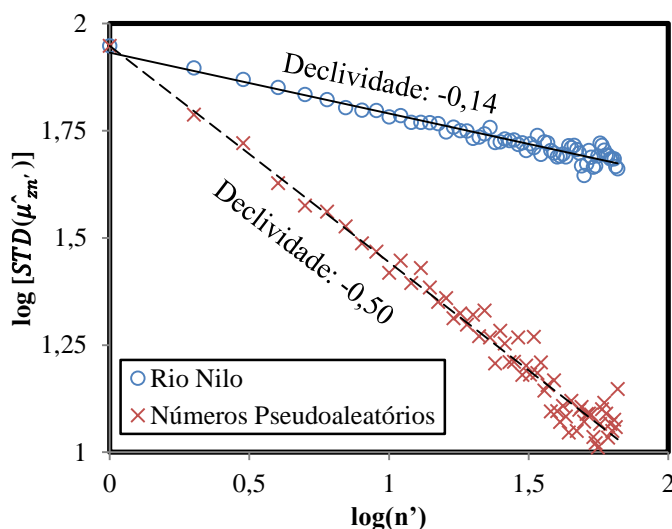
**Figura 4 – Série de níveis mínimos anuais do rio Nilo em Rhoda, com destaque ao período entre 680 a 780 [Fonte: adaptado de Koutsoyiannis (2006), com dados coletados em <http://lib.stat.cmu.edu/S/beran>]**

Aprofundando-se na caracterização das múltiplas escalas temporais, Koutsoyiannis (2006) prossegue sua análise baseando-se no cálculo de desvios padrão, estatísticas representativas do grau de variabilidade (ou mudança) de uma série temporal. Para isso, é utilizada a equação (1.42) que define o desvio padrão da distribuição amostral de médias, na estatística clássica:

$$STD(\hat{\mu}_{z_{n'}}) = \frac{\hat{\sigma}_z}{\sqrt{n'}} \tag{1.42}$$

onde  $STD[ ]$  é o operador desvio padrão e  $\hat{\mu}_{z_{n'}}$  é a média amostral estimada para a série  $z_t$  de tamanho  $n'$ , ou seja  $\hat{\mu}_{z_{n'}} = (z_1 + \dots + z_{n'})/n'$ . Considerando novamente a série de  $n$  níveis mínimos anuais do rio Nilo, o autor propõe o cálculo da equação (1.42) para diversos valores  $n'$ . Para  $n' = 1$ , chega-se à  $STD(\hat{\mu}_{z_{n'}}) = STD(z_t)$ , que é o estimador de  $\hat{\sigma}_z$  calculado utilizando todos os  $n$  elementos disponíveis na amostra. Para  $n' = 2$  (considerando a série com número par de elementos), pode-se obter  $n/2$  amostras  $\{\hat{\mu}_2^{(1)} = (z_1 + z_2)/2, \hat{\mu}_2^{(2)} = (z_3 + z_4)/2, \dots, \hat{\mu}_2^{(n/2)} = (z_{n-1} + z_n)/2\}$ , resultando na estimativa de  $STD(\hat{\mu}_2)$ . Koutsoyiannis (2006) prossegue o cálculo até que sejam obtidas 10 amostras  $STD(\hat{\mu}_{10})$  e exhibe os resultados em um

gráfico bilogárptico  $STD(\hat{\mu}_{zn'})$  versus  $n'$ , denominado climacograma. A título de comparação, o mesmo procedimento é feito para uma série de números pseudoaleatórios com estatísticas idênticas à série do rio Nilo em questão ( $\hat{\mu}_z = 1148$  mm e  $\hat{\sigma}_z = 89$  mm; para demais estatísticas e outras características ver seção 3.4 na página 127). O climacograma resultante é reproduzido na Figura 5.



**Figura 5 – Climacograma das séries de níveis mínimos anuais do rio Nilo e de números pseudoaleatórios com as mesmas estatísticas (Fonte: adaptado de Koutsoyiannis, 2006)**

A diferença entre os desvios padrão estimados pela equação (1.42) e os valores empíricos calculados é notável. A partir dessa evidência, Koutsoyiannis (2006) sugere uma generalização na lei clássica, propondo a equação (1.43):

$$STD(\hat{\mu}_{zn'}) = \frac{\hat{\sigma}_z}{n^{1-H}} \tag{1.43}$$

onde  $n$  é o tamanho da amostra e  $H$  é um expoente que varia entre 0,5 e 1. Valores fora desse intervalo acarretam uma improvável antipersistência ( $H < 0,5$ ) e a situação absurda de correlações com valores superiores à unidade ( $H > 1$ ) (Koutsoyiannis, 2013). No caso da série de níveis do rio Nilo  $H \sim 0,85$ , portanto a aplicação da equação (1.42) (e.g.  $H = 0,5$ ) fornece uma estimativa tendenciosa do desvio padrão e, conseqüentemente, de outras estatísticas dessa série.

O autor explica o comportamento estatístico mostrado na Figura 5 e na equação (1.43) como a manifestação do fenômeno de Hurst, discutido no subitem 1.3.1 (página 66). De fato, o expoente  $H$  é o próprio expoente de Hurst. Em trabalho mais recente, Koutsoyiannis (2013) denomina os fenômenos que seguem a equação (1.43) como processos de Hurst-Kolmogorov (HK), em referência ao próprio Hurst e ao matemático russo Andrei Nikolaevich Kolmogorov

que propôs um modelo semelhante a (1.43) para estudar a turbulência. Ainda de acordo com Koutsoyiannis, valores elevados de  $H$  sugerem mudanças de grande escala nas séries.

Em termos de modelagem, o autor trabalha com conceitos e equações adaptadas ao comportamento HK, focando nos chamados processos estocásticos agregados, cujas propriedades estatísticas são mostradas na próxima seção.

#### 1.4.1.1.1 Propriedades dos processos estocásticos agregados

A definição dos processos estocásticos agregados pode ser vista como uma generalização dos conceitos anteriormente apresentados na seção 1.1.2 (página 36). Parte-se do processo estocástico simples discreto estacionário  $Z_t$  ( $t = 1, 2, \dots, N$ ), com média  $\mu = E[Z_t]$  e covariância  $\gamma(t_1, t_2) = COV[Z_{t_1}, Z_{t_2}]$ . Diversas séries hidrológicas são modeladas utilizando esses processos, podendo-se citar o caso de séries de vazões médias anuais em um rio qualquer (ver seção 1.4.2).

Em muitos casos, todavia, a interação entre as realizações de um processo estocástico em diversas escalas temporais pode ser de grande valia. Utilizando novamente o caso das séries de vazões médias, a consideração conjunta de escalas anuais e maiores que um ano pode ser importante do ponto de vista da operação de um reservatório com capacidade de regularização plurianual. Nos processos estocásticos agregados isso é possível, tomando a escala anual como básica e as demais escalas como múltiplas inteiras da escala básica, tal como feito na montagem do climacograma da Figura 5. Dessa maneira, seja  $\eta t$  uma escala temporal maior do que  $t = 1$  (ano), na qual  $\eta$  representa um número inteiro positivo qualquer. Define-se uma realização de um processo estocástico agregado através da equação (1.44) (Koutsoyiannis, 2002):

$$Y_t^{(\eta)} = \sum_{l=(t-1)\eta+1}^{\eta t} Z_l \quad (1.44)$$

As realizações dos processos estocásticos agregados são definidas pela equação (1.45):

$$y_t^{(\eta)} = \sum_{l=(t-1)\eta+1}^{\eta t} z_l \quad (1.45)$$

Naturalmente, quando  $\eta = 1$  as realizações do processo estocástico agregado recaem nas de um processo estocástico simples ( $y_t^{(1)} = z_t$ ). Quando  $\eta = 2$ , tem-se  $\{y_1^{(2)} = z_1 + z_2, y_2^{(2)} = z_3 + z_4, \dots\}$  e assim por diante para qualquer valor de  $\eta$ . Da mesma forma, as propriedades

estatísticas básicas dos processos estocásticos agregados são uma extensão das definições dos processos estocásticos simples. Assim, a média, a autocovariância e a autocorrelação das realizações são dadas pelas equações (1.46), (1.47) e (1.48), respectivamente:

$$E[y_t^{(\eta)}] = \eta\mu \quad (1.46)$$

$$\gamma_k^{(\eta)} = COV[y_t^{(\eta)}, y_{t+k}^{(\eta)}] = \sum_{l=1}^{\eta} \sum_{m=k\eta+1}^{(k+1)\eta} \gamma_{m-l}, \quad j = 0, \pm 1, \pm 2, \dots \quad (1.47)$$

$$\rho_k^{(\eta)} = \frac{\gamma_k^{(\eta)}}{\gamma_0^{(\eta)}} \quad (1.48)$$

Diferenças nas abordagens de processos estocásticos simples e agregados podem ser avaliadas na prática através de um modelo AR(1) do tipo  $z_t = \phi z_{t-1} + a_t$ , sendo  $z_t$  a série temporal considerada,  $\phi$  o parâmetro do modelo e  $a_t$  a série de resíduos. Neste processo, a autocorrelação é definida por  $\rho_k = \phi^k$  (Box, Jenkins e Reinsel, 2008, p. 60). Estendendo essa autocorrelação para o caso agregado, Koutsoyiannis (2002) mostra que a equação resultante é mais complexa:

$$\rho_1^{(\eta)} = \frac{\phi(1 - \phi^{(\eta)})^2}{\eta(1 - \phi^2) - 2\phi(1 - \phi^{(\eta)})}, \rho_k^{(\eta)} = \rho_1^{(\eta)} \phi^{\eta(k-1)}, \quad j \geq 1 \quad (1.49)$$

Dessa maneira, um processo AR(1) só corresponde realmente a um AR(1) na escala básica  $t$ , se tornando outro processo, muito mais complexo, quando se consideram escalas temporais agregadas. Fisicamente, essa constatação pode ser entendida como uma limitação teórica do modelo AR(1) em representar processos que ocorram em outras escalas temporais diferentes da básica. Em outras palavras, se séries de vazões médias anuais forem modeladas com um modelo AR(1), ocorrências de menor frequência do que um ano (reflexos do fenômeno El Niño, por exemplo) poderão não ser bem captadas.

#### 1.4.1.1.2 Processo estocástico escalonado simples

Na intenção de superar as limitações de modelagem comentadas brevemente no item anterior, Koutsoyiannis (2003) propõe a utilização da relação (1.50):



$$(y_t^\eta - \eta\mu) \equiv \left(\frac{\eta}{\eta'}\right)^H (y_{t'}^{\eta'} - \eta'\mu) \quad (1.50)$$

onde o símbolo  $\equiv$  denota igualdade em termos de distribuição conjunta finito dimensional,  $H$  é o expoente de Hurst e  $\eta$  e  $\eta'$  são inteiros multiplicadores das escalas temporais  $t$  e  $t'$ , respectivamente. Essa equação é válida para qualquer valor inteiro de  $t$  e  $t'$  e para qualquer escala temporal  $\eta$  e  $\eta'$  maiores do que um.

O autor observa que a equação (1.50) nada mais é do que a definição do modelo Ruído Gaussiano Fractal (FGN) proposto por Mandelbrot e Wallis (1969a, 1969b, 1969c). No entanto, a denominação FGN foi inicialmente empregada para variáveis normalmente distribuídas (Ruído *Gaussiano* Fractal), o que raramente se observa em hidrologia. Além disso, os dois outros termos também remetem a conceitos não comuns à prática hidrológica. Por esse motivo, Koutsoyiannis (2003) modificou a denominação FGN para o anteriormente mencionado Processo Estocástico Escalonado Simples (SSS).

Fazendo  $t = t' = \eta' = 1$  na equação (1.50), chega-se à definição da variância do processo agregado, mostrada na equação (1.51):

$$\gamma_0^{(\eta')} = VAR[y_t^{(\eta')}] = \eta^{2H}\gamma_0 \quad (1.51)$$

Para escalas agregadas de ordens superiores  $t = t' = \eta' > 0$ , a função de autocovariância independe da escala e pode ser determinada de acordo com a equação (1.52):

$$\gamma_t^{(\eta')} = \gamma_t = \frac{1}{2}[(t+1)^{2H} + (t-1)^{2H}] - t^{2H}, \quad t > 0 \quad (1.52)$$

Assim como na modelagem clássica, os momentos estatísticos das realizações geradas por um processo SSS podem ser estimados a partir da amostra  $y_t$  ( $t = 1, \dots, n$ ), bastando que se usem as equações (1.53) e (1.54):

$$\hat{\mu}_y = \frac{1}{n} \sum_{i=1}^n y_i \quad (1.53)$$

$$\hat{\sigma}_y^2 = \frac{1}{n - n^{2H-1}} \sum_{i=1}^n (y_i - \hat{\mu}_y)^2 = \frac{n - 1}{n - n^{2H-1}} \hat{\sigma}_z^2 \quad (1.54)$$

Dessa maneira, a média amostral  $\hat{\mu}_y$  é estimada usando a mesma equação dos processos tradicionais. O mesmo não pode ser dito para a variância amostral  $\hat{\sigma}_y^2$ , que necessita de uma correção a fim de se tornar consistente com a abordagem SSS. Koutsoyiannis (2003) mostra um algoritmo eficiente para a estimação conjunta da variância amostral e do expoente de Hurst, baseado em um procedimento iterativo. Tal algoritmo será tema da seção 1.4.1.1.6.

Mesmo com as evidências que indicam o comportamento HK em diversas séries de fenômenos naturais, os modelos SSS (ou FGN) não são utilizados com frequência para a geração de séries sintéticas de variáveis hidrológicas. Koutsoyiannis (2003) alega que isso se deve parcialmente à complexidade de conceitos (e.g. amplitudes re-escaladas) e de algoritmos existentes para tal tarefa. Por esse motivo, ele propõe três algoritmos simplificados para gerar sequências sintéticas que obedecem aos princípios da abordagem SSS. Cada um mostra níveis diferentes de precisão, variantes de acordo com a distribuição de probabilidades marginal dos dados, assimetria e tempo computacional. Para esta tese, optou-se pelo uso do algoritmo de Médias Móveis Simétricas (do inglês *Symmetric Moving Average – SMA*), por ser uma abordagem aplicável a séries com qualquer distribuição marginal (gaussiana ou não), portanto simétricas ou assimétricas. Além disso, o SMA é suficientemente genérico para a geração de séries sintéticas em diversas localidades simultaneamente.

#### 1.4.1.1.3 Algoritmo médias móveis simétricas

A base do algoritmo de médias móveis simétricas (SMA) está na definição de um processo puro de médias móveis, mostrado na equação (1.55):

$$y_t = \sum_{j=-\infty}^0 \theta_{-j} a_{t+j} = \dots + \theta_2 a_{t-2} + \theta_1 a_{t-1} + \theta_0 a_t \quad (1.55)$$

onde  $a_t$  é a série de resíduos e  $\theta_j$  são coeficientes a serem estimados a partir da sequência de autocovariâncias  $\gamma_j$ , através da relação (1.56):

$$\gamma_j = \sum_{j=0}^{\infty} \theta_j \theta_{t+j} \tag{1.56}$$

O processo mostrado pela equação (1.55) é a forma pela qual é possível transformar uma série de resíduos independentes e identicamente distribuídos em uma série autocorrelacionada (Box, Jenkins e Reinsel, 2008, p. 9-10; Morettin e Toloi, 2004, p. 38). Koutsoyiannis (2000, 2002) propôs duas alterações neste processo: (i) limitar a quantidade de resíduos  $a_t$  do modelo a um número finito de elementos  $q$  e (ii) limitar o somatório para  $-q \leq t \leq q$ , de forma que o elemento  $t = 0$  fique centrado na série. Dessa maneira, tem-se o processo simétrico de médias móveis (SMA), definido pela equação (1.57).

$$y_t = \sum_{j=-q}^q \theta_{|j|} a_{t+j} = \theta_q a_{t-q} + \dots + \theta_1 a_{t-1} + \theta_0 a_t + \theta_1 a_{t+1} + \dots + \theta_q a_{t+q} \tag{1.57}$$

A justificativa dada pelo autor para a alteração (i) promovida no processo de médias móveis original é que a série de resíduos  $a_i$  possui um decaimento progressivo conforme  $j \rightarrow \infty$ , fazendo com que elementos com ordens superiores a  $q$  possam ser desprezados sem perdas de precisão. Ressalta-se que no caso do processo SMA, o número  $q$  de elementos é significativamente maior do que as ordens dos modelos estocásticos lineares (e.g. MA[ $q$ ], ver seção 1.4.3) tradicionalmente utilizados. A escolha do valor de  $q$  é feita de acordo com o número de termos da FAC a ser reproduzido pelo modelo. Koutsoyiannis (2002) propõe um critério matemático para sua seleção, mostrado pela equação (1.58):

$$q \geq \max \left[ m, \left( \frac{H^2 - 0,25}{2\beta} \right)^{1/(1,5-H)} \right] \tag{1.58}$$

onde  $m$  é, no mínimo, o número de termos significativos da FAC amostral e  $\beta$  é a tolerância aceita para o erro da estimativa de  $\theta_0$  e  $H$  é o expoente de Hurst. Tolerâncias de  $\beta \leq 0,001$  são consideradas suficientemente pequenas pelo autor que em seus ensaios chega a  $q = 250.000$  para  $H = 0,9$ . No outro extremo, para  $0,6 \leq H \leq 0,7$ , chega-se a  $q = 10.000$ .

Para a alteração (ii) a justificativa é mais detalhada e pode ser conferida em Koutsoyiannis (2000). Em suma, trabalhar com um modelo simétrico permite estimar os parâmetros  $\theta_j$  analiticamente, implicando em menores erros. Além disso, ele é capaz de

preservar o coeficiente de assimetria da série, o que o torna especialmente atraente para a modelagem de vazões médias e mensais, que são notadamente assimétricas.

A estimativa dos coeficientes  $\theta_q$  é feita através das equações (1.59), propostas por Koutsoyiannis (2002):

$$\begin{cases} \theta_0 = \frac{\sqrt{(2-2H)\gamma_0}}{1,5-H} \\ \theta_j = \frac{\theta_0}{2} [(j+1)^{H+0,5} + (j-1)^{H+0,5} - 2j^{H+0,5}], \quad j > 0 \end{cases} \quad (1.59)$$

Em aplicações práticas, a autocovariância  $\gamma_0$  pode ser substituída pela variância  $\hat{\sigma}_y^2$  estimada através da série.

Conforme subitem anterior, o algoritmo SMA pode trabalhar com séries assimétricas, o que é desejável quando as variáveis em questão são vazões médias anuais ou mensais. Para isso, é necessário aplicar as equações (1.60), que estabelecem as relações entre os três primeiros momentos da série modelada com os resíduos utilizados no modelo SMA (Koutsoyiannis, 2002):

$$\begin{cases} E[a_t] = \hat{\mu}_a = \hat{\mu}_y / \left( \theta_0 + 2 \sum_{j=1}^q \theta_j \right) \\ VAR[a_t] = \hat{\sigma}_a^2 = 1 \\ \hat{\xi}_a = \hat{\xi}_y \gamma_0^{3/2} / \left( \theta_0^3 + 2 \sum_{j=1}^q \theta_j^3 \right) \end{cases} \quad (1.60)$$

onde  $\hat{\xi}$  é o estimador amostral da assimetria de cada série; os demais símbolos foram anteriormente definidos. Da mesma forma que nas equações (1.59), a autocovariância  $\gamma_0$  pode ser substituída pela variância  $\hat{\sigma}_y^2$ .

Tendo-se os três primeiros momentos estatísticos para a série de resíduos, a assimetria pode ser modelada aplicando-se uma transformação numérica baseada em alguma distribuição de probabilidades conhecidamente assimétrica. Em hidrologia tradicionalmente empregam-se a distribuições gama ou log-normal a três parâmetros (LN3 – Loucks, Stedinger e Haith, 1981, p. 282); na presente tese, optou-se, pelo uso da distribuição LN3, cuja FDP é mostrada pela equação (1.61):

$$f_x(x) = \frac{1}{(x - \zeta)\sqrt{2\pi}\sigma_x} \exp\left\{-\frac{1}{2}\left[\frac{\ln(x - \zeta) - \mu_x}{\sigma_x}\right]^2\right\} \quad (1.61)$$

onde  $\zeta$  é o parâmetro de deslocamento da distribuição. Kelman (1987, p. 372) mostra a estimação dos parâmetros da distribuição obtida através do Método dos Momentos. Utiliza-se uma variável auxiliar  $\delta$ , mostrada na equação (1.62):

$$\delta = \left[\left(1 + \frac{\hat{\xi}_a^2}{2}\right) + \left(\frac{\hat{\xi}_a^2}{2} + \frac{\hat{\xi}_a^4}{4}\right)^{1/2}\right]^{1/3} + \left[\left(1 + \frac{\hat{\xi}_a^2}{2}\right) - \left(\frac{\hat{\xi}_a^2}{2} + \frac{\hat{\xi}_a^4}{4}\right)^{1/2}\right]^{1/3} - 1 \quad (1.62)$$

Os demais parâmetros são determinados pelas equações (1.63):

$$\begin{cases} \hat{\sigma}_a^2 = \ln(\delta) \\ \hat{\mu}_a = \frac{1}{2} \ln\left(\frac{\hat{\sigma}_a^2}{\delta^2 - \delta}\right) \\ \hat{\zeta} = \hat{\mu}_a - \left[\frac{\hat{\sigma}_a^2}{(\delta - 1)}\right]^{1/2} \end{cases} \quad (1.63)$$

A série  $a'_t$  com distribuição LN3 é obtida a partir da geração de uma série  $a_t$  com distribuição normal padrão. Na sequência, essa série é submetida à exponenciação e somada ao parâmetro de deslocamento  $\zeta$ , como mostra a equação (1.64):

$$a'_t = \hat{\zeta} + \exp(\hat{\mu}_a + \hat{\sigma}_a a_t) \quad (1.64)$$

A série  $a'_t$  está apta a ser utilizada no modelo (1.57) para geração de séries sintéticas univariadas. Na sequência será apresentada a extensão do algoritmo SMA para a geração de séries multivariadas.

#### 1.4.1.1.4 Generalização para o caso multivariado

A generalização do modelo SSS para o caso multivariado é mostrada em Koutsoyiannis (2000), o qual adapta a abordagem pioneira de Matalas (1967) para o algoritmo SMA. Inicialmente, define-se  $\mathbf{y}_t = [y_{t,1}, y_{t,2}, \dots, y_{t,l}]^T$  como um vetor de  $l$  séries correspondentes às localidades  $u = 1, 2, \dots, l$ . Os elementos da matriz de variância-covariância  $\boldsymbol{\gamma}$  das séries são definidos pela equação (1.65):

$$\gamma_{uv} = COV[y_{t,u}, y_{t,v}], \quad u, v = 1, 2, \dots, l \quad (1.65)$$

O modelo SMA multivariado é muito semelhante ao mostrado pela equação (1.57), diferenciando-se apenas pelos resíduos presentes na formulação. No caso, trabalha-se com  $\boldsymbol{\varepsilon}_{t,u}$  como séries de resíduos independentes no tempo e correlacionados espacialmente para um mesmo instante  $t$ . Escreve-se, portanto, a equação (1.66):

$$\mathbf{y}_{t,u} = \sum_{j=-q}^q \boldsymbol{\theta}_{|t|,u} \boldsymbol{\varepsilon}_{t+j,u} \quad (1.66)$$

Sendo  $\boldsymbol{\gamma}_{\varepsilon}$  a matriz de variância-covariância dos resíduos  $\boldsymbol{\varepsilon}_{t,u}$ , cujos elementos podem ser calculados através da equação (1.67):

$$\gamma_{\varepsilon,uv} = \gamma_{uv} / \sum_{j=-q}^q \theta_{|j|,u} \theta_{|j|,v} \quad (1.67)$$

onde os pesos  $\theta_j$  são calculados individualmente para cada localidade  $u$  e  $v$ . O modelo multivariado que transforma as séries de resíduos não correlacionados  $\mathbf{a}_t = [a_{t,1}, a_{t,2}, \dots, a_{t,l}]^T$  nos resíduos correlacionados  $\boldsymbol{\varepsilon}_{t,u} = [\varepsilon_{t,1}, \varepsilon_{t,2}, \dots, \varepsilon_{t,l}]^T$  é o conhecido modelo proposto por Matalas (1967):

$$\boldsymbol{\varepsilon}_{t,u} = \mathbf{B} \mathbf{a}_t \quad (1.68)$$

onde  $\mathbf{B}$  é uma matriz-parâmetro de dimensões  $l \times l$ , cuja estimação é feita através da busca por uma solução para a equação (1.69):

$$\mathbf{B} \mathbf{B}^T = \boldsymbol{\gamma}_{\varepsilon} \quad (1.69)$$

É sabido (Salas *et al.*, 1985, p. 86-87; Kelman, 1987, p. 388) que existem infinitas soluções que atendem à condição estabelecida pela equação (1.69). Um procedimento típico é aplicar a decomposição de Cholesky para obter uma matriz  $\mathbf{B}$  triangular inferior. Trata-se de uma técnica bastante difundida na literatura específica e que rende boas soluções, mas que está limitada a matrizes  $\boldsymbol{\gamma}_{\varepsilon}$  positivo-definida. Koutsoyiannis (2000) propõe outro algoritmo para o caso de matrizes que não atendem a essa condição.

Para completar a descrição do modelo SMA multivariado, é necessário reescrever as estatísticas (1.60) em termos dos resíduos correlacionados  $\boldsymbol{\varepsilon}_i$ . Para tanto, utilizam-se as equações (1.70):

$$\begin{cases} \hat{\boldsymbol{\mu}}_a = \mathbf{B}^{-1} \hat{\boldsymbol{\mu}}_\varepsilon \\ \hat{\boldsymbol{\sigma}}_a^2 = 1 \\ \hat{\boldsymbol{\xi}}_a = (\mathbf{B}^{(3)}) \hat{\boldsymbol{\xi}}_\varepsilon \end{cases} \quad (1.70)$$

onde  $\mathbf{B}^{(3)}$  é a matriz cujos elementos são os cubos da matriz  $\mathbf{B}$ .

#### 1.4.1.1.5 Algoritmo SMAD – uma proposta de adaptação para séries sazonais

A fim de adequar a abordagem SSS para séries sazonais, Langousis e Koutsoyiannis (2006) publicaram dois modelos capazes de reproduzir as múltiplas escalas temporais e o comportamento periódico da série simultaneamente. Tais formulações não são baseadas em técnicas de desagregação (e.g. Valencia e Schaake, 1973), pois segundo os autores, elas sofrem de limitações advindas da estimação de parâmetros, preservação da assimetria e tempo computacional. Ao invés disso, os modelos trabalham explicitamente as porções não sazonais e sazonais, considerando a última como um processo cicloestacionário de período  $\kappa$  (ver definições da seção 1.1.3, página 40).

A primeira abordagem é intitulada Modelo Periódico Autorregressivo com Filtro de Médias Móveis Simétricas (do inglês *multivariate periodic autoregressive model with symmetric moving average filter*) e, como o próprio nome sugere, trata a sazonalidade através de um modelo periódico autorregressivo de primeira ordem [PAR(1)], passando as saídas por um filtro SMA que preserva as múltiplas escalas temporais. Ao assumir o modelo PAR(1) para a sazonalidade, considera-se que as séries dos subperíodos  $\kappa$  são independentes entre si, ou seja, se  $\kappa = 12$  (séries mensais) o modelo considera que cada mês só recebe influência do mês de mesmo nome ocorrido no ano anterior. Por exemplo, a vazão média registrada em agosto de um ano é independente da que foi registrada em julho do mesmo ano, porém dependente do registrado em agosto do ano anterior. A segunda abordagem é mais sofisticada e se baseia na consideração de um processo estocástico com duas componentes, um para a sazonalidade e outro para as múltiplas escalas temporais. Através de um estudo de caso comparativo, os autores mostram que os modelos geram resultados equivalentes e indicam a utilização da primeira abordagem, por ser mais parcimoniosa.

Diferentemente dos estudos comentados, o autor desta tese propõe uma abordagem baseada na dessazonalização das séries. Dessa maneira, ao invés de se buscar soluções de

modelagem para a componente periódica das séries, propõe-se remover previamente sua sazonalidade e submeter os resíduos diretamente ao modelo SSS, através do algoritmo SMA explicado nas duas últimas subseções. Opta-se pela forma mais simples de remoção de sazonalidade, através da equação (1.71):

$$y'_{t,u} = \frac{y_{t,u} - \hat{\mu}_{y,u}}{\hat{\sigma}_{y,u}} \quad (1.71)$$

onde  $\hat{\mu}_{y,u}$  e  $\hat{\sigma}_{y,u}$  são as respectivas médias e desvios padrão das séries de cada uma das localidades  $u$  consideradas. Escreve-se o modelo multivariado de Médias Móveis Simétricas Dessazonalizadas (SMAD) através da equação (1.72):

$$y'_{t,u} = \sum_{j=-q}^q \theta_{|t|,u} \varepsilon_{t+j,u} \quad (1.72)$$

A estimação de  $\theta_{|t|,u}$  é feita individualmente para cada local  $u$  de acordo com a equação (1.59), com a única diferença que  $\gamma_0 = 1$  na estimativa de  $\theta_0$ , consequência da dessazonalização prévia das séries. Por sua vez, a obtenção dos resíduos espacialmente correlacionados  $\varepsilon_{t+j,u}$  é feita aplicando-se o mesmo procedimento explicado na subseção 1.4.1.1.4.

#### 1.4.1.1.6 Estimação do expoente de Hurst

O último parâmetro a ser considerado na abordagem SSS é, justamente, o expoente de Hurst. Diversos são os métodos existentes para sua estimação. Tyralis e Koutsoyiannis (2011) comparam o desempenho de nove estimadores conhecidos. Entretanto, os autores recomendam a utilização de métodos que condizem com a abordagem SSS.

Koutsoyiannis (2003) propõe um estimador baseado nos desvios padrão das séries agregadas, denominado Método de Mínimos Quadrados sobre o Desvio Padrão (do inglês *Least Squares Based on Standard Deviation* – LSSD). Considerando as múltiplas escalas temporais  $\eta$  de uma série com comprimento  $n$ , o estimador do desvio padrão para escalas  $1 \leq \eta \leq n/10$  é dado pela equação (1.73):



$$\hat{\sigma}^\eta \approx c_\eta(\hat{H})\eta^{\hat{H}}\hat{\sigma} \quad (1.73)$$

Onde  $\hat{H}$  é o estimador amostral do expoente de Hurst,  $\hat{\sigma}$  é o estimador amostral do desvio padrão e  $c_\eta(\hat{H})$  é um parâmetro auxiliar definido como:

$$c_\eta(\hat{H}) = \sqrt{\frac{n/\eta - (n/\eta)^{2\hat{H}-1}}{n/\eta - 1/2}} \quad (1.74)$$

O limite superior de  $\eta$  foi imposto pelo autor para que a estimativa de  $\hat{\sigma}^\eta$  fosse feita com um mínimo de 10 elementos em cada série agregada. Valores iniciais para  $\hat{H}$  e  $\hat{\sigma}$  podem ser obtidos ignorando-se o termo  $c_\eta(\hat{H})$  e aplicando-se logaritmos na equação (1.73). O que se obtém é uma equação de reta do tipo  $\ln \hat{\sigma}^\eta \approx \hat{H} \ln \eta + \ln \hat{\sigma}$  que nada mais é do que a reta ajustada no climacograma da série (ver Figura 5 na introdução da seção 1.4.1.1, página 74). Todavia, ao aplicar os logaritmos na equação (1.73) sem excluir o parâmetro auxiliar, pode-se obter uma equação em termos de erros quadráticos, tal como mostrado pela equação (1.75) (Koutsoyiannis, 2003a):

$$e^2(\hat{\sigma}, \hat{H}) = \sum_{\eta=1}^{n/10} \frac{[\ln c_\eta(\hat{H}) + \hat{H} \ln \eta + \ln \hat{\sigma} - \ln \hat{\sigma}^\eta]^2}{\eta^2} \quad (1.75)$$

A razão para a inclusão do peso  $\eta^{-2}$  está na compensação pela perda de precisão na estimativa dos parâmetros, à medida que a escala aumenta.

O mínimo erro de ajuste é determinado derivando-se a equação (1.75) em termos de  $\hat{\sigma}$  e  $\hat{H}$  e igualando as equações resultantes a zero. O procedimento é ilustrado em Koutsoyiannis (2003b), chegando-se à equação (1.76):

$$\hat{H} = \frac{a_{11}b_2(\hat{H}) - a_{21}(\hat{H})b_1(\hat{H})}{a_{11}a_{22}(\hat{H}) - a_{21}(\hat{H})a_{12}} \quad (1.76)$$

onde  $a_{ij}$  e  $b_i$  ( $i = j = 1,2$ ) são funções auxiliares definidas como:

$$a_{11} = \sum_{\eta=1}^{n/10} \frac{1}{\eta^2} \quad (1.77)$$

$$a_{12} = \sum_{\eta=1}^{n/10} \frac{\ln \eta}{\eta^2} \quad (1.77a)$$

$$a_{21}(\hat{H}) = \sum_{\eta=1}^{n/10} \frac{d_{\eta}(\hat{H})}{\eta^2} \quad (1.77b)$$

$$a_{22}(\hat{H}) = \sum_{\eta=1}^{n/10} \frac{d_{\eta}(\hat{H}) \ln \eta}{\eta^2} \quad (1.77c)$$

$$b_1(\hat{H}) = \sum_{\eta=1}^{n/10} \frac{\ln \hat{\sigma}^{\eta}}{\eta^2} - \sum_{\eta=1}^{n/10} \frac{\ln [c_{\eta}(\hat{H})]}{\eta^2} \quad (1.77d)$$

$$b_2(\hat{H}) = \sum_{\eta=1}^{n/10} \frac{d_{\eta}(\hat{H}) \ln \hat{\sigma}^{\eta}}{\eta^2} - \sum_{\eta=1}^{n/10} \frac{d_{\eta}(\hat{H}) \ln [c_{\eta}(\hat{H})]}{\eta^2} \quad (1.77e)$$

$$d_{\eta}(\hat{H}) = \ln \eta + \frac{\ln [n/\eta]}{1 - (n/\eta)^{2-2\hat{H}}} \quad (1.77f)$$

A estimação de  $H$  requer um procedimento iterativo, visto que o expoente aparece em ambos os lados da equação (1.76). Koutsoyiannis (2003b) sugere iniciar com  $\hat{H} = 0,5$ , calcular todas as equações (1.77) e obter uma estimativa atualizada de  $\hat{H}$ . O algoritmo prossegue até que se consiga convergência; uma vez finalizado, a variância na escala básica (e.g.  $\eta = 1$ ) da série pode ser estimada através da equação (1.78):

$$\hat{\sigma} = \exp\left(\frac{-a_{12}\hat{H} + b_1}{a_{11}}\right) \quad (1.78)$$

Ainda em Tyralis e Koutsoyiannis (2011), mostram-se outros dois estimadores: Método de Mínimos Quadrados baseados na Variância (do inglês *Least Squares based on Variance* – LSV) e o estimador de Máxima Verossimilhança. O método LSV é considerado como um aprimoramento do LSSD, pois torna possível exprimir uma relação exata entre a variância e o expoente de Hurst. Lembra-se que no LSSD essa relação é aproximada [equação (1.73)]. Por sua vez, o Método da Máxima Verossimilhança é apontado como mais preciso, embora sensivelmente mais complexo. Através dos resultados obtidos pelos autores, nota-se, entretanto,

que para  $H$  baixos ( $H \sim 0,6$ ) e amostras pequenas ( $n < 2000$ ) as diferenças entre LSSD, LSV e Máxima Verossimilhança são indistinguíveis.

#### 1.4.2 Modelo Markoviano de primeira ordem

O modelo Markoviano de primeira ordem [MAR(1)<sup>7</sup>] possivelmente é a abordagem estocástica mais conhecida para a geração de séries sintéticas hidrológicas. Inicialmente proposto por Matalas (1967), trata-se da extensão para o caso multivariado do modelo pioneiro de Thomas e Fiering (1962). É uma alternativa para a geração de séries não sazonais, com pequena dependência temporal e com comportamento estacionário, sendo frequentemente utilizado na modelagem de séries de vazões médias anuais.

Como propriedades, o MAR(1) preserva os dois primeiros momentos estatísticos das séries históricas, além das autocorrelações de primeira ordem e das correlações espaciais instantâneas (ordem zero). O fato do MAR(1) reproduzir apenas uma estrutura de persistência simples, leva alguns autores a caracterizá-lo como modelo de memória curta, apontando essa propriedade como uma limitação a ser considerada (Salas *et al.* 1985, p. 376; Koutsoyiannis, 2002). Detalhes de sua formulação e estimação de parâmetros são fornecidos no próximo subitem.

##### 1.4.2.1 Definição do modelo MAR(1)

Seja o vetor de  $l$  séries  $\mathbf{z}_t = [z_{t,1}, z_{t,2}, \dots, z_{t,l}]^T$ , cada um correspondente a uma localidade  $u$  ( $u = 1, 2, \dots, l$ ). O modelo MAR(1) é definido pela equação (1.79) (Matalas, 1967):

$$\mathbf{z}_t = \mathbf{A}\mathbf{z}_{t-1} + \mathbf{B}\mathbf{a}_t \quad (1.79)$$

onde  $\mathbf{A}$  e  $\mathbf{B}$  são matrizes-parâmetro de dimensões  $l \times l$ . No modelo MAR(1), a matriz  $\mathbf{A}$  é a responsável por modelar a dependência temporal entre as vazões propriamente ditas. Sua estimação via Métodos dos Momentos, é dada por:

---

<sup>7</sup> Tradicionalmente o modelo Markoviano de primeira ordem é abreviado como AR(1), como em Salas *et al.* (1985, p. 376) e Kelman (1987, p. 389). Porém, nesta tese será utilizada a sigla MAR(1), a fim de evitar confusões com os modelos autorregressivos de primeira ordem advindos da teoria de Box & Jenkins (Box, Jenkins e Reinsel, 2008, p. 59, descritos também na seção 1.4.3.1).

$$\mathbf{A} = \mathbf{M}_1 \mathbf{M}_0^{-1} \quad (1.80)$$

onde  $\mathbf{M}_0^{-1}$  é a inversa da matriz de variância-covariância de  $\mathbf{z}_t$  e  $\mathbf{M}_1$  é uma matriz cuja diagonal principal possui os coeficientes de autocorrelação de  $\mathbf{z}_t$  e os demais elementos são as correlações espaciais entre as séries  $z_{t+1}^u$  e  $z_t^v$  ( $u, v = 1, 2, \dots, l$ ).

Por sua vez, a matriz-parâmetro  $\mathbf{B}$  é responsável por modelar a dependência espacial entre as séries consideradas. Novamente utilizando a estimação via Método dos Momentos, chega-se à equação (1.81):

$$\mathbf{B}\mathbf{B}^T = \mathbf{M}_0 - \mathbf{M}_1 \mathbf{M}_0^{-1} \mathbf{M}_1^T \quad (1.81)$$

Matalas (1967) atesta que se as correlações espaciais de primeira ordem não foram relevantes para a modelagem, a matriz-parâmetro  $\mathbf{A}$  pode ser reduzida a uma matriz diagonal cujos elementos são os coeficientes de autocorrelação de  $\mathbf{z}_t$ . Loucks, Stedinger e Haith (1981, p. 299-300) complementam dizendo que um modelo com a matriz-parâmetro  $\mathbf{A}$  completa não deve ser chamado de Markoviano, pois sua estrutura teórica de covariâncias é diferente e, portanto, o modelo MAR(1) deve considerar a matriz-parâmetro  $\mathbf{A}$  diagonal. Nessa configuração, a estimação da matriz-parâmetro  $\mathbf{B}$  fica simplificada por:

$$\mathbf{B}\mathbf{B}^T = \mathbf{M}_0 - \mathbf{A}\mathbf{M}_0^{-1}\mathbf{A}^T \quad (1.82)$$

Solucionando-se a parte direita da igualdade da equação (1.82), chega-se a uma situação semelhante à mostrada pela equação (1.69), na qual pode-se aplicar a decomposição de Cholesky e obter uma matriz-parâmetro  $\mathbf{B}$  triangular inferior. Evidentemente, essa alternativa ainda está limitada a matrizes positivo-definidas.

Na construção de um modelo MAR(1), é válido apontar duas questões específicas: (i) tipicamente trabalha-se com séries padronizadas, tanto na equação principal do modelo quanto na estimação de seus parâmetros (Salas *et al.*, 1985, p. 353). Para isso, as séries históricas são subtraídas de suas médias e divididas por seus desvios-padrão respectivos, de forma análoga à mostrada pela equação (1.71); (ii) o modelo MAR(1) deve ser aplicado a séries com distribuição aproximadamente normal. Nesse caso, a utilização da transformação LN3 pode cumprir esse requisito, além de garantir a reprodução da assimetria das séries. O procedimento é descrito pelas nas equações (1.62) a (1.64), devendo-se adicionar o estimador para o coeficiente de correlação, dado por Kelman (1987, p. 372) e mostrado na equação (1.83):

$$\hat{\rho}_{z',1} = \frac{\ln[1 + \hat{\rho}_{z,1}(\zeta - 1)]}{\ln \zeta} \quad (1.83)$$

onde  $\hat{\rho}_{z,1}$  é o estimador do coeficiente de correlação de primeira ordem da série original e  $\zeta$  é o parâmetro de deslocamento da distribuição LN3, estimado de acordo com a equação (1.63). A transformação logarítmica é aplicada a cada série de acordo com a equação (1.84):

$$z' = \ln(z - \zeta) \quad (1.84)$$

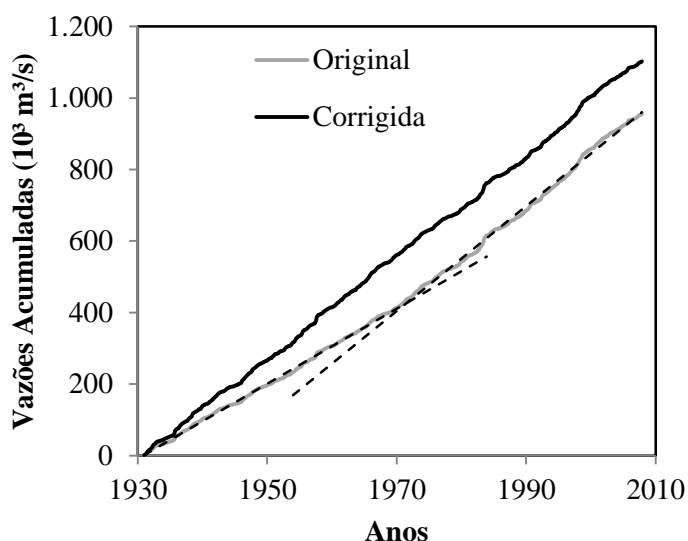
onde  $z'$  é a série transformada LN3.

#### 1.4.2.2 Remoção do comportamento não estacionário das séries históricas

Como dito na parte introdutória da seção 1.4.2, os modelo MAR(1) são formulações estacionárias e, portanto, devem ser aplicados a séries com comportamento estacionário. Para séries que não apresentam esse comportamento, pode-se aplicar um procedimento de correção antes de submetê-las ao modelo.

Batista et al. (2009) apresentaram um método gráfico no qual se plotam as séries acumuladas de vazão média *versus* o tempo. Séries com comportamento não estacionário apresentam uma (ou mais) quebra(s) de tendência em algum(ns) momento(s), sendo possível o ajuste de uma (ou mais) reta(s) sobre os dados. Posteriormente, Detzel *et al.* (2011) usaram esse método para remoção do comportamento não estacionário, dividindo a série acumulada em duas com base no ponto de quebra de tendência. Na sequência, os autores calcularam os coeficientes angulares ajustados a cada subperíodo. O procedimento de correção é completado quando se multiplica o coeficiente angular de um período com o outro, linearizando a tendência da série como um todo. A Figura 6 ilustra o procedimento elaborado para uma das séries de vazões médias anuais utilizadas no referido trabalho, na qual se percebe a quebra de tendência no início da década de 1970.

Tipicamente, multiplicam-se as vazões médias do período mais antigo pela relação entre os coeficientes angulares do período recente e do antigo, na tentativa de priorizar a porção atual da série. No entanto, o método possui a limitação de depender da subjetividade do analista quando da escolha dos períodos. Ademais, quando se trabalha com diversas séries esse método pode se tornar excessivamente laborioso.



**Figura 6 – Método gráfico para remoção do comportamento não estacionário utilizado em Detzel *et al.* (2011). As linhas tracejadas ajustadas sobre a série original evidenciam a quebra de tendências**

No caso em que o comportamento não estacionário da série possa ser representado por uma tendência linear (e.g. não estacionariedade na média), a alternativa é estimar seu coeficiente angular e aplica-lo à série original. Tal estimativa pode ser feita pelo método dos Mínimos Quadrados (ver subitem 1.2.1.1, página 50), ou pela estimativa não paramétrica proposta por Sen (1968). Esta última é tida como mais robusta, pois está imune à eventual não normalidade da série e é menos afetada por valores extremos (Zhang *et al.*, 2000). O estimador não paramétrico de Sen é dado pela equação (1.85):

$$b = \text{mediana} \left[ \frac{z_j - z_i}{t_j - t_i} \right], \quad 1 \leq i < j \leq n \tag{1.85}$$

onde  $z_{i,j}$  representa a série original e  $t_{i,j}$  os indexadores temporais nas posições  $i$  ou  $j$ . A remoção da tendência linear da série é feita através da equação (1.86) (Yue *et al.*, 2002):

$$z_t^* = z_t - bt, \quad t = 1, 2, \dots, n \tag{1.86}$$

Após o procedimento, a série  $z_t^*$  passa a ser considerada com comportamento estacionário, podendo ser submetida ao modelo MAR(1). Lembra-se que, ao final do processo de modelagem, a operação inversa deve ser realizada para reintrodução do coeficiente  $b$  nas séries sintéticas geradas.

### 1.4.3 Modelos Box & Jenkins

A aplicação de modelos estocásticos lineares a séries temporais se difundiu principalmente a partir do trabalho de George Box e Gwilym Jenkins (Box, Jenkins e Reinsel, 2008) sobre os modelos autorregressivos integrados e de médias móveis (ARIMA), também chamados de modelos Box & Jenkins. A base dessa formulação está na reprodução da estrutura de autocorrelação da série e, por esse motivo, sua aplicação sobre séries de vazões se tornou intensiva.

Os modelos ARIMA são uma combinação linear entre três elementos: componente autorregressivo (AR), fator de integração (I) e componente de médias móveis (MA). Dessa forma, uma série pode ser modelada pelos três elementos em conjunto, ou por apenas uma parte deles. Em termos gerais, o modelo não sazonal é representado por  $ARIMA(p, d, q)$ , onde as letras  $p$ ,  $d$  e  $q$  indicam a ordem do componente AR, o grau de diferenciação I e a ordem do elemento MA, respectivamente. Antes de passar para as definições matemáticas dos modelos ARIMA, introduz-se o operador defasagem, utilizado como ferramenta para escrever as equações de forma compacta. Define-se, portanto, o operador defasagem  $B$  aplicado sobre a série  $z_t$  através da equação (1.87):

$$z_{t-k} = B^k z_t \quad (1.87)$$

O correto entendimento do uso dessa classe de modelos frente a séries com comportamento estacionário ou não estacionário é de grande importância principalmente pelos assuntos discutidos no capítulo anterior. Por esse motivo, o detalhamento da formulação é feito em seções distintas.

#### 1.4.3.1 Modelos estocásticos lineares estacionários

Os modelos estacionários não possuem o fator de integração em seu equacionamento, reduzindo sua denominação para autorregressivos e de médias móveis e sua representação para  $ARMA(p, q)$ . O componente autorregressivo  $AR(p)$  é descrito pela equação (1.88):

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) z_t = \phi(B) z_t = a_t \quad (1.88)$$

onde  $\phi_p$  representa os  $p$  parâmetros autorregressivos,  $z_t$  a série temporal e  $a_t$  a série de resíduos (independentes e identicamente distribuídos) do modelo. Por outro lado, o componente de médias móveis  $MA(q)$  descreve a relação existente entre os resíduos do modelo, como mostra a equação (1.89):

$$(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) a_t = \theta(B) a_t = z_t \quad (1.89)$$

onde  $\theta_q$  representa os  $q$  parâmetros de médias móveis do modelo. A combinação entre as equações (1.88) e (1.89) forma o modelo ARMA( $p, q$ ), exibido pela equação (1.90):

$$\phi(B) z_t = \theta(B) a_t \quad (1.90)$$

Estes modelos possuem diversas propriedades teóricas relevantes, relacionadas às suas variâncias, funções de autocorrelações e espectros, todas exploradas em detalhes no trabalho de (Box, Jenkins e Reinsel, 2008, cap. 3). No entanto, duas dessas propriedades são mais bem explicadas aqui, devido à sua importância na estruturação do modelo ARMA. Ambas se refletem na formulação através de restrições impostas aos seus parâmetros  $\phi$  e/ou  $\theta$ .

A primeira propriedade é a condição de estacionariedade do modelo, cuja definição foi discutida na seção 1.1.3 (página 40). Em um modelo puramente AR, essa condição é garantida se as raízes do polinômio característico  $\phi(B) = 0$  estiverem fora do círculo unitário. Por exemplo, se uma série for representada por um modelo AR(1), deve-se impor uma restrição ao parâmetro para que respeite  $|\phi_1| < 1$ . Em um modelo puramente MA, nenhuma condição é necessária para garantir a estacionariedade.

A segunda propriedade é a invertibilidade do modelo e se refere à possibilidade de reescrever um processo MA de forma que se obtenha um modelo puramente AR. Box, Jenkins e Reinsel (2008, p. 52) exemplificam utilizando um modelo simples de primeira ordem MA(1):

$$z_t = (1 - \theta B) a_t \quad (1.91)$$

Invertendo-se a equação (1.91) pode-se escrever a série de resíduos em termos das defasagens de  $z_t$ :

$$a_t = (1 - \theta B)^{-1} z_t = (1 + \theta B + \theta^2 B^2 + \dots + \theta^k B^k) (1 - \theta^{k+1} B^{k+1})^{-1} z_t$$

$$z_t = -\theta z_{t-1} - \theta^2 z_{t-2} - \dots - \theta^k z_{t-k} + a_t - \theta^{k+1} a_{t-k-1} \quad (1.92)$$

Fazendo  $k \rightarrow \infty$ , chega-se a:

$$z_t = -\theta z_{t-1} - \theta^2 z_{t-2} - \dots - \theta^k z_{t-k} + a_t \quad (1.93)$$

A série expressa em (1.93) só será convergente caso  $|\theta| < 1$ ; se essa condição se verificar, diz-se que o processo é invertível. Dessa maneira, em um modelo puramente MA a



condição de invertibilidade é atendida caso as raízes do polinômio característico  $\theta(B) = 0$  fiquem fora do círculo unitário. Para modelos puramente AR, nenhuma restrição é imposta para garantir a invertibilidade.

Em um modelo ARMA, as propriedades de estacionariedade e invertibilidade são garantidas impondo-se as mesmas restrições mencionadas aos seus parâmetros. A combinação entre os componentes AR e MA não traz novas restrições aos parâmetros devido à independência entre as propriedades. Portanto, um modelo ARMA cujos parâmetros respeitem  $|\phi| < 1$  e  $|\theta| < 1$  atende às condições de estacionariedade e invertibilidade.

#### 1.4.3.2 Modelos estocásticos lineares não estacionários

Os modelos de Box & Jenkins são capazes de modelar séries que apresentam comportamento não estacionário homogêneo. Para tanto, inclui-se na formulação o fator de integração I, resultando no modelo ARIMA( $p, d, q$ ). Dessa maneira, o processo passa a ser definido por duas equações:

$$\phi(B)w_t = \theta(B)a_t \tag{1.94}$$

$$w_t = \nabla^d z_t \tag{1.95}$$

A equação (1.95) define a diferença de ordem  $d$  do processo. A denominação “fator de integração” pode ser entendida ao inverter essa equação, obtendo-se (1.96):

$$z_t = \Sigma^d w_t \tag{1.96}$$

Onde  $\Sigma$  representa um somatório de ordem  $d$ . Dessa maneira, um modelo não estacionário pode ser entendido como a soma (ou *integração*) de  $d$  modelos estacionários. Ressalta-se que na equação (1.94), tanto o operador  $\phi(B)$  quanto  $\theta(B)$  são estacionários.

As propriedades de estacionariedade e invertibilidade para o modelo ARIMA são mantidas, com uma sutil diferença para a condição da estacionariedade. Para séries que exibem comportamento não estacionário homogêneo, pelo menos uma das raízes do polinômio característico  $\phi(B) = 0$  resulta sobre o círculo unitário. Especificamente, em um modelo ARIMA( $p, d, q$ ),  $d \geq 1$  raízes de  $\phi(B) = 0$  estarão sobre o círculo unitário e as demais  $p - d$  raízes se manterão fora desse domínio. Lembra-se que um modelo sem componentes AR é estacionário e não precisa de restrições para atender a essa condição.

O operador diferença  $\nabla^d$  pode ser escrito em termos do operador defasagem  $B$  utilizando-se  $\nabla^d = (1 - B)^d$ . Dessa maneira, as equações (1.94) e (1.95) podem ser unidas, simplificando a formulação geral do modelo ARIMA:

$$\phi(B)(1 - B)^d z_t = \theta(B)a_t \quad (1.97)$$

Nota-se que o modelo passa a ser escrito somente em função de  $z_t$ , tornando-o mais conveniente na modelagem de séries temporais. Ao assumir essa estrutura, a diferenciação da série é feita diretamente no equacionamento, eliminando a necessidade de diferenciações prévias.

#### 1.4.3.3 Modelos estocásticos lineares sazonais

Os modelos ARIMA exploram a autocorrelação entre os valores da série em instantes sucessivos, mas quando os dados são observados em períodos inferiores a um ano a série também pode apresentar autocorrelação para um período de sazonalidade  $\kappa$ . Os modelos que contemplam as séries que apresentam autocorrelação sazonal são conhecidos como SARIMA, ou modelos ARIMA multiplicativos.

Os modelos SARIMA contêm uma parte não sazonal, com parâmetros  $(p, d, q)$ , e uma sazonal, com parâmetros  $(P, D, Q)_\kappa$ . O modelo mais geral é dado pela equação (1.98):

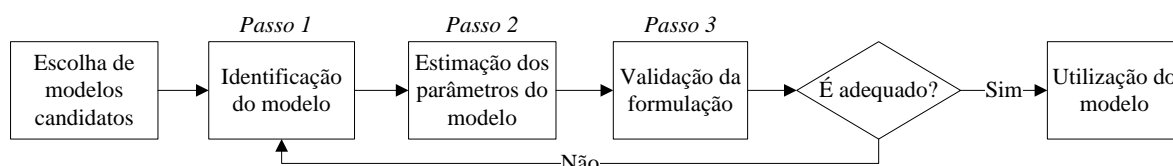
$$\phi(B)\Phi(B^\kappa)(1 - B)^d(1 - B^\kappa)^D z_t = \theta(B)\Theta(B^\kappa)a_t \quad (1.98)$$

onde os polinômios  $\Phi(B^\kappa)$ ,  $(1 - B^\kappa)^D$  e  $\Theta(B^\kappa)$  representam as porções autorregressivas, de integração e de médias móveis sazonais, para o período  $\kappa$ . A notação frequentemente utilizada para essa classe de modelos é  $SARIMA(p, d, q)(P, D, Q)_\kappa$ . Por fim, as propriedades de estacionariedade e invertibilidade da porção sazonal são extensões diretas das apresentadas para a porção não sazonal, ou seja, as raízes dos polinômios  $\Phi(B^\kappa)$  e  $\Theta(B^\kappa)$  devem se manter fora do círculo unitário.

Em casos nos quais as séries apresentem um comportamento estacionário em termos de médias e desvios-padrão dentre os períodos sazonais considerados, uma alternativa é o uso de modelos dessazonalizados. Nessa opção, a sazonalidade é previamente removida da série através da equação (1.71) exibida na seção 1.4.1.1.5 (página 83). Na sequência, a série resultante é modelada por um modelo ARIMA convencional. De acordo com Hipel e McLeod (1994, p. 463), essa é a grande vantagem em se trabalhar com modelos dessazonalizados, o que os torna flexíveis para aplicações práticas.

1.4.3.4 O procedimento iterativo de Box & Jenkins

A aplicação dos modelos ARIMA para geração de séries sintéticas pode ser esquematizada em três passos, propostos pelos próprios autores (Box, Jenkins e Reinsel, 2008, caps. 6, 7 e 8): identificação, estimação e validação. A Figura 7 exibe a estrutura para construção do modelo.



**Figura 7 – Procedimento iterativo de Box & Jenkins (Fonte: adaptado de Box, Jenkins e Reinsel, 2008)**

O primeiro passo consiste na definição da classe dos modelos autorregressivos (com ou sem médias móveis) e suas respectivas ordens. A técnica mais tradicional de se identificar um modelo ARIMA é através da comparação gráfica entre as funções de autocorrelação (FAC) e de autocorrelação parcial (FACP). A FAC, que considera a estrutura de dependência entre os elementos da amostra, foi definida pela equação (1.9) na seção 1.1.2 (página 40). A FACP, por sua vez, pode ser chamada de correlograma parcial e é representada por  $\varphi_{kk}$ . Diferentemente da função anterior, a FACP não tem uma interpretação física evidente, bastando dizer que é uma função auxiliar para o estudo da dependência entre os elementos da série (Salas *et al.*, 1985, p. 112). Sua determinação é feita a partir do ajuste de múltiplos modelos autorregressivos de ordens  $j$  ( $j = 1, 2, 3, \dots, k$ ), expressos genericamente pela equação (1.99):

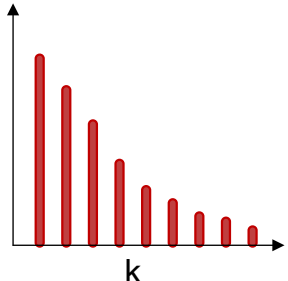
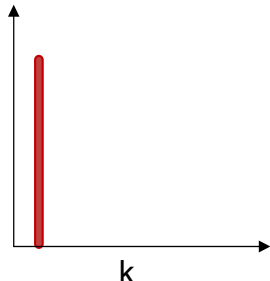
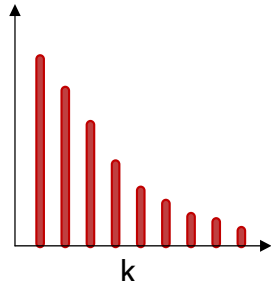
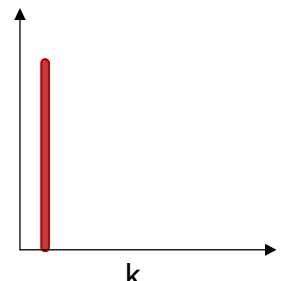
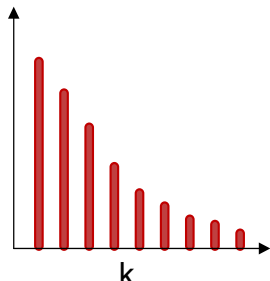
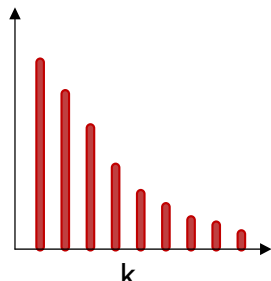
$$\rho_j = \varphi_{k1} \cdot \rho_{j-1} + \varphi_{k2} \cdot \rho_{j-2} + \dots + \varphi_{k(k-1)} \cdot \rho_{j-k+1} + \varphi_{kk} \cdot \rho_{j-k} \quad (1.99)$$

Os cálculos devem proceder de forma recursiva e os últimos coeficientes estimados para modelo ( $\varphi_{11}, \varphi_{22}, \varphi_{33}, \dots, \varphi_{kk}$ ) formam o conjunto de autocorrelações parciais.

Depois de estimadas através da amostra, as FAC e FACP devem ser grafadas e comparadas com seu comportamento teórico esperado. Diversas referências (Salas *et al.*, 1985; Hipel e McLeod, 1994; Souza e Camargo, 2004; Box, Jenkins e Reinsel, 2008) trazem a dedução das FAC e FACP teóricas para os modelos AR(p), MA(q) e ARMA(p,q). Souza e Camargo (2004, p. 69-73) traçam os gráficos resultantes dessas funções teóricas, cujos comportamentos esperados são expressos na Tabela 3. No processo de identificação dos modelos, o uso das funções mostradas deve ser conjunto. A forma gráfica da FAC revela se o

modelo possui porções autorregressivas, de médias móveis ou uma mistura das duas. A(s) ordem(ns)  $p$  e/ou  $q$  da formulação é(são) determinada(s) através da FACP.

**Tabela 3 – Comportamentos teóricos das FAC e FACP**

	AR(p)	MA(q)	ARMA(p,q)
<b>FAC</b> ( $\rho_k$ )	Infinita - exponenciais e/ou senóides amortecidas 	Finita – anula-se bruscamente após o lag $k$ (ex.: $k = 1$ ) 	Infinita - exponenciais e/ou senóides amortecidas para $k > q - p$ 
<b>FACP</b> ( $\varphi_{kk}$ )	Finita – anula-se bruscamente após o lag $k$ (ex.: $k = 1$ ) 	Infinita - exponenciais e/ou senóides amortecidas 	Infinita - exponenciais e/ou senóides amortecidas para $k > q - p$ 

Fonte: Adaptado de Souza e Camargo (2004).

Com relação à presença, ou não, do fator de diferenciação, é importante fazer uma ressalva: em uma série com a iminência de comportamento não estacionário, isto é, quando algum de seus parâmetros AR se localiza próximo ao limite do círculo unitário, Box, Jenkins e Reinsel (2008, p. 197) mostram que a FAC apresenta um decaimento lento e aproximadamente linear, independentemente de sua magnitude. Dessa forma, o procedimento tradicional seria aplicar a(s) diferenciação(ões) sempre que esse padrão for detectado, até que a FAC convirja para um dos comportamentos mostrados na Tabela 3. Portanto, considera-se conceitualmente equivocado tomar a decisão de incluir o fator de integração a partir de conclusões acerca do comportamento da série que tenham sido tiradas usando, por exemplo, inferências estatísticas. Em outras palavras, aplicar um teste de tendências em uma série e diferenciá-la se a hipótese de tendência não puder ser rejeitada, é um procedimento errôneo do ponto de vista tradicional dos modelos Box & Jenkins. Essa restrição é reforçada ao lembrar que muitos testes usam

premissas não condizentes com o que se observa em séries de fenômenos naturais, conforme discussão na seção 1.3. Ademais, Box, Jenkins e Reinsel (2008, p. 221) reforçam que em casos nos quais a escolha entre modelos estacionários e não estacionários não possa ser feita prontamente, deve-se dar preferência à classe de modelos não estacionários.

Na intenção de reduzir a subjetividade da análise gráfica, o procedimento de identificação pode ser realizado através de um método matemático denominado Critério de Informação Bayesiano (BIC, do inglês *Bayesian Information Criterion*, Schwartz, 1978). O BIC é dado pela equação (1.100):

$$\text{BIC} = -2 \cdot \ln L(\hat{\phi}, \hat{\theta}, \hat{\sigma}_a | z_t) + \ln n \cdot (p + q) \quad (1.100)$$

onde  $\ln L(\hat{\phi}, \hat{\theta}, \hat{\sigma}_a | z_t)$  é a função de log-verossimilhança, cuja definição será formalmente apresentada na sequência,  $n$  é o comprimento da série temporal  $z_t$  e  $p$  e  $q$  são as ordens dos modelos autorregressivos e de médias móveis, respectivamente. Este método consiste em aplicar a equação (1.100) para todos os modelos candidatos e escolher aquele que resultar no menor BIC. Além dele, existem outros critérios semelhantes, como o AIC (*Akaike Information Criteria*; Akaike, 1974) e suas extensões denominadas TIC (*Takeuchi Information Criteria*) e RIC (*Regularization Information Criteria*) (Hipel e McLeod, 1994, p. 215). Em particular, Shittu e Asemota (2009) apontaram que o critério AIC tende a superestimar as ordens de modelos predominantemente autorregressivos. Assim, optou-se nesta tese pelo uso apenas do critério BIC. A identificação dos modelos por meio dessa técnica é conveniente do ponto de vista computacional, pois possibilita a escolha automática de modelos para diversas séries. Entretanto, recomenda-se fortemente que a consulta às FAC e FACP das séries a serem modeladas não deva ser descartada.

Uma vez com o modelo  $\text{ARIMA}(p, q, d)$  identificado, o próximo passo é estimar seus parâmetros. A técnica mais indicada, principalmente por sua eficiência (ver comentários em Box, Jenkins e Reinsel, 2008, p. 231-232), é o Método da Máxima Verossimilhança. Mine (1984) descreve de forma clara a teoria envolvida neste método: seja uma amostra de tamanho  $n$ , na qual  $x$  são as observações. Pode-se associar uma distribuição de probabilidades do tipo  $f_X(x|\alpha)$  que depende de um grupo de parâmetros representados pelo vetor  $\alpha$ . Antes da realização do processo (que deu origem à série histórica), a função  $f_X(x|\alpha)$  poderia associar uma densidade de probabilidades a cada observação  $x$ , para um conjunto específico de parâmetros  $\alpha$ . Depois da realização do processo, a tarefa é encontrar qual a combinação de parâmetros que mais se aproxima do resultado observado. Esse objetivo é satisfatoriamente

atingido com o emprego da função de verossimilhança  $L(\alpha|x)$ . Nota-se que essa função é semelhante à distribuição de probabilidades  $f_X(x|\alpha)$ ; a única diferença reside no fato de que as incógnitas passam a ser os parâmetros e as ocorrências são os dados conhecidos.

Para o equacionamento de  $L(\alpha|x)$  é preciso adotar uma hipótese para a distribuição conjunta das observações  $x$ . A hipótese mais cômoda é que essa distribuição seja normal, podendo-se utilizar transformações numéricas [como Box-Cox, ver equação (1.11) na página 44] para aproximar os dados a uma normal, caso necessário. A função de verossimilhança, assumindo que a série de observações seja  $z_t$  ( $t = 1, \dots, n$ ) e os parâmetros do vetor  $\alpha$  sejam  $\phi$  e  $\theta$ , tem a forma:

$$L(\hat{\phi}, \hat{\theta}, \hat{\sigma}_a|z_t) = \frac{1}{(2\pi\hat{\sigma}_a)^{n/2}} \cdot \exp\left[-\frac{1}{2\hat{\sigma}_a} \cdot \sum_{t=1}^n \hat{a}_t(\hat{\phi}, \hat{\theta})^2\right] \quad (1.101)$$

onde  $a_t$  é a série de resíduos com desvio padrão amostral  $\hat{\sigma}_a$ . Em termos operacionais é comum tomar o logaritmo da função de verossimilhança, ressaltando-se que esse artifício não altera os pontos de mínimo ou máximo da função. A equação resultante, chamada de log-Verossimilhança, é dada pela equação (1.102):

$$\ln L(\hat{\phi}, \hat{\theta}, \hat{\sigma}_a|z_t) = -n \cdot \ln \hat{\sigma}_a - \frac{SSQ(\hat{\phi}, \hat{\theta})}{2\hat{\sigma}_a} \quad (1.102)$$

onde  $SSQ(\hat{\phi}, \hat{\theta})$  representa a função soma dos quadrados dos resíduos, definida pela equação (1.103):

$$SSQ(\hat{\phi}, \hat{\theta}) = \sum_{i=1}^n a_i^2 \quad (1.103)$$

Um conjunto de parâmetros que maximize a equação (1.102), ou minimize a soma dos quadrados dos resíduos (1.103), fornece a estimativa de máxima verossimilhança para o modelo.

Dois podem ser apontados como motivos complicadores no ajuste de um modelo ARIMA: (i) as equações do modelo [(1.90) e (1.97)] evidenciam que os resíduos possuem o mesmo índice temporal das observações. Dessa maneira, é necessário um procedimento iterativo, no qual devem ser fornecidas estimativas iniciais aos parâmetros, podendo-se adotar formulações específicas sugeridas por Box, Jenkins e Reinsel (2008, p. 213); (ii) os parâmetros devem respeitar as condições de estacionariedade e invertibilidade do modelo. Do ponto de

vista matemático, isso implica em otimizar uma função objetivo não linear (função *SSQ*) com restrições. Dependendo do tamanho da série considerada, esses dois complicadores podem interferir no tempo de processamento computacional.

Finalmente, o último passo do procedimento iterativo é a validação do modelo, através da investigação das propriedades da série de resíduos calculada com os parâmetros estimados. A primeira propriedade é a independência entre os elementos da série  $a_t$ , podendo-se avaliar através da aplicação do teste de Portmanteau (Li e McLeod, 1981). A segunda propriedade é a homocedasticidade dos resíduos, averiguada por meio do teste de Levene (Brown e Forsythe, 1974). A terceira e última verificação é quanto à distribuição de probabilidades de  $a_t$ , que deve se aproximar de uma normal. Diversos testes existem para essa tarefa, podendo-se indicar o teste de Shapiro-Wilk (Ferreira, 2008). Caso as propriedades dos resíduos não estejam de acordo com essas premissas, novo modelo deve ser selecionado e o procedimento iterativo é reiniciado. Vale ressaltar que, a exemplo do modelo MAR(1), toda a classe de modelos Box & Jenkins deve ser aplicada a variáveis com distribuição normal. Uma aproximação pode ser obtida através da transformação LN3 anteriormente explicada.

#### 1.4.3.5 Extensão para o caso multivariado – modelo CARIMA

A generalização da modelagem de Box & Jenkins para a geração de séries multivariadas é direta. Seja novamente o vetor de  $l$  séries  $\mathbf{z}_t = [z_{t,1}, z_{t,2}, \dots, z_{t,l}]^T$ , cada um correspondente a uma localidade  $u$  ( $u = 1, 2, \dots, l$ ). Seja também o vetor de  $l$  resíduos  $\boldsymbol{\varepsilon}_t = [\varepsilon_{t,1}, \varepsilon_{t,2}, \dots, \varepsilon_{t,l}]^T$ , independentes no tempo mas espacialmente correlacionados. O modelo ARIMA multivariado é definido pela equação (1.104) (Stedinger, Lettenmaier e Vogel, 1985):

$$\boldsymbol{\phi}_u(B)(1 - B)_u^d \mathbf{z}_{t,u} = \boldsymbol{\theta}_u(B)\boldsymbol{\varepsilon}_{t,u} \quad (1.104)$$

onde  $\boldsymbol{\phi}_u(B)$  e  $\boldsymbol{\theta}_u(B)$  são as matrizes-parâmetro das porções autorregressivas e de médias móveis para as  $u$  localidades consideradas, de ordens  $p = \max(p_1, p_2, \dots, p_l)$  e  $q = \max(q_1, q_2, \dots, q_l)$ . Nota-se a semelhança desse modelo com o univariado exibido na seção 1.4.3.2. De fato, a única diferença, além da notação matricial, é a porção aleatória do modelo representada pelos resíduos, cujos elementos são os responsáveis por preservar as correlações espaciais entre as localidades.

Muito embora a generalização teórica do modelo ARIMA para o caso multivariado seja intuitiva, seu uso pode se tornar laborioso devido à complexidade em se estimar todos os elementos das matrizes-parâmetro. Uma solução comumente adotada para contornar essa

questão é desconsiderar todos os elementos fora das diagonais principais das matrizes  $\phi_u(B)$  e  $\theta_u(B)$ , opção conhecida por CARIMA( $p, d, q$ ), ou modelo autorregressivo integrado de médias móveis contemporâneo, de ordens  $p, d$  e  $q$ . Dessa maneira, as matrizes-parâmetro são representadas por:

$$\phi_u = \begin{bmatrix} \phi_{11} & 0 & \dots & 0 \\ 0 & \phi_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_{ll} \end{bmatrix} \quad \theta_u = \begin{bmatrix} \theta_{11} & 0 & \dots & 0 \\ 0 & \theta_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \theta_{ll} \end{bmatrix}$$

Haltiner e Salas (1988) destacam as vantagens em se utilizar matrizes diagonais para representar os parâmetros e ressaltam que seu uso não resulta em perdas significativas de desempenho para o modelo como um todo, quando comparado a modelos ARIMA multivariados com matrizes de parâmetros não diagonais. A mesma conclusão foi observada anteriormente por Stedinger, Lettenmaier e Vogel (1985).

Fisicamente, o modelo contemporâneo respeita as autocorrelações de lags 1 e 2 em cada local e correlações cruzadas de lag zero. A estimação dos parâmetros segue o mesmo princípio apresentado na seção 1.4.3.4. Na realidade, Hipel e McLeod (1994, p. 784) atestam que o CARIMA não exige um procedimento intrincado para determinação dos parâmetros, pois a construção das matrizes diagonais se dá com as estimativas univariadas de cada localidade.

O campo espacial correlacionado para a formulação CARIMA pode ser obtido utilizando o mesmo modelo apresentado pela equações (1.68) e (1.69), bem como procedimento subsequente para a estimação da matriz triangular inferior  $B$ . Adicionalmente, é necessário calcular a série de resíduos para validação teórica do modelo, terceiro passo do processo iterativo de Box & Jenkins. Para tanto, inverte-se a equação (1.68), obtendo-se a expressão (1.105):

$$\hat{\mathbf{a}}_t = \hat{B}^{-1} \hat{\boldsymbol{\varepsilon}}_{t,u} \tag{1.105}$$

onde  $\hat{\boldsymbol{\varepsilon}}_{t,u} = [\hat{\varepsilon}_{t,1}, \hat{\varepsilon}_{t,2}, \dots, \hat{\varepsilon}_{t,l}]^T$  é o vetor de  $l$  séries de resíduos calculados com os parâmetros ajustados para cada localidade. O que se espera é que as séries  $\hat{\mathbf{a}}_t$  sejam independentes no tempo e no espaço, homocedásticas e com distribuição aproximadamente normal. Testes para independência temporal, homocedasticidade e normalidade foram indicados no final do subitem anterior; a independência espacial pode ser verificada através das correlações entre as séries consideradas, devendo estas resultar abaixo do limite de significância aceitável (ver Salas *et al.*, 1985, p. 91).



#### 1.4.4 Estado-da-arte da modelagem de séries de vazões médias

Um levantamento bibliográfico sobre trabalhos destinados à modelagem de séries de vazões médias publicados nos últimos cinco anos permitiu identificar algumas técnicas distintas das apresentadas nas seções anteriores. Muito embora trabalhos que utilizam modelos Box & Jenkins (Ghanbarpour *et al.*, 2010; Şengül e Can, 2011) e de Ruídos Gaussianos Fractais (Kirsch, Characklis e Zeff, 2012) ainda sejam encontrados, há de se apontar uma grande concentração de esforços em técnicas não paramétricas de reamostragem estatística. Dentre elas destaca-se o *bootstrap*, método no qual se considera que a amostra disponível possui uma distribuição análoga à da população a qual ela pertence. Tipicamente, trabalha-se com múltiplas reamostragens com reposição dos elementos da amostra para ser possível a construção de uma distribuição de frequências utilizada para as inferências desejadas.

O *bootstrap* em seu formato original possui limitações evidentes, principalmente quando utilizado para modelar séries com persistência. A impossibilidade de se gerar valores diferentes dos observados nas séries históricas e a dificuldade em preservar autocorrelações motivaram os estudos de Salas e Lee (2010) e Ndiritu (2011). Mais recentemente, Srivastav e Simonovic (2014) uniram o *bootstrap* a conceitos de entropia na elaboração de um modelo capaz de gerar séries sazonais e em múltiplas localidades simultaneamente. Hao e Singh (2011) também utilizaram a teoria da entropia, porém no intuito de obter a função densidade de probabilidades de séries de vazões médias e utilizá-la para a geração de séries sintéticas. Em ambos, a entropia é considerada como medida das dinâmicas presentes nas séries, entre elas a persistência temporal das vazões médias.

Luna *et al.* (2011) aliaram a lógica difusa ao *bootstrap* para a geração de séries sintéticas mensais de vazões médias em duas usinas hidrelétricas brasileiras. O modelo proposto é composto pela soma de uma porção determinística, modelada por lógica difusa, e uma porção estocástica, modelada por *bootstrap*. Os autores concluíram que a formulação foi capaz de preservar razoavelmente bem as estatísticas de média e desvio padrão das séries, porém identificaram-se dificuldades em reproduzir os coeficientes de assimetria em alguns meses.

Retornando para técnicas paramétricas, Guimarães e Santos (2011) e Silva e Portela (2012) propõem um modelo para geração de séries médias mensais através da desagregação de séries médias anuais. Os trabalhos são aplicados em séries de vazão média anual de rios portugueses que, devido à baixíssima estrutura de persistência anual (i.e., consideram-se as vazões como independentes), são modeladas através de uma simples amostragem aleatória da distribuição log-Pearson tipo III. A desagregação é realizada pelo método dos fragmentos, que

consiste em dividir as vazões médias mensais de um ano pela vazão total daquele mesmo ano, obtendo-se, assim, um conjunto com 12 vazões normalizadas. O procedimento é repetido para todos os anos da série histórica e cada fragmento é associado a uma vazão sintética média anual para completar a desagregação. A forma de seleção e aplicação desses fragmentos é o tema central dos citados trabalhos.

Ng e Vogel (2010) propõem modelar as vazões médias de bacias hidrográficas urbanas através de um modelo de regressão múltipla. O principal objetivo dos autores é contemplar os efeitos causados pela urbanização e, para isso, eles montam equações que relevam precipitação, uso da terra e uso da água. Os resultados são apresentados em termos de erros médios quadráticos e coeficientes de determinação, calculados entre os modelos ajustados e as séries históricas.

O último trabalho a ser citado é o de Lee (2012) que faz uso da análise de componentes independentes, técnica comumente aplicada nas áreas de telecomunicação, processamento de sinais e imagens. As séries de vazões médias são decompostas e os resíduos restantes são modelados através de uma formulação AR(1). O processo inverso é feito para obter as séries sintéticas.

## 1.5 Resumo do capítulo

Apresentou-se nesse capítulo a base teórica que fundamenta o assunto tratado na tese. Em um primeiro momento, definiram-se conceitos estatísticos importantes para o entendimento dos termos e premissas das principais formulações estocásticas destinadas a modelar séries hidrológicas. Mostrou-se que as flutuações encontradas em tais séries preocupam pesquisadores interessados em seu estudo. Em um primeiro momento, essas preocupações se limitavam ao fato de que os modelos tradicionais Markovianos requerem séries com comportamento estacionário. No entanto, a verificação dessa condição de equilíbrio estatístico nos registros hidrológicos guiou a elaboração de diversos outros estudos com propósito único de discutir as causas da variabilidade encontrada.

Dentre as principais técnicas utilizadas nas análises, destacaram-se testes de hipóteses paramétricos e não paramétricos, além de análises no domínio da frequência. Contudo, na grande maioria dos trabalhos é considerada a existência de uma tendência linear como forma de manifestação do comportamento não estacionário. Esse fato levou um grupo de pesquisadores a criticar tal abordagem, a começar pela própria classificação de séries em “estacionárias” ou “não estacionárias”. De acordo com esses autores, as denominações são aplicadas a modelos ou a processos estocásticos e não a séries temporais individuais (que

representam uma realização do processo estocástico). Além disso, consideram-se as flutuações observadas nas séries como seu comportamento natural esperado.

Na última parte do capítulo foram apresentadas as premissas e o equacionamento dos modelos estocásticos SSS, MAR(1) e ARIMA, visando a geração de séries sintéticas de vazões médias. Para o modelo SSS, foram expostas as considerações conceituais que o diferem das abordagens tradicionais. Introduziram-se os processos estocásticos agregados e suas propriedades estatísticas fundamentais, mostrando que o equacionamento tradicional apresenta um grau de tendenciosidade quando aplicado a séries que exibem o chamado comportamento HK. Na presente tese fez-se a opção pelo algoritmo SMA para a geração de séries sintéticas multivariadas, técnica que não requer que os dados sigam uma distribuição de probabilidades específica (e.g. normal). **Como proposta de inovação desta tese, sugeriu-se uma adaptação no algoritmo para aceitar a geração de séries sazonais, a qual se denominou SMAD.**

No caso do modelo MAR(1), mostrou-se a formulação desse tradicional modelo utilizado em hidrologia para a modelagem de séries não sazonais. Argumentou-se também que essa formulação possui algumas premissas, como distribuição normal para as séries consideradas e comportamento estacionário. Em situações nas quais esses requisitos não se aplicam, sugeriu-se a transformação LN3 para aproximar os dados de uma distribuição normal e a remoção da tendência de longo termo para corrigir eventuais comportamentos não estacionários.

Para o modelo ARIMA explicou-se a forma de estruturação dos componentes autorregressivos (AR), fator de integração (I) e de médias móveis (MA) na representação de séries temporais. Mostrou-se que essa abordagem necessita de restrições aos parâmetros para que sejam respeitadas as condições de estacionariedade e invertibilidade do modelo. Apresentou-se o procedimento iterativo de identificação do modelo, estimação de seus parâmetros e validação da formulação, de forma a obter o equacionamento apropriado para a modelagem das séries. Mostrou-se também a extensão para o caso multivariado através do modelo contemporâneo CARIMA.

Por fim, fez-se um levantamento do estado-da-arte da modelagem de séries de vazões médias, contemplando os principais trabalhos publicados nos últimos cinco anos. Além das técnicas paramétricas tradicionais, deu-se destaque a artigos que se basearam em técnicas não paramétricas de reamostragem (via *bootstrap*) e na teoria da entropia.



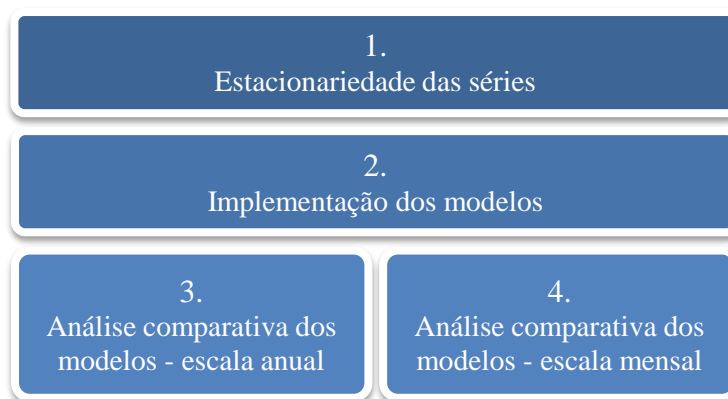
## 2 MÉTODO DE PESQUISA

Neste capítulo exibe-se a estratégia de pesquisa adotada para a investigação da hipótese levantada na introdução do trabalho. Depois da fundamentação teórica apresentada no último capítulo, mostra-se a forma com a qual os diversos conceitos foram vinculados e aplicados às séries hidrológicas consideradas. Estas, em específico, são detalhadas no próximo capítulo; por ora, é suficiente saber que foram consideradas séries de vazões médias nas escalas mensal e anual.

Nos itens subsequentes, comentários são feitos primeiramente a respeito da estruturação geral do estudo. Na sequência, elabora-se uma sistemática sugerida para a implementação computacional de cada um dos modelos estocásticos de geração. Por fim, as métricas e as estatísticas escolhidas para as análises comparativas são informadas.

### 2.1 Estrutura do estudo

Como mencionado no início deste capítulo, trabalha-se nesta tese com séries de vazões médias em duas escalas temporais: mensal e anual. Para tanto, foram preparados quatro grupos de experimentos numéricos, mostrados na Figura 8.



**Figura 8 – Experimentos numéricos desenvolvidos**

Comentários são feitos individualmente para cada grupo:

1. Foram aplicadas todas as análises detalhadas na seção 1.2.1 (Detecção de tendências em séries hidrológicas, página 48), a fim de discutir os métodos e os resultados produzidos. Além de utilizar esses resultados para adequar as séries históricas à formulação estacionária MAR(1), expõe-se também a subjetividade inerente à questão. As análises elaboradas para esse grupo fornecem subsídios para cumprir o quarto objetivo específico da tese.

2. Este grupo foi elaborado para mostrar detalhes específicos do ajuste dos modelos considerados. No caso do modelo SSS, o grupo 2 tem especial importância para validar a implementação computacional do algoritmo SMA. A averiguação é feita aplicando o modelo à série de níveis mínimos do rio Nilo na Ilha de Rhoda, a mesma utilizada em Koutsoyiannis (2002). Uma vez o modelo reproduzindo as estatísticas mostradas no citado artigo, considerou-se que ele estaria apto para aplicação em qualquer série temporal, incluindo as da bacia do rio Iguaçu utilizadas na presente tese. Para o modelo MAR(1), os exercícios do grupo objetivaram a estimação de seus parâmetros, enquanto que, do ponto de vista dos modelos CARIMA, os experimentos foram elaborados para explorar o processo iterativo de Box & Jenkins.
3. Foram geradas séries sintéticas em escala anual e confrontaram-se os resultados do modelo SSS, através do algoritmo SMA, com o modelo MAR(1). Especificamente para este último modelo, consideraram-se duas formas de entrada de séries: (i) MAR(1), com as séries originais e (ii) MAR(1)\*, com remoção do comportamento não estacionário das séries, de acordo com os resultados obtidos no primeiro grupo de experimentos numéricos. Esse procedimento colabora para cumprir o quarto objetivo específico da tese.
4. Geraram-se séries sintéticas em escala mensal e compararam-se os resultados do modelo SSS, através do algoritmo SMAD, com o modelo CARIMA. Este grupo é de grande interesse, pois exhibe a aplicação da principal contribuição desta tese. Os resultados conjuntos do terceiro com o quarto grupos completam o quinto objetivo específico da tese.

## 2.2 Sistemática para implementação dos algoritmos

Nesta seção é mostrado o método usado para a geração de séries sintéticas de vazões médias considerando os diversos conceitos e modelos apresentados no capítulo anterior. Apesar de bem difundida para os modelos MAR(1) e CARIMA, destaca-se que a sistemática para o modelo SSS é inédita.

Em todos os casos, um guia é apresentado para destacar os passos necessários. Na sequência é feita uma sucinta descrição em cada um dos elementos, sempre incluindo referências para as equações utilizadas.

### 2.2.1 Modelo SSS

O modelo SSS é utilizado para as duas escalas temporais consideradas, anual (algoritmo SMA) e mensal (algoritmo SMAD). Em termos de sistemática, a diferença entre os algoritmos consiste nos passos inicial e final, que devem ser considerados somente para séries com sazonalidade. A sistemática proposta possui 12 passos, listados abaixo:

#### Passo 1:

Com as  $\mathbf{y}_t$  séries históricas das  $l$  localidades consideradas, deve-se remover a sazonalidade de cada uma através da equação (1.71), obtendo-se as séries dessazonalizadas  $\mathbf{y}'_t$ .

#### Passo 2:

As séries  $\mathbf{y}'_t$  são submetidas ao algoritmo LSSD para estimação do expoente de Hurst. Envolvem-se as equações (1.74), (1.76) e o conjunto (1.77), que devem ser aplicados individualmente por série. As variâncias podem ser estimadas neste passo através da equação (1.78).

#### Passo 3:

Ainda individualmente por localidade, os coeficientes  $\theta_q$  devem ser estimados através das equações (1.59). O número de elementos  $q$  pode ser arbitrado, ou definido através do critério (1.58).

#### Passo 4:

As estatísticas  $\hat{\mu}_y$  e  $\hat{\xi}_y$  de cada série  $\mathbf{y}'_t$  são determinadas e substituídas nas equações (1.60) para obtenção de  $\hat{\mu}_a$  e  $\hat{\xi}_a$ .

#### Passo 5:

Calcula-se a matriz de variância-covariância  $\boldsymbol{\gamma}$  das séries  $\mathbf{y}'_t$ , submetendo-a à equação (1.67) para determinação da matriz de variância-covariância  $\boldsymbol{\gamma}_\varepsilon$  dos resíduos  $\boldsymbol{\varepsilon}_{t,u}$ .

#### Passo 6:

Aplica-se a decomposição de Cholesky sobre  $\boldsymbol{\gamma}_\varepsilon$ , obtendo-se a matriz-parâmetro  $\mathbf{B}$  triangular inferior. Atentar para que  $\boldsymbol{\gamma}_\varepsilon$  seja positivo-definida.

#### Passo 7:

Determinam-se as estatísticas dos resíduos espacialmente correlacionados  $\boldsymbol{\varepsilon}_{t,u}$ , por meio das equações (1.70).

#### Passo 8:

Gera-se um conjunto de variáveis pseudoaleatórias por localidade com distribuição normal padrão  $a_t \sim N(0,1)$ , submetendo-as à equação (1.64) para transformação LN3. O número de

variáveis pseudoaleatórias a ser gerado por local é  $2q + 1$ , equivalente à quantidade total de coeficientes  $\theta_q$ , além de  $\theta_0$ .

Passo 9:

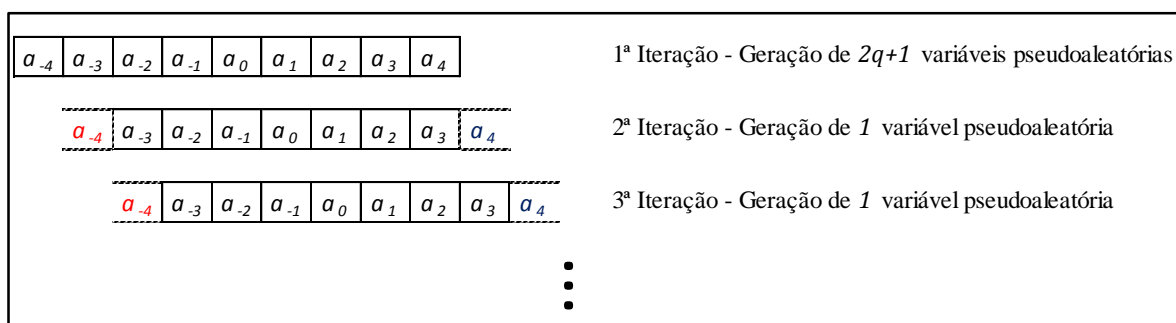
Aplica-se a equação (1.68) para obtenção da série de resíduos espacialmente correlacionados  $\epsilon_{t,u}$ .

Passo 10:

Utiliza-se a equação principal do algoritmo SMAD [equação (1.72)] para gerar um valor de vazão sintética, ainda dessazonalizada.

Passo 11:

Repetem-se os passos 8 a 10 até a obtenção do tamanho da série sintética desejada. É importante notar que a cada novo elemento da série sintética, apenas um novo valor de variável pseudoaleatória é gerado por local. Esse valor é adicionado ao final da série de números pseudoaleatórios utilizados na iteração anterior. O procedimento é mais bem explicado através da Figura 9, elaborada para um exemplo reduzido com  $q = 4$ .



**Figura 9 – Procedimento para geração de novos valores de série sintética, após a geração do primeiro elemento**

Como mostra a Figura 9, a cada nova iteração abandonam-se os valores iniciais das variáveis pseudoaleatórias, que dão lugar a uma nova variável.

Passo 12:

Por fim, uma vez gerada a série sintética de tamanho desejado, inclui-se a sazonalidade utilizando o processo inverso do mostrado no passo 1.

### 2.2.2 Modelo MAR(1)

O modelo MAR(1) é somente utilizado para a escala anual e possui uma etapa opcional relacionada à eventual correção das séries com comportamento não estacionário (passo 2). Considerando essa averiguação, a implementação do modelo MAR(1) pode ser feita em 11 passos:



Passo 1:

Com as  $\mathbf{z}_t$  séries históricas das  $l$  localidades consideradas, deve-se investigar a condição de estacionariedade, utilizando, por exemplo, algum dos procedimentos comentados na seção 1.2.1 (página 1.2.1).

Passo 2:

Caso seja detectado comportamento não estacionário, deve-se eliminar a tendência usando as equações (1.85) e (1.86), obtendo-se as séries  $\mathbf{z}_t^*$ . Caso contrário, prossegue-se ao passo 3.

Passo 3:

Calculam-se a média, desvio padrão e coeficientes de assimetria e de correlação de primeira ordem para cada série  $\mathbf{z}_t^*$ , ou  $\mathbf{z}_t$  para séries com comportamento estacionário.

Passo 4:

Usando as estatísticas calculadas no passo 3, estimam-se os parâmetros da transformação LN3 através das equações (1.63), para média, desvio padrão e deslocamento, e (1.83) para coeficiente de correlação de primeira ordem. Aplica-se a transformação utilizando a equação (1.84). Todo o procedimento é feito individualmente por série.

Passo 5:

Padronizam-se as séries, através da equação (1.71). No caso, a média e desvio padrão são únicos por série (e.g. média e desvio padrão de longo termo).

Passo 6:

Monta-se a matriz-parâmetro (diagonal)  $\mathbf{A}$ , de forma que sua diagonal principal seja preenchida com os coeficientes de correlação resultantes da equação (1.83).

Passo 7:

Determina-se a matriz de variância-covariância  $\mathbf{M}_0$ , utilizando as séries resultantes dos passos 2 ou 3.

Passo 8:

Estima-se a matriz-parâmetro  $\mathbf{B}$ , através da decomposição de Cholesky sobre a matriz resultante da operação do lado direito da equação (1.82).

Passo 9:

Geram-se as séries sintéticas através da equação (1.79), substituindo-se  $\mathbf{a}_t$  por uma matriz de números pseudoaleatórios com distribuição normal padrão. O tamanho dessa matriz deverá ser  $n \times l$ , onde  $n$  representa o tamanho desejado para as séries sintéticas. Na primeira iteração, faz-se  $\mathbf{z}'_{t-1} = 0$ , que equivale ao valor esperado da série LN3 padronizada.

Passo 10:

Incluem-se a média e desvio padrão nas séries sintéticas (processo inverso ao passo 4) e aplica-se a transformação inversa LN3 [invertendo a equação (1.84)], obtendo-se a matriz final de séries sintéticas.

Passo 11:

Caso o comportamento não estacionário tenha sido identificado no passo 1, é necessário reintroduzir a tendência linear nas séries sintéticas [invertendo a equação (1.86)].

Com relação ao passo 9, é recomendável fazer a geração de séries com tamanhos superiores ao desejado, retirando-se o excedente no final do processo. Esse procedimento objetiva eliminar a influência do valor inicial (zero) atribuído na primeira iteração.

### 2.2.3 Modelo CARIMA

Neste trabalho o modelo CARIMA é somente utilizado para a escala mensal. A sistemática a ser mostrada a seguir inclui o procedimento iterativo de Box & Jenkins, explicado através da Figura 7 (página 95). Na fase de identificação do modelo, é necessário ressaltar que se optou por formulações de ordem máxima dois, seguindo a recomendação de Box, Jenkins e Reinsel (2008). Além disso, não foram considerados modelos puramente de médias móveis (e.g. sem termos autorregressivos) devido às características das séries hidrológicas em questão. Dessa maneira, os modelos candidatos foram: CARIMA(1,0,0), CARIMA(2,0,0), CARIMA(1,0,1), CARIMA(2,0,1), CARIMA(2,0,2), CARIMA(1,1,0), CARIMA(2,1,0), CARIMA(1,1,1), CARIMA(2,1,1) e CARIMA(2,1,2), totalizando 10 formulações. A sistemática para o modelo CARIMA é composta por 12 passos.

Passo 1:

Com as  $z_t$  séries históricas das  $l$  localidades consideradas, calculam-se a média, o desvio padrão e o coeficiente de assimetria para cada série.

Passo 2:

Usando as estatísticas calculadas no passo anterior, estimam-se os parâmetros da transformação LN3 através das equações (1.63). Aplica-se a transformação utilizando a equação (1.84). Todo o procedimento é feito individualmente por série.

Passo 3:

Dessazonalizam-se as séries, através da equação (1.71). No caso, a média e desvio padrão são únicos por localidade (e.g. média e desvio padrão de longo termo).

Passo 4:

Identificação do modelo: obtém-se os gráficos da FAC [equação (1.9), página 40] e FACP [equação (1.99)]. Comparam-se os gráficos obtidos com os comportamentos teóricos dos modelos (ver Tabela 3, página 96) e obtém-se os componentes do modelo CARIMA. Alternativamente, determinam-se os critérios de informação BIC [equação (1.100)] para todos os modelos candidatos e escolhe-se a formulação cujo BIC seja mínimo.

Passo 5:

Estimação do modelo: trabalha-se na minimização da equação (1.103). Para tanto, a série de resíduos é obtida invertendo-se a equação principal do modelo [equação (1.97), no caso mais geral] de forma a ter a série  $a_t$  escrita em função de  $z_t$  [procedimento semelhante foi feito nas equações (1.92) e (1.93)]. A minimização pode ser feita através de um otimizador não linear restrito, a fim de respeitar as condições de estacionariedade e invertibilidade do modelo. Caso tenha sido feita a opção de uso do critério BIC no passo anterior, todo esse procedimento de estimação é feito junto com a identificação, economizando um passo na sistemática.

Passo 6:

Validação do modelo: com os parâmetros estimados, deve-se calcular a série de resíduos  $\hat{a}_t$  substituindo os valores obtidos e a série histórica (transformada LN3 e dessazonalizada) na equação principal do modelo [equação (1.97), no caso mais geral]. A série resultante é utilizada nas verificações teóricas de independência temporal, homocedasticidade e normalidade, ainda individualmente por série. Caso os resultados sejam satisfatórios, prossegue-se com o próximo passo; do contrário, retorna-se ao passo 4. O procedimento se repete até que se encontre um modelo apropriado.

Passo 7:

Novamente utilizando as séries de resíduos  $\hat{a}_t$  obtida no passo anterior, monta-se a matriz  $\hat{\boldsymbol{\epsilon}}_t$  cujas colunas representam as localidades e as linhas os resíduos dos modelos estimados para cada local. Caso as séries tenham tamanhos distintos, sugere-se selecionar o período histórico comum entre elas. Na sequência, padronizam-se as séries através da divisão pelos respectivos desvios padrão. Esse passo visa reduzir efeitos de escala provenientes das diferentes grandezas de cada série (rios com grandes vazões são modelados junto com rios de vazões pequenas, o que pode gerar problemas numéricos no passo seguinte).

Passo 8:

Calcula-se a matriz de correlações espaciais de  $\hat{\boldsymbol{\epsilon}}_t$  e, a seguir, aplica-se a decomposição de Cholesky, resultando na matriz-parâmetro  $\hat{\boldsymbol{B}}$  triangular inferior.

Passo 9:

Aplica-se a equação (1.105) para obter a matriz  $\hat{\mathbf{a}}_t$ , submetendo-a à verificação de independência espacial. Este é o último passo da verificação teórica do modelo.

Passo 10:

Uma vez com o modelo validado, passa-se à geração das séries sintéticas. Primeiramente geram-se números pseudoaleatórios espacialmente correlacionados  $\hat{\boldsymbol{\varepsilon}}_t$  invertendo-se a equação (1.105) e substituindo-se a matriz  $\hat{\mathbf{a}}_t$  por números pseudoaleatórios com distribuição normal padrão. Atenta-se que nesse ponto deve-se multiplicar as linhas da matriz resultante pelos respectivos desvios padrão, devolvendo a escala original das séries (procedimento inverso ao passo 7). O tamanho da matriz  $\hat{\boldsymbol{\varepsilon}}_t$  deverá ser  $n \times l$ , onde  $n$  representa o tamanho desejado para as séries sintéticas.

Passo 11:

Os números pseudoaleatórios são substituídos na equação do modelo CARIMA, obtendo-se, assim, as séries sintéticas transformadas LN3. Na primeira iteração, faz-se  $\mathbf{z}'_{t-1} = 0$ , que equivale ao valor esperado da série LN3 padronizada.

Passo 12:

Incluem-se a média e desvio padrão nas séries sintéticas (processo inverso ao passo 3) e aplica-se a transformação inversa LN3 [invertendo-se a equação (1.84)], obtendo-se a matriz final de séries sintéticas.

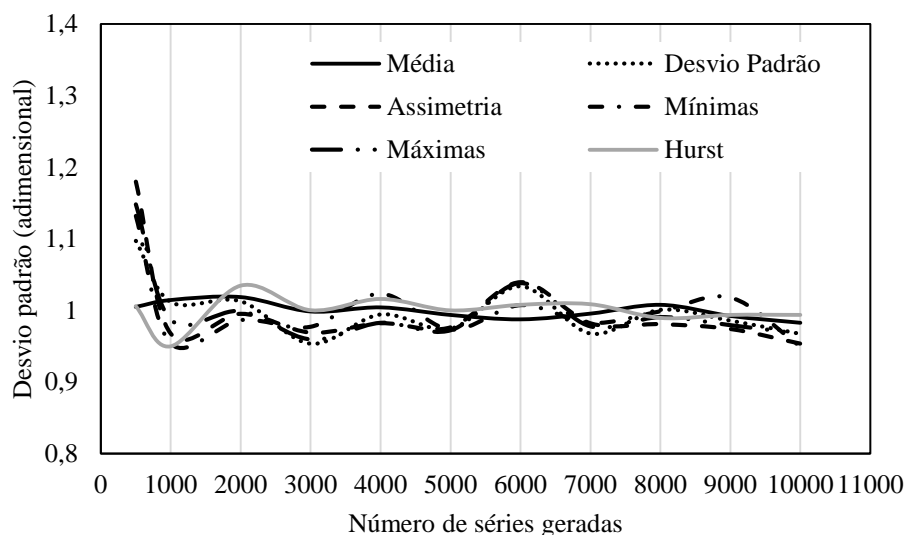
Da mesma forma que para o modelo MAR(1), é recomendável fazer a geração de séries com tamanhos superiores ao desejado, retirando-se o excedente no final do processo.

### 2.3 Métricas para análises comparativas

Depois de implementados os modelos, resta a geração das séries sintéticas propriamente dita. Neste ponto surgem as questões de quantas séries sintéticas devem ser geradas por localidade e qual o tamanho que elas devem assumir. Guimarães e Santos (2011) apresentam uma breve revisão bibliográfica a respeito, concluindo não haver nenhum consenso; os diversos estudos relatados usam de 500 a 10.000 séries sintéticas, de comprimentos que variam entre 10 e 1000 anos. Em suas próprias análises, os autores chegam a 1.200 séries sintéticas com comprimentos iguais aos das séries históricas utilizadas.

Na presente tese, fez-se uma análise com base nos desvios-padrão das estatísticas de curto termo, como forma de se medir as incertezas em suas estimações. As séries de vazões mensais em União de Vitória foram escolhidas para essa análise por terem o maior desvio

padrão entre todas as séries estudadas nesta tese. Foram gerados grupos de 500 e entre 1000 e 10.000 séries sintéticas, estes últimos espaçados em intervalos de 1000 séries cada. Na sequência, os desvios-padrão das estatísticas de curto termo foram determinados. Ao final, gerou-se a Figura 10 que contém todas as estatísticas grafadas (adimensionalizadas pela média).



**Figura 10 – Incertezas nas estatísticas de curto termo em função do número de séries sintéticas geradas para União da Vitória, considerando escala mensal**

Nota-se que com 500 séries sintéticas, o desvio padrão é elevado para a maioria das estatísticas, porém a partir de 1000 séries elas convergem a uma faixa mais estreita e oscilam levemente no entorno de um valor médio. Não houve ganho significativo na redução das incertezas com o aumento do número de séries sintéticas, portanto optou-se por trabalhar com a geração de 1000 séries sintéticas por local. Mantiveram-se os comprimentos iguais aos das maiores séries históricas do estudo, ou seja 77 anos ou 924 meses.

A validação das séries geradas seguiu a mesma métrica para os modelos SSS, MAR(1) e CARIMA, visando uma comparação direta de seus resultados. A avaliação é feita a partir de estatísticas diversas, as quais se agrupam em curto e de longo termos. A determinação de estatísticas de curto termo é a primeira forma de verificação do modelo baseado nas séries sintéticas. Nesse momento verifica-se a capacidade de reprodução de propriedades básicas do histórico observado, como uma forma de se medir a confiabilidade do modelo. Lembra-se que as próprias estatísticas de curto termo são utilizadas na construção do mesmo, fazendo com que a reprodução desses elementos somente confirme que a formulação foi corretamente implementada computacionalmente. As estatísticas de curto termo calculadas nesta fase são: média, desvio padrão, coeficiente de assimetria, vazão mínima, vazão máxima, FACs e o

expoente de Hurst para o modelo SSS. No caso das séries sazonais, calculam-se também as médias e desvios-padrão mensais. Para os modelos serem considerados apropriados, espera-se que esses parâmetros sejam semelhantes entre séries históricas e séries sintéticas.

As estatísticas de longo termo, por sua vez, foram determinadas com a intenção de avaliar as características dos modelos em épocas de secas e estiagens das séries. Sete foram as estatísticas de longo termo calculadas, podendo ser divididas em duas partes. A primeira, relacionada com períodos de estiagens, possui estatísticas calculadas conforme a chamada teoria das corridas (Haltiner e Salas, 1988). A partir de um valor de corte são contados quantos elementos em sequência estão abaixo dele. Cada conjunto de elementos com essa característica é chamado de corrida (adaptado do termo em inglês *run*). Nas análises, o valor de corte adotado foi a média de longo termo das respectivas séries históricas.

Sendo  $n_\Lambda$  o total de corridas observado em uma determinada série, pode-se trabalhar com o número de corridas e suas durações, tanto em termos temporais (meses) quanto em termos de afluência ( $\text{m}^3/\text{s}$ ) por corrida. Calculam-se:

$$\left\{ \begin{array}{l} \bar{\Lambda} = \frac{1}{n} \sum_{i=1}^n \Lambda_i \\ \Lambda_{\max} = \max(\Lambda_1, \Lambda_2, \dots, \Lambda_n) \end{array} \right. \quad (2.1)$$

onde  $\Lambda_i$  é a duração da  $i$ -ésima corrida,  $\bar{\Lambda}$  a duração média das corridas e  $\Lambda_{\max}$  a duração máxima da série de corridas. Para as estatísticas em termos de afluências a equação é a mesma, apenas considerando as vazões ocorridas em cada intervalo.

A segunda parte está relacionada com a regularização de reservatórios e se refere aos déficits acumulados. Esta propriedade pode ser determinada de diversas maneiras diferentes, dependendo da característica da série que se pretende verificar. No presente trabalho, o déficit é calculado com base em 80% da afluência média de uma série, de acordo com a seguinte formulação dada pela equação (2.2):

$$\Delta_t = \max \left\{ \begin{array}{l} 0 \\ \Delta_{t-1} - z_t + 0,8 \cdot \bar{z} \end{array} \right. \quad (2.2)$$

onde  $z_t$  é a afluência sintética no tempo  $t$  e  $\bar{z}$  é sua média de longo termo. Com a série  $\Delta_t$  podem ser calculados o máximo déficit acumulado  $\max[\Delta_t]$  (ou  $\Delta_{\max}$ ) e o déficit acumulado médio por série  $\bar{\Delta}$ . Fisicamente, essa grandeza representa o máximo déficit, em termos de afluências, necessário para o sistema suprir a regularização de um reservatório, assumindo que 80% da afluência média da série seja requerida para tal objetivo. Percebe-se que se trabalha com um

sistema sem consideração de falhas, mas, ainda assim, este índice é uma boa ferramenta de análise de desempenho do modelo.

Da mesma forma que as estatísticas de curto termo, calculam-se todos os indicadores apresentados tanto para a série histórica quanto para a série sintética e comparam-se os resultados. É este o ponto no qual se espera identificar diferenças entre os modelos estocásticos considerados.

## 2.4 Resumo do capítulo

Apresentaram-se nesse capítulo todas as premissas consideradas na aplicação dos modelos SSS e MAR(1) para a geração de séries sintéticas. Em um primeiro momento, mostrou-se que o estudo foi estruturado em quatro partes: (i) análise da condição de estacionariedade das séries históricas consideradas; (ii) implementação dos modelos; (iii) análises comparativas dos modelos para escala anual [SSS e MAR(1)] e (iv) análises comparativas dos modelos para escala mensal (SSS e CARIMA).

Na sequência foi descrito um passo-a-passo para implementação computacional de cada um dos modelos estocásticos. Entende-se essa etapa como uma ligação entre os diversos conceitos apresentados no capítulo anterior e a geração de séries sintéticas propriamente dita.

Por fim, foram estabelecidas as métricas de comparação entre os modelos considerados, todas baseadas em estatísticas sobre as séries geradas. Dividiu-se a investigação em duas partes: (i) estatísticas de curto termo, visando depurar a implementação computacional de cada método e (ii) estatísticas de longo termo, base para comparações mais criteriosas.





### 3 ÁREA DE ESTUDO

O local escolhido para aplicação dos métodos estudados se localiza na bacia hidrográfica do rio Iguaçu, uma das sub-bacias que contribuem para a bacia hidrográfica do rio Paraná. É também uma das bacias hidrográficas do Estado do Paraná (Figura 11). Além de possuir a maior área dentre as bacias paranaenses, a bacia do rio Iguaçu concentra grande parte da população e atividade econômica do Estado.



Figura 11 – Bacias hidrográficas do Estado do Paraná (Fonte: SEMA, 2010)

Foram selecionadas seis estações fluviométricas com históricos diversos que variam entre 59 e 77 anos de registros. Conhecendo o grande aproveitamento hidrelétrico existente na bacia, deu-se prioridade a postos hidrométricos cuja influência oriunda de barramentos a montante no rio fosse mínima. Por esse motivo, as estações fluviométricas selecionadas se concentram na cabeceira da bacia e em tributários importantes que deságuam no rio Iguaçu.

O presente capítulo tem por objetivo apresentar as características gerais da bacia do rio Iguaçu, em termos físicos e climáticos. São também detalhadas as informações relativas às estações fluviométricas escolhidas, mostrando suas localizações e características estatísticas relevantes para a modelagem.

### 3.1 Características gerais

A bacia hidrográfica do rio Iguaçu possui área total de 70.799 km<sup>2</sup>, dos quais 57.329 km<sup>2</sup> se localizam no Estado do Paraná e os 13.470 km<sup>2</sup> restantes abrangem o Estado de Santa Catarina (Maack, 2002, p. 361). Há ainda uma porção da bacia em território argentino, acrescentando cerca de 1830 km<sup>2</sup> de área. No Paraná, a bacia hidrográfica do rio Iguaçu ocupa 28% da área do Estado, contendo uma população superior a 4.400.000 habitantes (SEMA, 2010, p. 30 *apud* IBGE, 2004). Dentre as principais cidades contidas na área da bacia, citam-se: Curitiba e região metropolitana, Guarapuava, Cascavel (parcialmente), Foz do Iguaçu, Francisco Beltrão, Palmas e Pato Branco.

A origem do rio Iguaçu se dá na junção dos rios Iraí e Atuba, na divisa entre os municípios de Curitiba e Pinhais. Seu comprimento se estende por 910 km (Maack, 2002, p. 361), muito embora, se forem contabilizados os rios de origem, esse percurso chega a 1320 km (SEMA, 2010, p. 30). O curso do rio Iguaçu se dá no sentido leste-oeste até desaguar no rio Paraná, no município de Foz do Iguaçu, onde atinge uma vazão média de 1551 m<sup>3</sup>/s. Dentre os afluentes mais importantes, destacam-se: Negro, Potinga, da Areia, Iratim, Jordão, Cavernoso, Chopim, Guarani, São Salvador e Capanema. Dos tributários menores, mas também importantes, elencam-se: Barigui, Palmital, Atuba, Verde, Passaúna, Vargem, Claro, Passa Dois, Capivari, São Miguel e Timbó.

Geologicamente, Maack (2002, p. 362) classifica o rio Iguaçu como antecedente, ou antigo. Em seu trajeto atravessa os três planaltos paranaenses, sofrendo variações em sua declividade de acordo com as características de cada região. No primeiro planalto existem meandros vastos, por consequência das baixas quedas. Ao chegar na escarpa devoniana do segundo planalto, o rio ganha velocidade na forma de corredeiras que se estendem até as proximidades de Porto Amazonas. A partir desse ponto, novos meandros marcam o curso até o início da escarpa mesozoica do terceiro planalto.

As unidades aquíferas presentes na cabeceira da bacia dividem-se em Paleozóica Superior, Médio Superior, Pré-Cambriana, Guabirota e Karst. No restante, há ampla predominância da unidade aquífera Serra Geral Sul, com uma estreita faixa da unidade aquífera Guarani (SEMA, 2010, p. 31).

O uso do solo também apresenta uma distinção entre as regiões da cabeceira e o restante da bacia. Na porção inicial concentram-se grandes áreas urbanizadas, principalmente nas proximidades de Curitiba e região metropolitana. As atividades econômicas desenvolvidas são

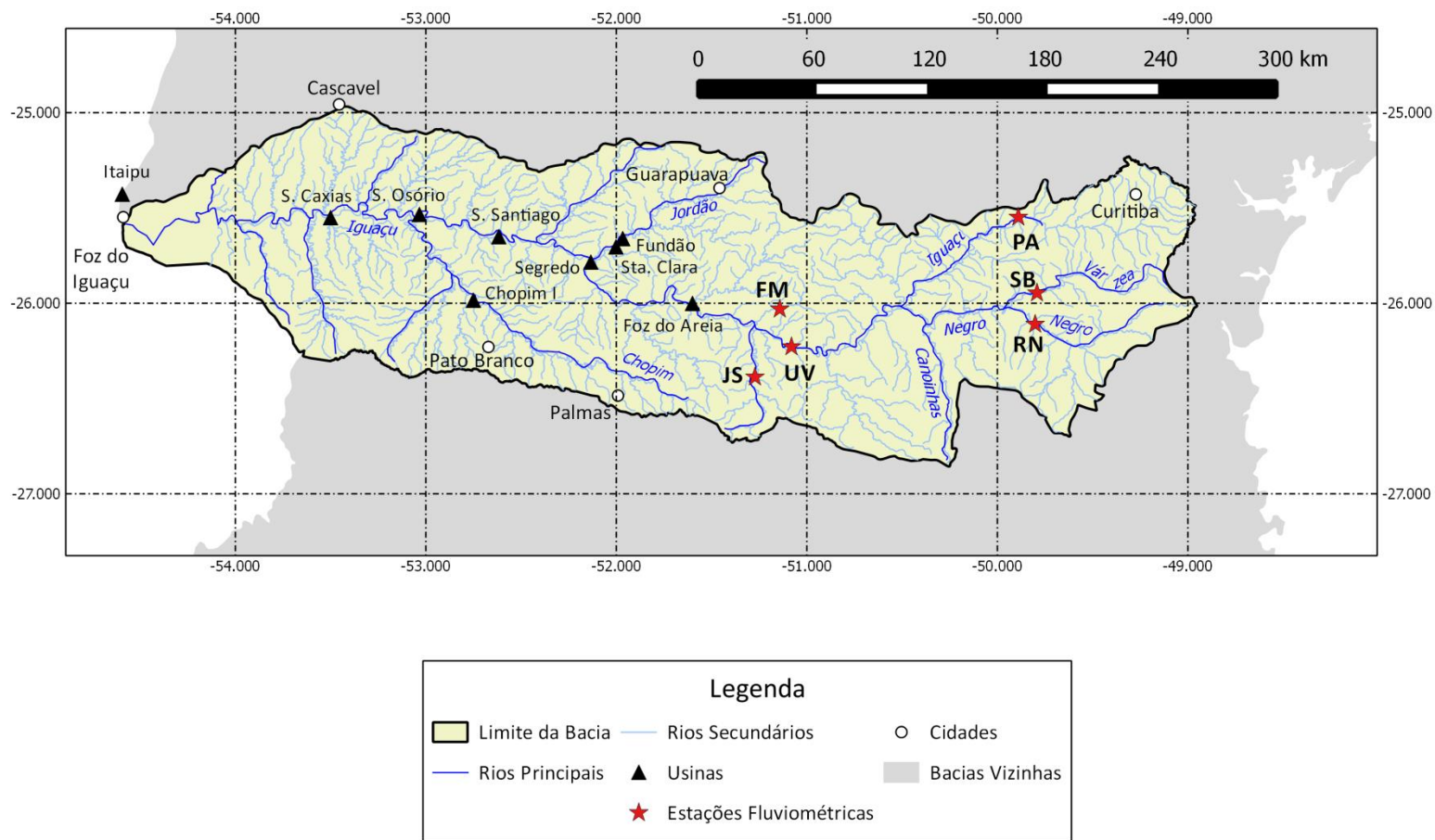
relacionadas a indústria, comércio e serviços. Na parte central e oeste da bacia, a atividade agrícola se destaca, dividindo-se entre culturas intensivas e usos mistos (SEMA, 2010, p. 32). O rio Iguaçu possui amplo aproveitamento hidrelétrico, contando com uma cascata formada por cinco grandes usinas que fornecem energia ao Sistema Interligado Nacional (SIN). De montante para jusante, operam Foz do Areia (1676 MW), Segredo (1260 MW), Salto Santiago (1420 MW), Salto Osório (1078 MW) e Salto Caxias (1240 MW). Outras duas usinas conectadas ao SIN operam no rio Jordão: Santa Clara (120 MW) e Fundão (120 MW). Juntas, todas essas usinas representam cerca de 8,5% da potência instalada de usinas hidrelétricas do SIN (ANEEL, 2014). No rio Chopim está instalada uma usina menor, Chopim I (1,98 MW), porém não está conectada ao SIN. Em termos hidrológicos, a presença dessas usinas é importante por alterar o regime dos rios onde estão instaladas, principalmente Foz do Areia, Segredo, Salto Santiago e Santa Clara, cujos reservatórios possuem capacidade de regularização de vazões. As usinas e outras características da bacia do rio Iguaçu apresentadas neste item estão mostradas na Figura 12.

### 3.2 Clima

A bacia do rio Iguaçu se localiza em uma região de clima subtropical úmido, com inverno fresco a frio na porção oeste e com inverno frio na porção leste, segundo classificação oferecida por Mendonça e Danni-Oliveira (2007, p. 151). Entre as massas de ar que atuam na região, estão a massa de ar polar atlântica (MPA), a massa de ar tropical continental (MTC) e a massa de ar tropical atlântica (MTA). Cavalcanti *et al.* (2009, p. 268) destacam a importância do fluxo de umidade na região principalmente nos meses de inverno, pois na porção sudeste da região sul do Brasil o balanço entre a precipitação e a evaporação é negativo. Esse transporte de umidade se dá pelas massas de ar marítimas MTA e MPA.

Utilizando a classificação de Köppen-Geider (Peel, Finlayson e McMahon, 2007), a bacia do rio Iguaçu está quase inteiramente inserida no tipo Cfa, ou seja, clima temperado sem estação seca e com verões quentes. Uma pequena porção da cabeceira da bacia está categorizada como do tipo Cfb, ou temperado sem estação seca e com verões amenos. A classificação de Köppen-Geider é feita utilizando séries históricas observadas de chuva e temperatura.

De uma maneira geral, a relação estabelecida entre a posição geográfica da bacia, o relevo existente e as massas de ar atuantes proporcionam uma distribuição de chuvas regular durante o ano, com médias entre 1250 mm e 2000 mm. Embora possam ser registradas máximas de precipitação em qualquer época do ano, nos meses de verão o fluxo de umidade em conjunto



**Figura 12 – Bacia do rio Iguazu e estações fluviométricas selecionadas**

com o aquecimento da superfície acaba por instabilizar a atmosfera e produzir mais convecção, provocando chuvas intensas (Cavalcanti *et al.*, 2009, p. 270).

Essa regularidade, no entanto, não é observada no regime de temperaturas, cuja sazonalidade é definida através de verões frescos a quentes e de invernos frios a frescos. O mês de janeiro é o que registra as maiores médias de temperaturas, que permanecem elevadas até o fim de fevereiro. Cavalcanti *et al.* (2009, p. 260) apontam que, além da própria radiação solar, a constante advecção de ar quente vinda do norte da região influencia a temperatura nos meses mais quentes. Por outro lado, julho é o mês com temperatura média mais fria, principalmente devido às geadas noturnas comuns no Estado do Paraná, consequência da ação da MPA durante os meses de inverno (Maack, 2002, p. 123). No verão, registram-se médias variando entre 20°C e 25°C, enquanto que no inverno as temperaturas médias caem para valores entre 11°C e 18°C.

De grande relevância ao clima do sul do Brasil e, conseqüentemente, da bacia do rio Iguaçu é o fenômeno El Niño. Trata-se do aquecimento anômalo das águas do oceano Pacífico próximo à América do Sul, podendo elevar de 1°C a 6°C as temperaturas da superfície do mar (Mendonça e Danni-Oliveira, 2007, p. 189). Embora com periodicidade discutível, em média o fenômeno ocorre a cada sete de um intervalo de 14 anos, iniciando sempre ao final de um ano e estendendo-se por todo o ano seguinte. Os reflexos do El Niño são sentidos no mundo todo, repercutindo em secas e inundações em diversas regiões. No sul do Brasil, em particular, são comuns o aumento da temperatura e do volume de precipitações. Como fato agravante tem-se que a intensidade do fenômeno varia a cada ocorrência, dificultando seu entendimento e, conseqüentemente, sua modelagem. O El Niño de 1982-1983, por exemplo, é apontado por Mendonça e Danni-Oliveira (2007, p. 192) como o mais intenso do século. No Brasil, suas conseqüências foram mais graves na região do vale do Itajaí, em Santa Catarina, manifestando-se através de uma cheia de enormes proporções. Da mesma forma, a bacia do rio Iguaçu também sofreu grandes enchentes, elevando os totais precipitados e afluências ocorridas em 1983 a valores extremos. Os municípios de União da Vitória e Porto Vitória ficaram inundados e houve problemas na operação da usina de Foz do Areia em julho de 1983, quando a vazão média atingiu 5150 m<sup>3</sup>/s (a média histórica das vazões na usina é aproximadamente 650 m<sup>3</sup>/s). Na investigação de eventuais quebras de tendência nas séries de vazão de algumas bacias brasileiras, Guetter e Prates (2002) removeram este ano de suas análises, pois seu valor sozinho causou uma elevação significativa nas médias das vazões pós década de 1970, influenciando os testes estatísticos empregados.

### 3.3 Estações fluviométricas selecionadas

De acordo com a Agência Nacional de Águas (ANA, 2014), a bacia do rio Iguaçu possui séries fluviométricas de 410 estações, a grande maioria monitorada por diferentes agentes como a própria ANA, o Instituto das Águas do Paraná, a Companhia de Saneamento do Paraná (SANEPAR) e a Companhia Paranaense de Energia (COPEL). No entanto, boa parte dessas estações hidrométricas não está ativa, ou possui histórico bastante reduzido. Dentre os postos fluviométricos ativos, podem-se destacar os que estão sob responsabilidade do Instituto das Águas do Paraná, totalizando 174 estações distribuídas por toda a bacia até a presente data.

Como mencionado na seção 3.1, o rio Iguaçu e alguns de seus afluentes possuem usinas hidrelétricas em operação. Evidentemente, a construção de barragens nos cursos dos rios causa alterações significativas em seus regimes, principalmente se o reservatório da usina possuir capacidade de regularização. Nessas regiões é muito comum a aplicação de técnicas específicas visando a reconstituição da vazão natural do rio, baseadas na incorporação de perdas devido à evaporação líquida dos reservatórios e usos consuntivos na bacia. Em muitos casos essas relações são lineares e dependem de informações provenientes de diversas fontes, como censos demográficos, órgãos econômicos e empresas de saneamento (Braga, Rocha e Gontijo, 2010).

Na tentativa de evitar tanto a influência de barragens nos rios quanto possíveis erros nas estimativas de vazões naturais, foram pesquisadas estações hidrométricas em locais que não possuem interferência dessas estruturas. Aliado a isso, buscou-se locais nos quais os dados de vazão fossem longos (mínimo de 50 anos), consistidos e possuíssem poucas falhas. Dessa maneira, foram selecionados seis postos fluviométricos, cujas principais características são descritas na Tabela 4 e com localização mostrada na Figura 12.

As séries históricas foram coletadas através do portal Hidroweb (ANA, 2014), que reúne dados hidrológicos fornecidos pelos diversos agentes responsáveis pelo monitoramento dos postos. Em estações hidrométricas que funcionam com equipamentos convencionais (régua limimétricas), são realizadas duas leituras diárias, às 07:00h e 17:00h. Nos postos com equipamentos automáticos (estações telemétricas), os dados são lidos continuamente e registrados em dispositivos específicos (*dataloggers*) em geral a cada 15 minutos (ANA e SGH, 2012). A análise de consistência dos dados é também de responsabilidade dos agentes e seguem recomendações de ANA e SGH (2012a). Ainda de acordo com esse documento, os operadores devem fornecer à ANA os dados hidrológicos em escala diária, independentemente da forma com a qual são adquiridos. Estes dados diários são, então, disponibilizados no portal Hidroweb.

**Tabela 4 – Estações fluviométricas selecionadas**

<b>Nome</b>	<b>Cód. ANA</b>	<b>Sigla</b>	<b>Data Início</b>	<b>Data Fim</b>	<b>Rio</b>	<b>Município</b>	<b>Latitude</b>	<b>Longitude</b>	<b>Altitude (m)</b>	<b>Área de drenagem (km<sup>2</sup>)</b>
<b>São Bento</b>	65155000	SB	01/05/1930	01/06/2005	Várzea	Lapa	-25:56:44	-49:47:29	750	2012
<b>União da Vitória</b>	65310000	UV	01/05/1930	01/12/2007	Iguaçu	União da Vitória	-26:13:41	-51:04:49	670	24.211
<b>Porto Amazonas</b>	65035000	PA	01/08/1935	01/12/2005	Iguaçu	Porto Amazonas	-25:32:53	-49:53:19	780	3662
<b>Jangada do Sul</b>	65370000	JS	01/11/1945	01/07/2005	Jangada	General Carneiro	-26:23:13	-51:16:19	800	1055
<b>Rio Negro</b>	65100000	RN	01/05/1930	01/12/2007	Negro	Rio Negro	-26:6:35	-49:48:04	766	3379
<b>Fazenda Maracanã</b>	65415000	FM	01/12/1945	01/05/2005	Palmital	União da Vitória	-26:01:49	-51:08:30	840	323

As vazões médias mensais são também disponibilizadas nos casos em que os registros diários não apresentem nenhuma falha dentro de um mesmo mês.

Para os trabalhos desenvolvidos nesta tese, coletaram-se diretamente vazões médias em escala mensal. Em um primeiro momento fez-se uma investigação acerca da presença de dados com falhas nas séries, ocorrências que poderiam prejudicar a determinação das vazões médias anuais e a modelagem como um todo. A Tabela 5 exibe a taxa de falhas encontradas, calculadas com referência ao tamanho de cada série.

**Tabela 5 – Taxa de falhas encontradas nos dados mensais coletados**

<b>Nome</b>	<b>Tamanho da Amostra (meses)</b>	<b>Falhas (meses)</b>	<b>Taxa (%)</b>
<b>São Bento</b>	902	1	0,11
<b>União da Vitória</b>	932	4	0,43
<b>Porto Amazonas</b>	845	0	0,00
<b>Jangada do Sul</b>	717	6	0,84
<b>Rio Negro</b>	932	4	0,43
<b>Fazenda Maracanã</b>	714	1	0,14

Todas as estações possuem taxas de falhas inferiores a 1%, distribuindo-se em diversos momentos dos respectivos históricos. Isso permitiu adotar um procedimento simples para seus preenchimentos, baseado em interpolação linear entre os meses de mesmo nome dos anos anterior e posterior à falha. É importante apontar, no entanto, que o preenchimento de falhas em séries hidrológicas constitui-se tarefa bastante complexa; para o leitor interessado em se aprofundar no assunto recomenda-se consulta a Elshorbagy, Panu e Simonovic (2000).

Depois de preenchidas as falhas, foi possível determinar as estatísticas descritivas e obter as séries em escala anual através das médias aritméticas das vazões médias mensais coletadas. A fim de caracterizar a hidrologia de cada uma das estações fluviométricas consideradas, foram elaborados resumos específicos para as escalas mensal e anual. Para a primeira, médias e desvios padrão foram grafados na Figura 13 e mostram grande aderência à descrição climática feita na seção 3.2. O que se percebe é uma regularidade na distribuição das vazões médias nos meses do ano, com as máximas ocorrendo em fevereiro (São Bento e Porto Amazonas) ou outubro (demais). Para os desvios-padrão, observa-se maior variabilidade no mês de julho em todas as estações, com exceção de Rio Negro que ocorre em agosto.

Nesse contexto, é útil a análise sobre o coeficiente de variação  $c_v$ , expresso como sendo a razão entre o desvio padrão e a média de cada série. Os  $c_v$  médios das séries mensais variam entre 0,61 (São Bento) e 0,74 (Jangada do Sul e Fazenda Maracanã), valores muito elevados



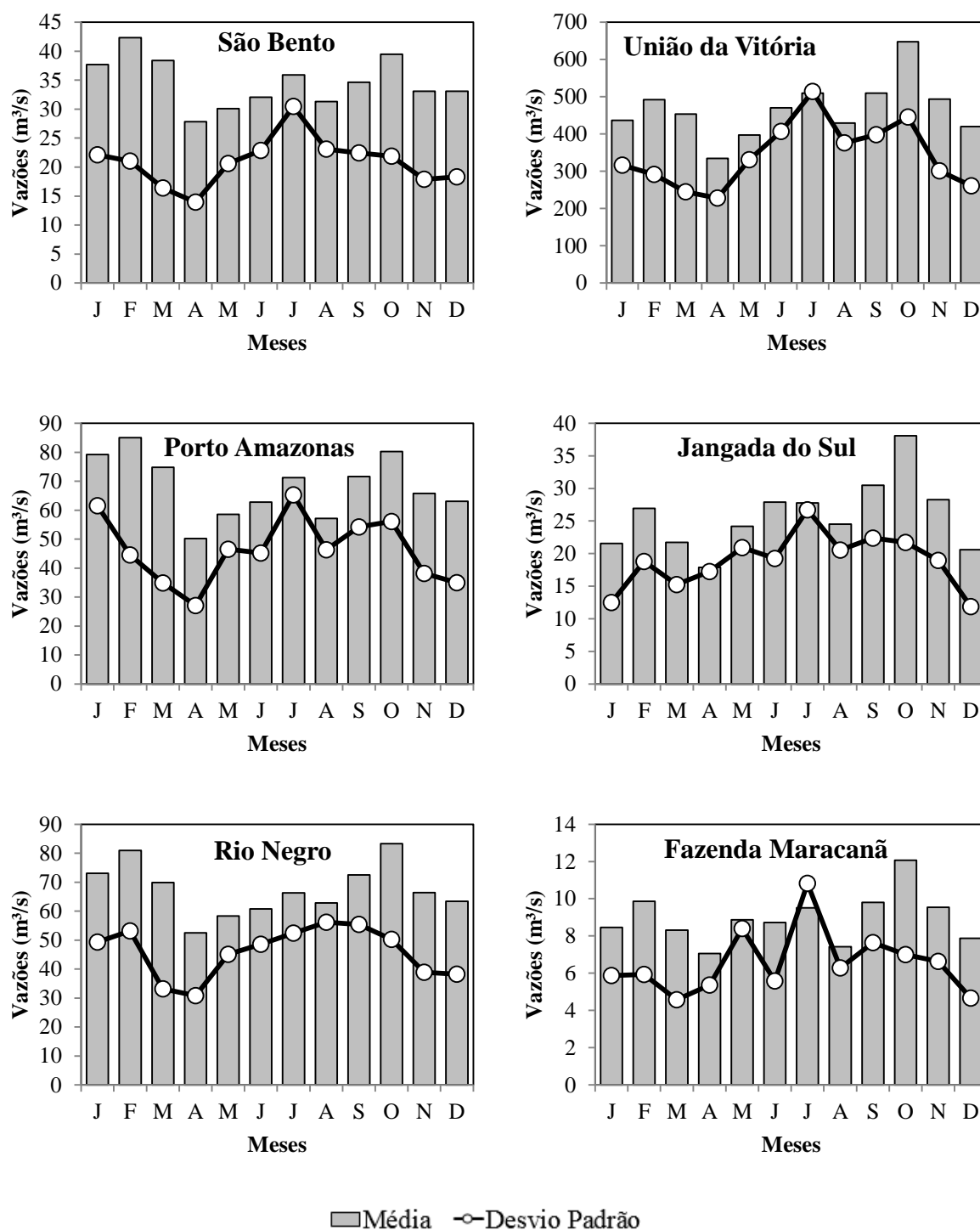


Figura 13 – Médias e desvios padrão mensais para as estações fluviométricas consideradas

quando comparados a outros rios brasileiros. Por exemplo, os rios Grande (em Minas Gerais), São Francisco (na Bahia), Tocantins (em Goiás, no trecho antes da fronteira com o Estado do Tocantins) e Paraná (no trecho fronteiro entre São Paulo e Mato Grosso do Sul) possuem  $c_v$

médios de 0,36, 0,37, 0,44 e 0,28, respectivamente<sup>8</sup>. Os valores altos de  $c_v$  na bacia do rio Iguaçu indicam um regime hidrológico relativamente disperso, caracterizado por desvios padrão de valores mais relevantes. Outro destaque ocorre especificamente no mês de julho, no qual os  $c_v$  superam a unidade em União da Vitória (1,01) e Fazenda Maracanã (1,14). Na comparação com as outras séries mencionadas neste parágrafo, os coeficientes de julho valem 0,27, 0,30, 0,36 e 0,17, respectivamente. Isso evidencia que a época mais fria do ano não necessariamente é seca na bacia do rio Iguaçu, mais uma vez corroborando com a descrição climática oferecida no item anterior.

Outras estatísticas descritivas para a escala mensal (sem remoção da sazonalidade) são mostradas na Tabela 6. Tratam-se do tamanho de cada série, coeficiente de assimetria, vazão mínima, vazão máxima e coeficiente de correlação de *lag* um. Destaca-se a elevada magnitude dos coeficientes de assimetria, sugerindo que a adoção de transformações numéricas (e.g. LN3) venha a ser necessária para a modelagem. A autocorrelação de primeira ordem, por sua vez, pode ser considerada apenas moderada.

**Tabela 6 – Estatísticas descritivas mensais das séries coletadas**

Nome	<i>n</i> (meses)	$\hat{\xi}$	Mín. (m <sup>3</sup> /s)	Máx. (m <sup>3</sup> /s)	$\hat{\rho}_1$
<b>São Bento</b>	901	1,94	7,3	218,9	0,55
<b>União da Vitória</b>	931	2,39	59,4	3641,6	0,51
<b>Porto Amazonas</b>	844	2,27	8,0	448,0	0,50
<b>Jangada do Sul</b>	715	2,07	2,8	183,6	0,44
<b>Rio Negro</b>	931	1,90	12,0	362,5	0,52
<b>Fazenda Maracanã</b>	713	2,75	0,9	77,9	0,41

*n* – tamanho da série;  $\hat{\xi}$  – coeficiente de assimetria;  $\hat{\rho}_1$  – coeficiente de correlação de *lag* um.

Passando para as estatísticas descritivas em escala anual, a Tabela 7 repete as informações anteriormente mostradas para a escala mensal, somente adicionando a vazão específica. Este último dado foi calculado dividindo-se as vazões médias históricas pelas respectivas áreas de drenagem (listadas na última coluna da Tabela 4), seguida das devidas adaptações nas unidades para se chegar a L/s/km<sup>2</sup>.

Das estações hidrométricas selecionadas, Fazenda Maracanã possui a menor afluência média, com 9,0 m<sup>3</sup>/s, e União da Vitória a maior com 468,4 m<sup>3</sup>/s. Em termos de tamanho do histórico, destacam-se São Bento, União da Vitória, Porto Amazonas e Rio Negro, com séries

<sup>8</sup> Para a determinação desses coeficientes, as séries históricas mensais dos citados rios foram consultadas no site do Operador Nacional do Sistema Elétrico (ONS) e estão disponíveis em [http://www.ons.org.br/operacao/vazoes\\_naturais.aspx](http://www.ons.org.br/operacao/vazoes_naturais.aspx) (acesso em 27/08/2015).

que contém 70 anos de registros ou mais. Novamente considerando os coeficientes de variação médios, eles possuem pouca variação entre os postos e valem 0,38. McMahon (1982) analisou vazões médias anuais de 126 rios ao redor do mundo, chegando a um  $c_v$  mediano de 0,25. Assim, diferentemente das séries em escala mensal, os coeficientes de variação não apresentam grandes distinções em comparação com outros rios.

**Tabela 7 – Estatísticas descritivas anuais das séries coletadas**

Nome	$n$ (anos)	$\hat{\mu}$ (m <sup>3</sup> /s)	Vazão específica (L/s/km <sup>2</sup> )	$\hat{\sigma}$ (m <sup>3</sup> /s)	$\hat{\xi}$	Mín. (m <sup>3</sup> /s)	Máx. (m <sup>3</sup> /s)	$\hat{\rho}_1$
São Bento	74	34,8	17,3	12,6	1,15	14,7	80,4	0,24
União da Vitória	77	468,4	19,3	184,4	1,08	183,6	1127,8	0,23
Porto Amazonas	70	67,9	18,5	25,4	1,02	27,3	151,9	0,26
Jangada do Sul	59	25,9	24,5	9,5	0,97	8,4	54,5	0,17
Rio Negro	77	67,7	20,0	25,5	0,90	25,5	154,4	0,29
Fazenda Maracanã	59	9,0	27,8	3,5	1,35	3,6	22,1	0,22

$n$  – tamanho da série;  $\hat{\mu}$  – média;  $\hat{\sigma}$  – desvio padrão;  $\hat{\xi}$  – coeficiente de assimetria;  $\hat{\rho}_1$  – coeficiente de correlação de lag um.

Os coeficientes de assimetria são menores do que os calculados para a escala mensal, contudo sua magnitude ainda sugere a adoção de transformações numéricas. Observa-se também que todos os postos apresentam uma correlação anual baixa, atingindo um máximo de 0,29 em Rio Negro.

### 3.4 Níveis mínimos anuais do rio Nilo

Previamente à aplicação do modelo SSS às estações fluviométricas supramencionadas, utilizou-se a série de níveis mínimos anuais do rio Nilo, medida na ilha de Rhoda em Cairo, no Egito, como auxiliar na depuração de sua implementação computacional. Como mencionado no capítulo anterior, a estratégia foi usar esta série devido à raridade de estudos que usam o modelo SSS em outras séries hidrológicas. Em compensação, existem estudos que aplicam o modelo SSS a essa série do rio Nilo (Koutsoyiannis, 2002), servindo de referência para a implementação do método.

As medições dos níveis do rio Nilo são provenientes de estruturas chamadas de nilômetros, construídas em épocas faraônicas e utilizadas até o início do século XIX. Existiam basicamente dois tipos de nilômetros, o mais simples consistindo de uma escada descendente até o leito do rio com graduações marcadas em suas paredes. O segundo tipo se trata de um pilar graduado submerso, construído dentro de uma estrutura maior que objetivava dispersar as flutuações naturais do rio. A conexão entre o poço onde ficava o pilar e o rio eram feitas por

túneis. Trazendo essas configurações para os dias de hoje, é possível fazer uma associação com régua linimétrica e linígrafos de boia (evidentemente sem o equipamento registrador) para o primeiro e o segundo tipos de nilômetro, respectivamente.

O nilômetro de Rhoda é do segundo tipo, com pilar octogonal construído em mármore e cuja graduação é dividida em 19 *cubits*. Cada *cubit* equivale a aproximadamente meio metro, assim o nilômetro tinha capacidade de medir variações de até aproximadamente 9,5 m (curiosamente, o nível considerado ideal pelos egípcios era de 16 *cubits*, ou 8 m). Havia três túneis de ligação com rio Nilo, construídos em diferentes níveis. As informações dadas neste parágrafo e no anterior foram coletadas de IWHA (2015), que oferece explicações adicionais sobre o funcionamento e a história do nilômetro de Rhoda.

A Figura 14 mostra três ilustrações acerca dos nilômetros. A primeira (Figura 14a) exibe um mapa da localidade do nilômetro de Rhoda, ao sul da ilha de mesmo nome. A segunda (Figura 14b) se trata de uma litogravura do artista escocês David Roberts, na qual é possível visualizar o nilômetro em operação. Finalmente, a terceira (Figura 14c) mostra o nilômetro nos dias de hoje, cuja estrutura é mantida para fins de turismo. Nessa fotografia é possível visualizar as graduações no pilar.

A série de níveis mínimos anuais do rio Nilo é o maior registro hidrológico existente no mundo, compreendendo o período entre os anos de 622 a 1922 (IWHA, 2015). No entanto, existem diversas falhas na série, de forma que sua maior sequência contínua varia de 622 a 1284, totalizando 663 anos (a série foi grafada anteriormente na Figura 4, página 73). Os dados foram compilados por Tousson (1935 *apud* IWHA, 2015), entretanto essa fonte está escrita em língua francesa e não foi encontrada pelo autor desta tese, impedindo análises mais aprofundadas sobre a forma de conversão dos dados do nilômetro de Rhoda. Felizmente, a série está disponível gratuitamente em <http://lib.stat.cmu.edu/S/beran> e suas estatísticas descritivas são mostradas pela Tabela 8.

Chama-se atenção à baixa magnitude do coeficiente de assimetria, sugerindo que a série possui uma distribuição próxima da normal. Dessa maneira, dispensa-se a aplicação de transformações numéricas para sua modelagem. Contrariamente, o coeficiente de autocorrelação de primeira ordem pode ser considerado significativo, principalmente por se tratar de uma série anual de níveis mínimos. Como será visto adiante, a estrutura de autocorrelação desta série é marcada por sua complexidade e persistência.

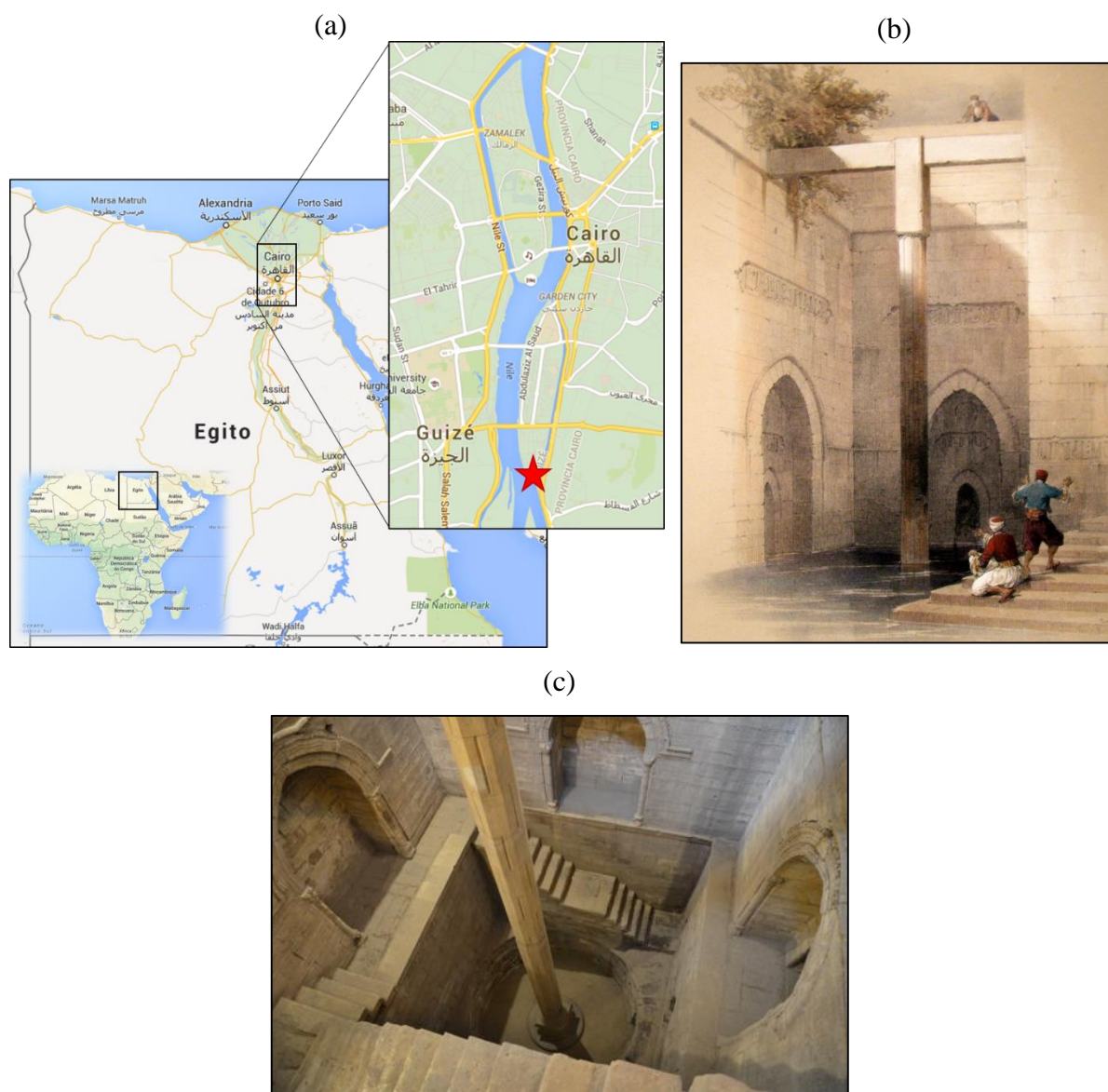


Figura 14 – Nilômetro da ilha de Rhoda, no Egito: (a) localização geográfica indicada pela estrela (Fonte: maps.google.com) (b) “The nilometer on the Isle of Rhoda, Cairo”, litogravura de David Roberts (série *Egypt & Nubia*, 1845-1849) que ilustra o funcionamento da estrutura (Fonte: albion-prints.com) (c) interior do nilômetro de Rhoda nos dias atuais (Fonte: cairobserver.com)

Tabela 8 – Estatísticas descritivas dos níveis mínimos anuais do rio Nilo, medidas no nilômetro da Ilha de Rhoda

Nome	$n$ (anos)	$\hat{\mu}$ (mm)	$\hat{\sigma}$ (mm)	$\hat{\xi}$	Mín. (mm)	Máx. (mm)	$\hat{\rho}_1$
Nilômetro de Rhoda	663	1148,1	88,7	0,24	935,0	1466,0	0,58

### **3.5 Resumo do capítulo**

Neste capítulo foram apresentados diversos aspectos relacionados à bacia hidrográfica do rio Iguaçu, escolhida como área de estudo desta tese. Iniciou-se com a caracterização geral em termos de áreas, principais rios, geologia, unidades aquíferas e uso do solo. Destacou-se que a bacia possui grandes barragens que interferem no regime dos rios, principalmente em reservatórios com capacidade de regularizar as vazões.

Na sequência, dedicou-se um item específico para descrever o clima local, dando ênfase aos regimes de chuva e temperatura. Mostrou-se que existe uma discrepância entre essas variáveis: enquanto as chuvas são regularmente distribuídas no ano, as temperaturas possuem sazonalidade bem definida. Alertou-se também para a influência do fenômeno El Niño na bacia.

Depois, elencaram-se as seis estações fluviométricas escolhidas para estudo, selecionadas de modo a não receber influência dos barramentos existentes e contendo poucas falhas em seus registros. Os dados de vazões médias se referem a históricos com comprimentos entre 59 e 77 anos e foram coletados em escala mensal. Exibiram-se estatísticas descritivas diversas para as escalas mensal e anual.

Por fim, foram dadas informações acerca da série de níveis mínimos anuais do rio Nilo, medida na ilha de Rhoda em Cairo, no Egito. Explicou-se o básico do funcionamento de um nilômetro, estrutura responsável pelo monitoramento do nível do rio Nilo desde épocas faraônicas até início do século XIX. Essa série foi utilizada no trabalho com a intenção de depurar o desenvolvimento computacional do modelo SSS.

## 4 ANÁLISE PRELIMINAR: ESTACIONARIEDADE DAS SÉRIES

No presente capítulo as técnicas descritas no primeiro capítulo para análise de tendências em séries hidrológicas foram aplicadas aos seis postos hidrométricos selecionados, com foco nas vazões médias em escala anual. Além do objetivo de estudar melhor as séries consideradas, buscou-se avaliar sua aplicabilidade ao modelo MAR(1). Como visto no subitem 1.4.2 (página 87), essa formulação é estacionária e requer séries com tal comportamento.

Em um primeiro momento as propriedades estatísticas das seis séries foram estudadas, principalmente no tocante à persistência. Na sequência, aplicaram-se os testes da regressão linear, t-Student, Wilcoxon, Pettitt, coeficiente de correlação de Spearman, Mann-Kendall e análises via transformada wavelet contínua, cujas definições matemáticas foram fornecidas na seção 1.2.1 (página 48). Os resultados apresentados estão acompanhados de comentários e justificativas.

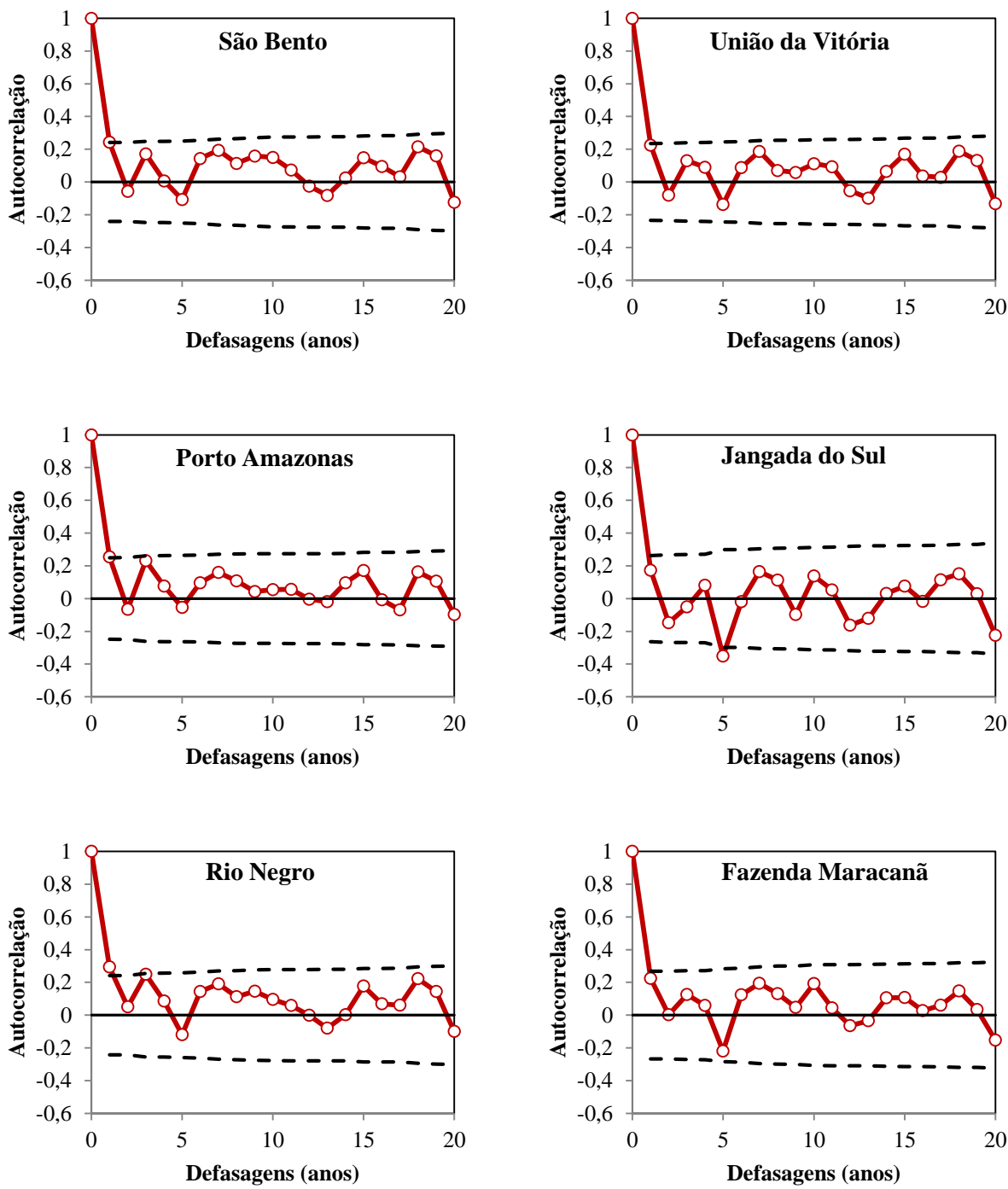
### 4.1 Persistência das séries

As estatísticas básicas das séries consideradas foram apresentadas anteriormente (Tabela 7 da página 127), contudo a estrutura de persistência das seis séries é explorada em maiores detalhes devido à importância para os testes estatísticos utilizados. Como explicado anteriormente, a presença de correlações significativas pode influenciar os resultados das inferências. Portanto, exibe-se a Figura 15 com as FAC das séries, limitando-se à defasagem máxima de 20 anos, na intenção de avaliar sua estrutura de persistência.

As linhas tracejadas horizontais representam o limite de 5% de significância, abaixo do qual a autocorrelação é considerada estatisticamente igual a zero. As seis séries apresentam um comportamento semelhante, o que era esperado por estarem localizadas em uma mesma região hidrológica. Com exceção de Rio Negro, todas apresentam uma FAC que se anula após o primeiro ano, evidenciando uma fraca estrutura de correlação interanual. A estação de Rio Negro, em particular, mostra uma correlação estatisticamente significativa na terceira defasagem, podendo sugerir que as vazões médias atuais recebem pequenas contribuições das vazões médias ocorridas três anos antes. O mesmo ocorre em Jangada do Sul para a defasagem de cinco anos. Nos dois casos, entretanto, essas correlações são muito fracas e podem ser consideradas espúrias.

Apesar de fracas, as autocorrelações com um ano de defasagem são estatisticamente significativas (estão acima das linhas que delimitam o nível de significância) para São Bento,

União da Vitória, Porto Amazonas e Rio Negro, exigindo que essa influência seja retirada da série previamente à aplicação dos testes. O próximo subitem explica esse e outros procedimentos efetuados para a realização dos estudos.



—○— Autocorrelação    - - - 5% Significância

**Figura 15 – FAC das seis séries estudadas**



#### 4.2 Premissas para execução das análises

A preparação das séries envolveu dois artifícios numéricos: (i) “branqueamento” das séries, através da aplicação da equação (1.22) (página 53), na intenção de retirar a influência da autocorrelação e (ii) transformação logarítmica convencional para testes que exigem a distribuição normal nos dados (regressão linear e t-Student). Cabe lembrar que a transformação logarítmica é um caso particular da transformação Box-Cox [equação (1.11), página 44] para quando  $\lambda = 0$ . Fez-se essa opção pois, quando  $\lambda \neq 0$ , a transformação inversa da série se torna imprecisa devido à exponenciação, podendo gerar valores errôneos. Em outras palavras, ao aplicar Box-Cox corre-se o risco de não ser possível obter a série original novamente, tornando-a inapta para a inserção nos modelos. A verificação da condição de normalidade das séries transformadas foi feita utilizando o teste de Lilliefors, que é baseado no teste de Kolmogorov-Smirnov (Lilliefors, 1967). Obtiveram-se os p-valores de 0,219 (SB); 0,261 (UV); 0,500 (demais postos), confirmando a hipótese de normalidade.

**Tabela 9 – Divisão das séries históricas para execução dos testes estatísticos**

Posto	Regressão Linear			t-Student e Wilcoxon		Spearman e Mann-Kendall
	Amostra 1	Amostra 2	Amostra 3	Divisão 1	Divisão 2	(amostra completa)
<b>SB</b>	1931-2004	1931-1975	1976-2004	1931-1967	1931-1975	1931-2004
				1968-2004	1976-2004	
<b>UV</b>	1931-2007	1931-1977	1978-2007	1931-1969	1931-1977	1931-2007
				1970-2007	1978-2007	
<b>PA</b>	1936-2005	1936-1975	1976-2005	1931-1970	1931-1975	1936-2005
				1971-2005	1976-2005	
<b>JS</b>	1946-2004	1946-1974	1975-2004	1946-1974	1946-1974	1946-2004
				1975-2004	1975-2004	
<b>RN</b>	1931-2007	1931-1977	1978-2007	1931-1969	1931-1977	1931-2007
				1970-2007	1978-2007	
<b>FM</b>	1946-2004	1946-1974	1975-2004	1946-1974	1946-1974	1946-2004
				1975-2004	1975-2004	

A divisão das subamostras para os testes t-Student e de Wilcoxon foi feita de duas formas: (i) central, na qual se confrontam duas subamostras de mesmo tamanho e (ii) separando-se os 30 últimos anos do histórico e testando-os contra a porção restante da série. A janela de 30 anos foi escolhida por ser considerada suficiente para a estabilização dos momentos

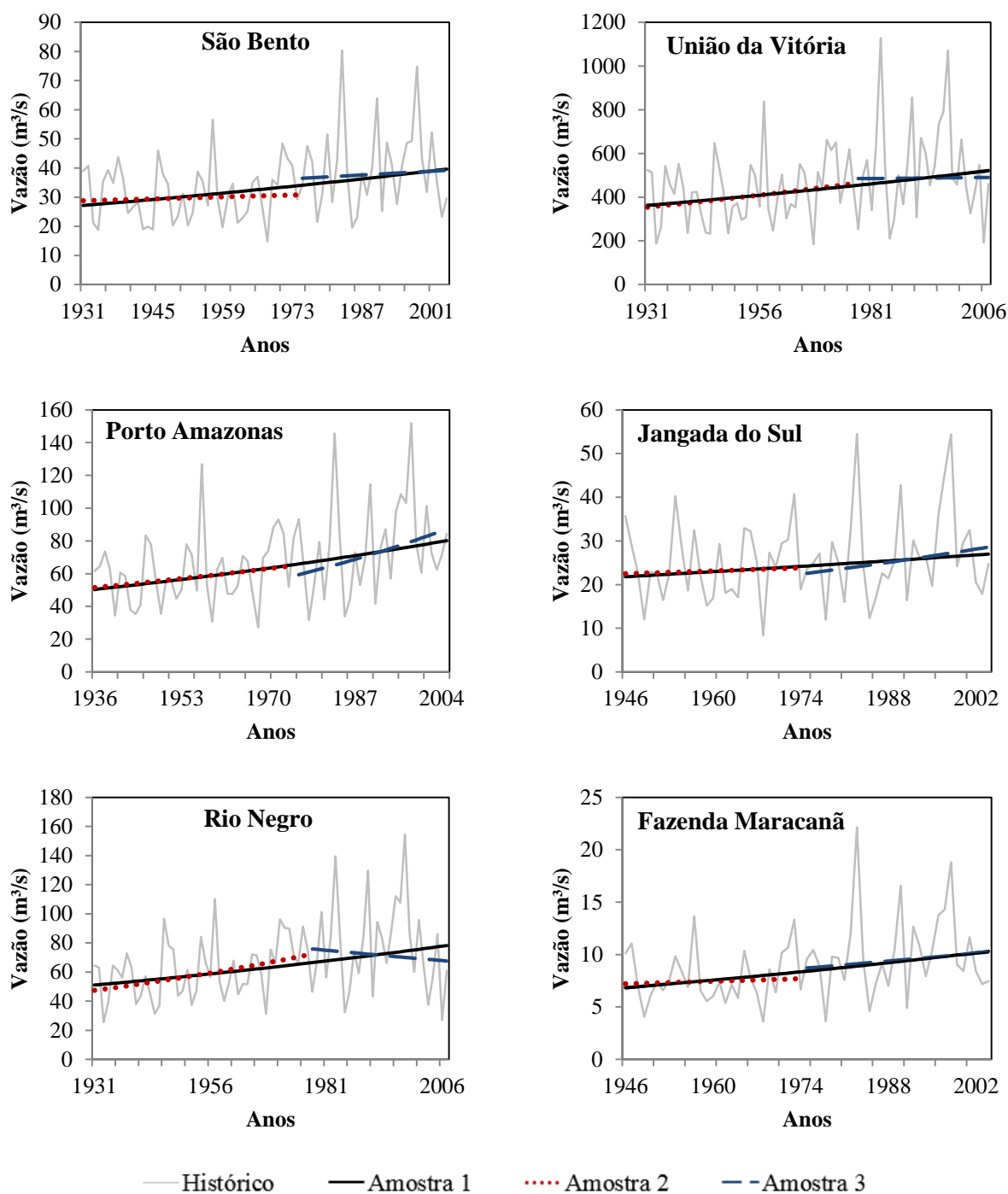
de primeira e segunda ordens de séries hidrometeorológicas e climáticas (Koutsoyiannis, 2006). Além disso, hipótese (ii) foi elaborada de forma a avaliar o possível efeito das alterações do uso do solo na vazão média. Segundo alguns trabalhos (Tucci, 2002; Batista *et al.*, 2009), tais alterações foram observadas no Estado do Paraná a partir da década de 1970. A fim de evidenciar graficamente essas possíveis alterações, os ajustes de regressão linear foram feitos também para a divisão de subamostras (ii), além da série completa. Vale mencionar que existem critérios alternativos aos adotados para dividir a amostra principal, como a técnica *Rescaled Adjusted Partial Sums* (Alexandre, Baptista e Naghettini, 2010), porém optou-se por manter somente as duas formas escolhidas nesses estudos. Evidentemente, para as séries de Jangada do Sul e Fazenda Maracanã as duas formas de divisão resultaram nas mesmas subamostras, pois as séries possuem histórico de 59 anos. Lembra-se que para os demais testes (Pettitt, coeficiente de correlação de Spearman e Mann-Kendall) a divisão em subamostras não é necessária, sendo que para o teste de Pettitt ela é um dos resultados obtidos. A Tabela 9 resume as amostras e subamostras consideradas em cada teste.

O teste de Mann-Kendall, em particular, foi aplicado tanto em sua forma tradicional quanto em formato sequencial, conforme explicado no subitem 1.2.1.6 (página 57). Todos os testes foram bilaterais, considerando como referência o nível de significância de 5%.

Além das inferências estatísticas, as séries também foram submetidas à análise espectral utilizando a *wavelet*-mãe de Paul, que possui boa resolução temporal (Rossi *et al.*, 2009). Para a obtenção dos diagramas, foi utilizado algoritmo elaborado por Torrence e Compo (1998), disponível em <http://paos.colorado.edu/research/wavelets/> (acesso em 05/02/2015).

### 4.3 Resultados

Os primeiros resultados a serem apresentados são os ajustes de tendências lineares nas séries, mostradas na Figura 16. É interessante observar que, apesar das estações hidrométricas estarem localizadas em uma mesma região hidrológica, as tendências ajustadas apresentam variações dentre as amostras utilizadas. Como elementos comuns a todas as séries, aponta-se apenas a aparente tendência ascendente da amostra 1, que engloba o histórico completo de cada uma das séries. Nota-se que os eventos hidrológicos de maior intensidade ocorridos nos anos de 1983 e 1998 foram responsáveis pela elevação no valor médio da porção recente da série, repercutindo na tendência positiva observada.

**Figura 16 – Regressões lineares ajustadas**

Entretanto, considerando as amostras 2 e 3 percebe-se algo bastante interessante principalmente na série de Rio Negro. Apesar de diversas referências apontarem um crescimento das vazões médias a partir da década de 1970, essa série apresenta uma situação oposta na amostra 3, indicando uma (fraca) tendência de decréscimo das vazões. O aumento das vazões pós década de 1970 só é evidente em Porto Amazonas e Jangada do Sul, cuja declividade do ajuste supera as retas das amostras 1 e 2. Para Porto Amazonas, o crescimento

das vazões poderia ser explicado pelo desenvolvimento urbano ocorrido nessa porção da bacia do rio Iguaçu, que fica próxima à capital e região metropolitana. Entretanto a grande área de drenagem desse posto (3662 km<sup>2</sup>, ver Tabela 4, página 123) não permite tirar conclusões definitivas acerca dessa questão, exigindo análises mais aprofundadas.

Em União da Vitória, o ajuste na amostra 3 não resulta em tendência alguma, o que pode ser explicado pelo comportamento oposto observado entre Porto Amazonas e Rio Negro, duas estações fluviométricas cujas contribuições chegam em União da Vitória. O aumento em Porto Amazonas é contrabalanceado pelo decréscimo em Rio Negro, eliminando indícios de tendência em União da Vitória. Cabe lembrar que São Bento também contribui para União da Vitória, porém sua tendência (positiva) é muito suave e exerce pouca influência a jusante.

Os resultados dos testes estatísticos são apresentados na Tabela 10 (exceto para o teste de Pettitt, apresentado separadamente), na qual os p-valores em negrito indicam significância estatística. Ainda considerando a regressão linear, nota-se que nos seis postos a presença de uma tendência (ascendente ou descendente) na amostra 3 é rejeitada, embora o mesmo não pode ser dito para a amostra 1 (exceto Jangada do Sul). Estes resultados evidenciam a dependência das análises de regressão linear da escolha das amostras.

**Tabela 10 – Resultados dos testes estatísticos em termos de p-valores**

Posto	Regressão Linear			t-Student/ Wilcoxon		Spearman/ Mann-Kendall
	Amostra 1	Amostra 2	Amostra 3	Divisão 1	Divisão 2	(amostra completa)
SB	<b>0,005</b>	0,670	0,746	<b>0,000/</b>	<b>0,000/</b>	0,051/
				<b>0,000</b>	<b>0,000</b>	<b>0,045</b>
UV	<b>0,016</b>	0,126	0,967	0,376/	0,261/	0,054/
				0,194	0,172	0,058
PA	<b>0,001</b>	0,191	0,103	0,374/	0,136/	<b>0,012/</b>
				0,179	0,094	<b>0,012</b>
JS	0,191	0,803	0,307	0,760/	0,760/	0,176/
				0,638	0,638	0,193
RN	<b>0,003</b>	<b>0,009</b>	0,661	<b>0,010/</b>	<b>0,003/</b>	<b>0,023/</b>
				<b>0,002</b>	<b>0,001</b>	<b>0,017</b>
FM	<b>0,011</b>	0,744	0,505	<b>0,015/</b>	<b>0,015/</b>	<b>0,033/</b>
				<b>0,005</b>	<b>0,005</b>	<b>0,043</b>

Os resultados dos testes t-Student e de Wilcoxon se mostraram consistentes, rejeitando a hipótese de igualdade de médias e medianas para as séries de Rio Negro, São Bento e Fazenda

Maracanã, respectivamente. De fato, ao analisar a Figura 16, pode-se notar uma quebra na tendência central das três séries. A mesma consistência foi observada nos testes não paramétricos de Spearman e Mann-Kendall, diferindo apenas na série de São Bento. Contudo, por considerar as amostras completas em suas formulações, ambos não captam a quebra de tendência observada em Rio Negro. Dessa maneira, um hidrólogo que baseie suas conclusões somente nesses testes não notaria essa mudança de padrão ocorrida no final da década de 1970. Essa situação pode ser considerada bastante preocupante ao lembrar que esses testes são os mais indicados pela literatura especializada (ver referências nas seções 1.2.1.4 e 1.2.1.6, páginas 55 e 57).

Os resultados do teste de Pettitt são mostrados na Tabela 11, que exhibe o ano no qual as quebras foram identificadas e a significância estatística dessas detecções.

**Tabela 11 – Resultados do teste de Pettitt**

<b>Nome</b>	<b>Ano de quebra</b>	<b>Significância</b>
<b>São Bento</b>	1968	<b>0,002</b>
<b>União da Vitória</b>	1968	<b>0,009</b>
<b>Porto Amazonas</b>	1968	<b>0,002</b>
<b>Jangada do Sul</b>	1964	0,281
<b>Rio Negro</b>	1968	<b>0,000</b>
<b>Fazenda Maracanã</b>	1970	<b>0,004</b>

Com exceção de Jangada do Sul, todas as estações tiveram quebras significativas detectadas em suas séries. Chama a atenção que o teste indicou o mesmo ano de 1968 para quatro dos cinco postos com quebras significativas, ficando Fazenda Maracanã com 1970. Comparando esses resultados com os dos outros testes baseados em subamostras, nota-se uma inconsistência nos postos de União da Vitória e Porto Amazonas, nas quais, diferentemente do teste de Pettitt, os testes t-Student e Wilcoxon falharam em rejeitar a hipótese nula. Esses resultados podem ser considerados contraditórios principalmente no caso da Divisão 1, cujas datas limites de separação das subamostras foram 1969 e 1970 para União da Vitória e Porto Amazonas, respectivamente. Claramente, esses anos estão muito próximos aos detectados pelo teste de Pettitt, evidenciando a citada contradição.

A Figura 17 exhibe os resultados do teste de Mann-Kendall, em sua forma sequencial. A curva MK1 corresponde ao cálculo da estatística do teste feita nas amostras ordenadas cronologicamente, enquanto que a curva MK2 é o mesmo cálculo empregado nas amostras com

ordem inversa. Exibem-se também os intervalos de confiança e as próprias séries históricas. Nesse teste, a variação no sentido das possíveis tendências é analisada através do sinal da estatística MK na curva MK1. A curva MK2, por sua vez, exerce um papel secundário servindo somente para determinação dos pontos críticos nas séries.

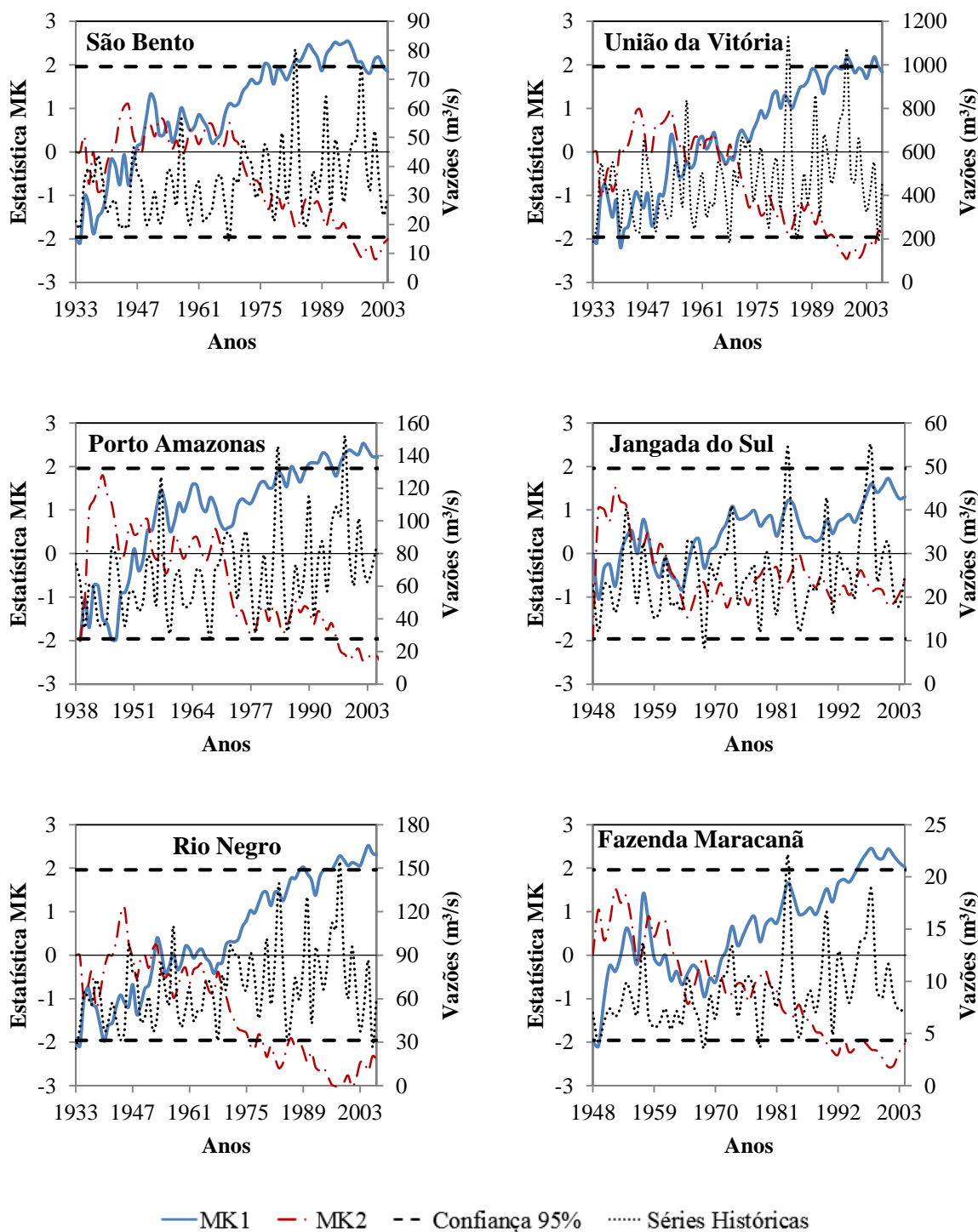


Figura 17 – Resultados do teste de Mann-Kendall sequencial

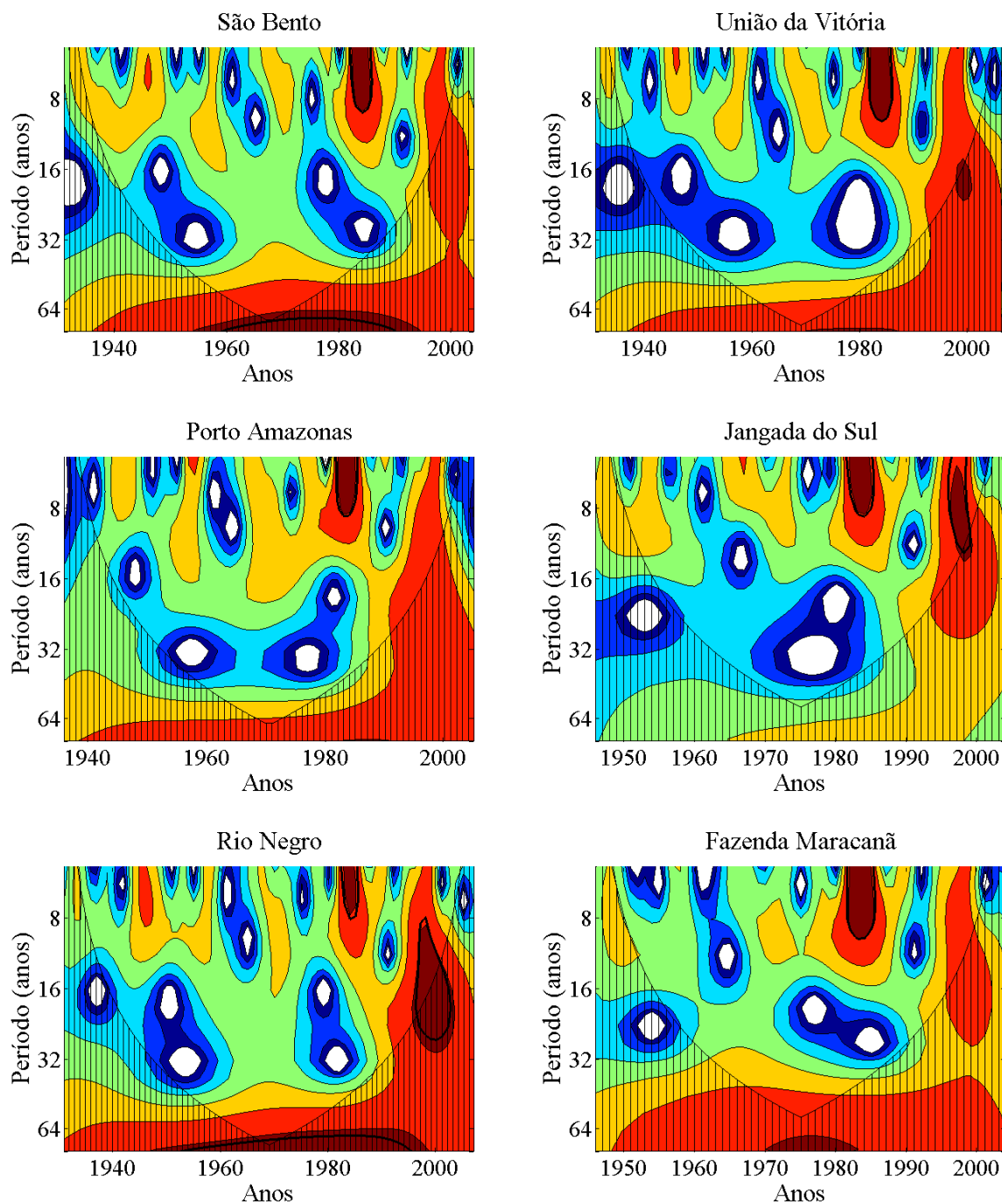
Dessa maneira, focando a análise na curva MK1, nota-se que todas as séries se iniciam com uma tendência negativa ( $MK < 0$ ). Em União da Vitória e Rio Negro, a tendência passa a ser positiva ( $MK > 0$ ) no início da década de 1950, enquanto que em São Bento e Porto Amazonas isso ocorre cerca de cinco anos antes. A partir desse ponto, as tendências positivas seguem até o fim das respectivas séries, tornando-se estatisticamente significativas no final da década de 1980, exceto em União da Vitória, na qual somente dois anos (1995 e 2005) ultrapassam em pouco o limite de significância. Em Jangada do Sul e Fazenda Maracanã, as tendências passam a ser positivas também no início da década de 1950, porém rapidamente voltam a ser negativas até nova ascensão no começo da década de 1970. Especificamente em Jangada do Sul nenhuma curva ultrapassa os limites de significância.

Os resultados mostrados, evidentemente, estão em concordância com os p-valores do teste de Mann-Kendall tradicional mostrados na Tabela 10. Com exceção de Jangada do Sul, o que se percebe claramente é que os picos de 1983, 1990 e 1999 (primeiro e o último associados ao fenômeno El Niño, ver Cavalcanti *et al.*, 2009, p. 379), são os motivadores dos valores positivos de MK1. Ainda assim, em todas as séries os cruzamentos das curvas MK1 e MK2 não são estatisticamente significativos, impossibilitando conclusões acerca de quebras nas séries.

Por fim, as transformadas wavelet contínuas são exibidas na Figura 18. Visando facilitar as análises, optou-se por exibir os eixos das ordenadas em termos de períodos (anos) ao invés de frequências (Hertz) como é comum. Nos diagramas, os sinais mais fortes são representados através de tons escuros, enquanto que os de menor influência recebem tons claros. Além disso, a área hachurada demarca o chamado cone de influência, abaixo do qual os resultados são distorcidos e devem ser desconsiderados ou analisados com cautela (Santos, Freire e Torrence, 2013). Por fim, os contornos em negrito apontam significância estatística (5%).

Os sinais predominantes para as seis séries se localizam no entorno do período de oito anos, manifestando-se com maiores intensidades uma vez a cada 10 anos. Os períodos de 16 e 32 anos, por outro lado, exibem os sinais mais fracos em todos os casos. Em termos estatísticos, os sinais significantes são associados ao ano de 1983 para todas as séries, manifestando-se nas baixas bandas de oito anos ou menos. Lembrando-se da relação inversa entre período e frequência, fica evidenciado que esse sinal é de alta frequência, o que segundo Rossi *et al.* (2009) é característico de eventos extremos. Dessa maneira, conclui-se que esse forte sinal em todas as séries está diretamente relacionado à grande cheia de 1983 observada na região. Ainda, Fazenda Maracanã e Rio Negro possuem sinais significantes também para o ano de 1998, porém ocorrendo na banda de 16 anos.

De um modo geral, em todas as séries a distribuição dos tons antes e depois do ano de 1970 possui certa distinção, com aumento de tons escuros na porção mais recente da série. Esse padrão pode ser interpretado como um crescimento na intensidade do sinal de vazão média, sugerindo uma tendência positiva (Pasquini e Depetris, 2007; Rossi *et al.*, 2009). É interessante notar que, apesar de terem abordagens diferentes, os resultados obtidos com o teste de Mann-Kendall sequencial e as transformadas wavelet contínuas são bastante consistentes.



**Figura 18 – Resultados da transformada wavelet contínua. Tons mais escuros indicam sinais mais fortes e vice-versa**



Relevando-se todos os resultados apresentados para as seis séries, tomar uma decisão quanto à condição de (não) estacionariedade das séries não é uma tarefa simples. Avaliando-se as FAC (Figura 15) e considerando-se a teoria dos modelos ARIMA, nenhuma série apresenta comportamento não estacionário, dispensando-se a exigência de diferenciações. No entanto, os resultados de alguns dos testes estatísticos mostram que as séries exibem uma tendência (linear), exigindo interferências do hidrólogo para torná-las próprias para a modelagem. Observou-se (Tabela 10) que o suposto comportamento não estacionário não é unânime para todos os testes, mesmo considerando apenas uma mesma série.

Em termos físicos, muito se disse sobre a influência das alterações no uso do solo pós década de 1970, no entanto sua conexão com o aparente aumento das vazões médias não pode ser feita diretamente, pois nos anos de 1982-83 e 1998-99 observaram-se manifestações de grande intensidade do fenômeno El-Niño (Cavalcanti *et al.*, 2009, p. 379). A coincidência dessas ocorrências com os picos de cheias registrados nesses anos é notável, impossibilitando apontar a alteração do uso do solo como causa única para as alterações no regime de vazões.

Em meio aos gráficos dos ajustes das regressões nas séries (Figura 16), Mann-Kendall sequencial (Figura 17) e transformada wavelet contínua (Figura 18), parece bastante plausível a hipótese de flutuações aleatórias (e.g. estocásticas) nas medidas centrais das séries, em detrimento de uma tendência linear com sentido definido. Nesse contexto, atribuir à série uma função determinística para representar uma suposta tendência linear não possui sentido físico, podendo ser considerada como uma simplificação para a modelagem.

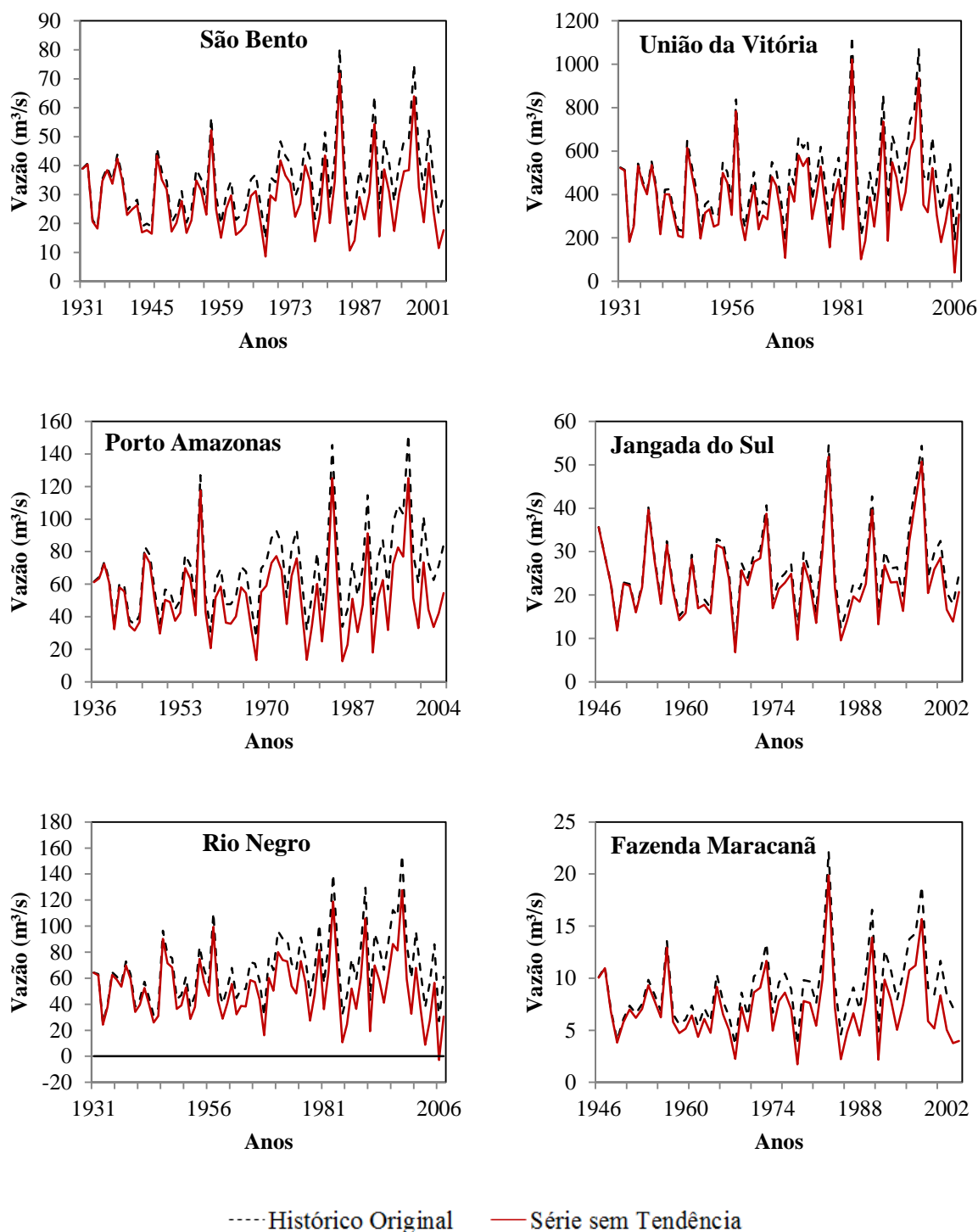
#### 4.4 Remoção do comportamento não estacionário

Em termos práticos os comportamentos não estacionários devem ser removidos das séries antes de sua aplicação em formulações estacionárias, como é o caso do modelo MAR(1). Essa tarefa pode ser resolvida através de algumas técnicas (ver seção 1.4.2.2, na página 89); para esta tese, optou-se pela remoção da tendência linear através do estimador não paramétrico de Sen. A Tabela 12 exhibe os coeficientes obtidos após aplicação da equação (1.85) (página 90).

**Tabela 12 – Coeficientes do estimador não paramétrico de Sen**

Nome	São Bento	União da Vitória	Porto Amazonas	Jangada do Sul	Rio Negro	Fazenda Maracanã
<b><i>b</i></b>	0,161	1,988	0,424	0,069	0,394	0,059

Um comparativo entre as séries originais e depois de terem sua tendência removida é mostrado na Figura 19. Das seis séries estudadas, Rio Negro apresentou um número negativo depois da remoção da tendência [aplicação da equação (1.86), página 90]. Observando a série



**Figura 19 – Séries com tendências lineares removidas. Em Rio Negro o procedimento resultou em vazão negativa em 2006**

original e as regressões mostradas na Figura 16, nota-se que o valor de vazão média anual no ano de 2006 é comparável ao mínimo da série inteira. Ao aplicar o coeficiente do estimador de Sen, obtém-se um valor negativo para esse ano.

Jangada do Sul pode ser considerada a série cuja remoção de tendência causou a menor diferença antes e depois do procedimento, mesmo que o valor de seu coeficiente  $b$  não tenha sido o mínimo. Fazenda Maracanã teve seu coeficiente menor, contudo a diferença entre as séries original e transformada é mais evidente. Essa aparente contradição pode ser explicada analisando novamente os resultados dos testes estatísticos; para Jangada do Sul, todas as inferências consideradas falharam em rejeitar a hipótese nula de comportamento estacionário, o que não ocorreu em Fazenda Maracanã. Por esse motivo, a remoção de tendência para Jangada do Sul poderia ser dispensada. Para as demais séries, a diferença antes e depois da transformação é notável.

#### 4.5 Resumo do capítulo

As técnicas estatísticas e de análise espectral detalhadas na seção 1.2.1 (página 48) foram colocadas em prática no presente capítulo. Além de ajustes de regressão simples com testes de significância sobre o coeficiente obtido, aplicaram-se os testes t-Student, Wilcoxon, Pettitt, Spearman e Mann-Kendall (tradicional e sequencial) e análises via transformada wavelet contínua sobre as séries consideradas no trabalho. Na literatura essa região é conhecida por sofrer alterações no uso do solo a partir de meados dos anos 1970 e, conseqüentemente, observar um suposto aumento nas vazões médias de seus rios.

Os resultados apresentados foram um tanto contraditórios, de forma que não se pode afirmar categoricamente que as vazões médias sofreram um aumento devido às alterações de uso do solo. Em todas as séries, os anos de 1983 e 1998 exibiram picos de cheias cuja intensidade elevou a média histórica, dando a impressão de uma tendência positiva nas vazões anuais. Porém, nesses anos ocorreram manifestações de grande intensidade do fenômeno El Niño, colocando em dúvida a real influência das alterações no uso do solo sobre as séries estudadas. Para complicar ainda mais, a série de Rio Negro apresentou uma tendência (não significativa) de redução nas vazões a partir do fim da década de 1970, ao contrário do que se esperava.

Argumenta-se que a tomada de decisões sobre a condição de estacionariedade das séries é uma tarefa de grande complexidade, mesmo sabendo que as técnicas utilizadas são altamente

recomendadas na literatura. A escolha do período histórico para análise exerce um papel determinante nos resultados obtidos.

O capítulo foi finalizado com a remoção do comportamento não estacionário das séries, visando sua aplicação em modelos estacionários como o MAR(1). Utilizou-se o estimador não paramétrico de Sen, através do qual Rio Negro obteve um valor negativo após o procedimento de remoção da tendência linear. Considerando todas as séries, apenas Jangada do Sul poderia ser submetida em sua forma original a modelos estacionários.

## 5 RESULTADOS E DISCUSSÕES

Neste capítulo são apresentados os resultados da aplicação dos modelos estocásticos na geração de séries sintéticas de vazões médias. Em um primeiro momento, serão mostrados detalhes considerados relevantes no ajuste de cada modelo. Na sequência são mostradas análises comparativas utilizando as métricas e critérios estatísticos definidos na seção 2.3 (página 112).

### 5.1 Detalhes do ajuste dos modelos estocásticos

#### 5.1.1 Modelo SSS

Levantaram-se alguns pontos de interesse no ajuste do modelo SSS via algoritmo SMA, os primeiros referentes aos seus parâmetros internos: a estimação do expoente de Hurst, por ser este o parâmetro principal do modelo, e o comportamento dos pesos  $\theta_j$  existentes em sua formulação, incluindo sua relação com a quantidade  $q$  de parâmetros utilizados.

Passando para a esfera prática do uso do modelo SSS e relevando sua limitada literatura, usa-se em um primeiro momento a série de níveis mínimos anuais do rio Nilo medida na Ilha de Rhoda para validar a implementação computacional. Ressalta-se que existem trabalhos (Koutsoyiannis, 2002) com estimativas de parâmetros específicos que auxiliaram na depuração da programação. Uma vez considerada correta, passa-se à aplicação para as séries da bacia do rio Iguaçu, sendo possível avaliar as diferenças devidas às séries de menor tamanho.

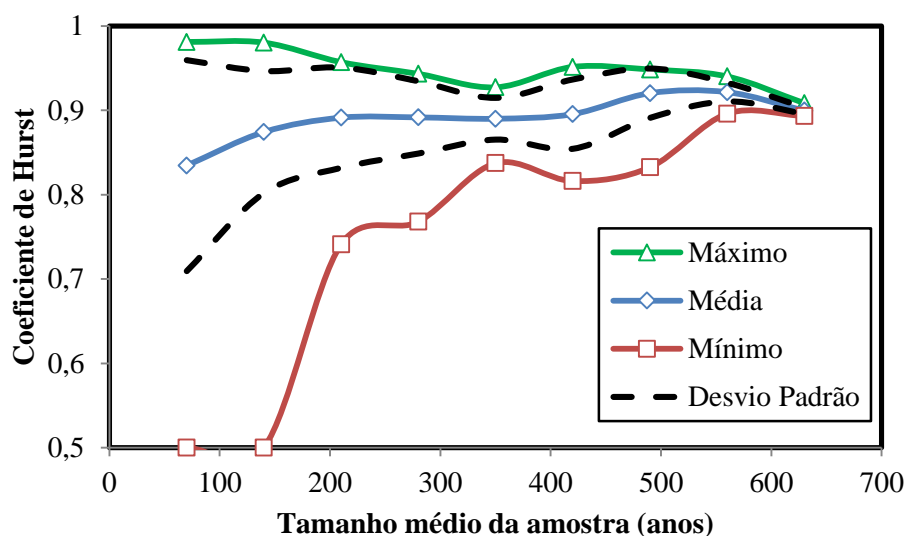
##### 5.1.1.1 Estimação do expoente de Hurst

A primeira preocupação no ajuste do modelo SSS é com a estimativa do expoente de Hurst. Koutsoyiannis (2002, 2006) apresenta o resultado para a série de níveis mínimos anuais do rio Nilo, no qual obtém-se a  $H = 0,85$  utilizando o método da Máxima Verossimilhança. Em Koutsoyiannis (2002), o valor estimado para a mesma série foi de  $H = 0,88$  através do método de mínimos quadrados baseados na variância (LSV). No presente trabalho, aplicando a mesma série ao algoritmo LSSD chegou-se ao valor de  $H = 0,89$  após oito iterações. Acredita-se que essa distinção possa ter ocorrido devido aos diferentes métodos de estimação utilizados. Dessa maneira, considera-se o valor encontrado neste trabalho como perfeitamente aceitável, validando a utilização do algoritmo LSSD para as demais séries.

Relevando-se o excepcional comprimento da série histórica do rio Nilo e a não existência de outra série hidrológica comparável, propôs-se investigar uma possível relação

entre o tamanho da amostra e a magnitude de  $H$ . Duas questões foram elaboradas: (i) o período histórico usado para estimação do  $H$  é importante? (ii) o tamanho da amostra e a magnitude de  $H$  apresentam uma relação monotônica crescente, ou seja, quanto maior a amostra maior o expoente  $H$ ? Evidentemente essa averiguação tem relação direta com a estimação de  $H$  para as séries de menor comprimento, como o caso da bacia do rio Iguaçu.

Para responder às questões levantadas, montou-se um experimento simples usando a série do rio Nilo, separando-a em subamostras com tamanhos 70, 140, 210, 280, 350, 420, 490, 560 e 633 anos, esta última sendo a série completa. A escolha do tamanho inicial de 70 anos foi feita por ser esta a ordem de grandeza das demais séries hidrológicas consideradas na tese. Para o cálculo dos expoentes de Hurst, trabalhou-se com janelas móveis da seguinte maneira: considerando o tamanho inicial de 70 anos, montaram-se séries dos anos 1 à 70, 2 à 71, 3 à 72 e assim por diante até os anos 563 à 663, totalizando 594 séries. Os expoentes de Hurst foram estimados para todas, extraíndo-se os valores mínimos, médios, máximos e os desvios-padrão. Esse procedimento foi repetido para os demais tamanhos de amostra definidos no início do parágrafo. Os resultados são mostrados na Figura 20.



**Figura 20 – Valores mínimos, médios, máximos e desvios-padrão do expoente de Hurst obtidos para diversas subamostras da série do Nilômetro de Rhoda**

A maior evidência revelada pelo gráfico é a progressiva redução da distância entre valores mínimos e máximos conforme o tamanho da subamostra cresce, o mesmo acontecendo com os desvios-padrão. Nos primeiros dois tamanhos de subamostra (70 e 140 anos), as estimativas mínimas chegaram ao limite inferior de  $H = 0,5$  e as máximas se aproximaram da unidade. A amplitude do intervalo delimitado pelos desvios-padrão sugere que para esses comprimentos há uma grande incerteza associada.

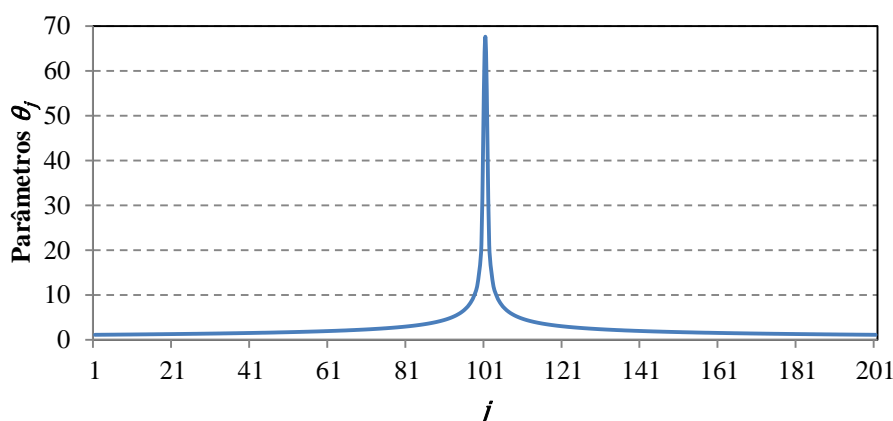
Com o aumento do tamanho das subamostras, as estimativas mínimas e máximas se aproximaram do valor médio, porém novo afastamento ocorreu para a subamostra de 420 anos. Depois disso, os valores convergiram para a média. A incerteza das estimativas seguiu padrão semelhante, reduzindo-se com o tamanho das subamostras até o tamanho de 350 anos, com leve aumento para 420 anos e subsequente redução até o comprimento final de 630 anos.

Essa análise provê informações suficientes para responder às duas questões levantadas. Dessa maneira, a resposta para a questão (i) é afirmativa, visto que grande variabilidade é observada principalmente entre as estimativas das subamostras de menor tamanho. Por outro lado, a resposta para a questão (ii) é negativa, pois não se observou um comportamento monotônico crescente em relação ao tamanho das subamostras. Isso significa que não se pode afirmar que amostras menores forneçam estimativas mais baixas do expoente de Hurst e vice-versa. A título de comparação, Szolgayova *et al.* (2014) estudaram possíveis variáveis que influenciam na magnitude de  $H$  utilizando análises via correlações. Uma dessas variáveis é, justamente, o tamanho das amostras consideradas; como conclusão, não foram encontradas correlações significativas entre tamanho das amostras e o expoente de Hurst.

Para as séries da bacia do rio Iguaçu, a resposta obtida para a questão (ii) é de grande importância, pois elas têm tamanhos reduzidos em comparação com a série de níveis mínimos anuais do rio Nilo. Dessa maneira, não há evidências de que os expoentes de Hurst estimados para as séries estudadas sejam tendenciosos devido aos seus menores históricos. Ainda assim, deve-se apontar que grandes incertezas foram identificadas nas estimativas do expoente de Hurst em subamostras de tamanho pequeno.

#### 5.1.1.2 Estrutura dos coeficientes do algoritmo SMA

Mesmo sendo componentes internos ao algoritmo SMA, é interessante analisar a estrutura dos coeficientes  $\theta_j$  a fim de melhor entender as características do modelo. Como o próprio nome do algoritmo sugere, a série  $\theta_j$  é simétrica em relação a um valor central  $\theta_0$  e decai progressivamente com o aumento dos valores de  $j$ . Utilizando novamente a série de níveis mínimos anuais do rio Nilo e arbitrando inicialmente um valor de  $q = 100$ , chega-se a uma série  $\theta_j$  com 201 elementos, cuja estrutura é mostrada na Figura 21.



**Figura 21 – Parâmetros  $\theta_j$  para os níveis mínimos anuais do rio Nilo, usando  $q = 100$**

No gráfico, o valor  $\theta_0$  equivale ao elemento  $j = 101$  e foi estimado em 67,7. O primeiro valor subsequente (e.g.  $\theta_{100} = \theta_{102}$ ) sofre uma queda significativa de valor, chegando a 21,2; o segundo valor chega a 12,4 e a partir do terceiro ponto o valor fica abaixo de 10. Desse modo, entende-se que os primeiros elementos da série de parâmetros são os que exercem a maior influência na determinação dos elementos das séries sintéticas. Essa visualização auxilia a determinação da quantidade  $q$  de pesos a ser considerada para a modelagem.

Outra forma de analisar a estrutura dos parâmetros em questão é através do último valor da série, ou seja,  $\theta_{-q} = \theta_{2q+1}$ . Quanto maior o valor de  $q$ , mais eles se aproximam de zero, sem, contudo, chegar a esse valor. Ainda assim, serve como uma informação da precisão desejada para o modelo. Fez-se novo ensaio usando a série do rio Nilo considerando diversos tamanhos  $\theta_{|q|}$ , com  $q$  variando entre 10 e  $10^6$ , com crescimento em ordens de 10. O resultado dos últimos valores das séries é mostrado na Tabela 13.

**Tabela 13 – Valores de  $\theta_{-q} = \theta_{2q+1}$  para diversas quantidades  $q$**

$q$	10	$10^2$	$10^3$	$10^4$	$0,325 \times 10^5$	$10^5$	$10^6$
$\theta_{-q} = \theta_{2q+1}$	4,5779	1,1302	0,2792	0,0690	0,0677	0,0170	0,0042

A partir de  $q = 10^4$  os valores de  $\theta_{-q} = \theta_{2q+1}$  se tornam mínimos, o que sugere que esse é um limite que pode ser considerado para o uso nessa série. A Tabela 13 ainda traz o valor para  $q = 0,325 \times 10^5$ , que é o resultado do critério matemático mostrado pela equação (1.58) (página 79), usando como tolerância o valor de  $\beta = 0,001$ . Dessa maneira, conclui-se que há uma harmonia no critério matemático em relação ao comportamento da série  $\theta_j$ .



Vale lembrar que, independentemente da quantidade  $q$ , os primeiros valores da série serão exatamente os mesmos. A diferença está somente na maior quantidade de termos a serem considerados a cada nova iteração do modelo.

### 5.1.1.3 Ajuste do modelo SSS à série de níveis mínimos anuais do rio Nilo

Ainda utilizando a série do rio Nilo, o modelo SSS foi estruturado e aplicado na geração de 1000 séries sintéticas, cada uma com os mesmos 663 anos de comprimento da série histórica. Este exercício trata da última etapa da implementação do algoritmo SMA antes de sua aplicação às séries da bacia do rio Iguaçu. Por se tratar de uma validação, os critérios empregados para análise foram somente as estatísticas de curto termo descritas na seção 2.3.

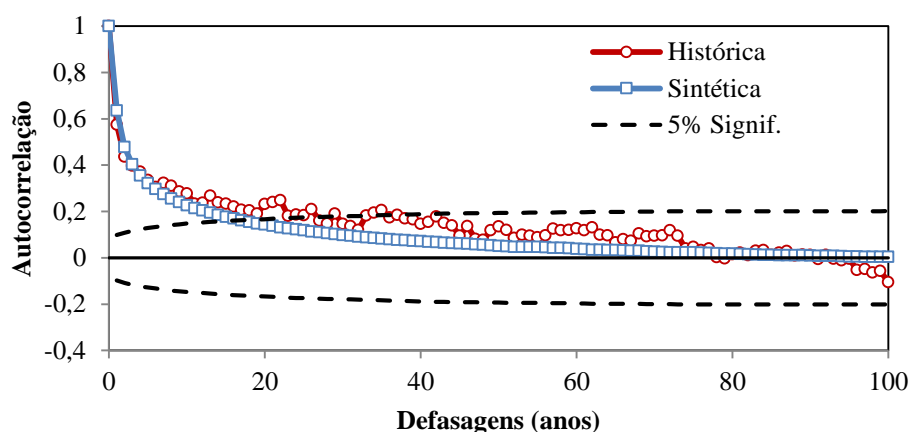
Os resultados obtidos são mostrados na Tabela 14. Para as séries sintéticas, as estatísticas se referem à média das séries geradas.

**Tabela 14 – Estatísticas de curto termo das séries sintéticas de níveis mínimos anuais do rio Nilo, geradas pelo modelo SSS**

Série	$\hat{\mu}$ (mm)	$\hat{\sigma}$ (mm)	$\hat{\xi}$	Mín. (mm)	Máx. (mm)	$\hat{H}$
<b>Histórica</b>	1148	89	0,24	935	1466	0,89
<b>Sintética</b>	1148	79	0,33	944	1432	0,89

A análise da Tabela 14 mostra que o modelo SSS foi bastante fiel às estatísticas históricas de curto termo para a série do rio Nilo. Algumas imprecisões foram anotadas para o desvio padrão (subestimação) e para o coeficiente de assimetria (superestimação), porém são ambas de baixa magnitude. Atenta-se à boa reprodução dos valores mínimos e máximos, estatísticas de difícil reprodução por modelos estocásticos. Por fim, destaca-se a perfeita reprodução do expoente de Hurst, principal objetivo do modelo SSS.

A estrutura de persistência é mostrada através da FAC mostrada na Figura 22, também de forma comparativa entre séries histórica e sintéticas. Novamente, as autocorrelações das séries sintéticas são as médias de todas as séries geradas.



**Figura 22 – FAC das séries sintéticas de níveis mínimos anuais do rio Nilo, geradas pelo modelo SSS**

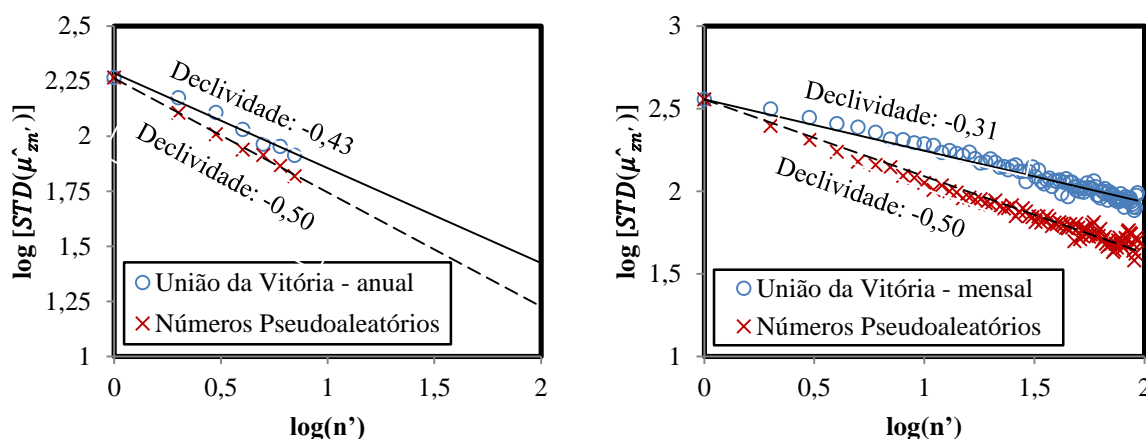
Chama-se a atenção ao decaimento extremamente lento da FAC dessa série, evidenciando a complexa estrutura de persistência existente. Considera-se que o modelo SSS foi capaz de reproduzir essa estrutura, com alguns desvios observados principalmente a partir da defasagem de 20 anos. Por outro lado, percebe-se também que esse é o momento no qual as FACs das duas séries se tornam significativamente nulas. Vale lembrar que o algoritmo SMA é uma simplificação do original FGN, (ver seção 1.4.1.1.2, página 76) e, portanto, as poucas imprecisões observadas na FAC das séries sintéticas podem ser reflexo dessa aproximação. Em termos práticos, esse resultado é suficiente.

Dado o exposto, considera-se o modelo SSS validado e pronto para aplicação nas séries da bacia do rio Iguaçu. Maiores detalhes da implementação para estas séries são dados na sequência.

#### 5.1.1.4 Parâmetros do modelo SSS para a bacia do rio Iguaçu

Sabendo do ineditismo da aplicação do modelo SSS na bacia do rio Iguaçu, é interessante estudar primeiramente as evidências da atuação e fenômenos de múltiplas escalas temporais nas séries. Um modo de se fazer isso é por meio os climacogramas, tais como mostrados na introdução da seção 1.4.1.1. A fim de evitar uma desnecessária exibição de gráficos para todas as séries coletadas, escolheu-se somente União da Vitória como representativa da área de estudo, por ter séries de maiores comprimento e magnitude.

Os climacogramas para as séries anuais e mensais em União da Vitória são mostradas na Figura 23.



**Figura 23 – Climacogramas de União da Vitória, nas escalas anual (direita) e mensal (esquerda)**

Para a escala anual, o climacograma não fornece informações suficientes para se tirar uma conclusão definitiva acerca da manifestação das múltiplas escalas temporais. Isso decorre do fato de que o comprimento da série é limitado, possibilitando a montagem de apenas subséries com escala máxima de  $n' = 7$  anos. A linha de tendência foi estendida para se tentar uma melhor visualização, porém as diferenças para com os números pseudoaleatórios são muito sutis. Por outro lado, na escala mensal as diferenças são mais evidentes. O maior tamanho do histórico permitiu montar subséries até a escala de  $n' = 93$  meses, evidenciando a distinção entre elas e a série de números pseudoaleatórios. Por meio dessa análise inicial, sugere-se que a escala mensal venha a fazer melhor proveito das premissas do modelo SSS.

O primeiro conjunto de parâmetros a ser exibido é a estimação do expoente de Hurst e do desvio padrão das séries, obtidos através do algoritmo LSSD. Os resultados, para ambas as escalas, são mostrados na Tabela 15.

**Tabela 15 – Expoentes de Hurst e desvios padrão para séries da bacia do rio Iguaçu**

Estação	Escala Anual			Escala Mensal		
	$\hat{H}$	$\hat{\sigma}$ (m <sup>3</sup> /s)	Iterações	$\hat{H}$	$\hat{\sigma}$ (m <sup>3</sup> /s)	Iterações
São Bento	0,67	9,1	6	0,78	13,1	6
União da Vitória	0,64	133,0	6	0,74	216,2	6
Porto Amazonas	0,67	18,2	6	0,75	29,2	6
Jangada do Sul	0,59	7,1	6	0,71	12,0	6
Rio Negro	0,73	18,9	7	0,76	28,5	6
Fazenda Maracanã	0,67	2,7	6	0,73	4,1	6

Para a escala anual Jangada do Sul e Rio Negro são as séries com menor e maior  $\hat{H}$ , respectivamente. Passando para a escala mensal, todos os expoentes tiveram um aumento em

seu valor e ficaram menos dispersos entre as estações fluviométricas. Jangada do Sul se manteve com o menor  $\hat{H}$ , no entanto São Bento passou a ser a série com o maior  $\hat{H}$ .

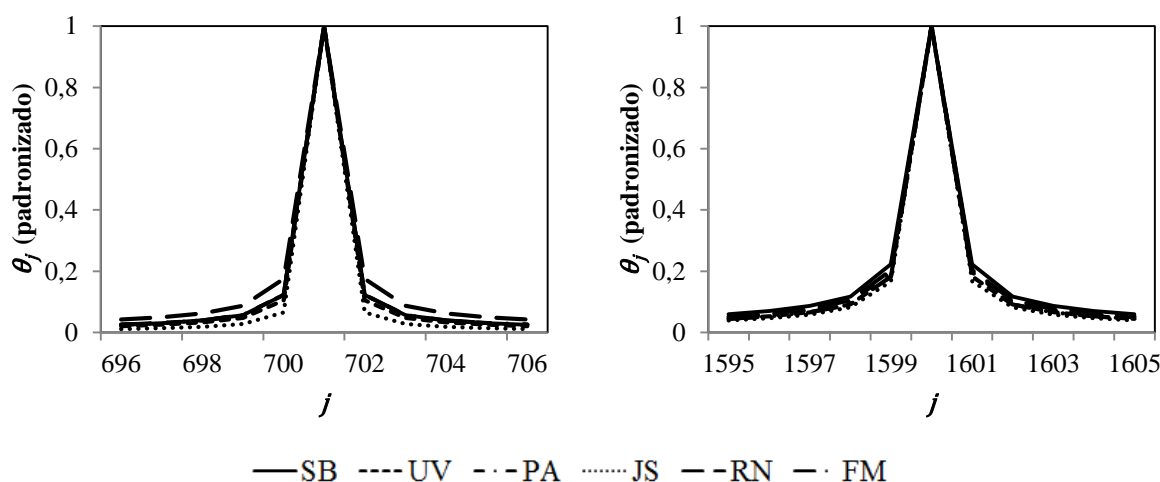
De uma maneira geral, o que se percebe é que os expoentes de Hurst têm valores significativos para as duas escalas. No caso da escala anual, deve-se lembrar que as estimativas de  $\hat{H}$  possuem incertezas devido ao tamanho das amostras (ver ensaios mostrados na Figura 20). Ainda assim, considera-se nesta tese que as flutuações em múltiplas escalas temporais estão presentes nas séries estudadas, justificando a utilização da modelagem SSS.

Os próximos parâmetros a serem exibidos se referem aos pesos  $\theta_j$  usados na formulação do modelo. Aplicou-se a equação (1.58) (página 79) para a determinação da quantidade  $q$  de pesos considerados em cada série. Usou-se a mesma tolerância  $\beta = 0,001$  adotada nos ensaios da seção 5.1.1.2. Os resultados para as escalas anual e mensal são mostrados na Tabela 16.

**Tabela 16 – Quantidades  $q$  de parâmetros  $\theta_q$  a serem usadas no modelo SSS**

<b>Estação</b>	<b>Escala Anual</b>	<b>Escala Mensal</b>
<b>São Bento</b>	243	1579
<b>União da Vitória</b>	174	748
<b>Porto Amazonas</b>	246	953
<b>Jangada do Sul</b>	74	555
<b>Rio Negro</b>	655	1103
<b>Fazenda Maracanã</b>	246	737

Como esperado, os resultados são reflexo direto da magnitude do expoente de Hurst em cada série. Assim, as estações com os maiores  $\hat{H}$  foram as que apresentaram as maiores estimativas para  $q$ . Na intenção de uniformizar o procedimento de geração para todas as séries, adotaram-se como padrão os maiores valores de  $q$  para todas as séries, nas respectivas escalas. Fez-se ainda um arredondamento para cima, resultando em  $q = 700$  para as séries em escala anual e  $q = 1600$  para as séries em escala mensal. Os parâmetros  $\theta_j$  resultantes para as duas escalas são mostrados na Figura 24.



**Figura 24 – Parâmetros  $\theta_j$  para séries da bacia do rio Iguazu, nas escalas anual (direita) e mensal (esquerda)**

Devido à diferença nas magnitudes das vazões das estações trabalhadas, foi necessário fazer uma padronização para tornar possível a análise comparativa. Essa padronização foi feita dividindo-se cada valor de  $\theta_j$  pelo seu respectivo valor central  $\theta_0$ . Também para facilitar a visualização, limitaram-se os intervalos  $j$  para  $\pm 5$  pontos antes e depois de  $\theta_0$ .

O que se observa é um formato semelhante para os pesos nas duas escalas temporais, com sutis distinções apenas na escala anual (gráfico da esquerda). A curva com decaimento mais lento se refere a Rio Negro, enquanto que a com decaimento mais brusco é a de Jangada do Sul. Tratam-se das séries com maior e menor  $\hat{H}$ , respectivamente, o que evidencia que a magnitude desse expoente tem influência direta na forma do decaimento da série de pesos  $\theta_j$ . Lembra-se que na escala mensal não houve diferenças significativas na magnitude de  $\hat{H}$  e, conseqüentemente, as curvas  $\theta_j$  não apresentaram distinções.

Com essa análise completa-se o detalhamento do ajuste do modelo SSS às séries consideradas na tese. Os resultados das séries sintéticas serão mostrados nas seções 5.2 e 5.3 para as escalas anual e mensal, respectivamente.

### 5.1.2 Modelo MAR(1)

Para o modelo MAR(1), o detalhamento do ajuste envolve a estimação das matrizes-parâmetro  $\hat{A}$  e  $\hat{B}$  utilizadas em sua equação principal. Considerando a matriz  $\hat{A}$  como diagonal, conforme descrito na seção 1.4.2.1 (página 87), os elementos das matrizes para as séries modeladas são exibidos na Tabela 17.

**Tabela 17 – Matrizes-parâmetros para modelo MAR(1)**

Postos	$\hat{A}$						$\hat{B}$					
	JS	FM	RN	SB	UV	PA	JS	FM	RN	SB	UV	PA
<b>JS</b>	0,181	0	0	0	0	0	0,980	0	0	0	0	0
<b>FM</b>	0	0,246	0	0	0	0	0,945	0,470	0	0	0	0
<b>RN</b>	0	0	0,302	0	0	0	0,789	0,226	0,389	0	0	0
<b>SB</b>	0	0	0	0,258	0	0	0,840	0,280	0,342	0,211	0	0
<b>UV</b>	0	0	0	0	0,237	0	0,886	0,244	0,225	0,048	0,162	0
<b>PA</b>	0	0	0	0	0	0,266	0,828	0,311	0,207	0,139	0,176	0,255

A título de comparação, a Tabela 18 exhibe as mesmas matrizes-parâmetro estimadas para as séries sem tendência, resultado do procedimento de remoção do comportamento não estacionário apresentado na seção 4.4 (página 141). Para distinguir os modelos, convencionou-se MAR(1)\* como a abreviação do modelo com séries sem tendência.

O que se percebe é que as diferenças notáveis são somente na matriz  $\hat{A}$ , responsável pela modelagem da persistência individual de cada série. No caso, todos os coeficientes sofreram redução, o que sugere que a estrutura de autocorrelação das séries é afetada pelo procedimento de remoção da tendência. A matriz  $\hat{B}$ , responsável pela modelagem da correlação espacial entre as séries não sofreu alterações significativas.

**Tabela 18 – Matrizes-parâmetros para modelo MAR(1)\***

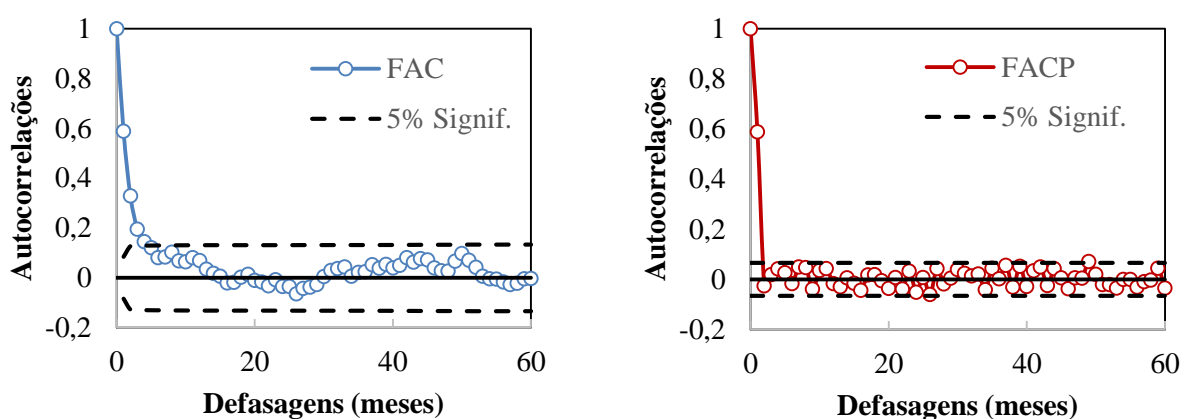
Postos	$\hat{A}$						$\hat{B}$					
	JS	FM	RN	SB	UV	PA	JS	FM	RN	SB	UV	PA
<b>JS</b>	0,148	0	0	0	0	0	0,978	0	0	0	0	0
<b>FM</b>	0	0,131	0	0	0	0	0,954	0,503	0	0	0	0
<b>RN</b>	0	0	0,185	0	0	0	0,851	0,280	0,427	0	0	0
<b>SB</b>	0	0	0	0,154	0	0	0,879	0,292	0,370	0,242	0	0
<b>UV</b>	0	0	0	0	0,157	0	0,913	0,247	0,227	0,075	0,182	0
<b>PA</b>	0	0	0	0	0	0,135	0,881	0,339	0,228	0,139	0,166	0,312

A fim de explorar melhor essas diferenças e verificar se elas são notadas também nas séries sintéticas, as análises comparativas apresentadas na seção 5.2 contemplarão as duas versões do modelo MAR(1).

### 5.1.3 Modelos CARIMA

As análises preliminares do modelo CARIMA estão relacionadas com o procedimento iterativo de Box & Jenkins descrito na Figura 7 (página 95). Como explicado na descrição

teórica do modelo (seção 1.4.3.4, página 95), a identificação da ordem das formulações foi feita através dos gráficos das funções FAC e FACP e também pelo cálculo do critério BIC. A Figura 25 exibe as funções obtidas para a série de União da Vitória, utilizando 60 meses como defasagem máxima. Deve-se ressaltar que, como está sendo utilizado um modelo não periódico os gráficos se referem às séries dessazonalizadas.



**Figura 25 – FAC e FACP para União da Vitória**

Nos gráficos da Figura 25 as linhas tracejadas horizontais representam o limite de significância abaixo do qual os valores das autocorrelações são estatisticamente nulos. O que se percebe é o comportamento de um modelo puramente autorregressivo, caracterizado pelo decaimento exponencial da FAC e com a repentina anulação da FACP após a primeira defasagem (ver comportamento teórico na Tabela 3, página 96). Desse modo, as funções sugerem que um modelo CARIMA(1,0,0) é apropriado para a modelagem. As funções FAC e FACP para as demais séries apresentaram um comportamento muito similar a União da Vitória, por isso são omitidos do texto e podem ser visualizadas no Apêndice A.1.

A Tabela 19 exibe os resultados do critério BIC calculado para as 10 variantes de modelo CARIMA consideradas no trabalho.

**Tabela 19 – Resultados do critério BIC para as variantes do modelo CARIMA**

	Ordens dos modelos CARIMA( <i>p, d, q</i> )									
	(1,0,0)	(2,0,0)	(1,0,1)	(2,0,1)	(2,0,2)	(1,1,0)	(2,1,0)	(1,1,1)	(2,1,1)	(2,1,2)
<b>SB</b>	<b>2133,3</b>	2139,9	2139,9	2146,6	2146,0	2288,8	2250,9	2138,3	2154,9	2166,2
<b>UV</b>	<b>2248,8</b>	2254,7	2255,1	2261,4	2267,6	2431,6	2402,1	2267,5	2263,3	2290,0
<b>PA</b>	<b>2015,0</b>	2018,2	2021,7	2024,4	2029,0	2169,7	2133,6	2063,3	2029,5	2053,6
<b>JS</b>	<b>1786,0</b>	1789,9	1792,4	1793,8	1800,7	1927,6	1902,2	1825,5	1833,4	1839,7
<b>RN</b>	<b>2219,4</b>	2226,1	2226,1	2232,6	2235,8	2385,0	2351,6	2221,9	2231,9	2250,7
<b>FM</b>	<b>1801,2</b>	1806,7	1807,1	1810,5	1819,1	1943,7	1925,1	1817,0	1852,2	1845,8

Como se nota, os mínimos BIC em todas as estações se deram para o modelo CARIMA(1,0,0), em perfeita concordância com as análises visuais das funções FAC e FACP. Há de se apontar que a diferenciação das séries não promoveu melhoras, podendo ser dispensada.

Seguindo o procedimento de Box & Jenkins, os próximos passos são a estimação e validação do modelo. Ambos estão reunidos na Tabela 20, que exhibe o único parâmetro autorregressivo para cada série e os p-valores para os testes de independência (teste de Portmanteau – Li e McLeod, 1981), homocedasticidade (teste de Levene – Brown e Forsythe, 1974) e normalidade (teste de Shapiro-Wilk – Ferreira, 2008).

**Tabela 20 – Parâmetro do modelo CARIMA(1,0,0) e p-valores da validação teórica**

	Parâmetro $\phi_1$	Independência (Portmanteau)	Homocedasticidade (Levene)	Normalidade (Shapiro-Wilk)
<b>São Bento</b>	0,615	0,912	0,821	0,000
<b>União da Vitória</b>	0,588	0,626	0,780	0,000
<b>Porto Amazonas</b>	0,604	0,998	0,980	0,021
<b>Jangada do Sul</b>	0,543	0,968	0,568	0,003
<b>Rio Negro</b>	0,605	0,903	0,962	0,000
<b>Fazenda Maracanã</b>	0,524	0,743	0,840	0,004

Como esperado, os parâmetros resultaram em valores muito próximos para todas as séries. Chama-se atenção que eles ficaram distantes da unidade (e.g. círculo unitário), o que confirma que diferenciações nas séries não são requeridas. Em outras palavras, não há necessidade de se empregar uma formulação não estacionária para estas séries, constatação que confronta os resultados de alguns testes estatísticos apresentados no capítulo anterior. Ressalta-se, portanto, que tomar a decisão de aplicar ou não as diferenciações com base em resultados de análises externas é uma prática equivocada do ponto de vista da teoria dos modelos Box & Jenkins.

A respeito dos testes de validação teórica, nota-se que os p-valores para os testes de independência e homocedasticidade não sugerem significância estatística. Por outro lado, os resultados para o teste de normalidade sugerem a rejeição da hipótese nula em todos os casos (considerando 5% de significância). A partir disso, fizeram-se extensivos testes com todas as demais nove formulações consideradas para o modelo CARIMA na intenção de identificar uma (ou mais) possível(is) configuração(ões) que resultasse(m) em resíduos normalmente distribuídos. Ainda assim, o teste de Shapiro-Wilk foi unânime em rejeitar a hipótese nula de normalidade. Por esse motivo, optou-se por manter o parcimonioso modelo CARIMA(1,0,0) para a geração de séries sintéticas nas seis estações hidrométricas estudadas.



## 5.2 Análise comparativa SSS versus MAR(1) para séries anuais

Depois de ajustados os modelos às séries anuais, geraram-se 1000 cenários sintéticos para cada formulação, cada uma com comprimento de 77 anos, equivalente à série de maior número de elementos. Os resultados foram submetidos às análises estatísticas descritas na seção 2.3 (página 112), as quais serão apresentadas neste item de diversas formas.

A apresentação dos resultados se inicia com as estatísticas de curto termo, exibidas na Tabela 21. Os resultados das séries sintéticas se referem à média dos cenários gerados. Ademais, a primeira linha em cada posto traz as estatísticas de curto termo observadas nas séries históricas, para referência. Nota-se que as estatísticas foram fielmente reproduzidas por todos os modelos e em todas as estações hidrométricas consideradas, com poucas exceções. As médias ( $\hat{\mu}$ ) foram respeitadas com precisão, a não ser pelo modelo MAR(1)\* que obteve maiores desvios tanto para mais (São Bento, Porto Amazonas, Jangada do Sul e Fazenda Maracanã) quanto para menos (demais estações). Para os desvios-padrão ( $\hat{\sigma}$ ), observou-se uma sutil subestimação em União da Vitória para os três modelos. Vazões mínimas e máximas apresentaram alguns desvios, entretanto não são significativos frente à dificuldade em se reproduzir estatísticas relacionadas a valores extremos. Rio Negro obteve maiores subestimações nas duas estatísticas, o que se acredita ser consequência da vazão negativa obtida após o processo de remoção da tendência linear (ver Figura 19, página 142) O expoente de Hurst ( $\hat{H}$ ) ficou limitado à formulação SSS devido à sua abordagem, tendo sido reproduzido com sucesso.

O coeficiente de assimetria ( $\hat{\xi}$ ) foi o parâmetro com os resultados mais interessantes e, por esse motivo, recebe comentários específicos. Os modelos SSS e MAR(1) o reproduziram com boa precisão, com pequenos desvios para as séries de Porto Amazonas e Jangada do Sul. Por outro lado, o modelo MAR(1)\* obteve severas subestimações em todos os postos hidrométricos considerados. Esse resultado chama a atenção por estar isolado em meio às demais estatísticas, as quais o modelo respeitou adequadamente. Dessa maneira, entende-se que o método de remoção da tendência linear utilizado, a fim de se obter séries com comportamento estacionário, causou também um abatimento na assimetria dos registros. Mesmo aplicando o processo inverso ao final do algoritmo de geração, o efeito não foi anulado e a assimetria das séries sintéticas foi subestimada. Ao analisar novamente a Tabela 21, percebe-se que os valores de vazões mínimas e máximas foram subestimados em muitos casos, podendo essa ser uma explicação do abatimento na assimetria. Considera-se esse um importante resultado correlato desta tese, por não ter sido anteriormente relatado na literatura.

**Tabela 21 – Médias das estatísticas de curto termo das séries sintéticas anuais**

Posto	Série	$\hat{\mu}$ (m <sup>3</sup> /s)	$\hat{\sigma}$ (m <sup>3</sup> /s)	$\hat{\xi}$	Mín. (m <sup>3</sup> /s)	Máx. (m <sup>3</sup> /s)	$\hat{H}$
<b>São Bento</b>	<i>Histórico</i>	34,8	12,6	1,15	14,7	80,4	0,67
	SSS	34,8	12,2	1,14	15,6	77,8	0,66
	MAR(1)	34,9	12,5	1,11	15,2	78,7	-
	MAR(1)*	35,2	12,8	0,82	12,2	76,6	-
<b>União da Vitória</b>	<i>Histórico</i>	468,4	184,4	1,08	183,6	1127,8	0,64
	SSS	468,4	180,6	1,03	172,7	1088,9	0,64
	MAR(1)	468,4	180,7	1,00	173,3	1079,7	-
	MAR(1)*	468,7	181,5	0,73	134,5	1038,6	-
<b>Porto Amazonas</b>	<i>Histórico</i>	67,9	25,4	1,02	27,3	151,9	0,67
	SSS	67,6	24,5	0,92	26,7	147,5	0,66
	MAR(1)	68,0	25,3	0,91	25,4	151,4	-
	MAR(1)*	69,7	26,4	0,63	19,3	150,3	-
<b>Jangada do Sul</b>	<i>Histórico</i>	25,9	9,5	0,97	8,4	54,5	0,59
	SSS	25,8	9,2	0,80	10,0	53,9	0,58
	MAR(1)	25,9	9,4	0,84	9,6	56,2	-
	MAR(1)*	26,4	9,2	0,71	9,2	55,4	-
<b>Rio Negro</b>	<i>Histórico</i>	67,7	25,5	0,90	25,5	154,4	0,73
	SSS	67,5	24,3	0,84	24,8	146,6	0,73
	MAR(1)	67,5	24,0	0,72	23,8	141,8	-
	MAR(1)*	66,1	25,4	0,29	10,7	134,6	-
<b>Fazenda Maracaná</b>	<i>Histórico</i>	9,0	3,5	1,35	3,6	22,1	0,67
	SSS	9,0	3,4	1,37	4,3	21,2	0,66
	MAR(1)	9,2	3,9	1,52	3,9	24,4	-
	MAR(1)*	9,6	3,9	1,15	3,3	23,6	-

SSS – modelo de múltiplas escalas temporais; MAR(1) – modelo Markoviano de primeira ordem; MAR(1)\* – modelo Markoviano de primeira ordem aplicado a séries com tendência removida;  $\hat{\mu}$  – média;  $\hat{\sigma}$  – desvio padrão;  $\hat{\xi}$  – coeficiente de assimetria;  $\hat{H}$  – expoente de Hurst.

A Tabela 22 exibe os desvios-padrão das estatísticas de curto termo, calculadas a partir de todo o conjunto de 1000 séries sintéticas geradas. Esses dados são exibidos de forma a mensurar a incerteza na determinação de cada uma das estatísticas mostradas na Tabela 21. Deve-se alertar, entretanto, que não se deve interpretar esses desvios como sendo intervalos equidistantes da média em cada estatística, pois a suas distribuições não necessariamente são normais. Assim, por exemplo, o coeficiente de assimetria médio em União da Vitória foi de 1,03 para o modelo SSS e a incerteza dessa estimativa foi de 0,49. Não se pode afirmar, contudo, que os coeficientes de assimetria variaram entre 1,52 e 0,54 ( $1,03 \pm 0,49$ ). O mesmo raciocínio é válido para as demais estatísticas.

**Tabela 22 – Incertezas das estatísticas de curto termo das séries sintéticas anuais**

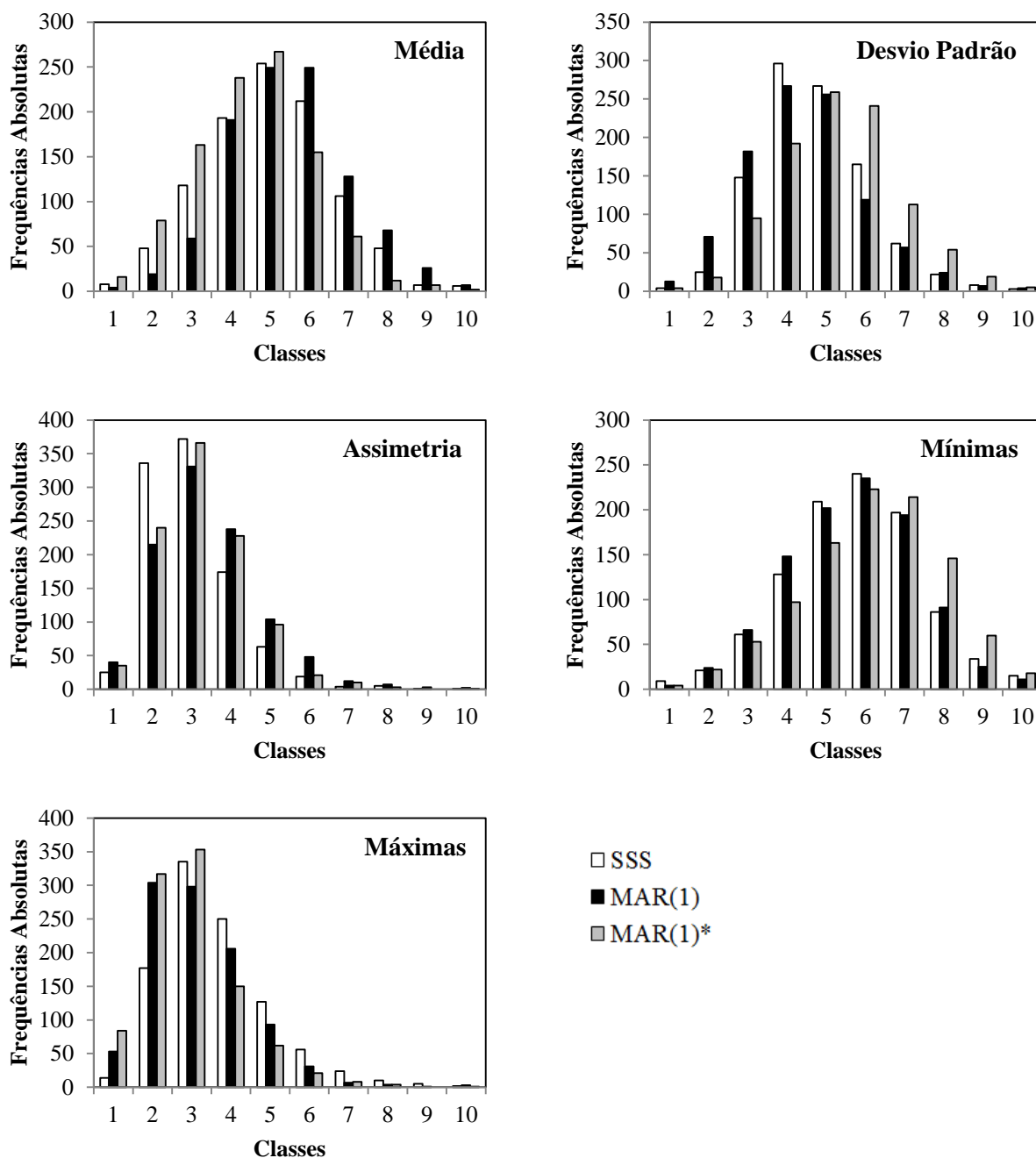
Posto	Série	$\hat{\mu}$ (m <sup>3</sup> /s)	$\hat{\sigma}$ (m <sup>3</sup> /s)	$\hat{\xi}$	Mín. (m <sup>3</sup> /s)	Máx. (m <sup>3</sup> /s)	$\hat{H}$
<b>São Bento</b>	SSS	2,9	1,6	0,50	2,7	14,2	0,09
	MAR(1)	1,8	1,6	0,47	2,0	13,3	-
	MAR(1)*	1,7	1,4	0,42	2,7	11,8	-
<b>União da Vitória</b>	SSS	40,2	23,2	0,49	38,2	202,9	0,09
	MAR(1)	25,8	21,7	0,44	31,4	179,1	-
	MAR(1)*	23,9	19,2	0,39	42,4	154,2	-
<b>Porto Amazonas</b>	SSS	6,1	3,1	0,46	6,1	27,4	0,10
	MAR(1)	3,7	2,9	0,43	4,9	24,3	-
	MAR(1)*	3,3	2,6	0,37	6,5	21,4	-
<b>Jangada do Sul</b>	SSS	1,6	1,1	0,42	2,0	8,9	0,09
	MAR(1)	1,3	1,0	0,39	1,9	8,2	-
	MAR(1)*	1,2	1,0	0,37	2,2	7,4	-
<b>Rio Negro</b>	SSS	7,6	2,9	0,41	7,8	24,8	0,10
	MAR(1)	3,7	2,5	0,37	5,6	19,6	-
	MAR(1)*	3,3	2,3	0,29	8,6	14,8	-
<b>Fazenda Maracaná</b>	SSS	0,9	0,6	0,67	0,6	5,5	0,09
	MAR(1)	0,6	0,6	0,63	0,4	5,5	-
	MAR(1)*	0,5	0,5	0,56	0,6	4,6	-

SSS – modelo de múltiplas escalas temporais; MAR(1) – modelo Markoviano de primeira ordem; MAR(1)\* – modelo Markoviano de primeira ordem aplicado a séries com tendência removida;  $\hat{\mu}$  – média;  $\hat{\sigma}$  – desvio padrão;  $\hat{\xi}$  – coeficiente de assimetria;  $\hat{H}$  – expoente de Hurst.

Os resultados das estatísticas de curto termo em função das médias das séries geradas são úteis para verificar a correta implementação computacional dos modelos. Entretanto, melhor visualização pode ser obtida através da Figura 26, que traz histogramas de frequências absolutas para todas as estatísticas apresentadas na Tabela 21, novamente limitadas à estação de União da Vitória. Adotaram-se 10 classes para distribuir as estatísticas das 1000 séries sintéticas de cada modelo. Sugere-se ao leitor que as análises devam ser feitas qualitativamente, pois não é possível associar valores às classes nos histogramas, visto que cada modelo obteve uma distribuição distinta. É suficiente saber que as classes com maior frequência em todos os casos se referem às estatísticas médias mostradas na Tabela 21. Os mesmos histogramas para as demais estações estão no Apêndice A.2 desta tese.

Os modelos apresentaram uma distribuição aproximadamente simétrica nas estatísticas de média, vazões mínimas e desvio padrão [somente MAR(1)], enquanto que para as demais estatísticas as distribuições foram assimétricas à esquerda. No primeiro caso, isso significa que as séries sintéticas têm uma representatividade uniforme de ocorrências com médias e vazões

mínimas acima e abaixo dos valores históricos. Para o segundo caso, há ocorrência de séries muito mais extremas, porém com pouca frequência.



**Figura 26 – Histogramas para as estatísticas de curto termo em União da Vitória, para séries sintéticas anuais**

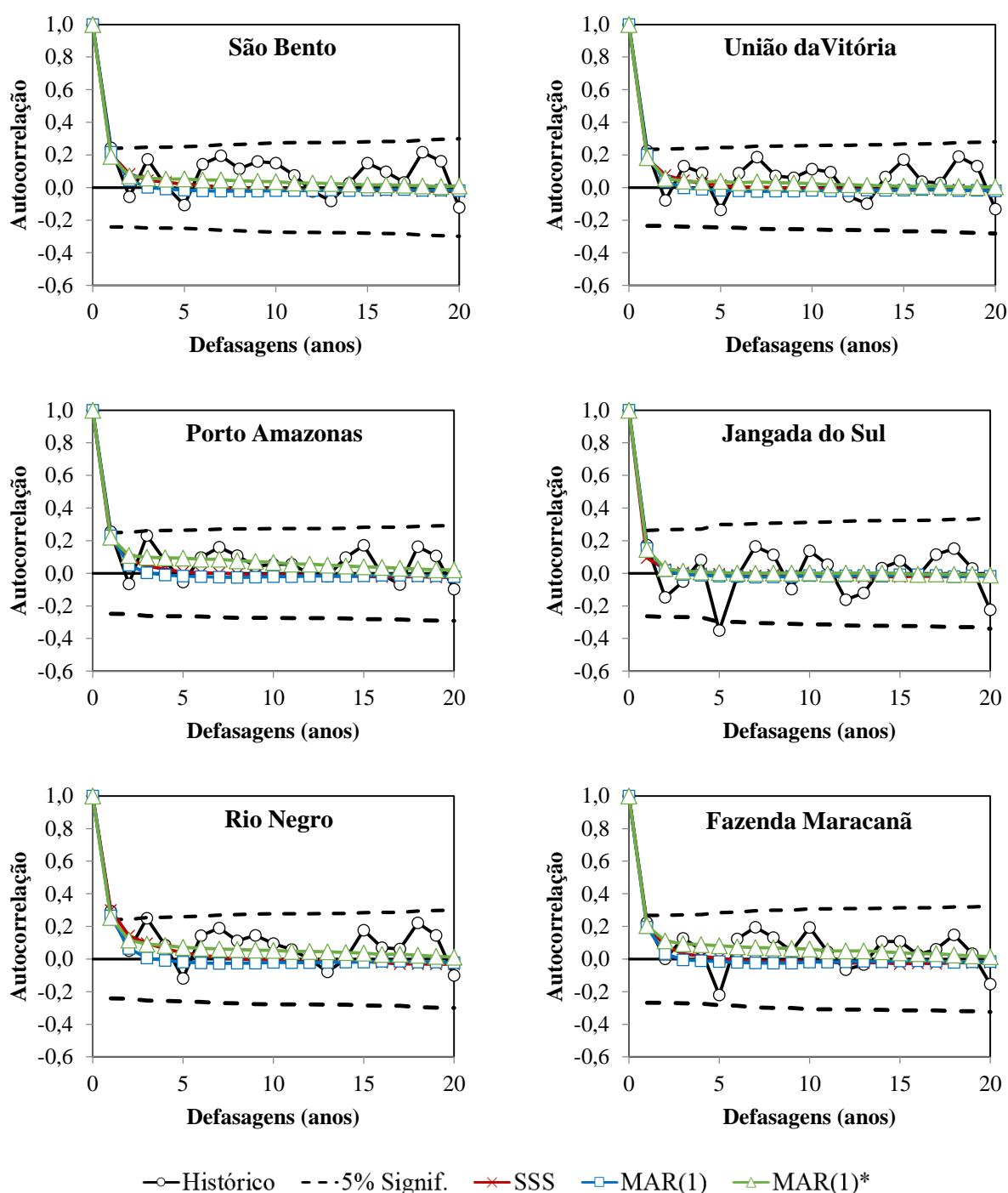
Muito embora os formatos dos histogramas tenham sido semelhantes entre os modelos, as distribuições de frequência em cada caso foram distintas. As maiores diferenças são observadas no desvio padrão, assimetria e vazões máximas. Para os dois primeiros, o modelo SSS concentra grande parte das ocorrências em torno do valor médio. Por sua vez, para vazões

máximas ele produz maiores frequências em classes mais elevadas quando comparado aos modelos Markovianos.

Lembrando a severa subestimação da assimetria por parte do modelo MAR(1)\*, é interessante apontar que as séries cujos coeficientes resultaram em valores equivalentes aos da série histórica se distribuem entre as classes quatro e cinco e totalizam 32,4% das séries geradas. Essa baixa representatividade confirma a deficiência do modelo em reproduzir a assimetria das séries estudadas. Entretanto detectou-se ainda um fato mais agravante para esse modelo: foram geradas séries com vazões médias negativas em alguns casos. Dentre os diversos ensaios de Monte Carlo realizados, União da Vitória apresentou vazões médias negativas em somente 0,40% das séries, porém em Rio Negro esse índice subiu para 10%. Tal fato exige ainda mais atenção na utilização da técnica de remoção da tendência linear das séries.

Neste ponto pode-se especular sobre a possível razão da geração de vazões médias negativas, juntamente com a subestimação de valores mínimos e máximos e, conseqüentemente, o abatimento na assimetria no modelo MAR(1)\*. Entende-se que isso tudo pode ser fruto do próprio método de remoção da tendência linear das séries. Essencialmente, a subtração do coeficiente linear dos registros causa uma redução das vazões médias observadas. Essa série modificada é a que o modelo utilizará para estimar os parâmetros necessários à geração dos cenários sintéticos e, naturalmente, todas as séries sintéticas serão estatisticamente semelhantes a essas séries modificadas. Através dos resultados mostrados, nem mesmo a reintrodução do coeficiente linear anteriormente estimado foi capaz de remover a tendenciosidade das séries sintéticas.

Avançando nos resultados, a Figura 27 exhibe as FAC comparativas para todos os modelos e postos hidrométricos, com defasagem temporal máxima de 20 anos. As únicas diferenças notáveis são para o modelo MAR(1)\* em todas as séries, exceto Jangada do Sul, nas quais as FAC tiveram decaimento um pouco mais lento. De certo modo isso é contraditório com a constatação feita na seção 5.1.2, quando se mostrou que os coeficientes da matriz responsável pela autocorrelação sofrem redução na modelagem MAR(1)\*. Entende-se, portanto, que esses fatos não estão relacionados.



**Figura 27 – FAC das séries sintéticas anuais**

A exibição dos resultados prossegue com as estatísticas de longo termo, exibidas na Tabela 23 em termos médios. Novamente, todos os modelos foram capazes de reproduzir adequadamente as estatísticas históricas, porém lembra-se que esse não é o objetivo principal da análise desses índices. Uma vez com os modelos validados por meio das estatísticas de curto termo, as reais características de cada formulação ficam evidenciadas através das estatísticas de longo termo. É justamente nesse ponto que se percebem maiores distinções entre as abordagens.

Tabela 23 – Médias das estatísticas de longo termo das séries sintéticas anuais

Posto	Série	$n_{\Lambda}$	$\bar{\Lambda}$	$\Lambda_{\max}$	$\bar{\Lambda}$ (m <sup>3</sup> /s)	$\Lambda_{\max}$ (m <sup>3</sup> /s)	$\bar{\Delta}$	$\Delta_{\max}$
<b>São Bento</b>	<i>Histórico</i>	9	3	7	86,2	187,1	4,3	29,9
	SSS	10	4	8	102,7	218,2	5,1	33,5
	MAR(1)	10	4	8	101,2	213,3	4,8	31,2
	MAR(1)*	10	4	8	96,3	205,9	6,9	42,1
<b>União da Vitória</b>	<i>Histórico</i>	11	3	7	1046,8	2531,8	64,7	370,6
	SSS	10	4	8	1294,2	2739,4	93,0	563,2
	MAR(1)	10	4	8	1284,6	2699,5	81,4	499,9
	MAR(1)*	10	4	8	1219,5	2568,1	106,0	620,8
<b>Porto Amazonas</b>	<i>Histórico</i>	8	4	7	183,1	352,2	9,1	58,0
	SSS	9	4	8	190,2	392,3	11,8	72,5
	MAR(1)	10	4	8	189,8	396,3	11,5	72,5
	MAR(1)*	9	4	8	188,4	400,2	19,6	106,8
<b>Jangada do Sul</b>	<i>Histórico</i>	8	4	6	72,1	122,1	2,7	12,3
	SSS	8	3	7	66,9	127,6	3,4	21,7
	MAR(1)	10	4	7	68,7	142,0	3,6	24,1
	MAR(1)*	10	4	7	69,0	143,5	3,9	25,9
<b>Rio Negro</b>	<i>Histórico</i>	11	3	7	170,4	382,5	10,2	72,5
	SSS	9	4	9	201,5	433,2	15,7	92,9
	MAR(1)	10	4	8	189,8	388,0	11,3	73,2
	MAR(1)*	9	4	8	168,5	351,3	21,3	114,3
<b>Fazenda Maracanã</b>	<i>Histórico</i>	7	4	7	26,4	44,6	1,7	7,1
	SSS	8	4	8	27,4	54,8	1,4	8,3
	MAR(1)	10	4	9	26,7	57,6	1,7	9,8
	MAR(1)*	10	4	9	27,7	60,9	2,8	15,2

SSS – modelo de múltiplas escalas temporais; MAR(1) – modelo Markoviano de primeira ordem; MAR(1)\* – modelo Markoviano de primeira ordem aplicado a séries com tendência removida;  $n_{\Lambda}$  – número de corridas;  $\bar{\Lambda}$  – durações médias das corridas;  $\Lambda_{\max}$  – durações máximas das corridas;  $\bar{\Delta}$  – déficits médios;  $\Delta_{\max}$  – máximo déficit acumulado.

Bons exemplos disso são as séries de União da Vitória e Rio Negro: para as corridas, todos os modelos subestimaram o número de ocorrências ( $n_{\Lambda}$ ) e superestimaram as durações médias ( $\bar{\Lambda}$ ) e máximas ( $\Lambda_{\max}$ ). Isso significa que eles trocaram um maior número de corridas de menor duração por menos corridas de maior duração. O modelo SSS deixa isso evidente, principalmente quando se analisam as vazões acumuladas médias (em m<sup>3</sup>/s) e máximas (em m<sup>3</sup>/s) em cada corrida. Ainda em relação às corridas, as únicas séries nas quais os modelos tiveram resultados semelhantes entre si foram São Bento e União da Vitória.

Com relação aos déficits, todos os modelos superestimaram as estatísticas históricas, com uma única exceção para o déficit médio ( $\bar{\Delta}$ ) do modelo SSS em Fazenda Maracanã. Deve-

se apontar para o desempenho muito semelhante entre os modelos SSS e MAR(1) principalmente em Porto Amazonas, na qual eles resultaram em valores praticamente idênticos.

A exemplo das estatísticas de curto termo, o modelo MAR(1)\* merece comentários específicos. Muito embora as métricas relacionadas com as corridas tenham ficado relativamente próximas aos demais modelos (e à série histórica), desvios maiores ocorreram para os déficits, sempre no sentido de superestimá-los. Os casos extremos são União da Vitória e Rio Negro, na qual o modelo estimou um máximo déficit acumulado ( $\Delta_{\max}$ ) quase duas vezes maior do que o observado no histórico. Mais uma vez, isso pode ser explicado como uma consequência direta do processo de remoção da tendência linear, que criou um viés nas séries estudadas no sentido de subestimar as vazões médias.

No caso do modelo SSS, a análise das estatísticas de longo termo permite notar claramente que as premissas teóricas da formulação estão sendo atendidas pelas séries sintéticas. Ao considerar fenômenos em múltiplas escalas temporais, era esperado que eventos de menor frequência de ocorrência tivessem melhor representatividade, o que se observou pelos resultados apresentados. Em quase todas as séries pode-se perceber um aumento das estatísticas relacionadas com vazões médias de maior e menor magnitude, representadas pelas corridas e déficits, respectivamente. Isso se percebe principalmente em Rio Negro, série com o maior expoente de Hurst estimado.

A Tabela 24 mostra as incertezas calculadas para as estatísticas de longo termo para cada modelo. Do mesmo modo que para as estatísticas de curto termo, esses valores de incerteza não devem ser confundidos com intervalos de confiança sobre as médias exibidas na Tabela 23. Em termos de números de corridas ( $n_{\Lambda}$ ), durações médias ( $\bar{\Lambda}$ ) e máximas ( $\Lambda_{\max}$ ), percebe-se que as incertezas foram igualmente baixas em todos os modelos. As maiores incertezas foram observadas nos valores máximos de corridas ( $\Lambda_{\max}$ ) em termos de vazões e no máximo déficit acumulado ( $\Delta_{\max}$ ). De certa forma isso é um resultado esperado por se tratarem de estatísticas que medem os extremos das séries. Ainda assim, as magnitudes das incertezas foram muito semelhantes entre os modelos, com um pequeno destaque ao MAR(1) que apresentou valores levemente inferiores nessas estatísticas.



**Tabela 24 – Incertezas das estatísticas de longo termo das séries sintéticas anuais**

Posto	Série	$n_{\Lambda}$	$\bar{\Lambda}$	$\Lambda_{\max}$	$\bar{\Lambda}$ (m <sup>3</sup> /s)	$\Lambda_{\max}$ (m <sup>3</sup> /s)	$\bar{\Delta}$	$\Delta_{\max}$
<b>São Bento</b>	SSS	1,8	0,7	2,8	21,2	73,6	3,1	15,4
	MAR(1)	1,6	0,7	2,4	18,7	65,0	2,3	12,1
	MAR(1)*	1,7	0,7	2,6	18,7	65,8	4,5	19,3
<b>União da Vitória</b>	SSS	1,7	0,7	2,7	273,9	914,4	57,0	253,0
	MAR(1)	1,7	0,6	2,3	233,4	812,9	35,8	182,9
	MAR(1)*	1,7	0,6	2,4	234,0	780,2	64,3	274,6
<b>Porto Amazonas</b>	SSS	1,8	0,7	2,7	41,1	133,2	8,0	37,4
	MAR(1)	1,7	0,7	2,5	36,9	125,2	5,2	26,8
	MAR(1)*	1,8	0,7	2,8	38,2	131,8	12,9	48,5
<b>Jangada do Sul</b>	SSS	1,7	0,6	2,1	12,3	41,9	1,6	9,3
	MAR(1)	1,7	0,6	2,2	12,0	43,9	1,5	8,5
	MAR(1)*	1,7	0,6	2,2	12,6	44,0	2,0	10,4
<b>Rio Negro</b>	SSS	1,8	0,9	3,2	50,1	160,5	11,7	47,5
	MAR(1)	1,6	0,7	2,4	36,1	122,4	5,4	28,0
	MAR(1)*	1,7	0,7	2,7	36,7	121,6	13,5	49,8
<b>Fazenda Maracaná</b>	SSS	1,7	0,7	2,8	5,7	19,1	1,1	4,7
	MAR(1)	1,6	0,6	2,6	4,6	17,2	0,9	4,1
	MAR(1)*	1,8	0,7	3,0	5,3	19,3	1,9	7,0

SSS – modelo de múltiplas escalas temporais; MAR(1) – modelo Markoviano de primeira ordem; MAR(1)\* – modelo Markoviano de primeira ordem aplicado a séries com tendência removida;  $n_{\Lambda}$  – número de corridas;  $\bar{\Lambda}$  – durações médias das corridas;  $\Lambda_{\max}$  – durações máximas das corridas;  $\bar{\Delta}$  – déficits médios;  $\Delta_{\max}$  – máximo déficit acumulado.

A exemplo das estatísticas de curto termo, a Figura 28 exhibe os histogramas das estatísticas de longo termo para União da Vitória. Os gráficos para as outras estações hidrométricas estão mostrados separadamente no Apêndice A.3 desta tese. Mesmo que as estatísticas em termos médios mostradas na Tabela 23 tenham sido semelhantes para este posto, é interessante perceber que os histogramas mostram comportamentos bastante distintos entre os modelos. De modo geral, somente o histograma para o número de corridas apresenta distribuição simétrica em relação à classe de maior frequência, enquanto as outras estatísticas exibem histogramas com assimetria à esquerda.

Há de se destacar os histogramas do número de corridas, duração média e de vazão média acumulada, para os quais os modelos MAR(1)\* e MAR(1) exibem grande acúmulo de frequências em uma mesma classe. O modelo SSS também mostrou ocorrência semelhante para o déficit médio, porém sem a mesma intensidade. No geral, este foi o modelo que mostrou distribuição mais uniforme entre as diversas estatísticas analisadas, confirmando a flexibilidade de valores extremos anteriormente mencionada.

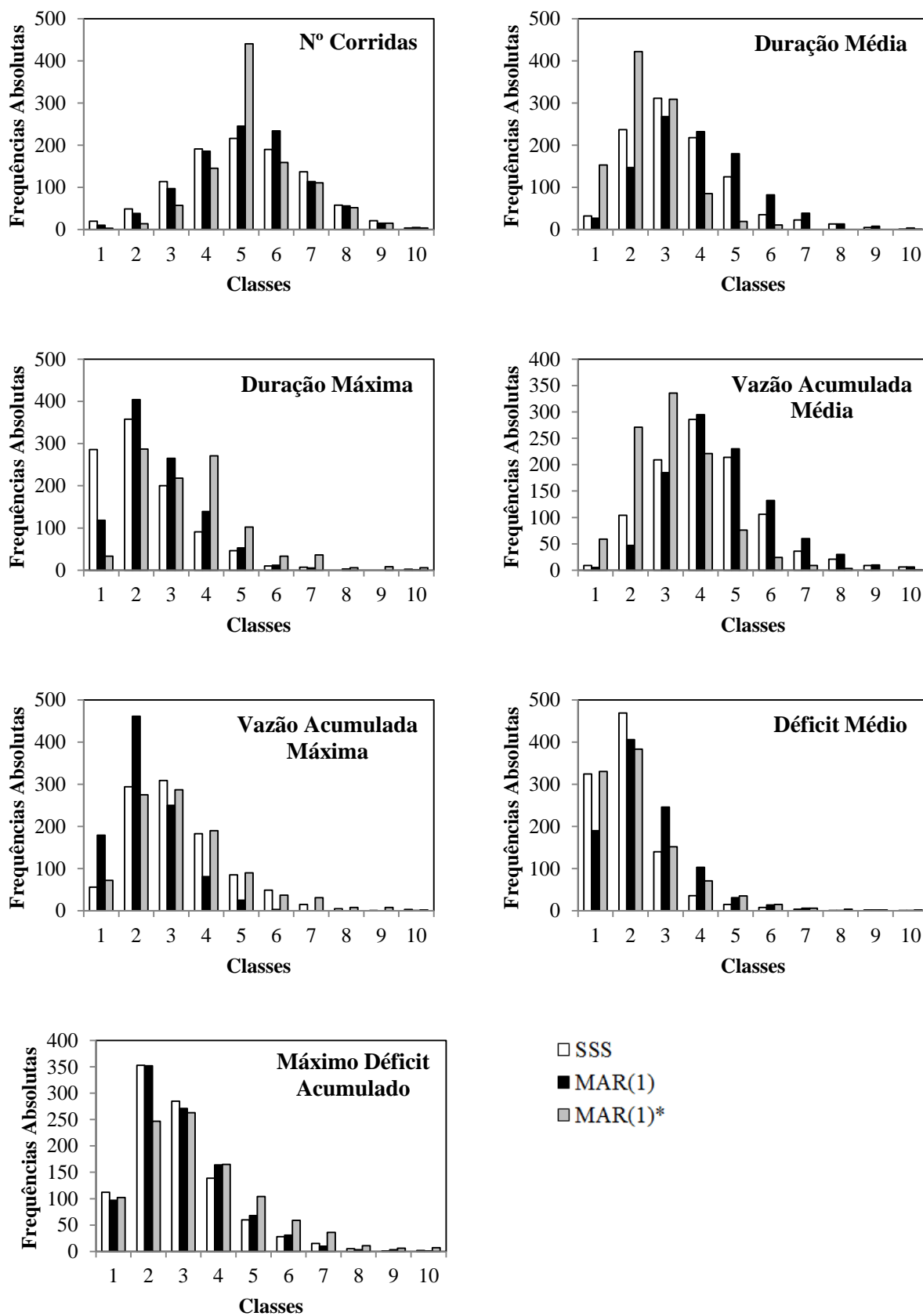
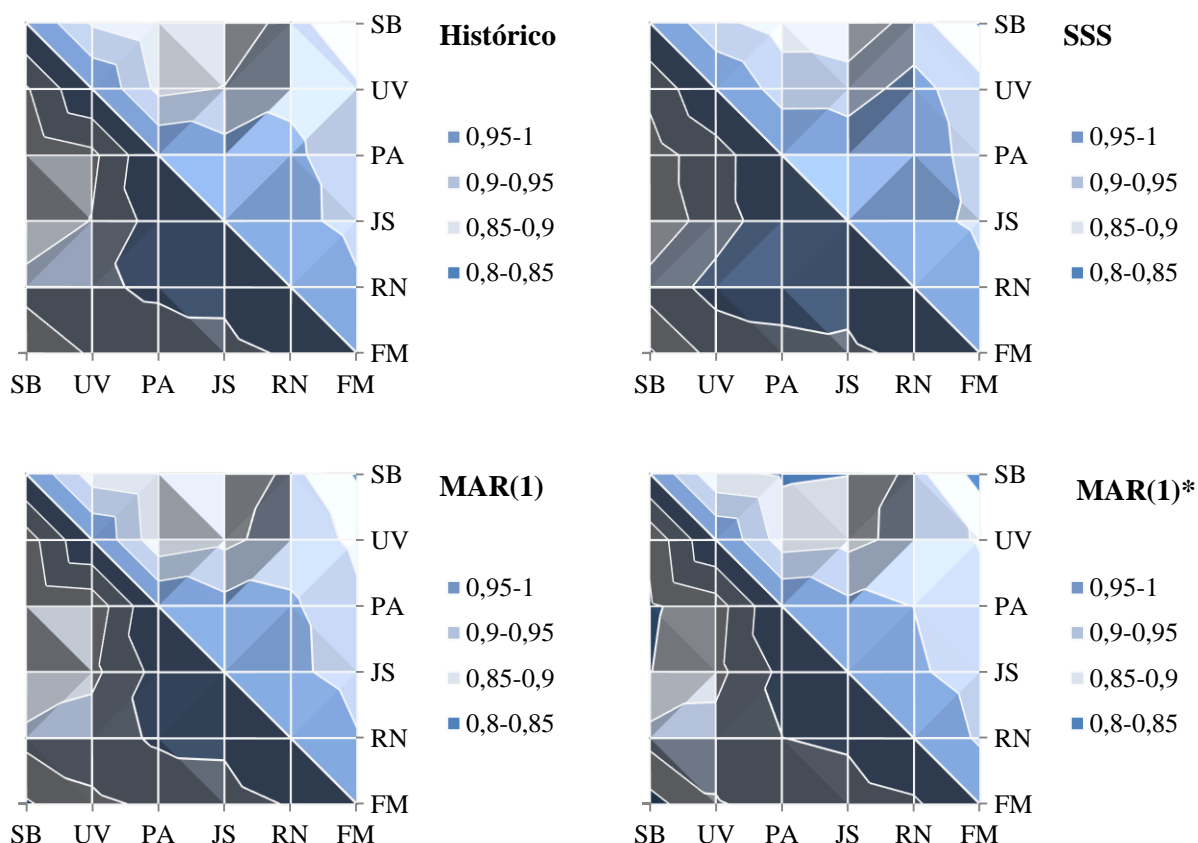


Figura 28 – Histogramas para as estatísticas de longo termo em União da Vitória, para séries sintéticas anuais

A última métrica a ser analisada pelos modelos anuais é a estrutura de correlações espaciais. Para tanto, elaboraram-se os mapas mostrados na Figura 29, nos quais se atribuiu uma escala monocromática para as diferentes faixas de correlação. As correlações históricas foram calculadas para o período comum entre as séries, de 1945 a 2005.



**Figura 29 – Mapas de correlações espaciais para as séries anuais**

A estrutura foi bem reproduzida por todos os modelos, com destaque para o MAR(1). O modelo SSS superestimou levemente as correlações entre União da Vitória e os demais postos, enquanto que o modelo MAR(1)\* subestimou as correlações entre Porto Amazonas e as outras estações hidrométricas. Ainda assim, considerando que a escala de correlações parte de 0,80, pode-se dizer nenhum desses desvios foi significativo.

Os resultados exibidos nesta seção confirmaram que os modelos SSS e MAR(1) são apropriados para aplicação nas séries anuais estudadas. O uso do modelo MAR(1)\* deve ser feito com cautela, visto que o processo de remoção do comportamento não estacionário causa diversas outras consequências nas séries sintéticas.

As diferenças apresentadas pelo modelo SSS foram muito sutis quando comparado às outras formulações e impedem conclusões definitivas acerca das vantagens de sua utilização.

Credita-se esse fato à fraca estrutura de persistência apresentada pelas séries da bacia do rio Iguaçu na escala anual (ver Figura 15, página 132). Na próxima seção serão mostrados os resultados deste modelo quando aplicado às séries em escala mensal.

### 5.3 Análise comparativa SSS versus CARIMA para séries mensais

Da mesma forma que para a escala anual, para a escala mensal também foram gerados 1000 cenários sintéticos de tamanho equivalente à maior série (932 elementos). Ressalta-se que para o modelo SSS, utilizou-se o algoritmo SMAD proposto pelo autor desta tese. Os resultados apresentados nesta seção comparam as diversas estatísticas dos modelos SSS e CARIMA com os registros históricos.

Seguindo o mesmo formato mostrado para a escala anual, inicia-se a exibição dos resultados com as estatísticas de curto termo, reunidas na Tabela 25. Mais uma vez, os resultados para as séries sintéticas se referem à média dos cenários gerados. Os resultados mostram uma aderência bastante fiel dos modelos quando comparados às estatísticas das séries históricas, principalmente em termos de médias ( $\hat{\mu}$ ) e desvios padrão ( $\hat{\sigma}$ ). Especificamente no caso do modelo SSS, além dessas estatísticas, a precisa reprodução dos expoentes de Hurst ( $\hat{H}$ ) valida a implementação do algoritmo SMAD. Outro destaque foi o bom desempenho das vazões máximas, apenas ficando levemente superestimada em União da Vitória, Porto Amazonas e Rio Negro. Para as vazões mínimas, o modelo SSS gerou valores inferiores ao modelo CARIMA e bem abaixo dos observados no histórico.

Com relação ao coeficiente de assimetria ( $\hat{\xi}$ ), observou-se uma superestimação em todas as estações e para os dois modelos, com exceção de Fazenda Maracanã para o modelo CARIMA. É interessante perceber que neste caso, em particular, todas as estatísticas ficaram próximas dos valores históricos. Por outro lado, observa-se que nos postos cujas assimetrias foram superestimadas as vazões mínimas e máximas das séries sintéticas ficaram abaixo e acima dos valores históricos, respectivamente, para ambos os modelos.

A Tabela 26 mostra as incertezas das estatísticas de curto termo para as séries sintéticas mensais. É interessante notar que para a média ( $\hat{\mu}$ ), desvio padrão ( $\hat{\sigma}$ ) e assimetria ( $\hat{\xi}$ ) as incertezas geradas pelo modelo SSS foram superiores ao modelo CARIMA, porém o mesmo não se pode dizer das estatísticas de vazões mínimas e máximas. Para essas últimas, os valores ficaram bastante próximos.

**Tabela 25 – Médias das estatísticas de curto termo das séries sintéticas mensais**

Posto	Série	$\hat{\mu}$ (m <sup>3</sup> /s)	$\hat{\sigma}$ (m <sup>3</sup> /s)	$\hat{\xi}$	Mín. (m <sup>3</sup> /s)	Máx. (m <sup>3</sup> /s)	$\hat{H}$
São Bento	Histórico	34,7	21,7	1,93	7,3	218,9	0,78
	SSS	34,7	21,9	2,34	0,9	217,0	0,78
	CARIMA	34,6	21,3	2,06	4,3	191,0	-
União da Vitória	Histórico	466,6	360,8	2,39	59,4	3641,6	0,74
	SSS	468,9	366,3	2,86	4,9	3822,9	0,73
	CARIMA	467,7	378,4	3,07	33,3	3912,3	-
Porto Amazonas	Histórico	68,3	48,6	2,26	8,0	448,0	0,74
	SSS	68,6	48,7	2,87	1,6	512,8	0,74
	CARIMA	68,8	51,1	2,77	5,8	517,5	-
Jangada do Sul	Histórico	25,8	20,0	2,07	2,8	183,6	0,71
	SSS	25,9	19,5	2,81	0,4	199,4	0,71
	CARIMA	25,9	22,0	3,07	1,9	225,4	-
Rio Negro	Histórico	67,5	47,5	1,90	12,0	362,5	0,76
	SSS	67,7	47,9	2,57	1,2	484,0	0,75
	CARIMA	66,7	45,7	2,38	6,4	429,2	-
Fazenda Maracaná	Histórico	9,0	6,9	2,74	0,9	77,9	0,73
	SSS	9,0	6,6	3,14	0,1	74,7	0,72
	CARIMA	8,9	6,8	2,78	0,7	67,7	-

SSS – modelo de múltiplas escalas temporais; CARIMA – modelo autorregressivo integrado e de médias móveis contemporâneo;  $\hat{\mu}$  – média;  $\hat{\sigma}$  – desvio padrão;  $\hat{\xi}$  – coeficiente de assimetria;  $\hat{H}$  – expoente de Hurst.

**Tabela 26 – Incertezas das estatísticas de curto termo das séries sintéticas mensais**

Posto	Série	$\hat{\mu}$ (m <sup>3</sup> /s)	$\hat{\sigma}$ (m <sup>3</sup> /s)	$\hat{\xi}$	Mín. (m <sup>3</sup> /s)	Máx. (m <sup>3</sup> /s)	$\hat{H}$
São Bento	SSS	4,8	1,9	0,82	2,1	61,4	0,04
	CARIMA	1,3	1,7	0,68	1,0	53,5	-
União da Vitória	SSS	55,0	35,3	1,15	9,2	1254,2	0,03
	CARIMA	22,5	42,6	1,25	9,7	1431,0	-
Porto Amazonas	SSS	8,6	4,8	1,10	2,9	163,1	0,03
	CARIMA	3,1	4,9	1,11	1,6	165,5	-
Jangada do Sul	SSS	2,9	1,8	0,89	0,4	56,5	0,04
	CARIMA	1,2	2,4	1,17	0,5	79,6	-
Rio Negro	SSS	8,8	4,4	0,90	2,5	147,3	0,03
	CARIMA	2,9	4,1	0,87	1,6	136,1	-
Fazenda Maracaná	SSS	1,2	0,8	1,19	0,2	23,7	0,04
	CARIMA	0,4	0,6	0,91	0,2	20,7	-

SSS – modelo de múltiplas escalas temporais; CARIMA – modelo autorregressivo integrado e de médias móveis contemporâneo;  $\hat{\mu}$  – média;  $\hat{\sigma}$  – desvio padrão;  $\hat{\xi}$  – coeficiente de assimetria;  $\hat{H}$  – expoente de Hurst.

A Figura 30 exibe os histogramas das estatísticas de curto termo para União da Vitória. Os mesmos gráficos elaborados para as demais estações hidrométricas consideradas no trabalho estão exibidos no Apêndice A.4. Seguindo as recomendações dadas quando da apresentação dos histogramas na comparação dos modelos em escala anual, sugere-se que o estudo dos gráficos tenha conotação qualitativa, devido às diferentes vazões obtidas em cada classe e para os dois modelos. É válido afirmar, contudo, que em todos os histogramas as classes de maiores frequências equivalem às estatísticas médias mostradas na Tabela 25. Por exemplo, para o histograma das médias a classe seis do modelo SSS equivale ao valor de 468,9 m<sup>3</sup>/s, enquanto que a classe cinco do modelo CARIMA se refere ao valor de 467,7 m<sup>3</sup>/s.

Ainda relevando o pequeno exemplo supramencionado, nota-se de imediato que as classes com maiores frequências absolutas se distinguem entre os modelos. Isso ocorreu para todas as estatísticas, com exceção do coeficiente e assimetria cuja distribuição foi semelhante. Outra diferença está no desvio padrão e na vazão máxima, para os quais o modelo SSS ofereceu maior uniformidade na distribuição de frequências. O CARIMA, obteve séries sintéticas com maior concentração de ocorrências na classe dois para desvio padrão e nas duas primeiras classes para a vazão máxima. Por outro lado, o SSS distribuiu melhor essas estatísticas, tendo representatividade em classes de maior ordem.

Evidentemente, a maior distinção entre os modelos ficou para as vazões mínimas. A análise inicial da Tabela 25 permitiu observar que o modelo SSS gerou valores de vazões mínimas bem abaixo do histórico. O histograma dessa estatística mostra que mais de 70% das séries geradas possuem valores mínimos locados na circunvizinhança de 4,9 m<sup>3</sup>/s, enquanto que a média histórica é de 59,4 m<sup>3</sup>/s. Além da diferença com relação ao histórico, essa pode ser apontada como a primeira grande diferença entre os resultados dos modelos SSS e CARIMA.

O fato do modelo SSS também gerar vazões máximas acima do histórico, porém respeitando as demais estatísticas apropriadamente, leva a crer que a geração desses valores extremos seja a confirmação de um dos propósitos de sua formulação, que é o de gerar eventos com baixa probabilidade de ocorrência (ver discussão na seção 1.4.1, página 70).

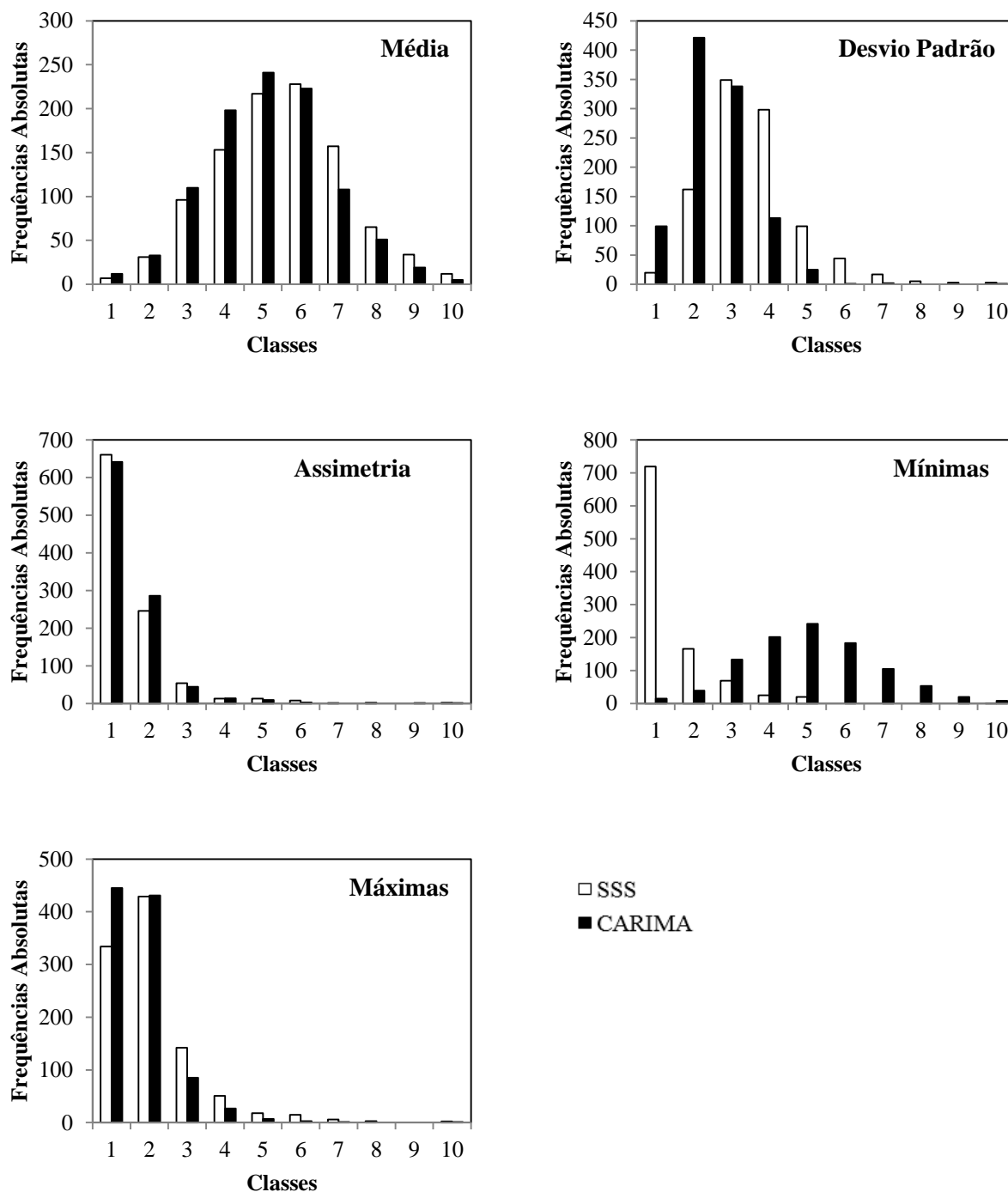


Figura 30 – Histogramas para as estatísticas de longo termo em União da Vitória, para séries sintéticas mensais

Como análise exclusiva da escala mensal, têm-se as estatísticas de média e desvio padrão mensais, exibidos na Figura 31. Nota-se que para o modelo SSS todas as estatísticas foram preservadas, com subestimações muito sutis somente em Jangada do Sul e Fazenda Maracanã para o desvio padrão. No entanto, o modelo CARIMA obteve maiores diferenças nos desvios padrão em todas as estações. O pior caso foi em Jangada do Sul, na qual essas estatísticas alternaram períodos acima e abaixo do histórico. Essa pode ser apontada a segunda diferença significativa entre os modelos SSS e CARIMA para a escala mensal.

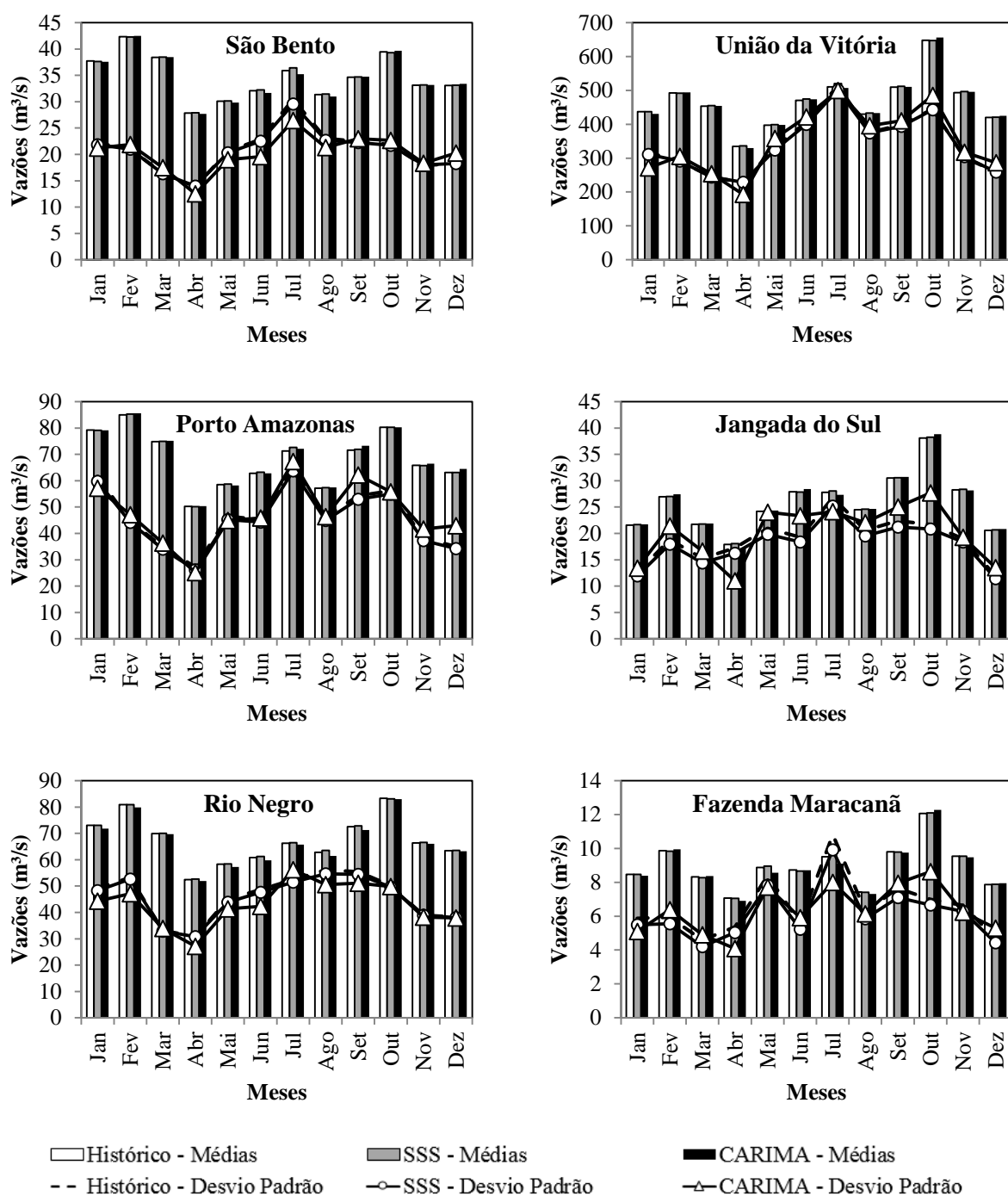
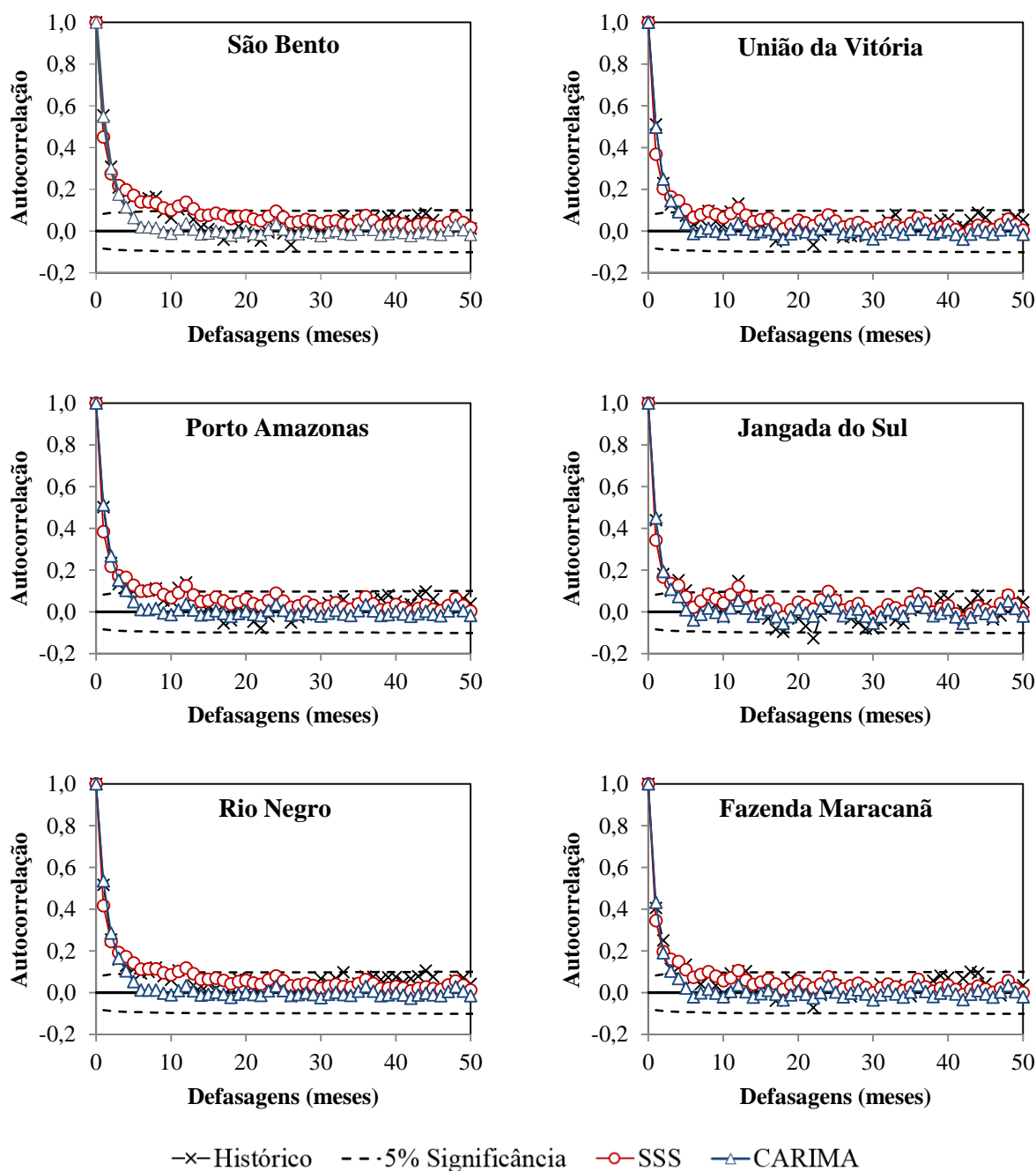


Figura 31 – Estatísticas mensais das séries sintéticas



A Figura 32 mostra a comparação entre as FAC das séries históricas e sintéticas obtidas pelos modelos, utilizando defasagem máxima de 50 meses. Diferentemente da escala anual, é evidente que as autocorrelações apresentam decaimento mais lento; São Bento e Rio Negro, por exemplo, mostram autocorrelações estatisticamente nulas somente depois de 12 meses e todas as estações exibem autocorrelações significativas no 24° mês. Todos esses comportamentos foram muito bem reproduzidos pelo modelo SSS, com destaque a São Bento, cuja FAC tem o decaimento mais lento dentre as séries estudadas. O modelo SSS foi capaz de reproduzir perfeitamente esse comportamento.



**Figura 32 – FAC das séries sintéticas mensais**

Sobre as defasagens sazonais, as séries sintéticas geradas pelo modelo SSS exibiram valores significativos para 12 e 24 meses, mostrando persistência de até dois anos, com exceção de Fazenda Maracanã que se limitou ao primeiro ano. Em Jangada do Sul, as defasagens sazonais significativas chegaram a quatro anos. Ainda, São Bento, União da Vitória, Porto Amazonas e Jangada do Sul exibem concentrações de autocorrelações históricas negativas na circunvizinhança dos 24 meses que não foram reproduzidas pelas séries sintéticas. Porém estas, na maioria, são estatisticamente nulas e não influenciam na modelagem.

Por outro lado, o modelo CARIMA não mostrou a mesma precisão na reprodução das FAC históricas. No primeiro mês de defasagem os valores de autocorrelação obtidos por ele foram melhores do que os resultantes do modelo SSS, entretanto todos valores das defasagens superiores foram subestimados. Em outras palavras, os decaimentos das FAC obtidas pelo modelo CARIMA foram mais rápidos, evidenciando seu comportamento teórico criticado por Koutsoyiannis (2002) (ver comentários no final da seção 1.4.1.1.1, página 75). Tem-se, com isso, a terceira diferença no comparativo entre os modelos SSS e CARIMA, com clara superioridade do primeiro.

As estatísticas de longo termo para os modelos estão mostradas na Tabela 27. Deve-se alertar para análises sobre o número de corridas ( $n_{\Lambda}$ ) nas séries mais curtas (Jangada do Sul e Fazenda Maracanã) quando comparadas ao histórico, pois, como mencionado anteriormente, todas as séries sintéticas foram geradas com comprimento igual ao da série mais longa. Dessa maneira o número de corridas fica naturalmente superestimado. Ainda assim é possível perceber distinções entre os modelos: em todos os postos o SSS gerou mais corridas do que o CARIMA, entretanto os comprimentos médios ( $\bar{\Lambda}$ ) e máximos ( $\Lambda_{\max}$ ) de cada uma (tanto em termos absolutos quanto em termos de vazões acumuladas) ficaram próximos. Em outras palavras, as séries sintéticas do modelo SSS apresentam um maior número de conglomerados de meses consecutivos com vazões médias acima ou abaixo da média de longo termo, o que pode ser consequência da estrutura de persistência mostrada nas FACs da Figura 32. Lembre-se que esse tipo de comportamento é característico do Efeito Joseph (ver seção 1.3, página 64), outro elemento que, por definição, é contemplado pelo modelo SSS. Essa é a quarta distinção apontada entre os modelos.

A respeito dos déficits, as estimativas das séries sintéticas do modelo SSS ficaram maiores em magnitude do que as do modelo CARIMA e também do que as das séries históricas, tanto para o déficit médio ( $\bar{\Delta}$ ) quanto para o máximo déficit acumulado ( $\Delta_{\max}$ ). Devido à relação dessas estatísticas com períodos de estiagens, entende-se que os resultados são reflexo direto

dos baixos valores de vazão mínima obtidos. As maiores diferenças foram observadas em União da Vitória, para a qual os déficits estimados foram significativamente maiores. Por sua vez, os resultados do modelo CARIMA ficaram próximos das estatísticas históricas.

**Tabela 27 – Médias das estatísticas de longo termo das séries sintéticas mensais**

Posto	Série	$n_{\Lambda}$	$\bar{\Lambda}$	$\Lambda_{\max}$	$\bar{\Lambda}$ (m <sup>3</sup> /s)	$\Lambda_{\max}$ (m <sup>3</sup> /s)	$\bar{\Delta}$ (m <sup>3</sup> /s)	$\Delta_{\max}$ (m <sup>3</sup> /s)
<b>São Bento</b>	<i>Histórico</i>	97	5	23	113,7	439,7	60,7	389,0
	SSS	99	5	26	110,8	452,1	120,9	670,0
	CARIMA	95	6	24	120,8	467,1	48,1	299,0
<b>União da Vitória</b>	<i>Histórico</i>	102	5	24	1424,3	5973,0	936,0	4712,3
	SSS	107	5	24	1291,7	4958,9	1877,9	9955,7
	CARIMA	97	6	24	1526,7	5928,4	1043,6	5586,1
<b>Porto Amazonas</b>	<i>Histórico</i>	91	5	23	207,9	777,4	140,2	732,0
	SSS	104	5	25	210,1	829,0	243,5	1331,7
	CARIMA	96	6	24	228,2	884,5	135,7	753,9
<b>Jangada do Sul</b>	<i>Histórico</i>	82	5	23	73,2	314,9	46,0	185,1
	SSS	112	5	22	71,9	276,1	75,8	435,4
	CARIMA	103	5	23	78,4	300,1	55,9	301,2
<b>Rio Negro</b>	<i>Histórico</i>	104	5	21	216,3	764,2	142,5	881,3
	SSS	102	5	25	203,1	794,1	264,8	1431,5
	CARIMA	95	6	24	226,9	867,1	113,3	658,7
<b>Fazenda Maracaná</b>	<i>Histórico</i>	81	5	24	26,3	105,3	22,0	86,6
	SSS	108	5	24	25,6	102,9	29,4	164,3
	CARIMA	105	5	21	26,8	102,0	14,9	85,2

SSS – modelo de múltiplas escalas temporais; CARIMA – modelo autorregressivo integrado e de médias móveis contemporâneo;  $n_{\Lambda}$  – número de corridas;  $\bar{\Lambda}$  – durações médias das corridas;  $\Lambda_{\max}$  – durações máximas das corridas;  $\bar{\Delta}$  – déficits médios;  $\Delta_{\max}$  – máximo déficit acumulado.

A Tabela 28 exibe as incertezas das estatísticas de longo termo para o conjunto das 1000 séries sintéticas geradas. A exemplo dos resultados da escala anual, maiores incertezas foram identificadas nas estatísticas relacionadas aos eventos extremos, como para durações máximas das corridas ( $\Lambda_{\max}$ ), principalmente em termos de vazões acumuladas, e para o máximo déficit acumulado ( $\Delta_{\max}$ ). Nesse caso, o modelo SSS apresentou os maiores valores de incertezas, o que permite concluir que as séries sintéticas geradas por ele são mais distintas entre si quando comparadas às séries geradas pelo modelo CARIMA.

**Tabela 28 – Incertezas das estatísticas de longo termo das séries sintéticas mensais**

Posto	Série	$n_{\Lambda}$	$\bar{\Lambda}$	$\Lambda_{\max}$	$\bar{\Lambda}$ (m <sup>3</sup> /s)	$\Lambda_{\max}$ (m <sup>3</sup> /s)	$\bar{\Delta}$ (m <sup>3</sup> /s)	$\Delta_{\max}$ (m <sup>3</sup> /s)
<b>São Bento</b>	SSS	7,1	0,4	8,6	23,9	155,5	71,9	300,4
	CARIMA	6,0	0,4	5,3	9,5	102,4	13,3	90,3
<b>União da Vitória</b>	SSS	6,8	0,3	6,6	256,0	1559,3	1133,0	4171,9
	CARIMA	6,0	0,4	5,5	133,1	1319,3	317,5	1720,8
<b>Porto Amazonas</b>	SSS	6,5	0,4	6,6	40,5	265,5	145,9	543,5
	CARIMA	5,8	0,4	5,6	18,6	201,2	38,1	227,1
<b>Jangada do Sul</b>	SSS	5,7	0,3	5,5	12,4	76,4	49,0	187,5
	CARIMA	6,0	0,4	5,1	6,1	67,6	14,9	86,0
<b>Rio Negro</b>	SSS	7,1	0,4	7,2	43,1	251,0	170,2	666,8
	CARIMA	5,8	0,4	5,3	18,0	196,3	31,6	188,9
<b>Fazenda Maracanã</b>	SSS	6,1	0,4	6,2	5,5	30,8	16,6	66,6
	CARIMA	6,0	0,3	5,2	2,1	24,7	4,2	25,1

SSS – modelo de múltiplas escalas temporais; CARIMA – modelo autorregressivo integrado e de médias móveis contemporâneo;  $n_{\Lambda}$  – número de corridas;  $\bar{\Lambda}$  – durações médias das corridas;  $\Lambda_{\max}$  – durações máximas das corridas;  $\bar{\Delta}$  – déficits médios;  $\Delta_{\max}$  – máximo déficit acumulado.

A Figura 33 mostra os histogramas obtidos para as estatísticas de longo termo em União da Vitória. Da mesma forma, os histogramas para as demais séries são mostrados no Apêndice A.5 do trabalho. O que se percebe é um comportamento similar entre os modelos para as estatísticas de duração média e vazão acumulada média nas corridas. No entanto, para a duração máxima, déficit médio e máximo déficit acumulado, o modelo SSS resultou em um agrupamento de ocorrências nas primeiras classes, o que se distinguiu do modelo CARIMA. Esse comportamento é, de certa forma, um reflexo do que já foi anteriormente comentado na análise da Tabela 27.

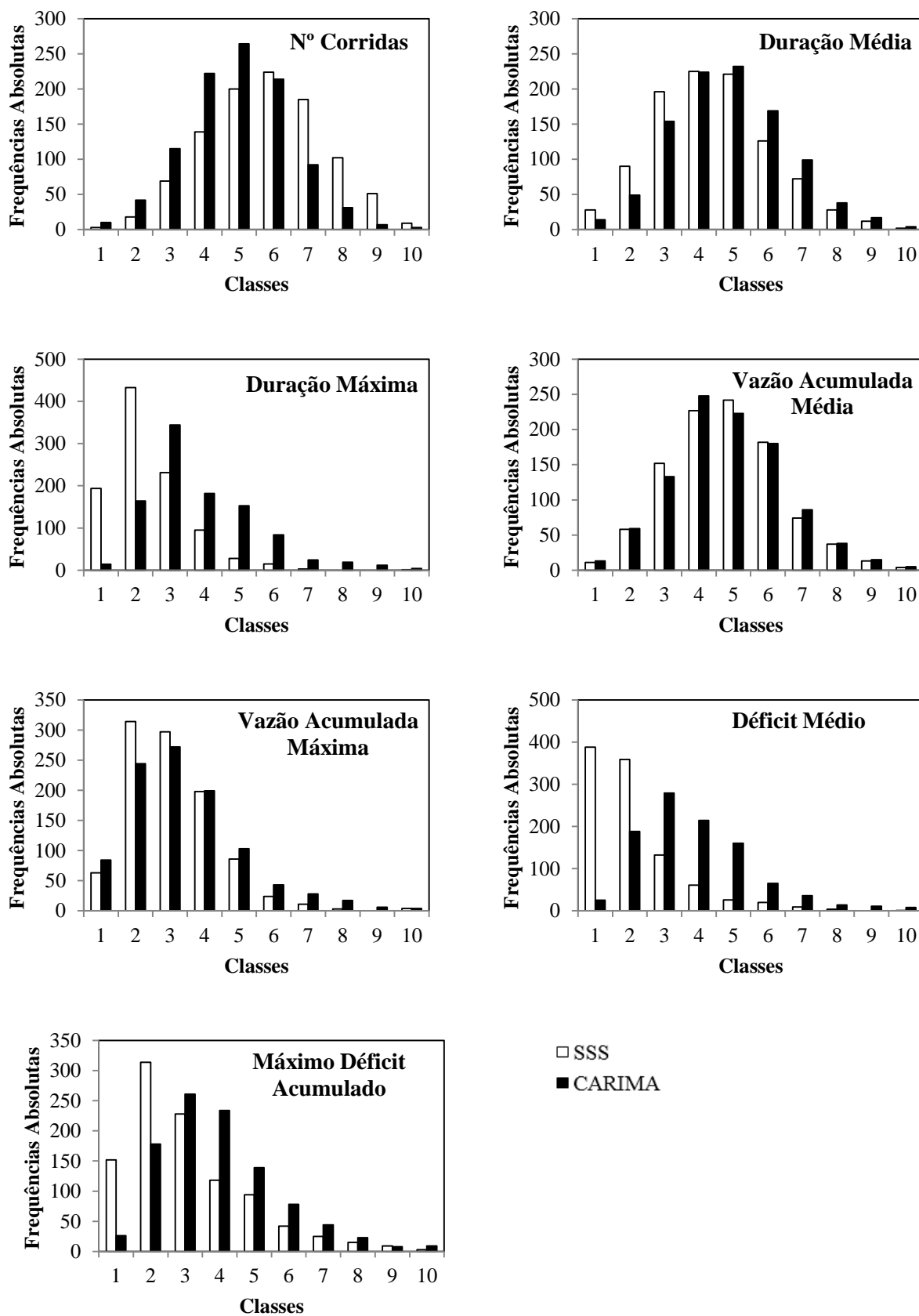
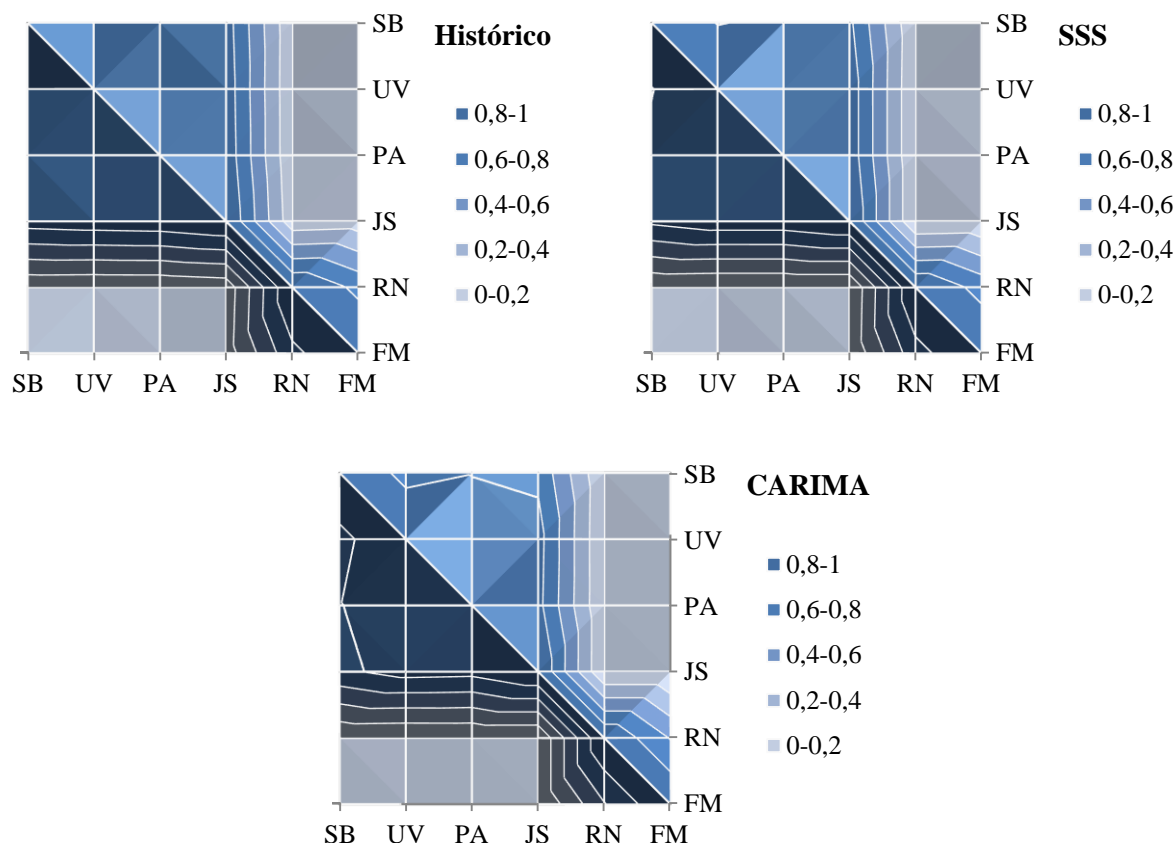


Figura 33 – Histogramas das estatísticas de longo termo em União da Vitória, para séries sintéticas mensais

Por fim, o último resultado a ser mostrado são as correlações espaciais calculadas pelos modelos. Novamente, as correlações históricas foram determinadas para o período em comum a todas as séries. Os diagramas resultantes para a análise mensal são mostrados na Figura 34.



**Figura 34 – Mapas de correlações espaciais para as séries mensais**

Primeiramente, nota-se que a matriz de correlações mensais é bastante diferente da anual, sendo caracterizada por valores baixos em Rio Negro e Fazenda Maracanã. São Bento, União da Vitória, Porto Amazonas e Jangada do Sul, por outro lado, formam um núcleo de correlações fortes na bacia estudada. Essa estrutura heterogênea é interessante do ponto de vista da validação do modelo multivariado, pois agrega uma maior complexidade ao seu ajuste. Mesmo assim, a estrutura foi bem modelada pela formulação SSS, que apresentou apenas uma pequena diferença entre São Bento e União da Vitória. Em valores, essa diferença foi de 0,07 (histórico 0,86 e modelado 0,79) e não representa preocupações. Contudo, maiores imprecisões foram anotadas pelo modelo CARIMA, principalmente entre São Bento, União da Vitória, Porto Amazonas e Jangada do Sul. É importante apontar que as correlações entre esses postos são altas (superiores a 0,70) e, portanto, sua reprodução é de grande importância. Dessa maneira, considera-se essa a quinta e última distinção entre os modelos.

Depois de apresentadas e discutidas as diversas características obtidas com as séries sintéticas geradas pelos dois modelos estudados, entende-se que o principal resultado obtido foi a viabilidade da utilização do modelo SSS sazonal através do algoritmo proposto SMAD. Os índices estatísticos empregados nas análises foram preservados e, além disso, as características teóricas inerentes do modelo foram evidenciadas nas séries geradas, fazendo com que esse modelo apareça como alternativa às modelagens existentes.

#### **5.4 Resumo do capítulo**

Apresentaram-se neste capítulo diversos resultados numéricos obtidos com os modelos de geração de séries sintéticas trabalhados nesta tese. Em um primeiro momento, foram mostrados detalhes acerca da implementação e ajuste dos modelos, com foco na formulação SSS. A intenção foi entender o comportamento dos parâmetros do modelo e, conseqüentemente, seu funcionamento na geração de séries sintéticas. Da mesma forma, apresentou-se o detalhamento das formulações MAR(1) e CARIMA, utilizadas como modelos alternativos, cujos resultados serviram como base para comparações.

Na sequência o foco recaiu sobre as séries sintéticas em escala anual, as quais foram geradas utilizando três abordagens: SSS, MAR(1) e MAR(1)\*, o último como uma variante do modelo original, entretanto considerando a remoção da tendência linear de crescimento das séries históricas. Mostrou-se que os resultados das formulações SSS e MAR(1) foram relativamente próximos, porém algumas diferenças puderam ser notadas no modelo SSS, principalmente relacionadas à melhor representação dos dados extremos de cheias e secas. Por outro lado, mostrou-se que a utilização do modelo MAR(1)\* deve ser feita com cuidado, pois foram detectadas algumas distorções na reprodução do coeficiente de assimetria e em outras estatísticas. Foram também detectadas vazões médias negativas em alguns casos.

Por fim, mostraram-se os resultados do ajuste dos modelos mensais SSS e CARIMA, o primeiro através do algoritmo SMAD proposto pelo autor desta tese. De um modo geral, ambos foram capazes de preservar as estatísticas históricas, porém identificaram-se cinco distinções entre eles: (i) o modelo SSS gerou vazões mínimas bem abaixo do CARIMA (e do histórico); (ii) as estatísticas mensais, principalmente de desvio padrão, foram mais bem reproduzidas com o modelo SSS; (iii) a estrutura de persistência das séries foi melhor modelada com a formulação SSS; (iv) as séries sintéticas do modelo SSS apresentaram um maior número de conglomerados de meses consecutivos com vazões acima ou abaixo da média de longo termo e (v) a estrutura de correlações espaciais foi mais bem reproduzida pelo modelo SSS.





## CONCLUSÃO

O propósito deste trabalho consistiu no estudo do modelo de múltiplas escalas temporais (SSS), visando seu emprego em séries de vazões médias. A motivação para a consideração desse modelo extrapolou as peculiaridades matemáticas ou estatísticas, tendo nascido da sua própria filosofia de modelagem. Ao ponderar sobre a influência de eventos em diversas escalas temporais nas séries de vazões médias e considerar variabilidades (ou comportamentos não estacionários) como características esperadas de um fenômeno natural, o modelo SSS utiliza premissas que rumam na direção oposta dos principais trabalhos da hidrologia estocástica atual.

Como explicado em diversos momentos da tese, aplicações do modelo SSS estavam limitadas à excepcional série de níveis mínimos anuais do rio Nilo, portanto uma aplicação em séries comuns à hidrologia operacional exigiu uma contextualização completa do problema. Ademais, o interesse não estava somente em aplicar o modelo, mas em comparar os resultados com formulações conhecidas e tradicionalmente utilizadas. Com esse raciocínio foram elaborados os objetivos principal e específicos descritos na introdução do trabalho, os quais, depois dos resultados alcançados no trabalho, podem ser considerados plenamente atingidos.

Contudo, entende-se que as contribuições desta tese foram além das análises comparativas mostradas, podendo-se destacar:

- A fundamentação teórica mostrada no capítulo 1 que contemplou, além de outros assuntos, o estado-da-arte das técnicas utilizadas para a análise de séries hidrológicas.
- A sistemática elaborada para implementação computacional dos três modelos estocásticos considerados, incluindo um método inédito para o modelo SMAD proposto como inovação da tese.
- A profunda análise no comportamento estacionário das séries consideradas, mostrando que mesmo utilizando as técnicas mais recomendadas pela literatura os resultados podem ser contraditórios, dificultando o julgamento do hidrólogo.
- O detalhamento acerca do modelo SSS, incluindo o comportamento de seus parâmetros, podendo ser considerado como outro ponto de ineditismo do trabalho.
- A consequência da aplicação do estimador não paramétrico de Sen para a remoção da tendência linear das séries anuais, que resultou em uma severa subestimação do coeficiente de assimetria das séries sintéticas geradas.

- A comprovação da aplicabilidade do modelo SSS nas séries hidrológicas da região Sul do Brasil, através dos algoritmos SMA e SMAD para séries anuais e mensais, respectivamente.

Os argumentos supramencionados juntamente com as análises comparativas dos modelos SSS *versus* Markoviano e SSS *versus* CARIMA permitem uma sólida conclusão acerca da hipótese levantada na introdução do trabalho. A partir das duas comparações, fica evidenciado que os modelos de múltiplas escalas temporais são, de fato, viáveis para utilização na hidrologia operacional. Entretanto, a fraca estrutura de persistência existente nas séries anuais da bacia do rio Iguaçu não permitiu a identificação de uma melhoria significativa por parte do modelo SSS. Por outro lado, a aplicação nas séries mensais gerou cinco pontos de distinção entre as formulações SSS e CARIMA, três deles representando uma melhora real na representatividade das séries sintéticas geradas pelo modelo SSS: as estatísticas mensais, a estrutura de persistência e a matriz de correlações espaciais. Dada a importância desses elementos, considera-se o modelo SSS uma válida alternativa para a modelagem de séries de vazões médias.

O estudo dos modelos de múltiplas escalas temporais apresentado nesta tese pode ser entendido como uma abordagem inicial ao problema e que possibilita outras investigações. Dessa maneira, recomenda-se como trabalhos futuros a aplicação do modelo SSS, anual e mensal, a séries hidrológicas com uma estrutura de persistência mais relevante, como é o caso de séries de vazões médias de rios das regiões Sudeste e Centro-Oeste brasileiras. Da mesma forma, pode-se estender as métricas de análise para outros índices relacionados ao armazenamento e/ou déficits de água em reservatórios.

Do ponto de vista teórico, o modelo SSS pode ser aperfeiçoado através da utilização de métodos de estimação distintos dos apresentados neste trabalho. Em específico, o expoente de Hurst e a variância das séries podem ser estimados com maior precisão através do método da máxima verossimilhança proposto em Tyralis e Koutsoyiannis (2011). Da mesma forma, a matriz-parâmetro **B** responsável pela geração de resíduos independentes no tempo, mas correlacionados no espaço, pode ser estimada pelo método alternativo mencionado em Koutsoyiannis (2000), que se distingue da tradicional decomposição de Cholesky.

Ainda em relação ao modelo SSS, sugere-se que ele seja ampliado para propósitos de previsão de séries hidrológicas (ver Koutsoyiannis, 2000) e de análises que envolvam a incerteza. De acordo com Koutsoyiannis (2006), o modelo de múltiplas escalas temporais oferece uma nova percepção dessa questão, significativamente distinta das estimativas existentes até então.

Outra sugestão seria a comparação entre os modelos SSS e ARFIMA (autorregressivo com integração fracionada e de médias móveis). O AFIRMA é, como o próprio nome sugere, uma variante do modelo ARIMA, porém com o parâmetro de diferenciação situado entre  $0 \leq d \leq 1$ . Trata-se de uma formulação utilizada para modelar processos de longa memória, ou com forte persistência (Baillie e Chung, 2002). Por esse motivo, a comparação com o modelo SSS pode gerar resultados interessantes.

Por fim, o estudo específico das séries históricas hidrológicas pode ser ampliado com o uso de técnicas distintas das apresentadas nesse trabalho. Cita-se como exemplo o modelo GAMLSS (do inglês *Generalized Additive Models for Location, Scale and Shape*), ferramenta que permite um melhor entendimento das periodicidades e outras características presentes nas séries. Zhang *et al.* (2014) apresentam um trabalho recente utilizando essa técnica.



## REFERÊNCIAS BIBLIOGRÁFICAS

ABI-ZEID, I., PARENT, É., BOBÉE, B. The stochastic modeling of low flows by the alternating point processes approach: methodology and application. **Journal of Hydrology**, v. 285, n. 1-4 , p. 41–61, 2004.

ADAMOWSKI, J. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. **Journal of Hydrology**, v. 353, n. 3-4, p. 247– 266, 2008.

ADAMOWSKI, K. Spectral density of a river flow time series. **Journal of Hydrology**, v. 14, p. 43–52, 1971.

AGÊNCIA NACIONAL DE ÁGUAS (ANA). **Hidroweb** - Sistema de Informações Hidrológicas. Disponível em <http://hidroweb.ana.gov.br>. Acesso em 05/02/2014.

AGÊNCIA NACIONAL DE ÁGUAS (ANA), SUPERINTENDÊNCIA DA GESTÃO DA REDE HIDROMETEOROLÓGICA (SGH). **Orientações para consistência de dados fluviométricos**. Brasília, 2012a. Disponível em: <http://www2.ana.gov.br/Paginas/servicos/informacoeshidrologicas/monitoramentohidro.aspx>. Acesso em 05/02/2014.

AGÊNCIA NACIONAL DE ÁGUAS (ANA), SUPERINTENDÊNCIA DA GESTÃO DA REDE HIDROMETEOROLÓGICA (SGH). **Orientações para Operação de Estações Hidrométricas**. Brasília, 2012b. Disponível em: <http://www2.ana.gov.br/Paginas/servicos/informacoeshidrologicas/monitoramentohidro.aspx>. Acesso em 05/02/2014.

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, p. 716–723, 1974.

ALEXANDRE, G. R., BAPTISTA, M. B., NAGHETTINI, M. Estudo para Identificação de Tendências do Regime Pluvial na Região Metropolitana de Belo Horizonte a Partir de Métodos Estatísticos. **Revista Brasileira de Recursos Hídricos**, v. 15, n. 2, p. 115–126, 2010.

ANCTIL, F., COULIBALY, P. Wavelet Analysis of the Interannual Variability in Southern Québec Streamflow. **Journal of Climate**, v. 17, n. 1, p. 163–173, 2004.

ANDREO, B., JIMENEZ, P., DURAN, J., CARRASCO, F., VADILLO, I., MANGIN, A. Climatic and hydrological variations during the last 117–166 years in the south of the Iberian Peninsula, from spectral and correlation analyses and continuous wavelet analyses. **Journal of Hydrology**, v. 324, n.1-4, p. 24–39, 2006.

AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA (ANEEL). **BIG - Banco de Informações de Geração**. Disponível em: <http://www.aneel.gov.br/aplicacoes/capacidadebrasil/capacidadebrasil.cfm>. Acesso em: 05/02/2014.

ASSANI, A. A., QUESSY, J.-F., MESFIOUI, M., MATTEAU, M. An example of application: The ecological “natural flow regime” paradigm in hydroclimatology. **Advances in Water Resources**, v. 33, n. 5, p. 537–545, 2010.

- BAILLIE, R. T., CHUNG, S.-K. Modeling and forecasting from trend-stationary long memory models with applications to climatology. **International Journal of Forecasting**, v. 18, n. 2, p. 215–226, 2002.
- BATISTA, A. L., FREITAS JR., S. A. de, DETZEL, D. H. M., MINE, M. R. M. FILL, H. D. O. A., FERNANDES, C. KAVISKI, E. Verificação da estacionariedade de séries hidrológicas no Sul-Sudeste do Brasil. In SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 18., 2009. Campo Grande. **Anais...** Porto Alegre: ABRH, 2009, p. 1-19.
- BOES, D. C., SALAS, J. D. Nonstationarity of the mean and the Hurst phenomenon. **Water Resources Research**, v. 14, n. 1, p. 135–143, 1978.
- BORMANN, H., PINTER, N., ELFERT, S. Hydrological signatures of flood trends on German rivers: Flood frequencies, flood heights and specific stages. **Journal of Hydrology**, v. 404, n. 1-2, p. 50–66, 2011.
- BOX, G. E. P., COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society, Series B**, v. 26, p. 211–243, 1964.
- BOX, G. E. P., JENKINS, G. M., REINSEL, G. C. **Time Series Analysis Forecasting and Control** 4ª ed. New Jersey: John Wiley & Sons, 2008.
- BRAGA, R. S., ROCHA, V. F., GONTIJO, E. A. Revisão das séries de vazões naturais nas principais bacias hidrográficas do Sistema Interligado Nacional. In SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 18., 2009. Campo Grande. **Anais...** Porto Alegre: ABRH, 2009, p. 1-20.
- BREINHOLT, A., MOLLER, J. K., MADSEN, H., MIKKELSEN, P. S. A formal statistical approach to representing uncertainty in rainfall – runoff modelling with focus on residual analysis and probabilistic output evaluation – Distinguishing simulation and prediction. **Journal of Hydrology**, v. 472-473, p. 36–52, 2012.
- BROWN, M. B., FORSYTHE, A. B. Robust tests for the equality of variances. **Journal of the American Statistical Association**1, v. 69, n. 346, p. 364–367, 1974.
- CAVALCANTI, I. F. A., FERREIRA, N. J., DIAS, M. A. F. da S., SILVA, M. G. A. J. da. **Tempo e clima no Brasil** São Paulo: Oficina de Textos, 2009.
- CELKA, P., VETTER, R., GYSELS, E., HINE, T. J. Dealing with Randomness in Biosignals. In: SCHELTER, B., WINTERHALDER, M., TIMMER, J. (Org.), **Handbook of Time Series Analysis**. Weinheim: Wiley-VCH, 2006, cap. 5.
- CHEBANA, F., OUARDA, T. B. M. J., DUONG, T. C. Testing for multivariate trends in hydrologic frequency analysis. **Journal of Hydrology**, v. 486, p. 519–530, 2013.
- CHOE, G. H. **Computational Ergodic Theory**. Berlin Heidelberg: Springer-Verlag, 2005.
- CHU, P.-S., KATZ, R. W. Spectral estimation from time series models with relevance to the Southern Oscillation. **Journal of Climate**, v. 2, n. 86-90, 1989.
- CLARKE, R. T. Hydrological prediction in a non-stationary world. **Hydrology and Earth System Sciences**, v. 11, n. 1, p. 408–414. 2007.

- CLUIS, D., LABERGE, C. Climate Change and Trend Detection in Selected Rivers within the Asia-Pacific Region Climate. **Water International**, v. 26, n. 3, p. 411–424, 2001.
- CRAIGMILE, P. F., GUTTORP, P., PERCIVAL, D. B. Trend assessment in a long memory dependence model using the discrete wavelet transform. **Environmetrics**, v. 15, n. 4, p. 313–335, 2004.
- CUNDERLIK, J. M., BURN, D. H. Linkages between Regional Trends in Monthly Maximum Flows and Selected Climatic Variables. **Journal of Hydrologic Engineering**, v. 9, n. 4, p. 246–256, 2004.
- DELGADO, J. M., APEL, H., MERZ, B. Flood trends and variability in the Mekong river. **Hydrology and Earth System Sciences**, v. 14, n. 3, p. 407–418, 2010.
- DETZEL, D. H. M., BESSA, M. R., VALLEJOS, C. A. V, SANTOS, A. B., THOMSEN, L. S. Estacionariedade das Afluências às Usinas Hidrelétricas Brasileiras. **Revista Brasileira de Recursos Hídricos**, v. 16, n. 3, p. 95–111, 2011.
- ELSHORBAGY, A. A., PANU, U. S., SIMONOVIC, S. P. Group-based estimation of missing hydrological data : I. Approach and general methodology. **Hydrological Sciences Journal**, v. 45, n. 6, p. 849–866, 2000.
- FERREIRA, D. F. **Estatística Multivariada**. Lavras: UFLA, 2008.
- FILL, H. D. Análise da estacionariedade das vazões do rio iguaçu em União da Vitória. In: SIMPÓSIO BRASILEIRO DE RECURSOS HÍDRICOS, 19., 2011. Maceió. **Anais...** Porto Alegre: ABRH, 2011.
- FLEMING, S. W., WEBER, F. A. Detection of long-term change in hydroelectric reservoir inflows: Bridging theory and practise. **Journal of Hydrology**, v. 470-471, p. 36–54, 2012.
- FRAEDRICH, K., JIANG, J., GERSTENGARBE, F.-W., WERNER, P. C.. Multiscale detection of abrupt climate changes: application to Nile river flood levels. **International Journal of Climatology**, v. 17, p. 1301–1315, 1997.
- GAUTAM, M. R., ACHARYA, K., TULADHAR, M. K. Upward trend of streamflow and precipitation in a small, non-snow-fed, mountainous watershed in Nepal. **Journal of Hydrology**, v. 387, n. 3-4, p. 304–311, 2010.
- GENTA, J. L., PEREZ-IRIBARREN, G., MECHOSO, C. R. A Recent Increasing Trend in the Streamflow of Rivers in Southeastern South America. **Journal of Climate**, v. 11, n. 11, p. 2858–2862, 1998.
- GERSTENGARBE, F., WERNER, P. C. Estimation of the beginning and end of recurrent events within a climate regime. **Climate Research**, v. 11, p. 97–107, 1999.
- GHANBARPOUR, M. R., ABBASPOUR, K. C., JALALVAND, G., MOGHADDAM, G. A. Stochastic modeling of surface stream flow at different time scales: Sangsoorakh karst basin, Iran. **Journal of Cave and Karst Studies**, v. 72, n. 1, p. 1–10, 2010.
- GOMIDE, F. L. S. Markovian inputs and the Hurst phenomenon. **Journal of Hydrology**, v. 37, p. 23–45, 1978.

- GUETTER, A. K., PRATES, J. E. Degrau climático nas séries de vazões das bacias brasileiras. In: CONGRESSO BRASILEIRO DE METEOROLOGIA, 12., Foz do Iguaçu. **Anais...** Rio de Janeiro: SBMET, 2002.
- GUIMARÃES, R. C., SANTOS, E. G. Principles of Stochastic Generation of Hydrologic Time Series for Reservoir Planning and Design : Case Study. **Journal of Hydrologic Engineering**, v. 16, n. 11, p. 891–898, 2011.
- HALTINER, J. P., SALAS, J. D. Development and testing of a multivariate, seasonal ARMA(1,1) model. **Journal of Hydrology**, v. 104, p. 247–272, 1988.
- HAMILTON, J. D. **Time series analysis**. New Jersey: Princeton University Press, 2004.
- HAO, Z., SINGH, V. P. Single-site monthly streamflow simulation using entropy theory. **Water Resources Research**, v. 47, n. W09528, 2011.
- HIPEL, K. W., McLEOD, A. I. **Time Series Modelling of Water Resources and Environmental Systems**, 1994. Disponível em: <http://www.stats.uwo.ca/faculty/aim/1994Book/>. Acesso em: 05/02/2014.
- HURST, H. E. Long term storage capacities of reservoirs. **Transactions ASCE**, v. 116, p. 776–808, 1951.
- INTERNATIONAL WATER HISTORY ASSOCIATION (IWHA). **The Nilometer in Cairo**. Disponível em: <http://www.waterhistory.org/histories/cairo/>. Acesso em: 04/03/2015.
- KAHYA, E., KALAYCI, S. Trend analysis of streamflow in Turkey. **Journal of Hydrology**, v. 289, n. 1-4, p. 128–144, 2004.
- KELMAN, J. Modelos Estocásticos no Gerenciamento de Recursos Hídricos. In: BARTH, F. T., POMPEU, C. T., FILL, H. D., TUCCI, C. E. M., KELMEN, J., BRAGA, B. P. F. **Modelos para Gerenciamento de Recursos Hídricos I**. São Paulo: Nobel/ABRH, 1987, cap. 4.
- KENDALL, M. G. **Rank Correlation Methods**. Londres: Griffin, 1975.
- KIRSCH, B. R., CHARACKLIS, G. W., ZEFF, H. B. Evaluating the Impact of Alternative Hydro-Climate Scenarios on Transfer Agreements: A Practical Improvement for Generating Synthetic Streamflows. **Journal of Water Resources Planning and Management**, v. 139, p. 396-406, 2012.
- KITE, G. Use of time series analysis to detect climatic change. **Journal of Hydrology**, v. 111, p. 259–279, 1989.
- KLEMEŠ, V. The Hurst Phenomenon: A Puzzle? **Water Resources Research**, v. 10, n. 4, p. 675–688, 1974.
- KOTTEGODA, N. T., NATALE, L., RAITERI, E. Simulation of climatic series with nonstationary trends and periodicities. **Journal of Hydrology**, v. 398, n. 1-2, p. 33–43, 2011.
- KOUTSOYIANNIS, D. A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series. **Water Resources Research**, v. 36, n. 6, p. 1519–1533, 2000.



- KOUTSOYIANNIS, D. The Hurst phenomenon and fractional Gaussian noise made easy. **Hydrological Sciences Journal**, v. 47, n. 4, p. 573–596, 2002.
- KOUTSOYIANNIS, D. Climate change, the Hurst phenomenon, and hydrological statistics. **Hydrological Sciences Journal**, v. 48, n. 1, p. 3–24, 2003a.
- KOUTSOYIANNIS, D. Climate change, the Hurst phenomenon, and hydrological statistics – supplementary material. **Hydrological Sciences Journal**, v. 48, n. 1, p. 3–24, 2003b.
- KOUTSOYIANNIS, D. Nonstationarity versus scaling in hydrology. **Journal of Hydrology**, v. 324, n. 1-4, p. 239–254, 2006.
- KOUTSOYIANNIS, D. Hurst-Kolmogorov Dynamics and Uncertainty. **Journal of the American Water Resources Association**, v. 47, n. 3, p. 481–495, 2011.
- KOUTSOYIANNIS, D. Hydrology and change. **Hydrological Sciences Journal**, v. 58, n. 6, p. 1177–1197, 2013.
- KUNDZEWICZ, Z. W., ROBSON, A. J. Change detection in hydrological records — a review of the methodology. **Hydrological Sciences**, v. 49, n. 1, p. 7–20, 2004.
- LANGOUSIS, A., KOUTSOYIANNIS, D. A stochastic methodology for generation of seasonal time series reproducing overyear scaling behaviour. **Journal of Hydrology**, v. 322, n. 1-4, p. 138–154, 2006.
- LEE, T. Serial dependence properties in multivariate streamflow simulation with independent decomposition analysis. **Hydrological Processes**, v. 26, n. 7, p. 961–972, 2012.
- LI, H., ZHANG, Y., VAZE, J., WANG, B. Separating effects of vegetation change and climate variability using hydrological modelling and sensitivity-based approaches. **Journal of Hydrology**, v. 420-421, p. 403–418, 2012.
- LI, K.-J., McLEOD, A. I. Distribution of the residual autocorrelations in multivariate ARMA time series models. **Journal of the Royal Statistical Society**, v. 43, n. 2, p. 231–239, 1981.
- LIANG, L., LI, L., LIU, Q. Precipitation variability in Northeast China from 1961 to 2008. **Journal of Hydrology**, v. 404, n. 1-2, p. 67–76, 2011.
- LILLIEFORS, H. On the Kolmogorov–Smirnov test for normality with mean and variance unknown. **Journal of the American Statistical Association**, v. 62, p. 399–402, 1967.
- LINDSTRÖM, G., BERGSTRÖM, S. Runoff trends in Sweden 1807–2002. **Hydrological Sciences Journal**, v. 49, n. 1, p. 69–83, 2004.
- LINS, H. F., SLACK, J. R. Streamflow trends in the US. **Geophysical Research Letters**, v. 26, n. 2, p. 227–230, 1999.
- LOUCKS, D. P., STEDINGER, J. R., HAITH, D. A. **Water resource systems planning and analysis**. Englewood Cliffs: Prentice-Hall, 1981.
- LUNA, I., BALLINI, R., SOARES, S., SILVA FILHO, D. Fuzzy inference systems for synthetic monthly inflow time series generation. In: CONFERENCE OF THE EUROPEAN

SOCIETY FOR FUZZY LOGIC AND TECHNOLOGY, 7., Aix-les-Bains. **Anais...** pp.1060–1065. Aix-les-Bains: Atlantis Press, 2011.

MAACK, R. **Geografia física do Estado do Paraná**. 3ª ed. Curitiba: Paraná, Imprensa Oficial do Paraná, 2002.

MAKRIDAKIS, S., WHEELWRIGHT, S. C., HYNDMON, R. J. **Forecasting methods and applications**. 3ª. ed. New Jersey: John Wiley & Sons, 1998.

MANDELBROT, B. B., WALLIS, J. R. Noah , Joseph , and operational hydrology. **Water Resources Research**, v. 4, n. 5, p. 909–918, 1968.

MANDELBROT, B. B., WALLIS, J. R. Computer experiments with fractional Gaussian noises. Part I: Sample graphs, averages and variances. **Water Resources Research**, v. 5, p. 228–241, 1969a.

MANDELBROT, B. B., WALLIS, J. R. Computer experiments with fractional Gaussian noises. Part 2: Rescaled bridge range and “pox diagrams”. **Water Resources Research**, v. 5, p. 228–241, 1969b.

MANDELBROT, B. B., WALLIS, J. R. Computer experiments with fractional Gaussian noises. Part 3: Mathematical appendix. **Water Resources Research**, v. 5, p. 260–267, 1969c.

MANN, H. B. Nonparametric tests against trend. **Econometrica**, v. 13, p. 245–259, 1945.

MARENGO, J. A., ALVES, L. M. Tendências hidrológicas da bacia do rio Paraíba do Sul. **Revista Brasileira de Meteorologia**, v. 20, n. 2, p. 215–226, 2005.

MATALAS, N. C. Mathematical assessment of synthetic hydrology. **Water Resources Research**, v. 3, n. 4, p. 937–945, 1967.

MATALAS, N. C. Stochastic hydrology in the context of climate change. **Climatic Change**, v. 37, p. 89–101, 1997.

McCABE Jr., G. J., WOLOCK, D. M. Climate change and the detection of trends in annual runoff. **Climate Research**, v. 8, p. 129–134, 1997.

McMAHON, T. A. **Hydrological characteristics of selected rivers of the world**. Paris: UNESCO, 1982.

MENDONÇA, F., DANNI-OLIVEIRA, I. M. **Climatologia: noções básicas e climas do Brasil**. São Paulo: Oficina de Textos, 2007.

MILLY, P. C. D., BETANCOURT, J., FALKENMARK, M., HIRSCH, R. M., ZBIGNIEW, W., LETTENMAIER, D. P., STOUFFER, R. J. Stationarity Is Dead: Whither Water Management? **Science**, v. 319, p. 573–574, fev. 2008.

MINE, M. R. M. **Modelos Estocásticos Lineares para Previsão de Cheias em Tempo Real**, 127 f. 1984. Dissertação (Mestrado em Engenharia Hidráulica). Escola Politécnica, Universidade de São Paulo, São Paulo, 1984.

- MORETTIN, P. A., TOLOI, C. M. C. **Análise de séries temporais**. São Paulo: Edgard Blücher, 2004.
- MÜLLER, I. I., KRÜGER, C. M., KAVISKI, E. Análise de estacionariedade de séries hidrológicas na bacia incremental de Itaipu. **Revista Brasileira de Recursos Hídricos**, v. 3, n. 4, p. 51–71, 1998.
- MULVANY, T. J. On the use of self-registering rain and flood gauges in making observations of the relations of rain fall and flood discharges in a given catchment. **Transactions of the Institution of Civil Engineers of Ireland**, v. 4, n. 2, p. 18–33, 1850.
- NAGHETTINI, M., PINTO, E. J. de A. **Hidrologia Estatística**. Belo Horizonte: CPRM, 2007.
- NAIK, P. K., JAY, D. A. Distinguishing human and climate influences on the Columbia River: Changes in mean flow and sediment transport. **Journal of Hydrology**, v. 404, n. 3-4, p. 259–277, 2011.
- NDIRITU, J. A variable-length block bootstrap method for multi-site synthetic streamflow generation. **Hydrological Sciences Journal**, v. 56, n. 3, p. 362–379, 2011.
- NG, M., VOGEL, R. Multivariate Non-stationary Stochastic Streamflow Models for Two Urban Watersheds. In: WORLD ENVIRONMENTAL AND WATER RESOURCES CONGRESS 2010: CHALLENGES OF CHANGE, Providence, Rhode Island. **Anais...** pp. 2550–2561, Virginia: ASCE, 2010.
- PAPOULIS, A. **Probability, random variables and stochastic processes**. Tokyo: McGraw-Hill Kogakusha LTD, 1965.
- PARTAL, T., KUCUK, M. Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region (Turkey). **Physics and Chemistry of the Earth, Parts A/B/C**, v. 31, n. 18, p. 1189–1200, 2006.
- PASQUINI, A. I., DEPETRIS, P. J. Discharge trends and flow dynamics of South American rivers draining the southern Atlantic seaboard: An overview. **Journal of Hydrology**, v. 333, n. 2-4, p. 385–399, 2007.
- PEEL, M. C., FINLAYSON, B. L., McMAHON, T. A. Updated world map of the Köppen-Geiger climate classification. **Hydrology and Earth System Sciences**, v. 11, p. 1633–1644, 2007.
- PETTITT, A. N. A Non-Parametric Approach to the Change-Point Problem. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, v. 28, n. 2, p. 126–135, 1979.
- PINTER, N., van der PLOEG, R. R., SCHWEIGERT, P., HOEFER, G. Flood magnification on the River Rhine. **Hydrological Processes**, v. 20, n. 1, p. 147–164, 2006.
- PINTO, N. L. de S., HOLTZ, A. C. T., MARTINS, J. A., GOMIDE, F. L. S. **Hidrologia Básica**. 2ª ed. São Paulo: Edgard Blücher, 1976.
- POTTER, K. W. Evidence for nonstationarity as a physical explanation of the Hurst phenomenon. **Water Resources Research**, v. 12, n. 3, p. 1047–1052, 1976.

POTTER, K. W. Annual precipitation in the Northeast United States: long memory, short memory, or no memory? **Water Resources Research**, v. 15, n. 2, p. 340–346, 1979.

POVEDA, G. Fin al diletantismo sobre el calentamiento global y su origen antrópico. In: Congreso Latinoamericano de Hidráulica, 25, San Jose. **Anais...** . p.1–2, 2012.

POVEDA, G., ÁLVAREZ, D. El colapso de la hipótesis de estacionariedad por cambio y variabilidad climática: implicaciones para el diseño hidrológico en ingeniería. **Revista de Ingeniería**, v. 36, p. 65–76, 2012.

POVEDA, G., MESA, O. J. Acerca de la existencia del fenomeno de Hurst. **Avances en Recursos Hidraulicos**, n. 0, p. 5–19, 1991.

GHANBARPOUR, R. M., ABBASPOUR, K. C., JALALVAND, G., MOGHADDAM, G. A. Stochastic modeling of surface stream flow at different time scales: Sangsoorakh karst basin, Iran. **Journal of Cave and Karst Studies**, v. 72, n. 1, p. 1–10, 2010.

ROSSI, A., MASSEI, N., LAIGNEL, B., SEBAG, D., COPARD, Y. The response of the Mississippi River to climate fluctuations and reservoir construction as indicated by wavelet analysis of streamflow and suspended-sediment load, 1950–1975. **Journal of Hydrology**, v. 377, n. 3-4, p. 237–244, 2009.

ROUGÉ, C., GE, Y., CAI, X. Detecting gradual and abrupt changes in hydrological records. **Advances in Water Resources**, v. 53, n. 33–44, 2013.

SACO, P., KUMAR, P. Coherent modes in multiscale variability of streamflow over the United States. **Water Resources Research**, v. 36, n. 4, p. 049–1067, 2000.

SALAS, J. D., DELLEUR, J. W., YEVJEVICH, V., LANE, W. L. **Applied Modeling of Hydrologic Time Series**. 2ª ed. Chelsea: Water Resources Publications, 1985.

SALAS, J. D., LEE, T. Nonparametric Simulation of Single-Site Seasonal Streamflows. **Journal of Hydrologic Engineering**, v. 15, n. 4, p. 284–296, 2010.

SALAS, J. D., RAJAGOPALAN, B., SAITO, L., BROWN, C. Special Section on Climate Change and Water Resources: Climate Nonstationarity and Water Resources Management. **Journal of Water Resources Planning and Management**, v. 138, n. 5, p. 385–388, 2012.

SANTOS, C. A. G., FREIRE, P. K. de M. M., TORRENCE, C. A Transformada Wavelet e sua Aplicação na Análise de Séries Hidrológicas. **Revista Brasileira de Recursos Hídricos**, v. 18, n. 3, p. 271–280, 2013.

SCHWARTZ, G. Estimating the Dimension of a Model. **The Annals of Mathematical Statistics**, v. 6, n. 2, p. 461–464, 1978.

SEMA – SECRETARIA ESTADUAL DO MEIO AMBIENTE. **Bacias Hidrográficas do Paraná**, 2010. Disponível em: [www.meioambiente.pr.gov.br](http://www.meioambiente.pr.gov.br). Acesso em: 05/02/2014.

SEN, P. K. Estimates of the regression coefficient based on Kendall's tau. **Journal of the American Statistical Association**, v. 63, n. 324, p. 1379–1389, 1968.

- ŞENGÜL, S., CAN, I. Stochastic modelling of mean monthly flows of Karasu River, in Turkey. **Water and Environment Journal**, v. 25, p. 31–39, 2011.
- SERINALDI, F., KILSBY, C. G. Stationarity is undead: Uncertainty dominates the distribution of extremes. **Advances in Water Resources**, v. 77, p. 17–36, 2015.
- SHARMA, K. P., MOORE III, B., VOROSMARTY, C. J. Anthropogenic, climatic, and hydrologic trends in the Kosi basin, Himalaya. **Climatic Change**, v. 47, n. 141–165, 2000.
- SHITTU, O. I., ASEMOTA, M. J. Comparison of criteria for estimating the order of autoregressive process: a Monte Carlo approach. **European Journal of Operational Research**, v. 30, n. 3, p. 409–416, 2009.
- SILVA, A. T., PORTELA, M. M. Disaggregation modelling of monthly streamflows using a new approach of the method of fragments. **Hydrological Sciences Journal**, v. 57, n. 5, p. 942–955, 2012.
- SMITH, L. C., TURCOTTE, D. L., ISACKS, B. L. Stream flow characterization and feature detection using a discrete wavelet transform. **Hydrological Processes**, v. 12, n. 2, p. 233–249, 1998.
- SOUZA, R. C., CAMARGO, M. E. **Análise e previsão de Séries Temporais: os modelos ARIMA**. 2<sup>a</sup> ed., Rio de Janeiro: Regional, 2004.
- SRIVASTAV, R. K., SIMONOVIC, S. P. An analytical procedure for multi-site, multi-season streamflow generation using maximum entropy bootstrapping. **Environmental Modelling and Software**, v. 59, p. 59–75, 2014
- STEDINGER, J. R., LETTENMAIER, D. P., VOGEL, R. M. Multisite ARMA(1,1) and disaggregation models for annual streamflow generation. **Water Resources Research**, v. 21, n. 4, p. 497–509, 1985.
- SZOLGAYOVA, E., LAAHA, G., BLÖSCHL, G., BUCHER, C. Factors influencing long range dependence in streamflow of European rivers. **Hydrological Processes**, v. 28, n. 4, p. 1573–1586, 2014.
- THOMAS, B. E. Climatic Fluctuations and Forecasting of Streamflow in the Lower Colorado River Basin. **Journal of the American Water Resources Association**, v. 43, n. 6, p. 1550–1569, 2007.
- THOMAS, H. A., FIERING, M. B. Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation. In Maass A. *et al.* (Org.), **Design of Water Resources Systems**. Cambridge: Harvard University Press, 1962. P. 459-493.
- TODINI, E. Hydrological catchment modelling: past, present and future. **Hydrology and Earth System Sciences**, v. 11, n. 1, p. 468–482, 2007.
- TORRENCE, C., COMPO, G. P. A Practical Guide to Wavelet Analysis. **Bulletin of the American Meteorological Society**, v. 79, n. 1, p. 61–78, 1998.

- TUCCI, C. E. M. **Impactos da variabilidade climática e do uso do solo nos recursos hídricos**, 2002. Disponível em: <http://www.eclac.cl/samtac/noticias/documentosdetrabajo/6/23336/InBr02902.pdf>. Acesso em 05/02/2014.
- TUCCI, C. E. M. **Modelos Hidrológicos**. 2ª ed. Porto Alegre: Editora da UFRGS, 2005.
- TUCCI, C. E. M., BRAGA, B. **Clima e Recursos Hídricos no Brasil**. Porto Alegre: ABRH, 2003.
- TYRALIS, H., KOUTSOYIANNIS, D. Simultaneous estimation of the parameters of the Hurst–Kolmogorov stochastic process. **Stochastic Environmental Research and Risk Assessment**, v. 25, n. 1, p. 21–33, 2011.
- UNITED STATES DEPARTMENT OF ENERGY (U.S. DOE). Scientific discovery through advanced computing. 2000.
- VALENCIA, D., SCHAAKE, J. C. Disaggregation processes in stochastic hydrology. **Water Resources Research**, v. 9, n. 3, p. 211–219, 1973.
- VENEMA, H. D., SCHILLER, E. J., ADAMOWSKI, K., THIZY, J.-M. A Water Resources Planning Response to Climate Change in the Senegal River Basin. **Journal of Environmental Management**, v. 49, n. 1, p. 125–155, 1997.
- VILLARINI, G., SMITH, J. A., BAECK, M. L., KRAJEWSKI, W. F. Examining Flood Frequency Distributions in the Midwest U.S. **Journal of the American Water Resources Association**, v. 47, n. 3, p. 447–463, 2011.
- VON STORCH, H., NAVARRA, A. **Analysis of Climate Variability: applications of statistical techniques**, 2ª ed. Berlin: Springer, 1999.
- WANG, W., CHEN, X., SHI, P., van GELDER, P. H. A. J. M. Detecting changes in extreme precipitation and extreme streamflow in the Dongjiang River Basin in southern China. **Hydrology and Earth System Sciences**, v. 12, p. 207–221, 2008.
- WEI, W. W. S. **Time Series Analysis: Univariate and Multivariate Methods**. 2ª ed. São Francisco: Pearson Addison Wesley, 2006.
- WILKS, D. S. Multisite generalization of a daily stochastic precipitation generation model. **Journal of Hydrology**, v. 210, n. 1-4, p. 178–191, 1998.
- WILKS, D. S. **Statistical Methods in the Atmospheric Sciences**. 2ª ed. New York: Academic Press, 2006.
- XU, Z. X., TAKEUCHI, K., ISHIDAIRA, H. Monotonic trend and step changes in Japanese precipitation. **Journal of Hydrology**, v. 279, n. 1-4, p. 144–150, 2003.
- YEVJEVICH, V., OBEYSEKERA, J. T. B. Effects of incorrectly removed periodicity in parameters on stochastic dependence. **Water Resources Research**, v. 21, n. 5, p. 685–690, 1985.

YUE, S., PILON, P., CAVADIAS, G. Power of the Mann–Kendall and Spearman’s rho tests for detecting monotonic trends in hydrological series. **Journal of Hydrology**, v. 259, n. 1-4, p. 254–271, 2002.

YUE, S., PILON, P., PHINNEY, B., CAVADIAS, G. The influence of autocorrelation on the ability to detect trend in hydrological series. **Hydrological Processes**, v. 16, n. 9, p. 1807–1829, 2002.

ZHANG, Q.; GU, X.; SINGH, V. P.; XIAO, M.; XU, C-Y. Stationarity of annual flood peaks during 1951–2010 in the Pearl River basin, China. **Journal of Hydrology**, v. 519, part D, p. 3263-3274, 2014.

ZHANG, Q., LIU, C., XU, C., XU, Y., JIANG, T. Observed trends of annual maximum water level and streamflow during past 130 years in the Yangtze River basin, China. **Journal of Hydrology**, v. 324, n. 1-4, p. 255–265, 2006.

ZHANG, X., VINCENT, L. A., HOGG, W. D., NIITSOO, A. Temperature and precipitation trends in Canada during the 20th century. **Atmosphere-Ocean**, v. 38, n. 3, p. 395–429, 2000.

ZHANG, W., YAN, Y., ZHENG, J., LI, L., DONG, X., CAI, H. Temporal and spatial variability of annual extreme water level in the Pearl River Delta region, China. **Global and Planetary Change**, v. 69, n. 1-2, p. 35–47, 2009.





## APÊNDICES

### A.1. Funções de autocorrelação (FAC) e de autocorrelação parcial

Na Figura 35 a Figura 39 são mostradas as funções FAC e FACP para as cinco séries não mostradas anteriormente. Em cada gráfico, as linhas tracejadas horizontais são os limites de significância estatística dentro dos quais os valores das funções são considerados nulos.

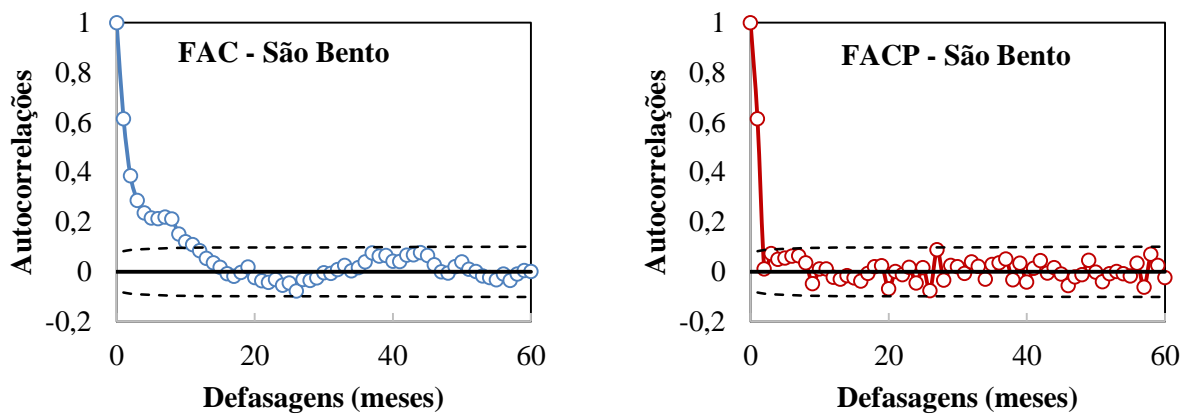


Figura 35 – FAC e FACP para São Bento

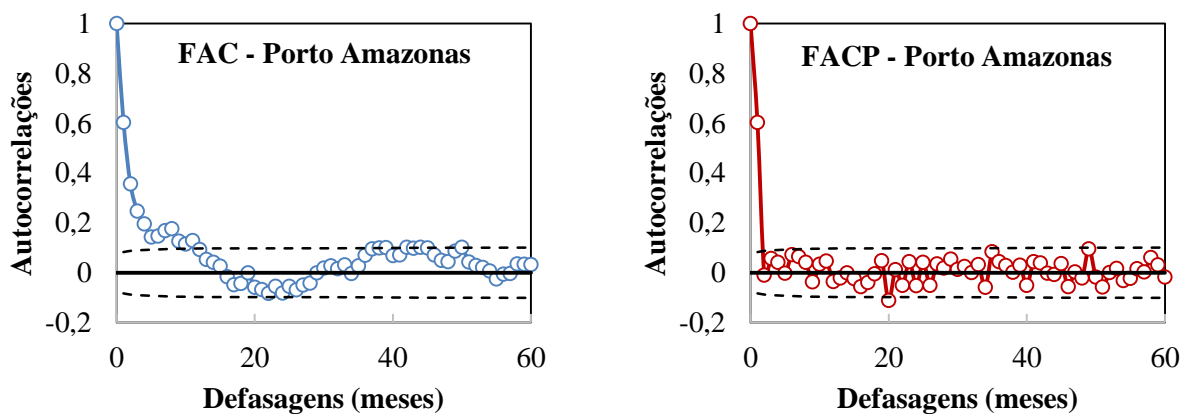
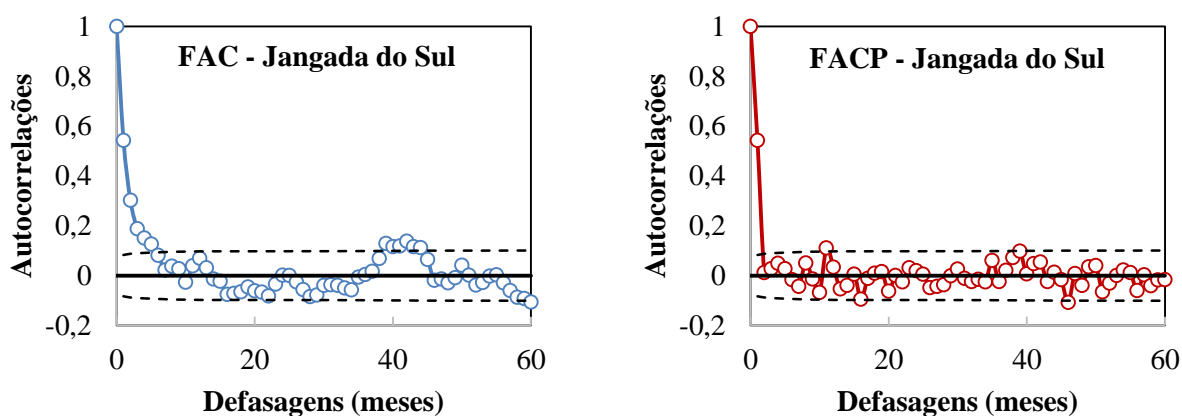
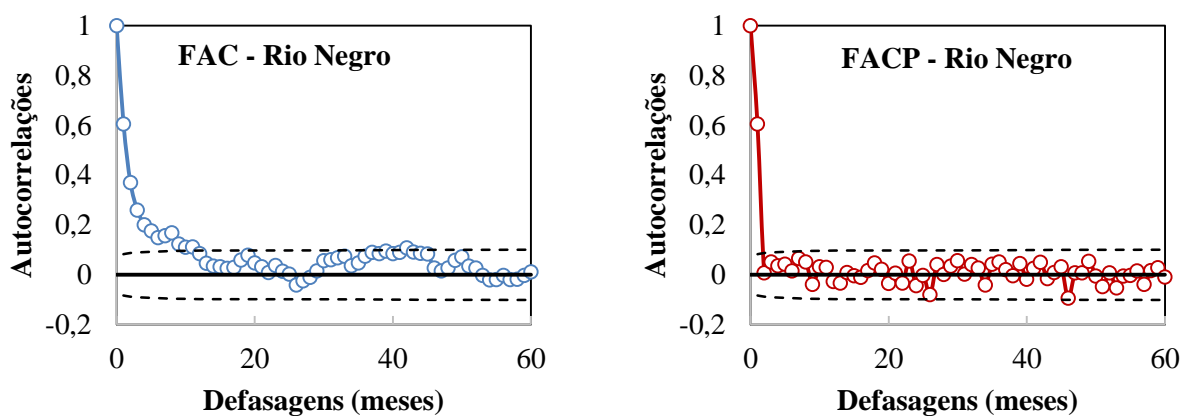


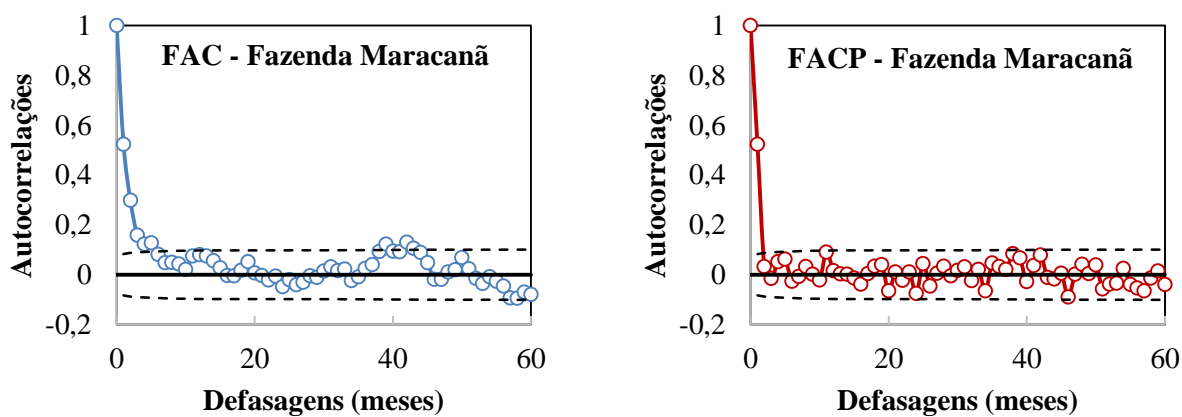
Figura 36 – FAC e FACP para Porto Amazonas



**Figura 37 – FAC e FACP para Jangada do Sul**



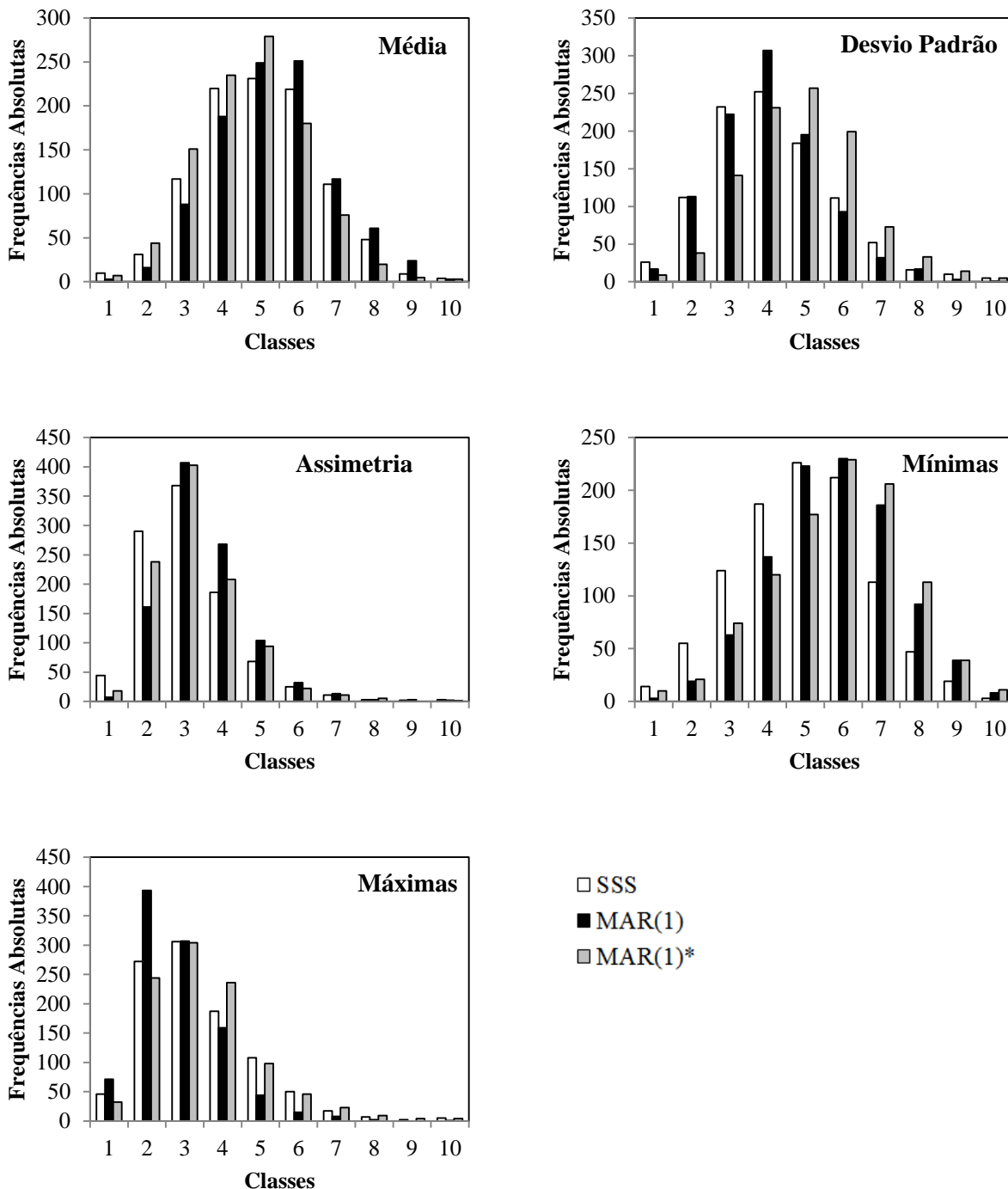
**Figura 38 – FAC e FACP para Rio Negro**



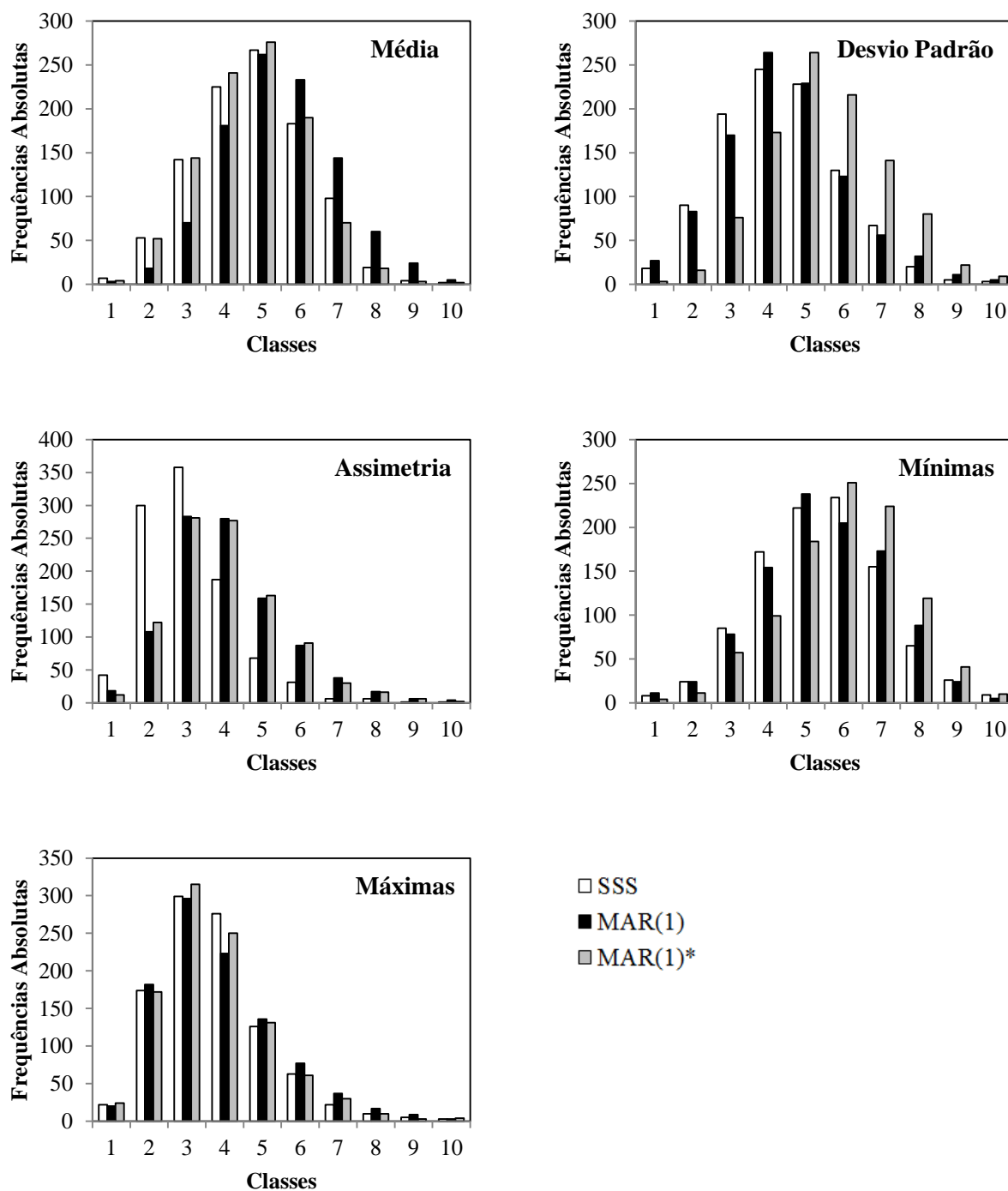
**Figura 39 – FAC e FACP para Fazenda Maracanã**

**A.2. Histogramas das estatísticas de curto termo para séries anuais**

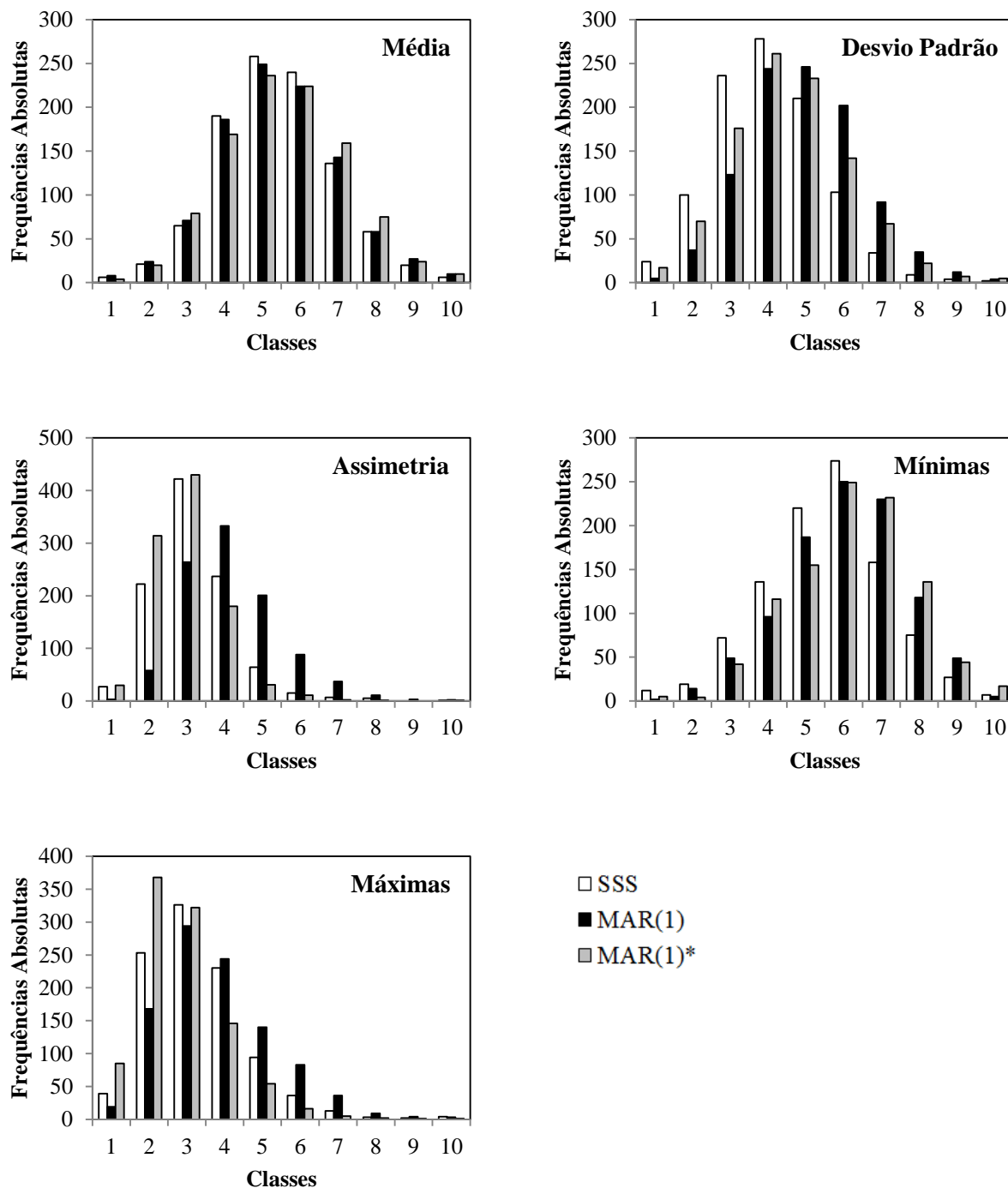
Na Figura 40 a Figura 44 são exibidos os histogramas das estatísticas de curto termo obtidas para as cinco séries anuais não exibidas no corpo do trabalho.



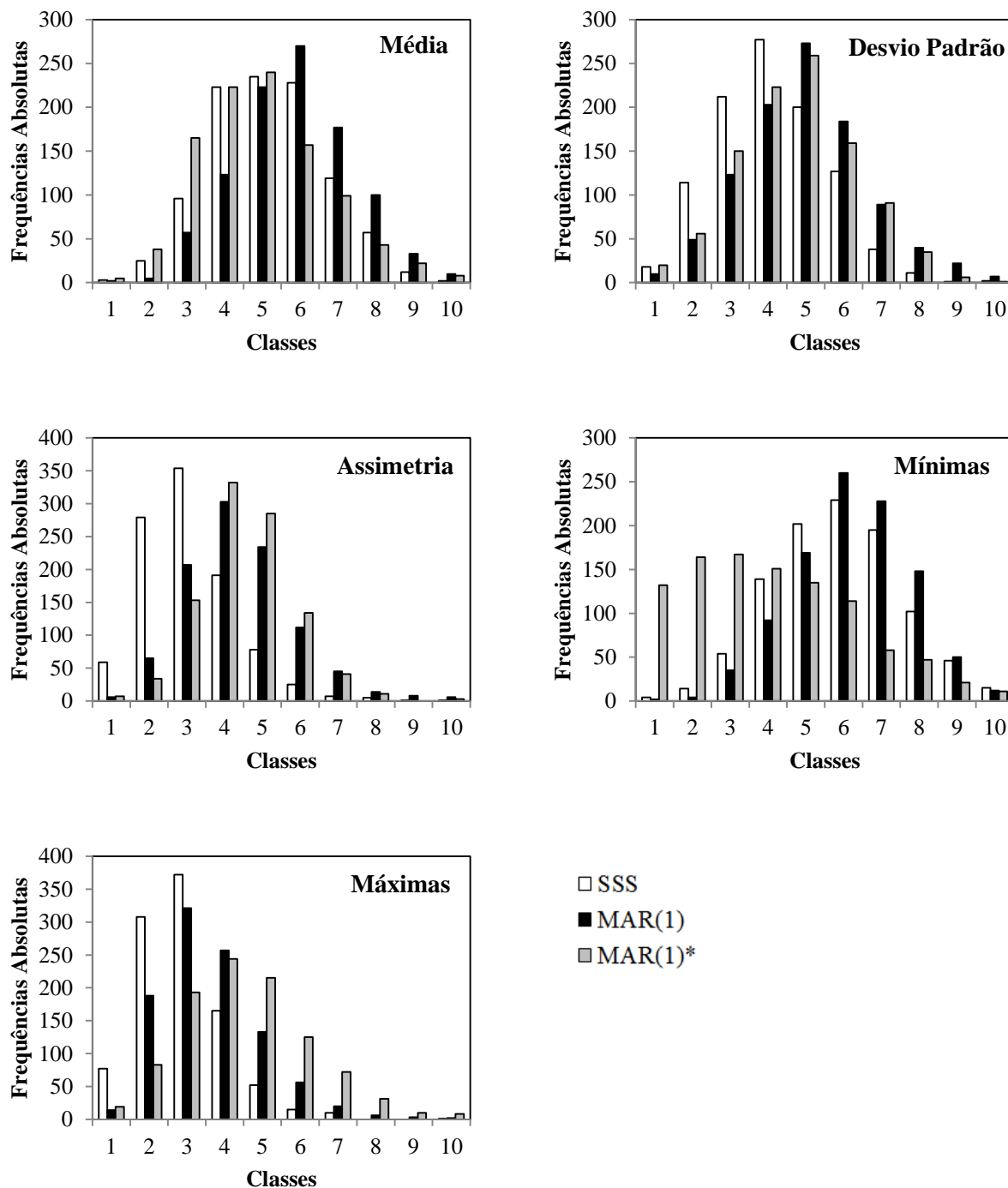
**Figura 40 – Histogramas para as estatísticas de curto termo em São Bento, para séries sintéticas anuais**



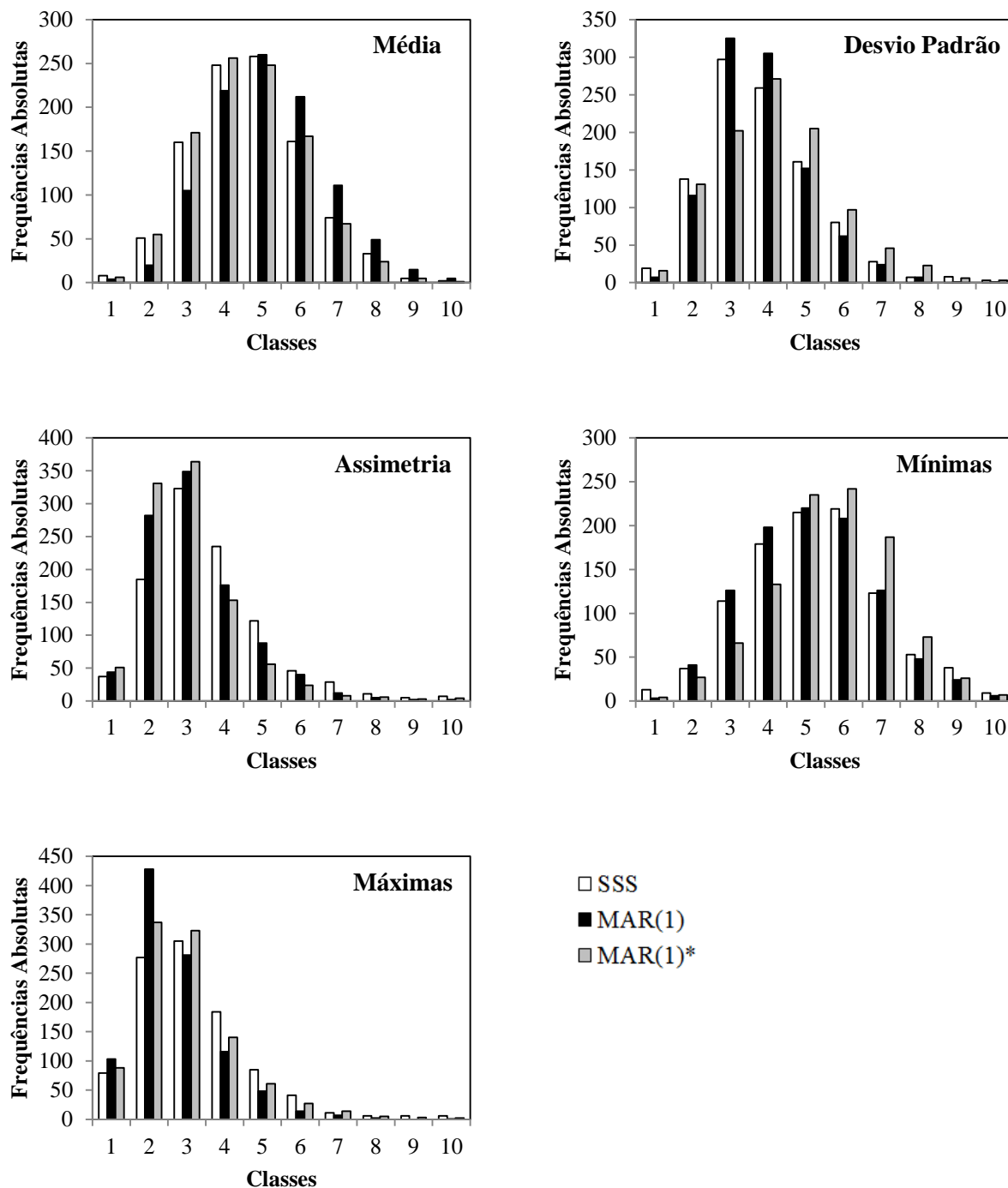
**Figura 41 – Histogramas para as estatísticas de curto termo em Porto Amazonas, para séries sintéticas anuais**



**Figura 42 – Histogramas para as estatísticas de curto termo em Jangada do Sul, para séries sintéticas anuais**



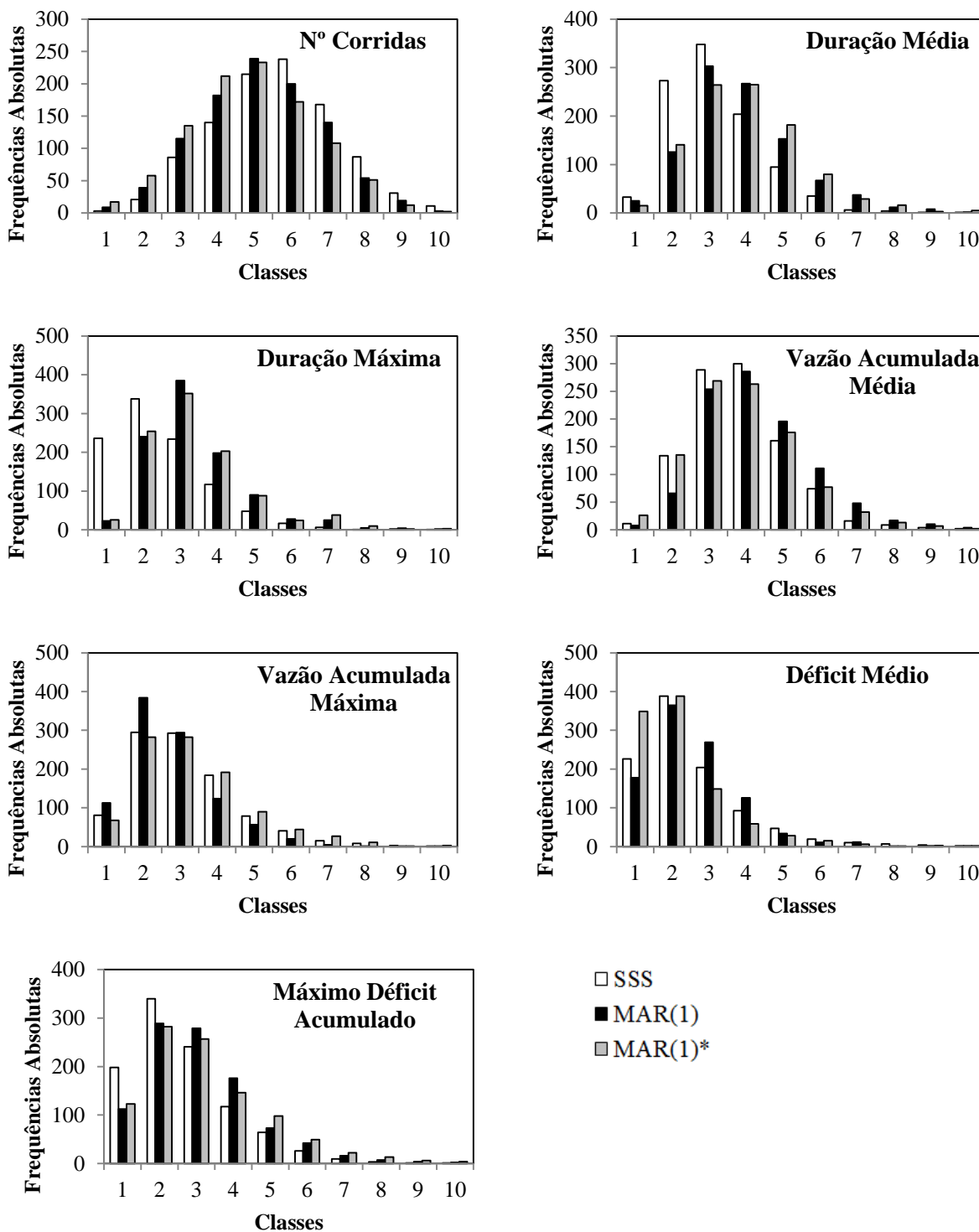
**Figura 43 – Histogramas para as estatísticas de curto termo em Rio Negro, para séries sintéticas anuais**



**Figura 44 – Histogramas para as estatísticas de curto termo em Fazenda Maracanã, para séries sintéticas anuais**

**A.3. Histogramas das estatísticas de longo termo para séries anuais**

Na Figura 45 a Figura 49 são exibidos os histogramas das estatísticas de longo termo obtidas para as cinco séries anuais não exibidas no corpo do trabalho.



**Figura 45 – Histogramas para as estatísticas de longo termo em São Bento, para séries sintéticas anuais**



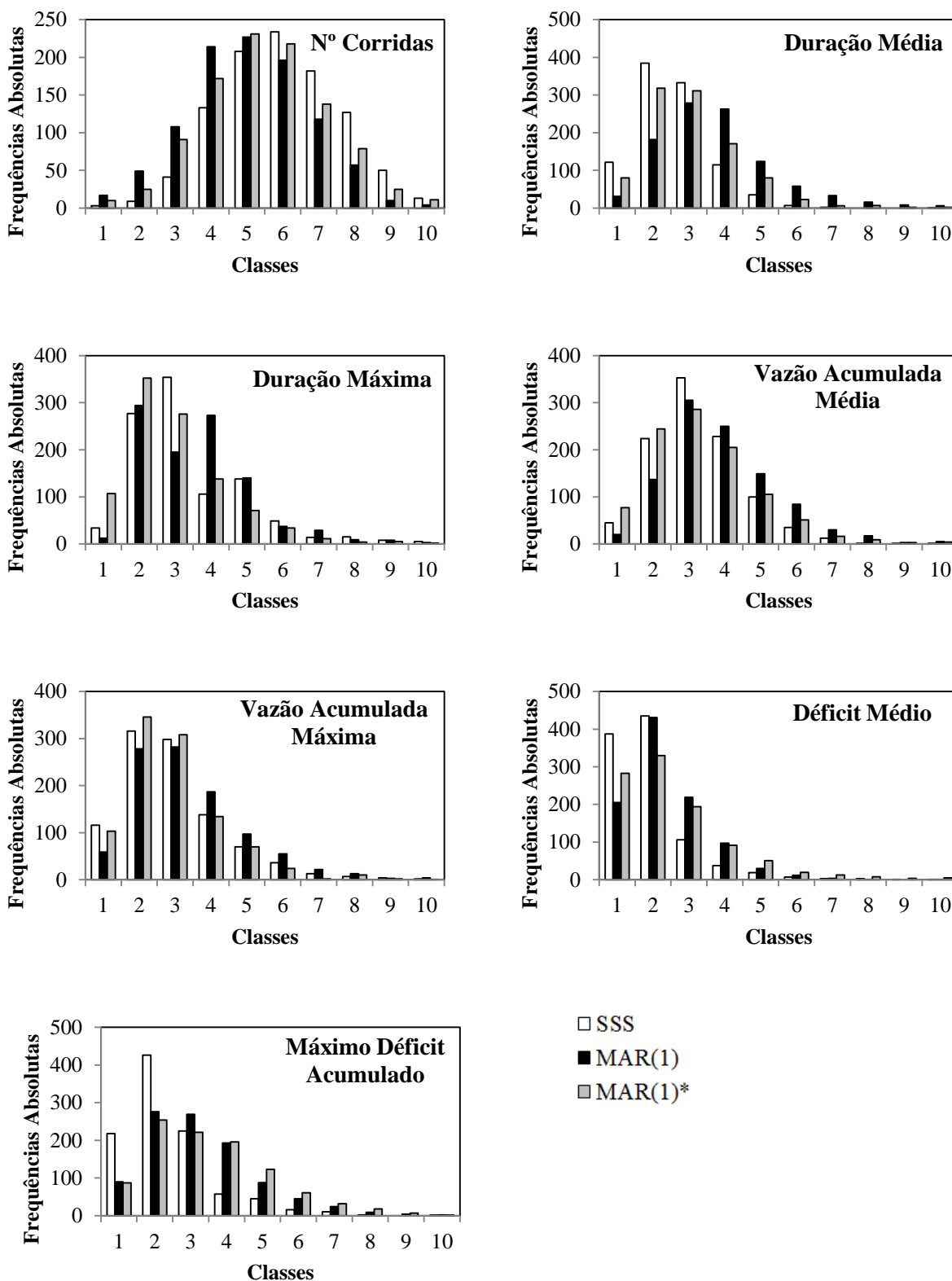


Figura 46 – Histogramas para as estatísticas de longo termo em Porto Amazonas, para séries sintéticas anuais

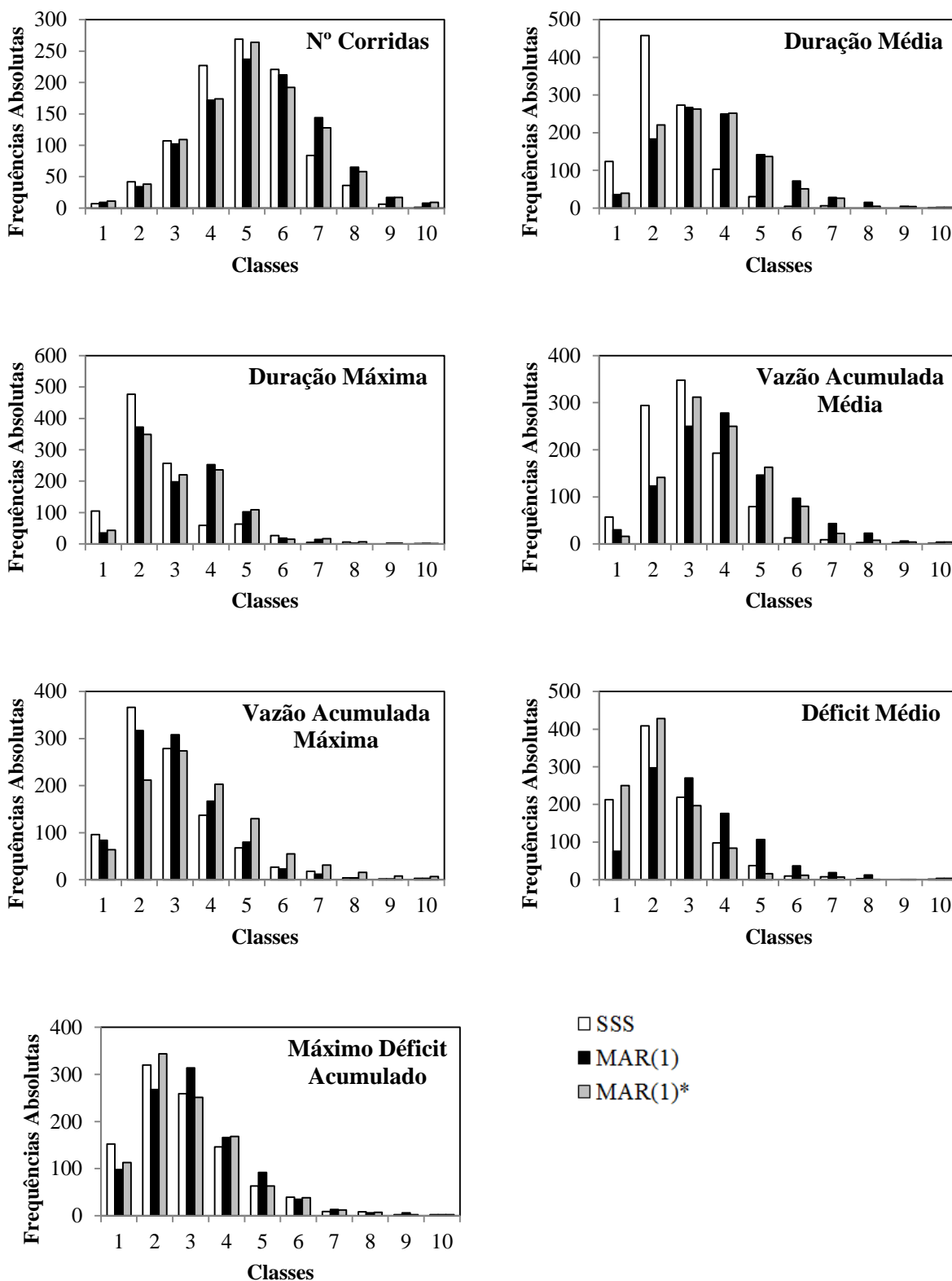


Figura 47 – Histogramas para as estatísticas de longo termo em Jangada do Sul, para séries sintéticas anuais

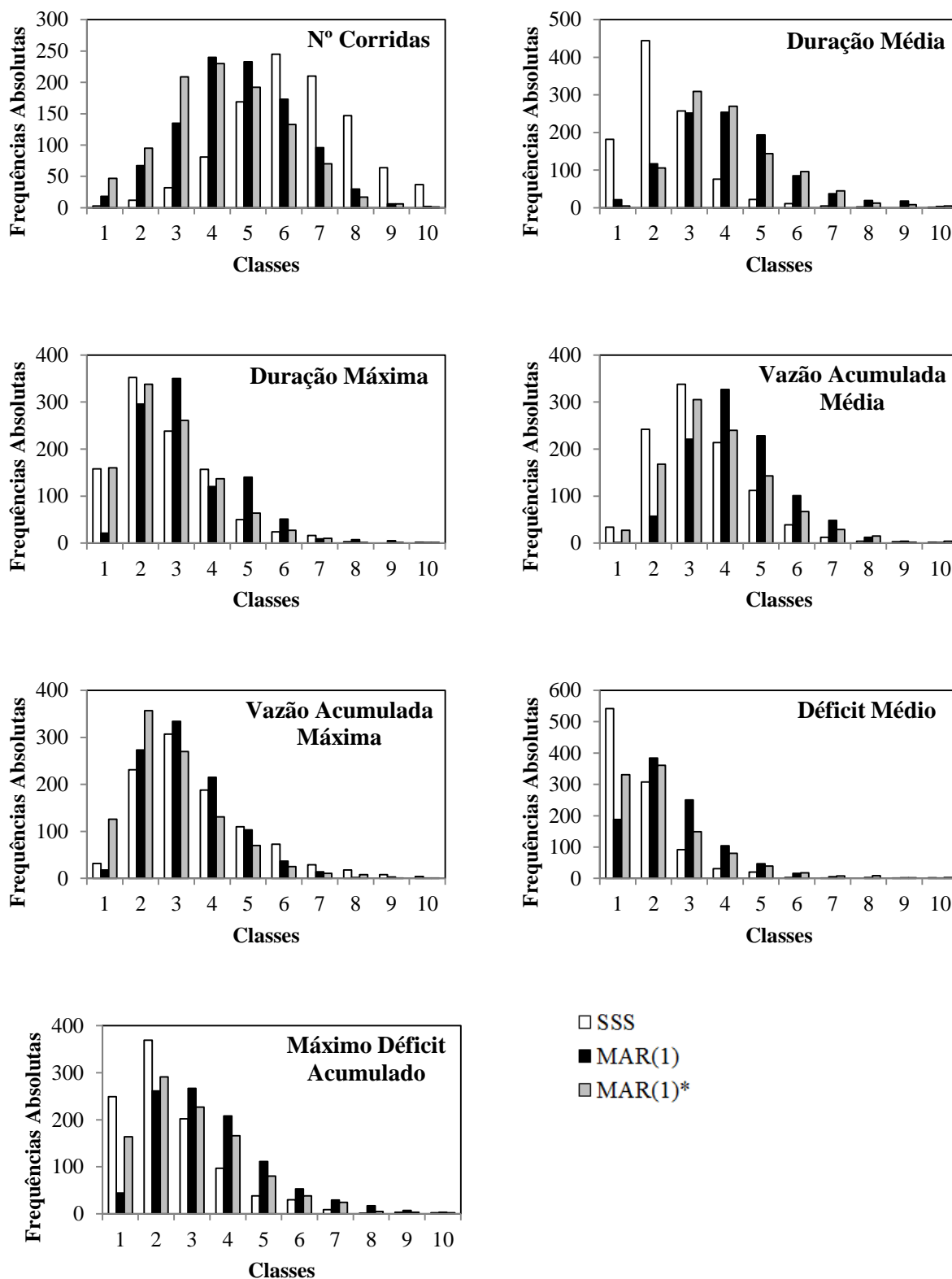


Figura 48 – Histogramas para as estatísticas de longo termo em Rio Negro, para séries sintéticas anuais

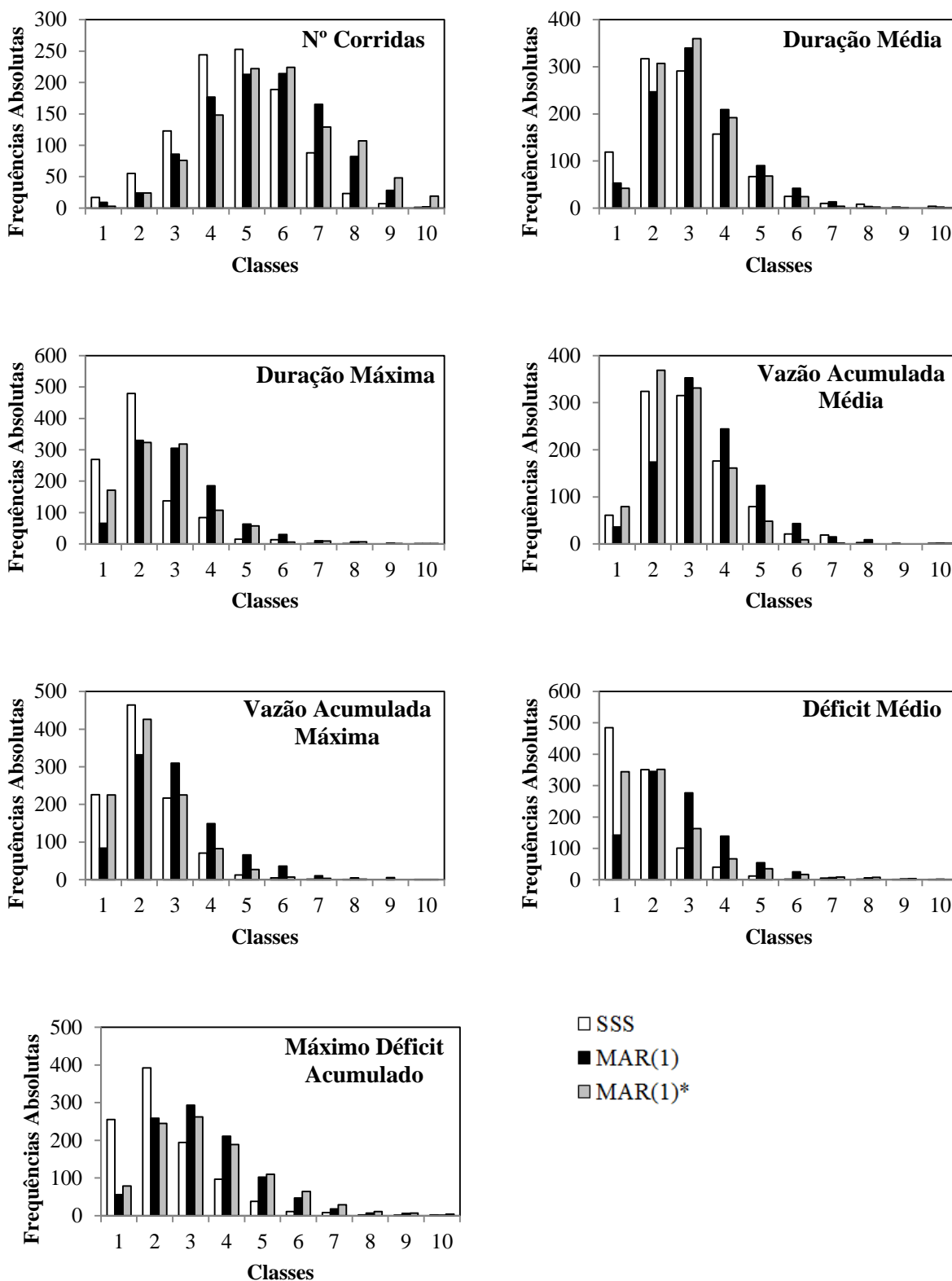
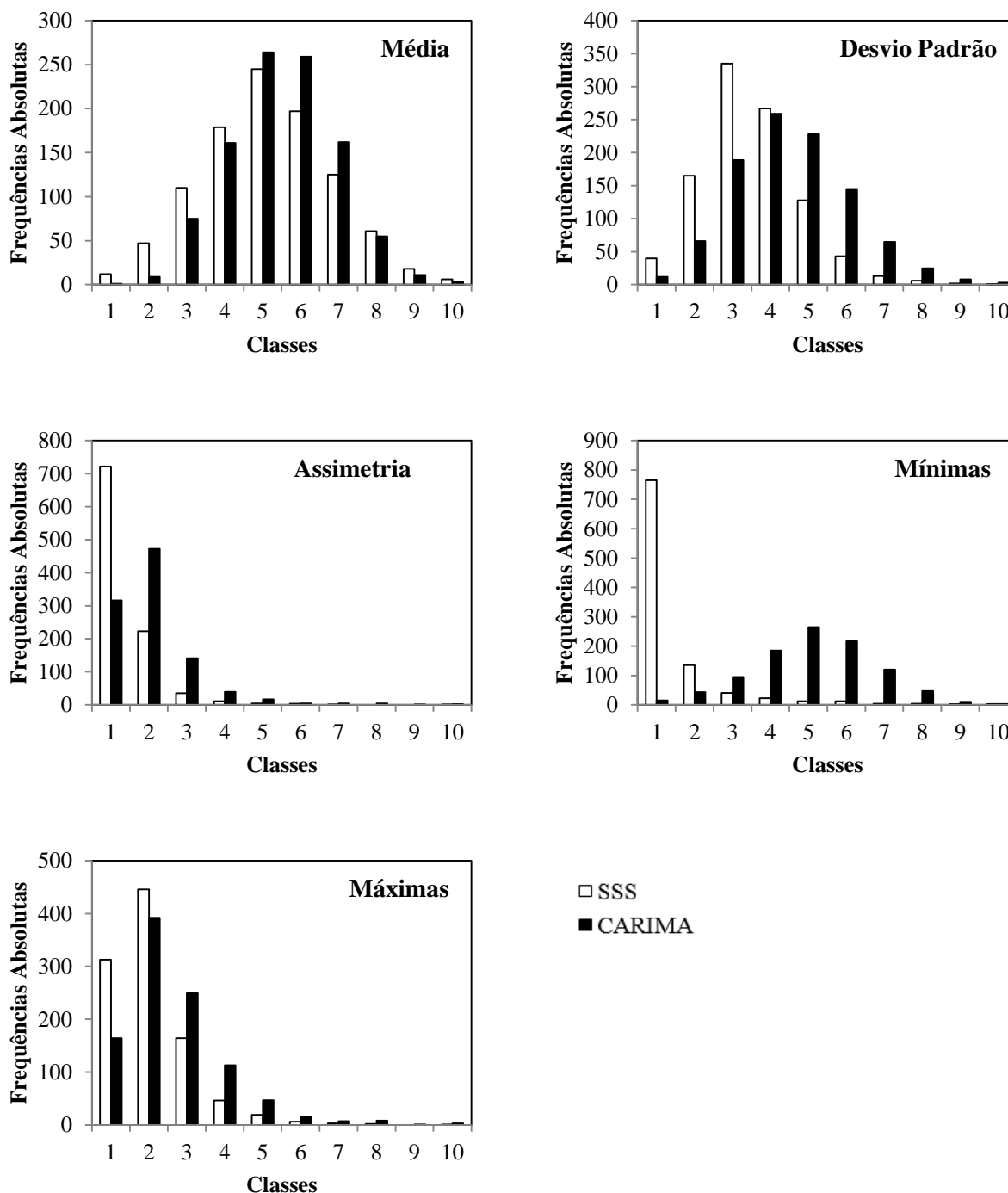


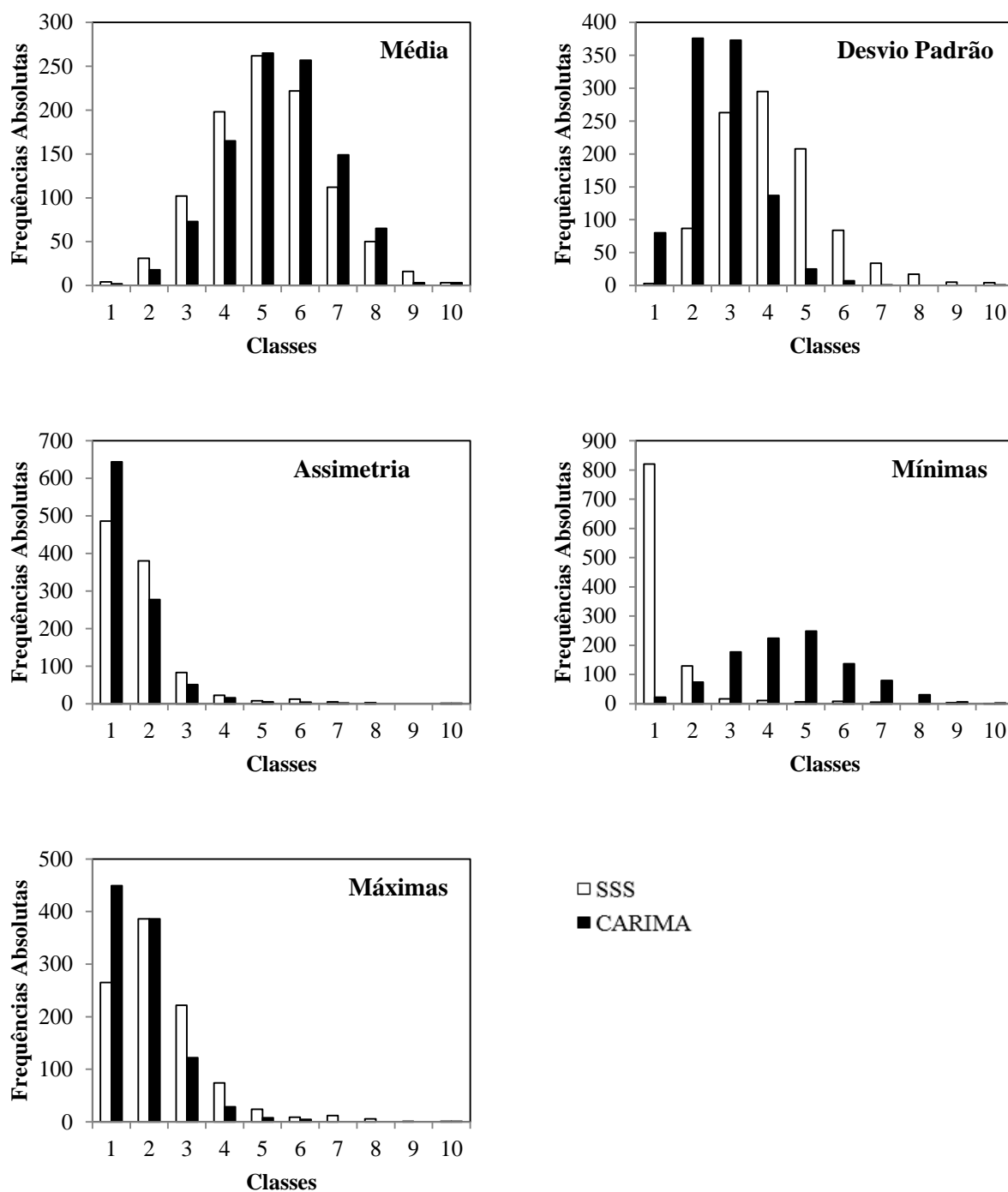
Figura 49 – Histogramas para as estatísticas de longo termo em Fazenda Maracanã, para séries sintéticas anuais

**A.4. Histogramas das estatísticas de curto termo para séries mensais**

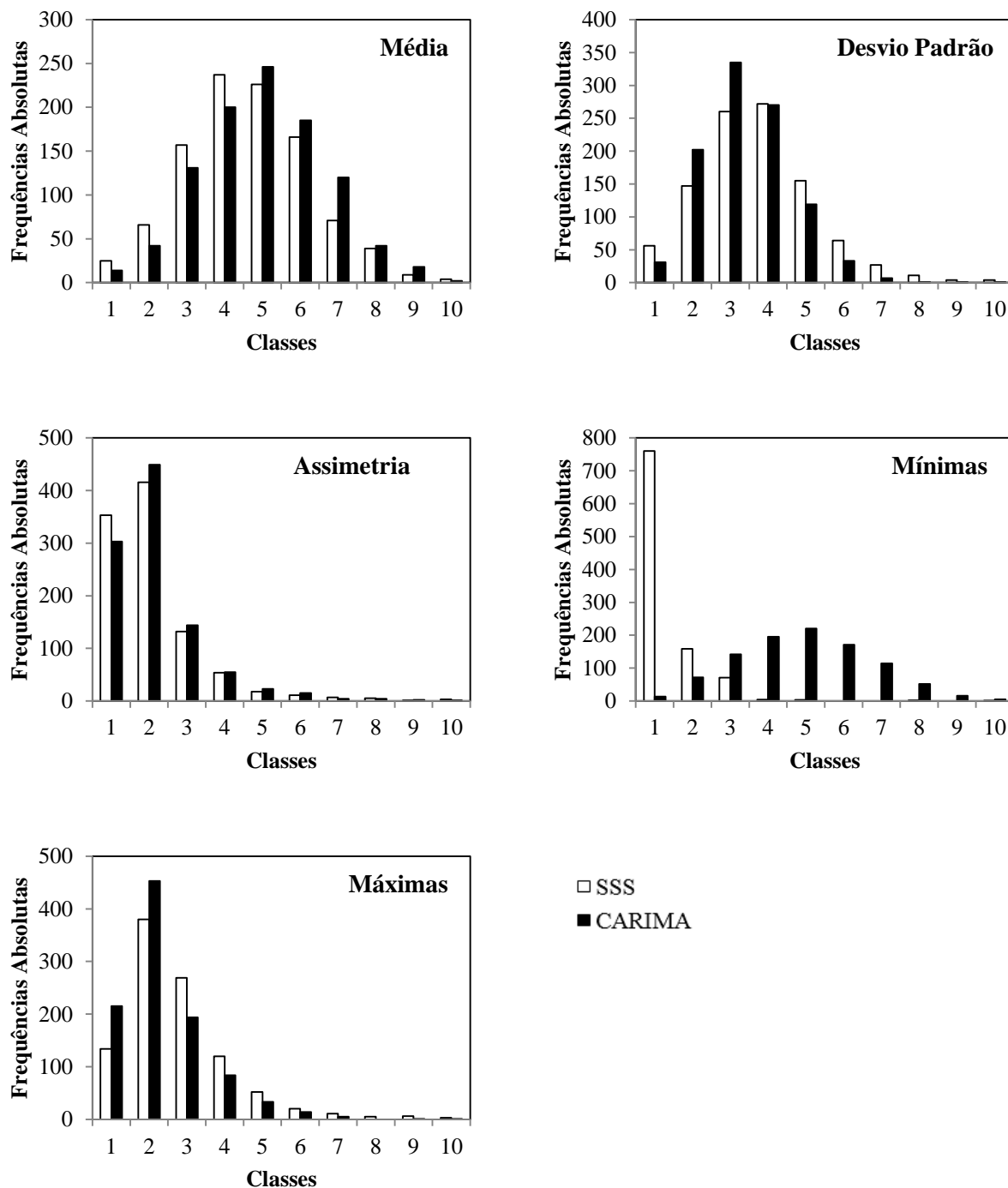
Na Figura 50 a Figura 54 são exibidos os histogramas das estatísticas de curto termo obtidas para as cinco séries mensais não exibidas no corpo do trabalho.



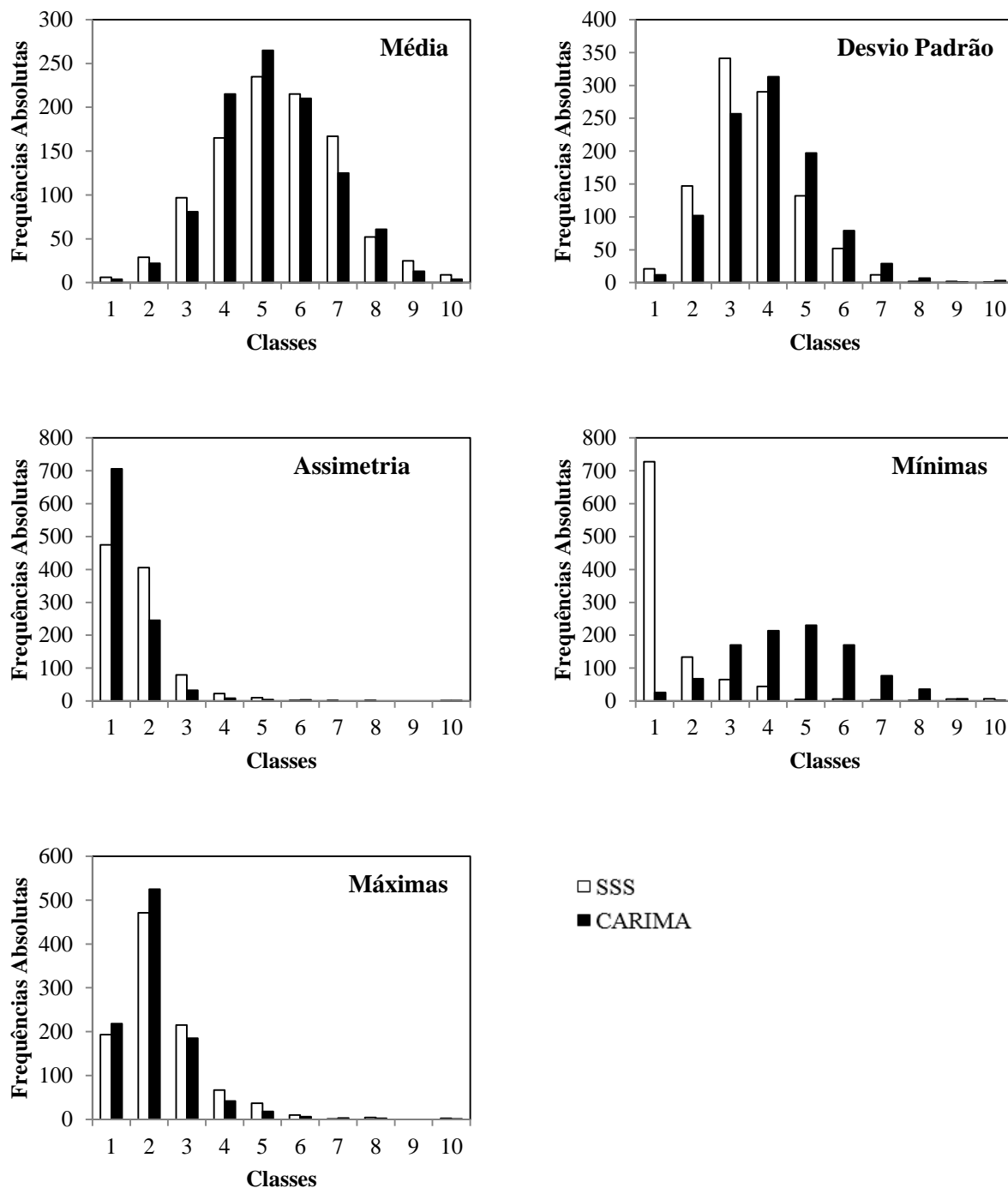
**Figura 50 – Histogramas para as estatísticas de curto termo em São Bento, para séries sintéticas mensais**



**Figura 51 – Histogramas para as estatísticas de curto termo em Porto Amazonas, para séries sintéticas mensais**

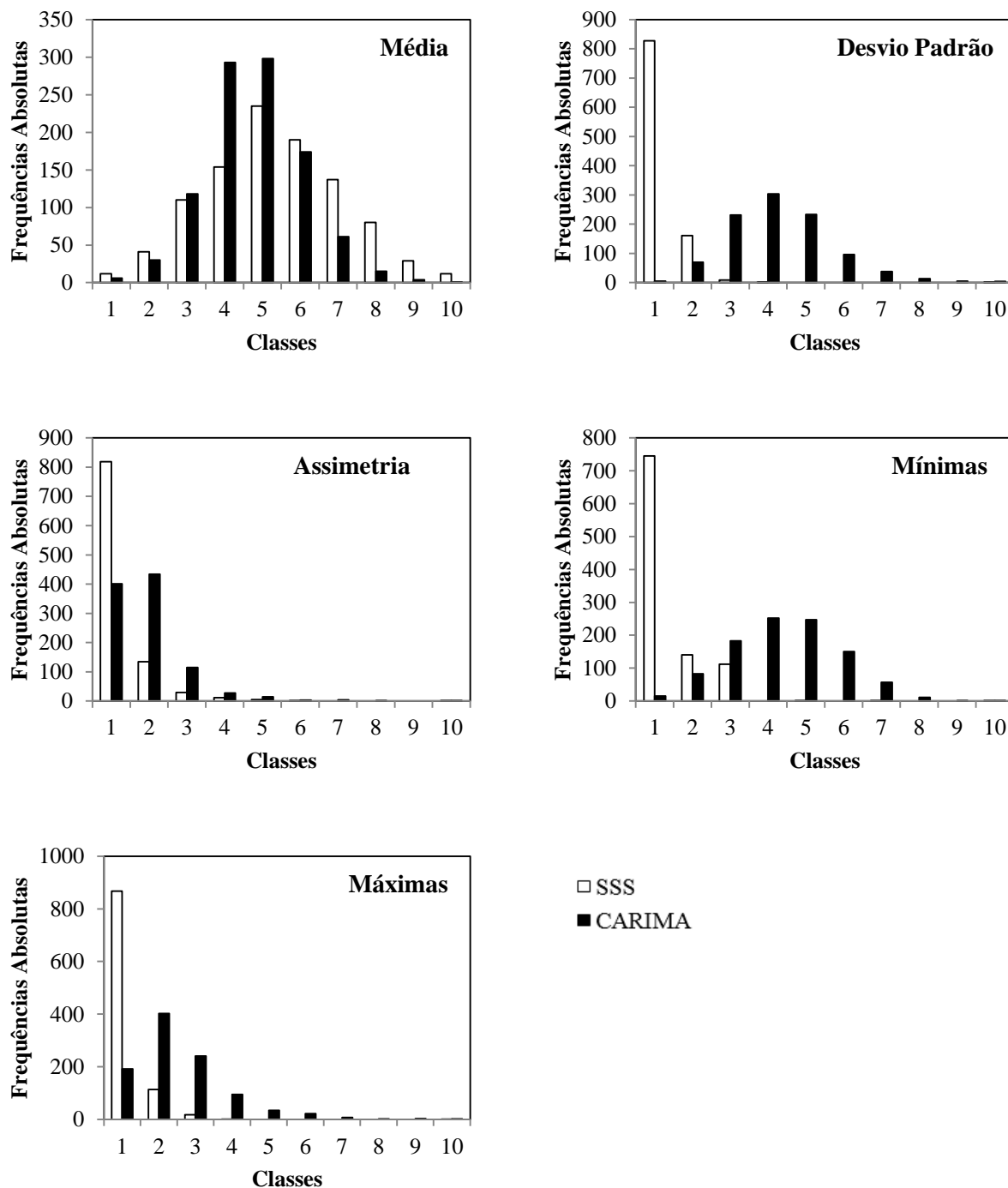


**Figura 52 – Histogramas para as estatísticas de curto termo em Jangada do Sul, para séries sintéticas mensais**



**Figura 53 – Histogramas para as estatísticas de curto termo em Rio Negro, para séries sintéticas mensais**

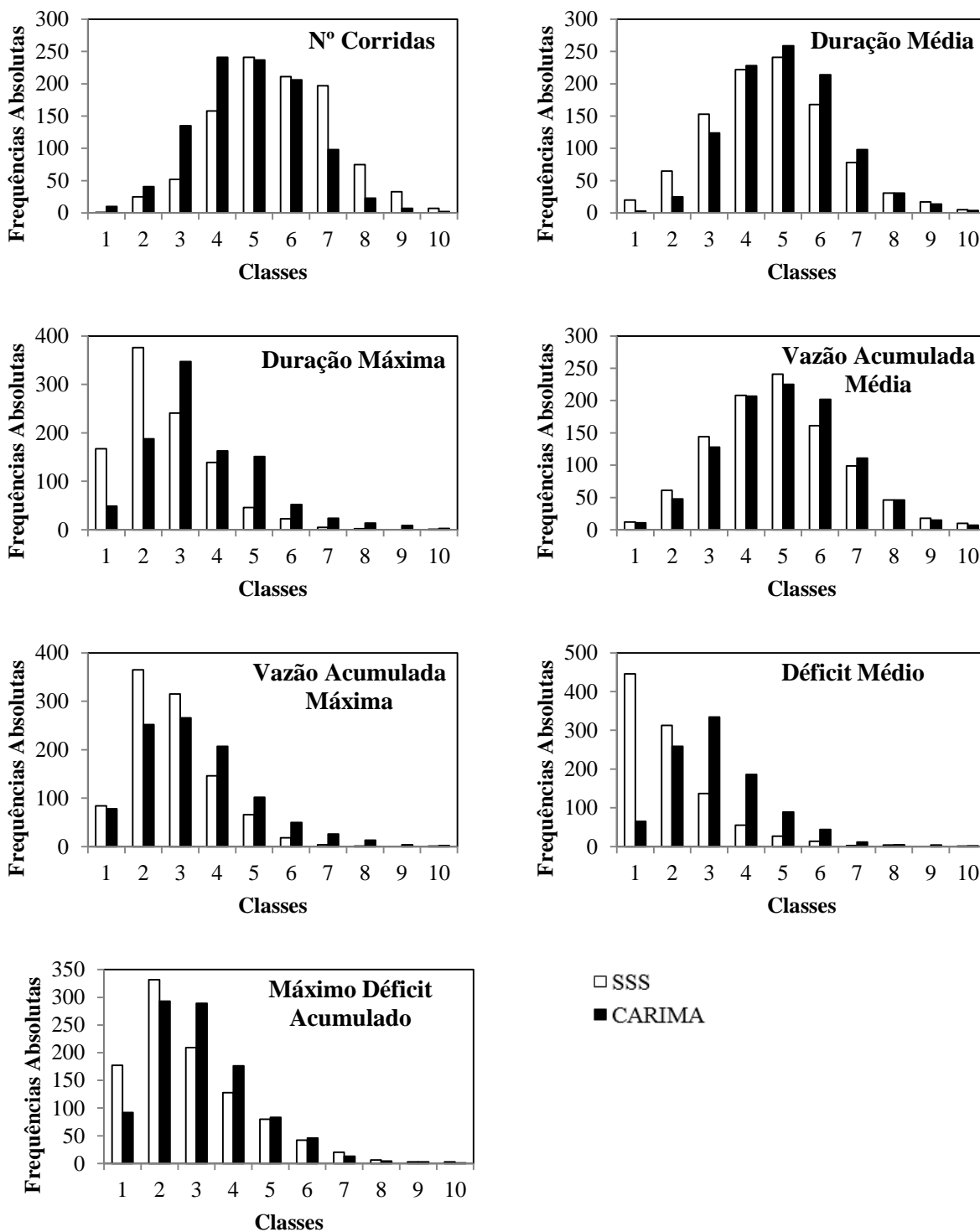




**Figura 54 – Histogramas para as estatísticas de curto termo em Fazenda Maracanã, para séries sintéticas mensais**

**A.5. Histogramas das estatísticas de longo termo para séries mensais**

Na Figura 55 a Figura 59 são exibidos os histogramas das estatísticas de longo termo obtidas para as cinco séries mensais não exibidas no corpo do trabalho.



**Figura 55 – Histogramas para as estatísticas de longo termo em São Bento, para séries sintéticas mensais**

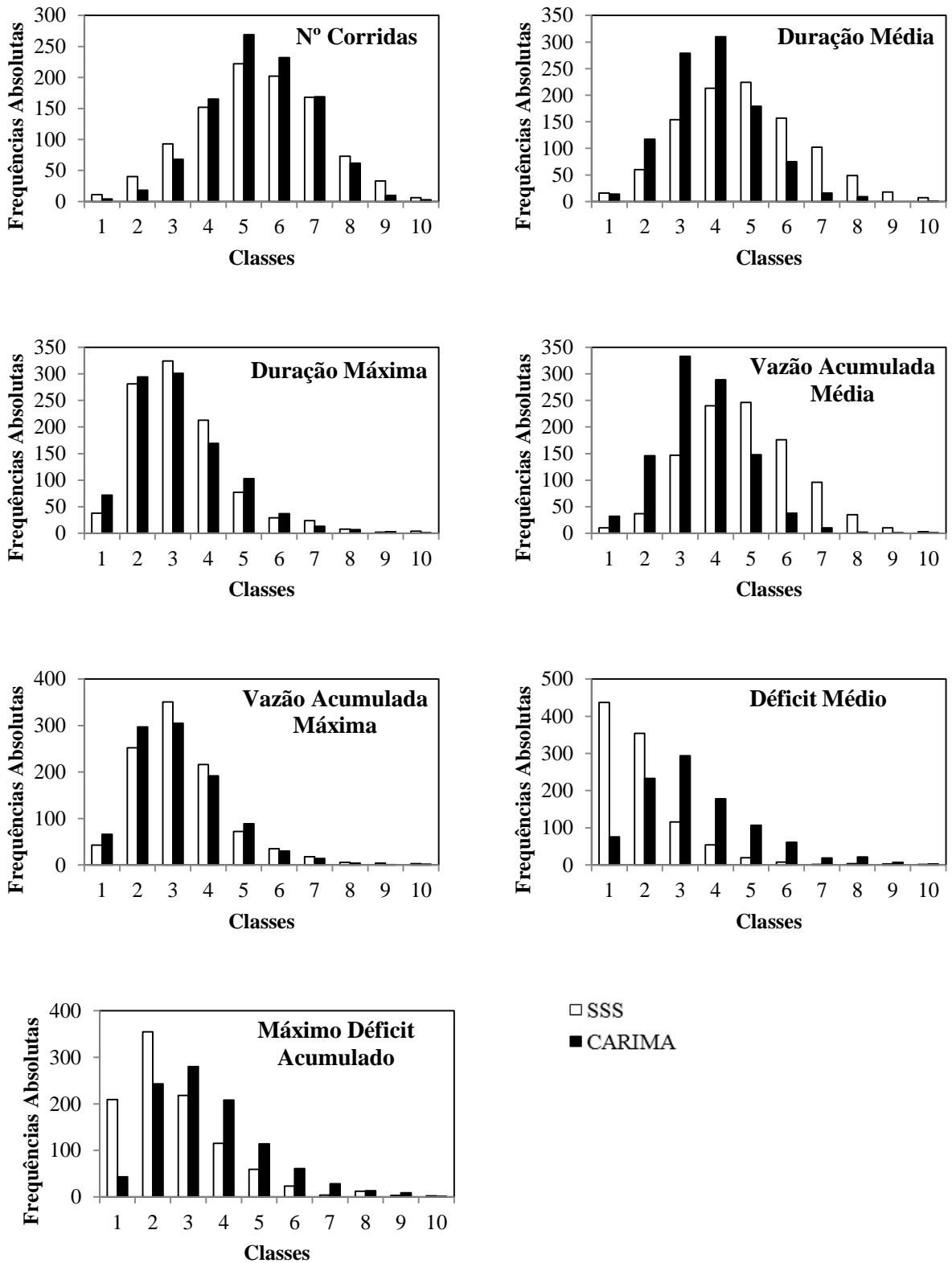


Figura 56 – Histogramas para as estatísticas de longo termo em Porto Amazonas, para séries sintéticas mensais

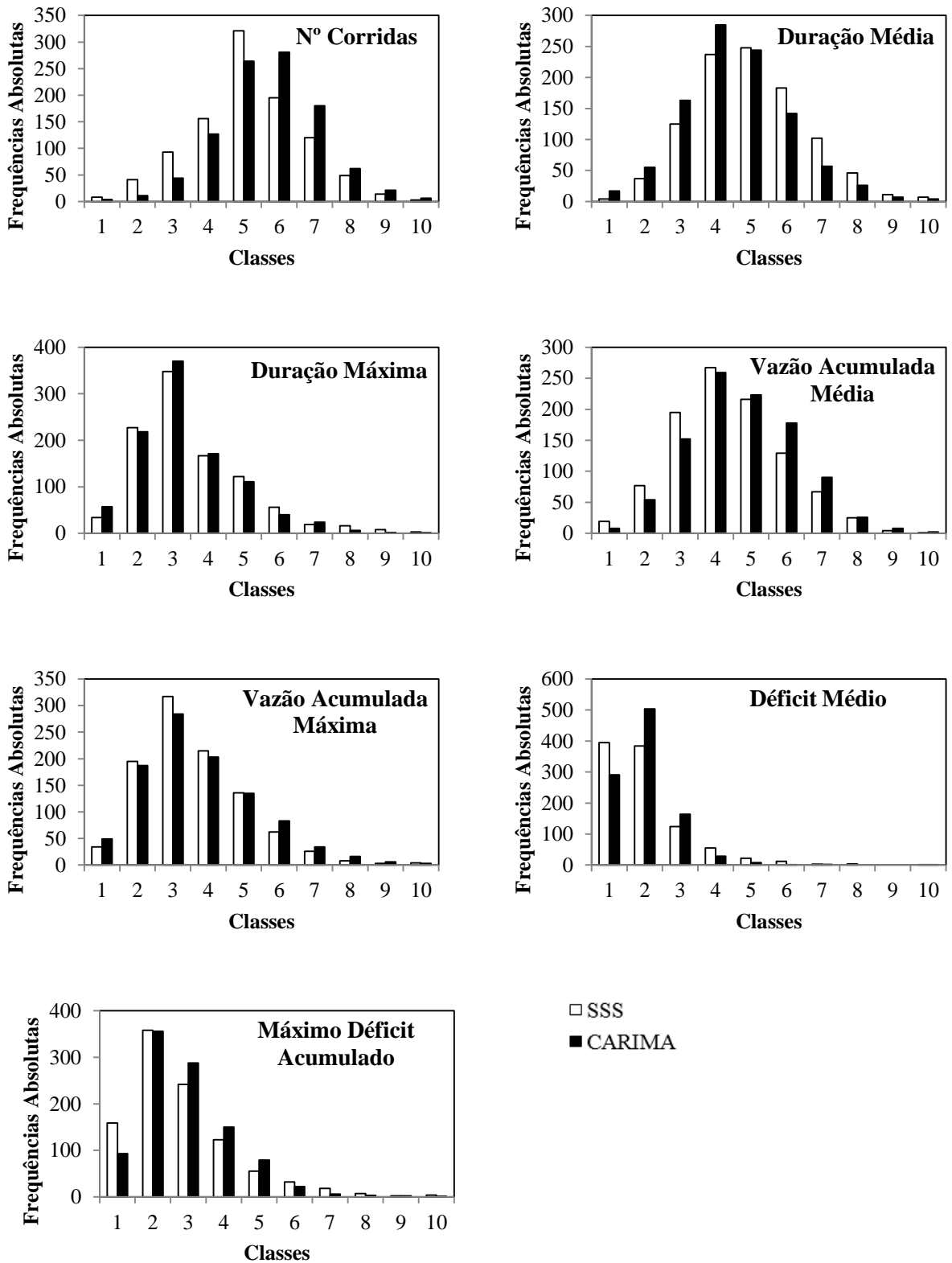
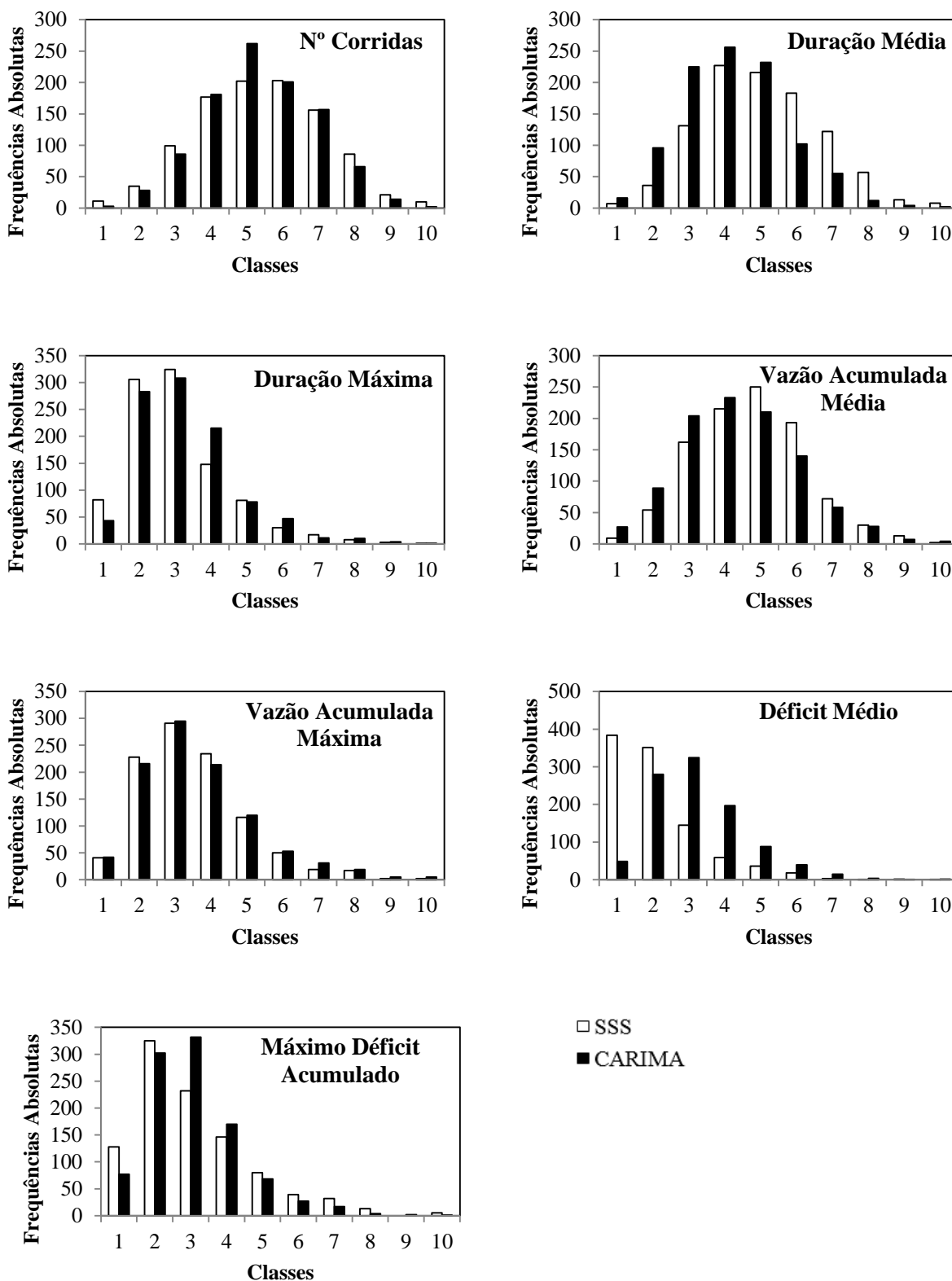
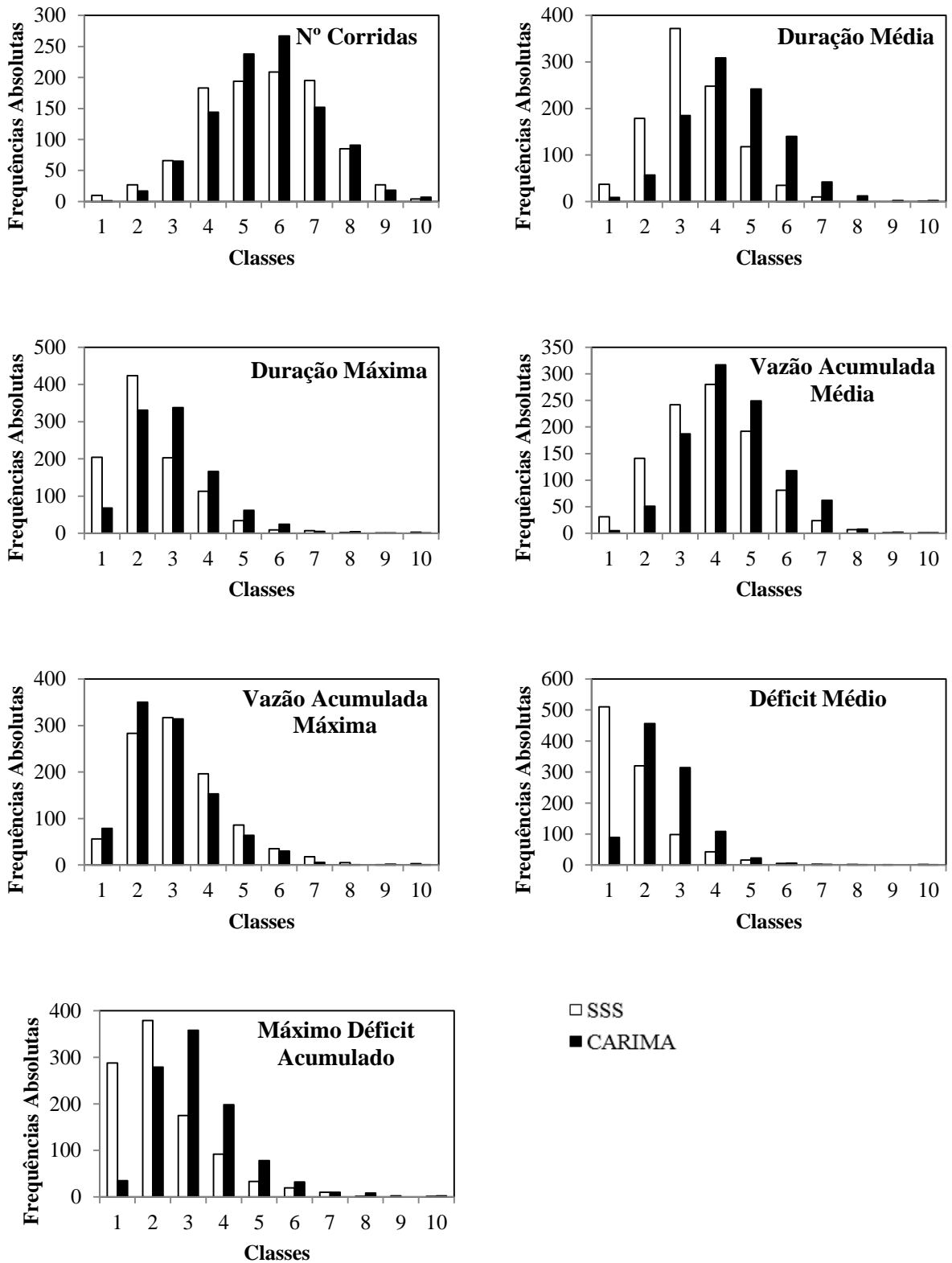


Figura 57 – Histogramas para as estatísticas de longo termo em Jangada do Sul, para séries sintéticas mensais



**Figura 58 – Histogramas para as estatísticas de longo termo em Rio Negro, para séries sintéticas mensais**



**Figura 59 – Histogramas para as estatísticas de longo termo em Fazenda Maracanã, para séries sintéticas mensais**