



Paula Medina Maçaira Louro

**Modelos de Simulação Estocástica de Cenários
de Vazão com Incorporação de Variáveis
Climáticas**

Tese de Doutorado

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-graduação em Engenharia de Produção da PUC-Rio.

Orientador: Prof. Fernando Luiz Cyrino Oliveira

Rio de Janeiro
Dezembro de 2018



Paula Medina Maçaira Louro

**Modelos de Simulação Estocástica de Cenários de Vazão
com Incorporação de Variáveis Climáticas**

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-Graduação em Engenharia de Produção da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Fernando Luiz Cyrino Oliveira

Orientador

Departamento de Engenharia Industrial - PUC-Rio

Prof. Reinaldo Castro Souza

Departamento de Engenharia Industrial - PUC-Rio

Prof. André Luís Marques Marcato

Universidade Federal de Juiz de Fora - UFJF

Prof. Guilherme Armando de Almeida Pereira

Fundação Getúlio Vargas - FGV

Prof. Daniel Takata Gomes

Fundação Instituto Brasileiro de Geografia e Estatística

Prof. Soraida Aguilar Vargas

Departamento de Engenharia Industrial - PUC-Rio

Prof. Márcio da Silveira Carvalho

Coordenador Setorial do Centro

Técnico Científico - PUC-Rio

Rio de Janeiro, 07 de dezembro de 2018

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Paula Medina Maçaira Louro

Graduou-se em Estatística pela Escola Nacional de Ciências Estatísticas em 2012. Mestrado em Engenharia Elétrica, com ênfase em Métodos de Apoio a Decisão, pela Pontifícia Universidade Católica do Rio de Janeiro em 2015. Durante o Doutorado atuou em diversos projetos de consultoria e P&D para o mercado de energia elétrica.

Ficha Catalográfica

Louro, Paula Medina Maçaira

Modelos de simulação estocástica de cenários de vazão com incorporação de variáveis climáticas / Paula Medina Maçaira Louro ; orientador: Fernando Luiz Cyrino Oliveira. – 2018.
98 f. : il. color. ; 30 cm

Tese (doutorado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2018.
Inclui bibliografia

1. Engenharia Industrial – Teses. 2. Vazão. 3. Modelos híbridos. 4. SSA. 5. PARX. 6. GAM. I. Oliveira, Fernando Luiz Cyrino. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. III. Título.

CDD: 658.5

Agradecimentos

Gostaria de agradecer primeiramente a Deus e meus guias espirituais, por me conduzirem no caminho do bem.

Aos meus pais, Leila e Arnaldo, por serem meus guias e exemplos aqui na Terra.

As minhas irmãs, Nanda e Joana, por serem minha fonte de inspiração diária.

Ao meu marido, Marcinho, por estar ao meu lado incondicionalmente e por me lembrar que a vida pode ser tão feliz quanto se deseja.

Aos meus sobrinhos e afilhados, pela doação de energia sempre que preciso recarregar.

Aos meu avós, tão importantes para minha formação de caráter e ideais.

A todos das minhas famílias, Medina, Maçaira e Louro. Sou privilegiada só por carregar o mesmo nome que vocês.

Aos amigos Lu e Victinho, pelas risadas a qualquer momento e também pelas conversas tão profundas que atravessam madrugada a dentro.

Aos amigos Fê e Xandy, pela cumplicidade em qualquer situação.

Ao meu orientador, Fernando, por ter se tornado um grande amigo nesta longa jornada.

Ao Reinaldo, por sempre ter se mantido por perto e me ensinando.

Aos colegas da PUC, principalmente a Yasmin, Bruno, Danilo, Felipe, Plutarcho, Soraida, Gheisa, Hugo, Juliana e Margarete por terem tornado um pouco mais leve o dia-a-dia.

Aos funcionários da PUC e do DEI, principalmente a Cláudia, sempre tão solícita e a Graça por manter tudo impecável.

A qualquer um que de alguma forma contribuiu nesse meu percurso.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001

Resumo

Louro, Paula Medina Maçaira; Cyrino Oliveira, Fernando Luiz. **Modelos de Simulação Estocástica de Cenários de Vazão com Incorporação de Variáveis Climáticas**. Rio de Janeiro, 2018. 98p. Tese de Doutorado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

Apesar do crescimento exponencial da instalação de novas usinas eólicas nos últimos anos, a matriz energética Brasileira é composta, principalmente, por usinas hidrelétricas. Uma das principais características dos sistemas de geração com predominância hidráulica é a forte dependência dos regimes hidrológicos. Atualmente, o setor elétrico brasileiro utiliza a Energia Natural Afluente para gerar cenários hidrológicos a partir de um modelo PAR. Tal modelo é ajustado a partir dos parâmetros estimados do histórico da série temporal, isto é, não considera quaisquer informações exógenas que possam afetar os regimes hidrológicos e, conseqüentemente, a produção de energia. Estudos recentes identificaram que o uso de variáveis climáticas na modelagem de séries de aflúências nas bacias brasileiras pode servir como fator de diminuição de incertezas devido a existência de correlação entre essas variáveis. Também foram identificados benefícios ao decompor as séries hidrológicas em sinal e ruído e utilizar somente o sinal para a modelagem. Neste contexto, o desenvolvimento de modelos híbridos que combinem técnicas de decomposição das séries hidrológicas e modelos de séries temporais com variáveis exógenas são objetos de estudo deste trabalho, assim como o desenvolvimento de modelos que associem tais variáveis de forma não-linear e periódica. Essas novas abordagens contemplam o uso das técnicas de decomposição SSA e MSSA em combinação com PAR, a aplicação do modelo PARX e o desenvolvimento do modelo PGAM. Como conclusão tem-se que os modelos aplicados se mostraram eficientes para os objetivos propostos e também apresentaram melhor performance, em alguns casos, quando comparados com modelos já publicados na literatura.

Palavras-chave

Vazão; Modelos híbridos; SSA; PARX; GAM.

Abstract

Louro, Paula Medina Maçaira; Cyrino Oliveira, Fernando Luiz (Advisor). **Stochastic Simulation Models of Inflow Scenarios with Incorporation of Climate Variables**. Rio de Janeiro, 2018. 98p. Tese de doutorado – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

Despite the exponential growth of wind farms in recent years, the Brazilian energy matrix is mainly composed of hydroelectric plants. One of the main characteristics of hydroelectric generation systems is the strong dependence on hydrological regimes. Currently, the Brazilian electric sector uses the Natural Energy Inflow to generate hydrological scenarios from a PAR model. Such model is adjusted from the estimated parameters of the time series history, that is, it does not consider any exogenous information that could affect the hydrological regimes and, consequently, the energy production. Recent studies indicate that the use of climatic variables in the modeling of inflow series in the Brazilian basins may serve as a factor to reduce uncertainties due to the existence of correlation between these variables. It was also identified benefits by decomposing hydrological series into signal and noise and using only the signal for modeling. In this context, the development of hybrid models that combine techniques of decomposition of the hydrological series and time series models with exogenous variable are study objects of this work, as well as the development of models that associate such variables in a non-linear and periodic way. These new approaches contemplate the use of SSA and MSSA decomposition techniques in combination with PAR, the application of the PARX and the development of the PGAM model. As conclusion, the applied models were efficient for the proposed objectives and also presented better performance, in some cases, when compared with models already published in the literature.

Keywords

Inflow; Hybrid models; SSA; PARX; GAM.

Sumário

1	Introdução	13
1.1	Considerações Iniciais	13
1.2	Motivação	14
1.3	Principais produtos desta Tese	16
2	Revisão da Literatura	19
2.1	Modelos univariados	19
2.2	Modelos com variáveis exógenas	21
2.3	Métodos de decomposição de dados	23
3	Metodologia	25
3.1	<i>Singular Spectrum Analysis</i> (SSA)	25
3.1.1	Algoritmo da técnica SSA	26
3.1.1.1	Primeiro estágio: Decomposição	26
3.1.1.2	Segundo estágio: Reconstrução	27
3.1.2	<i>Multichannel Singular Spectrum Analysis</i> (MSSA)	28
3.1.3	Escolha dos parâmetros L e m	29
3.2	Modelo periódico autorregressivo (PAR)	31
3.3	Modelo periódico autorregressivo com uma variável exógena (PARX)	33
3.4	Modelo Aditivo Generalizado (GAM)	35
3.4.1	GAM Periódico (PGAM)	37
3.5	Geração de Cenários Sintéticos de Vazão	38
4	Análise descritiva dos dados	41
4.1	Variáveis de vazão	41
4.1.1	Principais bacias hidrográficas Brasileiras	43
4.1.2	Séries temporais das vazões por bacia	46
4.2	Variáveis climáticas	53
4.2.1	Precipitação	53
4.2.2	O Fenômeno ENSO	56
4.2.2.1	Southern Oscillation Index (SOI) e Equatorial SOI	58
4.2.2.2	Sea Surface Temperature (SST) e Oceanic Niño Index (ONI)	63
4.3	Relação entre as vazões e as variáveis climáticas	68
5	Resultados alcançados	73
5.1	Melhores modelos	75
5.2	Resultados comparados ao artigo Lima et al.	83
6	Conclusões	87
	Referências bibliográficas	89

Lista de figuras

Figura 1.1	Alocação de capacidade instalada no Brasil. Fonte: [1].	13
Figura 4.1	Fragmento do arquivo VAZOES.DAT após usar o VazEdit.exe. Fonte: [99].	42
Figura 4.2	Fragmento do arquivo CONFHD.DAT. Fonte: [99].	43
Figura 4.3	Grandes bacias hidrográficas. Fonte: [100].	44
Figura 4.4	Esquema das usinas hidrelétricas por bacia. Fonte: [100].	45
Figura 4.5	Vazões naturais em um trecho da Bacia do Paranaíba (janeiro de 1931). Fonte: Elaboração própria.	47
Figura 4.6	Séries temporais de vazão das bacias. Fonte: [99].	49
Figura 4.7	Função de autocorrelação das séries temporais de vazão das bacias. Fonte: Elaboração própria.	52
Figura 4.8	Séries temporais de precipitação. Fonte: [103].	55
Figura 4.9	Circulação de Walker sob condições normais. Fonte: [105].	57
Figura 4.10	Circulação de Walker durante El Niño. Fonte: [105].	57
Figura 4.11	Circulação de Walker durante La Niña. Fonte: [105].	58
Figura 4.12	Localização do Southern Oscillation Index (SOI). Fonte: [107].	59
Figura 4.13	Localização do Equatorial SOI. Fonte: [107].	60
Figura 4.14	Série temporal do SOI. Fonte: [108].	61
Figura 4.15	Série temporal do Equatorial SOI. Fonte: [108].	61
Figura 4.16	Série temporal do SOI acumulado. Fonte: [108].	62
Figura 4.17	Série temporal do Equatorial SOI acumulado. Fonte: [108].	62
Figura 4.18	Localização do SST e ONI. Fonte: [107].	63
Figura 4.19	Série temporal do Niño 1+2. Fonte: [108].	64
Figura 4.20	Série temporal do Niño 3. Fonte: [108].	64
Figura 4.21	Série temporal do Niño 4. Fonte: [108].	65
Figura 4.22	Série temporal do Niño 3.4. Fonte: [108].	65
Figura 4.23	Série temporal do ONI. Fonte: [108].	66
Figura 4.24	Série temporal dos índices SST e ONI acumulados. Fonte: [108].	67
Figura 4.25	Correlação de Pearson e Spearman. Fonte: Elaboração própria.	68
Figura 4.26	Função de correlação cruzada vazões <i>versus</i> precipitação. Fonte: Elaboração própria.	70
Figura 4.27	Função de correlação cruzada vazões <i>versus</i> ENSO. Fonte: Elaboração própria.	71
Figura 4.28	Média de vazões por evento ENSO. Fonte: Elaboração própria.	72
Figura 5.1	MSRE obtido para todos os 48 modelos (cinza), com o melhor modelo que usa precipitação (azul), melhor modelo que usa um índice ENSO (verde) e o PAR (vermelho) para cada bacia.	76

Figura 5.2	MARE obtido para todos os 48 modelos (cinza), com o melhor modelo que usa precipitação (azul), melhor modelo que usa um índice ENSO (verde) e o PAR (vermelho) para cada bacia.	78
Figura 5.3	NSC obtido para todos os 48 modelos (cinza), com o melhor modelo que usa precipitação (azul), melhor modelo que usa um índice ENSO (verde) e o PAR (vermelho) para cada bacia.	79
Figura 5.4	Vazão observada (preto), previsão obtida via modelo PAR (vermelho), modelo com precipitação (azul) ou com índice ENSO (verde).	82
Figura 5.5	Comparação do MSRE entre os modelos de Lima et al. e da Tese.	85
Figura 5.6	Comparação do MARE entre os modelos de Lima et al. e da Tese.	86
Figura 5.7	Comparação do NSC entre os modelos de Lima et al. e da Tese.	86

Lista de tabelas

Tabela 1.1	Composição da matriz energética brasileira em outubro de 2018.	14
Tabela 4.1	Capacidade instalada por bacia hidrográfica.	46
Tabela 4.2	Exemplo de procedimento de cálculo das vazões incrementais	47
Tabela 4.3	Estatísticas descritivas das séries de vazão por bacia.	51
Tabela 4.4	P-valor dos testes de estacionariedade.	53
Tabela 4.5	Coordenadas das bacias para obtenção das precipitações.	54
Tabela 4.6	Estatística descritivas das séries de precipitação por bacia.	56
Tabela 5.1	Janelas de ajuste e previsão.	73
Tabela 5.2	Resumo das variáveis a serem utilizadas.	75
Tabela 5.3	Melhores modelos segundo a métrica MSRE.	77
Tabela 5.4	Melhores modelos segundo a métrica MARE.	78
Tabela 5.5	Melhores modelos segundo a métrica NSC.	80
Tabela 5.6	Resumo dos modelos mais escolhidos como melhores.	80
Tabela 5.7	Resumo das variáveis climáticas mais escolhidas como melhores.	81
Tabela 5.8	Comparação entre a métrica MSRE do PAR (<i>benchmark</i>) e do modelo proposto.	83
Tabela 5.9	Comparação entre a métrica MARE do PAR (<i>benchmark</i>) e do modelo proposto.	84
Tabela 5.10	Comparação entre a métrica NSC do PAR (<i>benchmark</i>) e do modelo proposto.	84
Tabela 5.11	Resumo dos melhores resultados encontrados.	85

Lista de Abreviaturas

ACF	–	Função de Autocorrelação
ADF	–	<i>Augmented Dickey-Fuller</i>
AIC	–	Critério de Informação de Akaike
AR	–	Modelo Autorregressivo
ARMA	–	Modelo Autorregressivo de Médias Móveis
ARX	–	Modelo Autorregressivo com Variáveis Exógenas
BIC	–	Critério de Informação Baiano
CCEE	–	Câmara de Comercialização de Energia Elétrica
CCF	–	Função de Correlação Cruzada
EARM	–	Energia Armazenada
EMD	–	<i>Empirical Mode Decomposition</i>
ENA	–	Energia Natural Afluente
ENSO	–	<i>El Niño-Southern Oscillation</i>
FT	–	Função De Transferência
GAM	–	Modelo Aditivo Generalizado
GLM	–	Modelo Linear Generalizado
GP	–	Programação Genética
MARE	–	<i>Mean Absolute Relative Error</i>
MSRE	–	<i>Mean Square Relative Error</i>
MSSA	–	<i>Multichannel Singular Spectrum Analysis</i>
NAO	–	<i>North Atlantic Oscillation</i>
NASA	–	Função de Autocorrelação Parcial
NOAA	–	<i>National Oceanic and Atmospheric Administration</i>
NSC	–	Critério de Nash-Sutcliffe
ONI	–	<i>Oscillation Niño Index</i>
ONS	–	Operador Nacional do Sistema
PAR	–	Modelo Periódico Autorregressivo
PARMA	–	Modelo Periódico Autorregressivo de Médias Móveis
PARX	–	Modelo Periódico Autorregressivo com uma Variável Exógena
PCA	–	Análise de Componentes Principais
PDO	–	<i>Pacific Decadal Oscillation</i>
PEACF	–	Função de Autocorrelação Periódica
PEPACF	–	Função de Autocorrelação Parcial Periódica

PGAM	–	Modelo Periódico Aditivo Generalizado
SAR	–	Modelo Autorregressivo Sazonal
SEB	–	Setor Elétrico Brasileiro
SIN	–	Sistema Interligado Nacional
SOI	–	<i>Southern Oscillation Index</i>
SPAR	–	Modelo Autorregressivo Periódico Espacial
SSA	–	<i>Singular Spectrum Analysis</i>
SST	–	<i>Sea Surface Temperature</i>
SVD	–	<i>Singular Value Decomposition</i>
SVM	–	<i>Support Vector Machine</i>

1

Introdução

1.1

Considerações Iniciais

O sistema de produção e transmissão de energia elétrica do Brasil, chamado Sistema Interligado Nacional (SIN), é um sistema hidro-termo-eólico de grande porte e com múltiplos proprietários. O SIN é formado pelas empresas das regiões Sul, Sudeste, Centro-Oeste, Nordeste e parte da região Norte. As demais localidades da região Norte são atendidas por sistemas isolados e apresentam um consumo que representa menos de 1% da carga total do país [1]. A maior concentração de capacidade de geração de energia está no subsistema Sudeste/Centro-Oeste e a menor no Norte, excluindo os Sistemas Isolados, como pode ser visto na Figura 1.1.

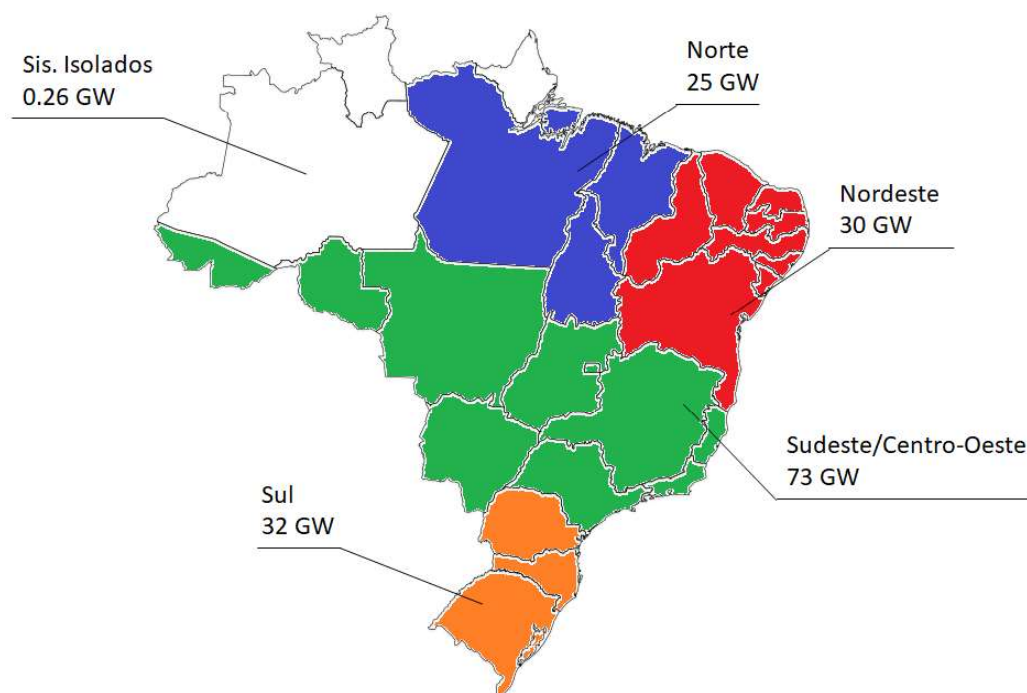


Figura 1.1: Alocação de capacidade instalada no Brasil. Fonte: [1].

Apesar do crescimento exponencial da instalação de novas usinas eólicas nos últimos anos, a capacidade instalada do SIN é composta, principalmente,

por usinas hidrelétricas distribuídas em doze bacias hidrográficas. As fontes hidrológicas representam 64% da capacidade de geração em outubro de 2018 com aproximadamente 102 GW de potência instalada (Tabela 1.1).

Tabela 1.1: Composição da matriz energética brasileira em outubro de 2018.

Tipo	Quantidade	Potência	%
		Fiscalizada (MW)	
Central Geradora Hidrelétrica	693	690,13	0,43
Central Geradora Undi-Elétrica	1	0,05	0,00
Central Geradora Eólica	545	13400,34	8,34
Pequena Central Hidrelétrica	427	5129,18	3,19
Central Geradora Solar Fotovoltaica	2257	1406,37	0,88
Usina Hidrelétrica	218	97056,84	60,39
Usina Termelétrica	3001	41040,69	25,54
Usina Termonuclear	2	1990,00	1,24
Total	7144	160713,61	100,00

Planejar o setor energético brasileiro significa, basicamente, tomar decisões acerca do despacho de usinas hidrelétricas e térmicas. As consequências de decisão equivocadas pode gerar prejuízo financeiro e até riscos de períodos de racionamento [2], como o que aconteceu em 2001 afetando quase todas as regiões do país. Mesmo com o crescimento substancial da participação das usinas eólicas no suprimento de energia, que desde junho de 2016 são a principal fonte de geração do Nordeste [4], é importante para o setor elétrico investir em modelos mais apurados para a geração de cenários de vazão com o objetivo de evitar um racionamento no futuro.

1.2 Motivação

O modelo utilizado no planejamento da operação de médio prazo, chamado Newave, consiste em determinar metas de geração para as usinas hidrelétricas e termelétricas para cada estágio ao longo do horizonte de estudo, atendendo à demanda de energia elétrica, às restrições operativas das usinas e às restrições elétricas do sistema [2]. Usar muita água para a geração de energia hoje pode levar a altos custos operacionais ou, ao racionamento elétrico, em um período seco não previsto no futuro. Por outro lado, manter uma quantidade excessiva de água nos reservatórios hoje pode levar ao vertimento em um período úmido não previsto no futuro. A decisão sobre o uso deste recurso é realizada através de um processo de otimização, chamado Programação Dinâmica Dual Estocástica, que busca minimizar os custos da operação ao ponderar o custo de todo o horizonte de planejamento. Este processo é diretamente de-

pendente dos regimes hidrológicos, sendo gerados diversos cenários futuros com o objetivo de cobrir as realizações mais prováveis dos regimes e de diminuir os riscos da decisão em cada período. Em suma, a decisão da quantidade de água que será despachada a cada período em cada um dos reservatórios é tomada a partir da consideração dos custos de operação para todo o horizonte de planejamento levando em conta a quantidade de água esperada no futuro.

Assim, a incerteza associada ao planejamento energético no Brasil exige a modelagem estocástica das séries hidrológicas (e de energia) da maneira mais eficiente possível. Percebe-se, portanto, a importância dos modelos de geração de cenários hidrológicos com vistas à otimização do despacho hidrotérmico, com consequente aumento de benefícios e confiabilidade e, sobretudo, redução de custos da operação [6]. Atualmente, tal processo de otimização tem como variável estocástica a Energia Natural Afluente (ENA), que é a energia que pode ser produzida com a vazão de água de um determinado rio a um reservatório de uma usina hidrelétrica. Os cenários de ENA são gerados a partir de um modelo Periódico Autorregressivo (PAR), ajustado com base nos parâmetros estimados do histórico da mesma [7]. Esse modelo de geração de cenários assume que as séries hidrológicas são estacionárias, lineares e seguem distribuição Normal, dentre outras características [8]. Além disso, é um modelo univariado, isto é, não considera quaisquer informação exógena que possa afetar os regimes hidrológicos e, consequentemente, a produção de energia.

Estudos realizados recentemente apontam que o uso de variáveis climáticas na modelagem de séries de afluências nas bacias brasileiras pode servir como fator de diminuição de incertezas devido aos altos níveis de correlação entre essas variáveis [9, 10, 11, 12, 13, 14, 15, 16]. As informações climáticas mais utilizadas estão relacionadas à ocorrência do fenômeno El Niño, que é o aquecimento das águas do Oceano Pacífico, e também ao fenômeno La Niña, o oposto do primeiro. Além da consideração de informações climáticas, alguns autores identificaram benefícios ao realizar uma decomposição nas séries hidrológicas, isto é, modelos híbridos que combinam técnicas de decomposição de dados e modelos de série temporal, principalmente os de inteligência artificial, tem mostrado serem um caminho para a melhora na modelagem [17, 18, 19, 20]. Tal melhoria encontrada nos modelos de inteligência artificial pode ser proveniente da possibilidade de modelar a relação entre as séries hidrológicas de forma não linear. Maçaira et al. [21] realizou uma Revisão Sistemática da literatura e identificou que modelos de regressão, Redes Neurais, ARIMAX, *Support Vector Machine* e modelos estruturais figuram entre os mais utilizados em estudos que desejam a incorporação de variáveis exógenas. Todavia, não foi encontrada a aplicação de um modelo estatístico que relacione as variáveis resposta

e regressoras de forma não-linear e com equações periódicas.

Neste contexto, o desenvolvimento de modelos híbridos que combinem técnicas de decomposição das séries hidrológicas e modelos de séries temporais, de forma que a informação de variáveis de fenômenos físicos e meteorológicos seja levada em consideração, são objetos de estudo deste trabalho, assim como o desenvolvimento de modelos que associem tais variáveis de forma não-linear e periódica. A expectativa é que as novas abordagens propostas tenham capacidade de modelagem e simulação superior a atualmente empregada.

Assim, o principal objetivo desta Tese é a proposição de uma nova abordagem metodológica para a geração de séries sintéticas que seja capaz de capturar as características particulares das séries de vazão e também incorporar variáveis climáticas de forma a aperfeiçoar a modelagem e simulação. Os objetivos secundários podem ser enumerados como:

1. Aplicar a técnica *Singular Spectrum Analysis* (SSA) e *Multichannel-SSA* para decompor a série temporal em componentes observáveis e não observáveis antes de aplicar o modelo PAR;
2. Adicionar uma variável explicativa ao PAR utilizando o modelo Periódico Autorregressivo com variável exógena (PARX);
3. Relacionar as variáveis climáticas e hidrológicas de forma não-linear através do modelo Periódico Aditivo Generalizado (PGAM); e
4. Gerar cenários sintéticos das séries hidrológicas garantindo a não ocorrência de valores negativos.

1.3

Principais produtos desta Tese

Além dos diversos artigos e resumos publicados e apresentados em congressos nacionais e internacionais, esta Tese resultou, até o momento, na publicação de três artigos em periódicos nacionais e internacionais:

1. Maçaira, P.M.; Cyrino Oliveira, F.L.; Ferreira, P.G.C.; de Almeida, F.V.N; Souza, R.C. **Introducing causal PAR(p) model to evaluate the influence of climate variables in reservoir inflows: a brazilian case.** Pesquisa Operacional, v. 37, p. 107-128, 2017.
 - Em Maçaira et al. [15] a inclusão de variáveis climáticas foi realizada através dos resíduos do modelo PAR com o modelo de Regressão Dinâmica. As séries temporais utilizadas foram as de vazão natural

de oito usinas hidrelétricas brasileiras e sete índices relacionados ao ENSO e um relacionado as Manchas Solares. A partir das métricas de erro MASE, MAPE e RMSE e permitindo que mais de uma variável climática seja utilizada por vez, concluiu-se que, em média, o causal PAR(p) melhorou os resultados do PAR em 3%.

2. Maçaira, P.M.; Thomé, A.M.T.; Cyrino Oliveira, F.L.; Ferrer, A.L.C. **Time series Analysis with explanatory variables: A systematic literatura review**. Environmental Modelling & Software, v. 107, p. 199-209, 2018.

– Após constatar um *gap* na literatura em relação a revisão de modelos de séries temporais que consideram variáveis exógenas, decidiu-se desenvolver o trabalho contido em Maçaira et al. [21]. Uma rigorosa e reprodutível revisão sistemática da literatura foi empregada em conjunto com uma análise bibliométrica a fim de estudar a evolução dos métodos de séries temporais com variáveis explanatórias. Foi possível identificar que o maior número de artigos está relacionado a área de ciências ambientais, seguido de economia e saúde. O modelo de regressão é o método com o maior número de aplicações, seguido por Redes Neurais e *Support Vector Machine*, que tiveram crescimento rápido e recente.

3. Huang, X.; Maçaira, P.M.; Hassani, H.; Cyrino Oliveira, F.L.; Dhesi, G. **Hydrological natural inflow and climate variables: Time and frequency causality analysis**. Physica A, v. 516, p. 480-495, 2019.

– A partir de uma parceria internacional estabelecida durante o *International Symposium on Forecasting*, o trabalho Huang et al. [22] foi desenvolvido com o objetivo de analisar a influência das variáveis climáticas nas principais bacias brasileiras através de cinco métodos de detecção de causalidade que abrangem abordagens empíricas bem estabelecidas e também inovadoras. Tanto os testes de causalidade no domínio do tempo como no domínio da frequência obtêm evidências válidas de causalidade unidirecional para um grupo de séries, isto é, conclui-se que algumas variáveis climáticas influenciam significativamente bacias brasileiras.

Esta Tese está organizada em seis capítulos, iniciados por esta Introdução, que aborda as considerações iniciais, motivação e objetivos.

No Capítulo 2 é realizada uma revisão da literatura acerca de modelos de séries temporais univariados e aqueles que incorporam variáveis exógenas, além de métodos de decomposição de dados aplicados à área de hidrologia. A seguir, no Capítulo 3, a metodologia e embasamento matemático das técnicas e modelos empregados na pesquisa são descritos.

No Capítulo 4 uma análise descritiva das variáveis hidrológicas e climáticas é realizada. Em seguida, no Capítulo 5, são apresentados os resultados alcançados através das metodologias descritas e também a comparação com os resultados obtidos através de um modelo já consolidado na literatura.

Finalmente, no Capítulo 6, as conclusões são desenhadas juntamente com um resumo de todas as melhorias alcançadas.

2

Revisão da Literatura

Os métodos para previsão/simulação de vazão podem ser divididos em dois grupos: conceituais hidrológicos (ou físicos) e baseados nos dados históricos (ou *data-driven*). Os modelos conceituais hidrológicos visam modelar os processos físicos subjacentes ao sistema para explicar as vazões resultantes, por exemplo, em um modelo hidrológico do tipo *rainfall-runoff*, que é regido por variáveis como evaporação, umidade do solo, temperatura etc. Já os modelos do tipo *data-driven* utilizam as relações empíricas observadas em séries temporais históricas para calibrar os parâmetros do modelo e por isso são capazes de representar processos complexos baseados em critérios matemáticos sem um conhecimento mais profundo dos processos físicos subjacentes. Uma revisão da literatura de ambas as categorias de métodos pode ser encontrada em Zhang et al. [23]. Como o foco desta Tese é utilizar modelos do tipo *data-driven*, a revisão da literatura será baseada no segundo tipo de modelos.

2.1

Modelos univariados

Modelos de séries temporais são populares e úteis para previsão de longo prazo e geração de dados sintéticos. Existe uma grande variedade de métodos que atendem a esse propósito e a seleção de um método adequado para modelagem e posterior previsão e/ou simulação de um problema particular depende de muitos fatores, como, quantidade de séries, precisão necessária, tempo disponível, interpretabilidade dos resultados etc. Dentre os métodos univariados, isto é, métodos que levam em consideração somente o histórico da própria série de tempo para modelagem, os mais utilizados são pertencentes à família Box & Jenkins [24, 25], pois segundo Salas et al. [26] os processos naturais são, em geral, estacionários, condição necessária para aplicação de tal metodologia. Na área de hidrologia, mais especificamente para dados de vazão, os modelos mais aplicados possuem a característica de serem modelos periódicos, isto é, modelos que possuem a capacidade de capturar a dependência não somente do intervalo do tempo entre as observações, mas também do período dos dados [27], sendo os que mais se destacam o modelo Periódico Autorregressivo (PAR) e o modelo Periódico Autorregressivo de Médias Móveis (PARMA).

Em hidrologia, o modelo PAR surgiu a partir do trabalho de Thomas & Fiering [28], de acordo com Hipel & McLeod [7], e recebe essa nomenclatura, pois aplica para cada período das observações um modelo Autorregressivo (AR). De maneira similar, o modelo PARMA pode ser considerado uma extensão periódica do modelo Autorregressivo de Médias Móveis (ARMA), onde a cada período dos dados é ajustado um ARMA. Até o início dos anos 90 os trabalhos voltados para os modelos PAR e PARMA objetivaram, em sua maioria, o desenvolvimento dos mesmos, como por exemplo, a análise das propriedades assintóticas [29], condições necessárias e suficientes para estacionariedade periódica [30]. Destaca-se também outros trabalhos que foram fundamentais para o aperfeiçoamento e popularização deste tipo de modelo: Gladyshev [31] e Jones & Brelsford [32] estudam as propriedades dos modelos periódicos; Pen-Chih [33] aplica os modelos ARMA sazonais e não sazonais em séries hidrológicas; Croley et al. [34] apresentam procedimentos para a análise da estrutura estocástica de séries periódicas com aplicação em séries temporais hidrológicas; Tiao e Gruppe [35] desenvolvem a relação entre os modelos periódicos e os modelos múltiplos autorregressivos de média móvel; Dunsmuir [36] discute a especificação de parâmetros de séries periódicas. Em 1985, Cipra [37] e Cipra & Thlusty [38] sugerem procedimentos para estimação do modelo PARMA; Noakes et al. [8] e Vecchia [39] utilizam os modelos periódicos para prever a série mensal de vazão de rios; e Vecchia [40] desenvolve o estimador de máxima verossimilhança para o modelo PARMA. Uma revisão detalhada dos modelos periódicos aplicados a séries hidrológicas pode ser encontrada em Hipel & McLeod [7].

No quesito de aplicação de modelos periódicos no ramo da hidrologia destacam-se os trabalhos que utilizam os modelos periódicos como *benchmark*, isto é, quando as métricas de erro obtidas ao aplicar o PAR ou PARMA são dadas como valores a serem superados pelas abordagens desenvolvidas. Thompstone et al. [41] utiliza a série trimestral de vazão do reservatório Lac St. Jean, localizado em Québec, para mostrar que as metodologias desenvolvidas por eles são melhores, nesse caso, do que ao utilizar o modelo PAR. Em Eshete & Vandewiele [42] o modelo multicomponente não-Gaussiano e o modelo PAR são utilizados para a previsão e simulação da vazão de rios localizados na Bélgica e na França, com periodicidade semanal e mensal, concluindo que o modelo não-Gaussiano possui melhores resultados.

Para complementar a revisão de modelos periódicos aplicados à hidrologia, destacam-se os trabalhos de Wang et al. [43] que combinam o modelo PAR com a técnica de Monte Carlo para gerar previsões de vazão do rio Zaire, em Kinshasa, com intervalo de confiança; Srinivasan & Philipose [44] usam

o PAR para gerar séries temporais de vazão e avaliar o efeito de três indicadores no desempenho de um reservatório no sudoeste da Índia. Maceira & Damázio [45] descrevem um problema ocorrido em setembro de 2000, período pré-acionamento obrigatório no Brasil, ao aplicar o modelo PAR e apresentam a solução ao reduzir a ordem do modelo. Mondal & Wasimi [46] desenvolvem novas técnicas para gerar cenários a partir do PAR e testam com dados de vazão mensal do rio Ganger obtendo uma boa representatividade da variação sazonal desta série tanto a curto quanto a longo prazo. Shao [47] desenvolve um procedimento heurístico para estimar de forma robusta os parâmetros do modelo PAR enquanto Cyrino Oliveira et al. [48] propõem uma nova abordagem via simulação *Bootstrap* para tal estimação. Lohmann et al. [2] apresentam um novo modelo para previsão das séries de vazão baseados no modelo PAR com informação espaço-temporal, chamado de SPAR, e mostram que o modelo proposto performa melhor para previsões pontuais, porém adiciona complexidade na parte de otimização do problema.

2.2

Modelos com variáveis exógenas

Com os avanços da ciência da computação, métodos que incorporam informação externa para melhorar a previsão e/ou simulação de séries temporais ganharam espaço. Estudos recentes confirmam a aplicabilidade de tais modelos para previsão das mais diversas variáveis, mostrando que o uso de variáveis exógenas apropriadas tornam os modelos de previsão mais robustos com ampla possibilidade de representar eventos futuros com características diferentes das que aconteceram no passado. Na área de hidrologia, mais especificamente vazão de rios, é comum que as variáveis utilizadas sejam relacionadas a informações climáticas.

Amarasekera et al. [9] mostram a correlação entre a temperatura da superfície do mar (SST) e as vazões dos rios Amazonas, Congo, Paraná e Nilo. Uvo & Graham [10] verificam as altas correlações entre a SST sobre os oceanos Atlântico e Pacífico com a disponibilidade de água na região amazônica desenvolvendo, assim, modelos de previsão trimestral a partir de Redes Neurais. Dettinger & Diaz [49] estudam a série de vazão de 1345 localizações e concluem que as variações do fenômeno El Niño (ENSO) são correlacionadas com as vazões da América Tropical, Europa e Austrália. Kelman et al. [50] utilizam a informação El Niño para estimar os parâmetros em cada etapa de um modelo autorregressivo para a série de energia mensal da Colômbia. Berri et al. [11] concluem que em períodos de El Niño a vazão média mensal do rio Paraná é sempre superior a vazão média constatada durante o

La Niña. Foley et al. [51] analisam a relação entre os eventos El Niño e La Niña e os rios Amazonas e Tocantins, mostrando que durante o primeiro evento há uma diminuição na vazão dos rios dessas bacias e um aumento durante o La Niña. Huang et al. [22] aplicam cinco métodos de causalidade no domínio do tempo e da frequência e concluem que alguns índices climáticos influenciam significativamente as séries de vazão das principais bacias brasileiras.

Souza Filho & Lall [52] realizam a previsão da série de vazão de seis reservatórios no Ceará usando como precursor dois índices que medem o evento El Niño e mostram que a correlação existente entre as variáveis climáticas e hidrológicas é benéfica para a previsão. Cardoso & Silva Dias [12] mostram que há melhoria na previsão da vazão do rio Paraná ao inserir a SST como variável explicativa. Soares et al. [53] analisam o efeito da variabilidade inter-anual associada aos eventos El Niño e La Niña na resposta hidrológica (relação entre vazão específica e a precipitação) das oito grandes bacias hidrológicas do Brasil. Stuck et al. [54] mostram a influência da variabilidade dos eventos El Niño nos reservatórios da América do Sul.

Fu et al. [55] calculam as elasticidades entre vazão, temperatura e precipitação da bacia dos rios Yellow, na China, e Spokane, nos Estados Unidos. Kumar & Maity [56] aplicam um modelo de regressão dinâmica para prever a vazão do rio Devil, nos Estados Unidos, inserindo o índice de Oscilação do Atlântico Norte (NAO)¹ como variável exógena. García & Mechoso [57] verificam a existência de periodicidades similares ao El Niño nos grandes rios da América do Sul. Boyer et al. [58] geram cenários de vazão do rio St. Lawrence, no Canadá, a partir da variação da temperatura e precipitação. Lima & Lall [13] incluem informação climática, traduzida por índices de temperatura da superfície do mar, em um modelo periódico autorregressivo para prever a vazão mensal de 54 usinas hidrelétricas no Brasil. Lima et al. [14] aplicam modelos lineares dinâmicos para prever a vazão das quinze principais bacias do Brasil usando precipitação e um índice El Niño.

Córdoba-Machado et al. [59] apresentam melhorias na previsão da vazão do rio Magdalena, na Colômbia, ao usar a SST e a precipitação no modelo de previsão, notando uma melhoria ainda maior quando a defasagem entre a vazão e as variáveis climáticas aumenta. Maçaira et al. [15] desenvolvem um modelo causal periódico autorregressivo para estimar a influência entre diversos índices El Niño e as séries de vazão de algumas localizações do Brasil. Silveira et al. [16] propõem o modelo periódico autorregressivo com uma variável exógena (PARX) para prever simultaneamente todas as vazões naturais do Sistema

¹O NAO caracteriza-se por uma variação em oposição de fase entre as anomalias de pressão registradas na região da Islândia e as registradas na região dos Açores

Interligado Nacional, mostrando que para o período seco das regiões Norte e Nordeste o modelo com variável exógena apresenta melhor desempenho do que o atualmente empregado. Yang et al. [60] desenvolvem previsões mensais de vazão usando Redes Neurais e informação climática do fenômeno ENSO e do índice Oscilação Decadal do Pacífico (PDO). He et al. [61] utiliza quatro índices climáticos para avaliação de sua influência na vazão do rio Yangtze, na China.

Em suma, inicialmente os trabalhos identificaram a correlação entre os regimes hidrológicos e as variáveis climáticas, principalmente as associadas a precipitação e aos fenômenos El Niño e La Niña. Posteriormente, as pesquisas aplicaram diferentes técnicas para modelar, simular e/ou prever as séries de vazão utilizando as variáveis identificadas anteriormente, mostrando que as novas abordagens superam as atualmente empregadas na maioria dos casos. Em relação a aplicação de modelos que incorporem variáveis exógenas foram identificados dois *gaps*. Em primeiro lugar, não foram encontrados trabalhos que verifiquem a correlação e/ou influência de todos os possíveis índices ENSO para cada região/bacia Brasileira. Em segundo lugar, pela relação entre variáveis hidrológicas e climáticas possuir um comportamento não-linear e apresentar uma periodicidade bem marcada, não foram encontrados trabalhos que utilizem ambas as características, isto é, modelos que admitem dados não-lineares e periódicos.

2.3

Métodos de decomposição de dados

No campo da hidrologia, modelos híbridos que combinem técnicas de decomposição de dados com modelos de séries temporais e inteligência artificial têm ganhado força por melhorar o desempenho de métodos do tipo *data-driven* [19]. Tais modelos têm como vantagens a capacidade de descrever estatisticamente as observações (*data-driven*) e também informações que podem estar ocultas nos dados (decomposição). Os métodos mais utilizados de decomposição de dados aplicados a séries hidrológicas são: Wavelet e *Singular Spectrum Analysis* (SSA), pois possuem a capacidade de separar as componentes não estacionárias e não lineares da série temporal em um conjunto de componentes mais simples e conseqüentemente mais fáceis de serem modelados [23].

Wang & Ding [62] usam um modelo de Redes Neurais e Wavelet para gerar previsões mensais do nível de água subterrânea e também valores diários de vazão; enquanto Anctil & Tape [63] investigam o desempenho da previsão de séries de vazão um dia a frente com uma estrutura similar. Zhou et al. [64] combinam Wavelet e modelo ARMA para a simulação e previsão de

séries mensais de disponibilidade de água. Adamowski [18] desenvolve modelos baseados em Wavelet e Cross-wavelet para prever a série diária de vazão. Mais estudos que aplicam modelos híbridos baseados em Wavelet podem ser vistos em Sang [65] e Nourani et al. [66].

Wu & Chau [67] desenvolveram três modelos híbridos ao combinar Redes Neurais e Regressão Linear com SSA para um modelo conceitual hidrológico. Zhang et al. [68] propõem a combinação de SSA e um modelo autorregressivo de média móvel para a previsão anual de vazão. Wang et al. [69] combinam Support Vector Machine (SVM), Programação Genética (GP) e o modelo autorregressivo sazonal (SAR) com a técnica SSA para provar que os modelos que não consideram uma técnica de decomposição apresentam informação hidrológica nos resíduos; já Yu et al. [70] combinam SSA e SVM para realizar previsões mensais de vazão e aplica na operação do reservatório das Três Gargantas, na China.

Todos os estudos aqui apresentados mostram que os modelos híbridos mostram maior acurácia do que os aplicados diretamente aos dados hidrológicos, porém não foi identificada a aplicação no Brasil, um dos objetos de estudo desta Tese de Doutorado.

Recapitulando, a revisão da literatura mostrou que, tradicionalmente, são utilizados modelos estatísticos periódicos - PAR e PARMA - para a previsão da vazão. No entanto, esses modelos não consideram qualquer informação exógena e já foi mostrado em diversos estudos que as variáveis climáticas tem correlação e influenciam as séries de vazão, sendo importante a inclusão deste tipo de conhecimento nos modelos. Mais recentemente, com o objetivo de extrair mais informações dos dados, as técnicas de decomposição são eficientemente usadas em combinação com modelos do tipo *data-driven*, aumentando a performance na modelagem e conseqüentemente, na previsão. Outro ponto importante é a ausência de aplicação de modelos estatísticos que relacionem as variáveis dependente e independentes de forma não-linear e também com equações periódicas.

No próximo capítulo serão apresentados os detalhes matemáticos necessários para compreender as propostas metodológicas desta Tese.

3

Metodologia

Neste capítulo será apresentado o arcabouço metodológico necessário para compreender as metodologias aplicadas nessa pesquisa. Na seção 3.1 as técnicas de decomposição, SSA e *Multichannel*-SSA (MSSA) são exploradas, descrevendo, em detalhes, o algoritmo e explicando como serão definidos seus parâmetros. Na seção 3.2 é apresentada a formulação padrão do modelo periódico autorregressivo (PAR) juntamente com as equações para estimação dos parâmetros e métodos para identificação estrutural, enquanto na seção 3.3 o modelo periódico autorregressivo com uma variável exógena (PARX) é explorado. Na seção 3.4, o modelo aditivo generalizado (GAM) é apresentado em conjunto com seus detalhes matemáticos assim como a proposta do GAM periódico (PGAM). Na última seção a abordagem adotada para geração de cenários sintéticos de vazão é detalhada para cada modelo.

3.1

Singular Spectrum Analysis (SSA)

Singular Spectrum Analysis (SSA) é uma ferramenta poderosa de análise de séries temporais que fundamenta-se na decomposição de uma dada série em seus componentes constituintes, que são independentes (ou separáveis) e interpretáveis [71] ao incorporar as fundamentações teóricas da análise clássica de séries temporais, estatística multivariada, geometria multivariada, sistemas dinâmicos e processamento de sinais [72]. Existem vários livros dedicados à SSA ([73],[72],[74]), bem como muitos trabalhos relacionados à teoria da SSA e especialmente as várias aplicações da SSA [75, 76, 77].

Em linhas gerais, o objetivo principal da SSA é filtrar uma dada série temporal e reconstruí-la de maneira “menos ruidosa”, que pode então ser utilizada para a previsão e/ou simulação. Para isso a SSA particiona uma série temporal em séries defasadas e posteriormente, ao fazer uso de conceitos e manipulações matemáticas, obtém-se uma aproximação da série inicial.

3.1.1

Algoritmo da técnica SSA

Considere uma série temporal com valores reais de tamanho N , $X = (x_1, \dots, x_N)$. Assuma que $N > 2$ e X é uma série de valores diferentes de zero, isto é, existe pelo menos um i para o qual $x_i \neq 0$.

A técnica SSA básica fundamenta-se em dois estágios complementares e cada estágio é composto por dois passos, constituindo os quatro passos da técnica:

1. Decomposição
 - (a) Incorporação
 - (b) Decomposição em Valor Singular (SVD)
2. Reconstrução
 - (a) Agrupamento
 - (b) Média Diagonal

3.1.1.1

Primeiro estágio: Decomposição

No passo de incorporação, a série temporal original é mapeada em uma sequência de vetores defasados de tamanho L formando assim $K = N - L + 1$ vetores defasados

$$X_i = (x_i, \dots, x_{i+L-1})^T, 1 \leq i \leq K. \quad (3-1)$$

O L é denominado comprimento da janela e é o único parâmetro do estágio da decomposição, representando a quantidade de componentes em que a série original será decomposta. Tal parâmetro deve ser um valor inteiro entre $1 < L < N$, no entanto segundo o estudo de Esquivel et al. [78], valores entre $[2; \frac{N}{2}]$ são suficientes para a realização da decomposição da série temporal.

Todavia, em Golyandina & Stepanov [79] e Hassani et al. [75] é chamado atenção na escolha do L ao trabalhar com séries temporais periódicas. Para alcançar separabilidade suficiente dos componentes, sugere-se utilizar um comprimento L proporcional ao período de sazonalidade dos dados. De qualquer forma, indica-se que o L seja suficientemente grande para que cada vetor L -defasado incorpore essencialmente o comportamento da série inicial.

A série temporal multidimensional, que é uma sequência de vetores constituídos por elementos da série X , forma a matriz apresentada na Equação 3-2,

denominada como matriz de trajetória. Essa matriz é o resultado deste primeiro passo.

$$\mathbf{X} = [X_1 : \dots : X_K] = (x_{ij})_{i,j=1}^{L,K} = \begin{bmatrix} x_1 & x_2 & x_3 & \dots & x_K \\ x_2 & x_3 & x_4 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & x_N \end{bmatrix} \quad (3-2)$$

Existem duas importantes propriedades da matriz de trajetória:

1. Ambas as linhas e colunas de \mathbf{X} são subséries da série original; e
2. \mathbf{X} tem elementos iguais nas anti-diagonais e, portanto, a matriz de trajetória é Hankel, isto é, uma matriz quadrada com a anti-diagonal constante.

Em suma, o passo de incorporação é considerado como um mapeamento que transforma uma série unidimensional X em uma série multidimensional formada pelas subséries x_1, \dots, x_K ; em que $x_i = [x_i, \dots, x_{i+L-1}]^T$, sendo $1 \leq i \leq K$.

No segundo passo, de Decomposição em Valor Singular (SVD), seja $\{P_i\}^L$ uma base ortonormal no \mathbb{R}^L . Considere a seguinte decomposição da matriz de trajetória:

$$\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_L = \sum_{i=1}^L P_i Q_i^T, \quad (3-3)$$

onde $Q_i = \mathbf{X}^T P_i$ e defina $\lambda_i = \|\mathbf{X}_i\|^2 = \|Q_i\|^2$. Assim, a base $\{P_i\}_{i=1}^L$ pode ser considerada os autovetores de $\mathbf{X}\mathbf{X}^T$ e é justamente a SVD de \mathbf{X} , isto é, $\mathbf{X} = \sum_i \sqrt{\lambda_i} U_i V_i^T$, onde $P_i = U_i$ é o vetor singular a esquerda de \mathbf{X} , $Q_i = \sqrt{\lambda_i} V_i$, V_i é o vetor singular a direita, λ_i são os autovalores de $\mathbf{X}\mathbf{X}^T$ e portanto $\lambda_1 \geq \dots \lambda_L \geq 0$. A coleção $(\sqrt{\lambda_i}, U_i, V_i^T)$ será chamada de i -ésima autotripla da SVD da matriz de trajetória \mathbf{X} .

3.1.1.2

Segundo estágio: Reconstrução

O principal objetivo do passo de agrupamento é a distinção dos componentes aditivos da série temporal em termos de matrizes separáveis. Em outras palavras, neste passo, identifica-se as componentes mais correlacionadas entre si, com o intuito de organizá-las em um mesmo grupo.

Seja $d = \max\{j : \lambda_j \neq 0\}$. Uma vez que a expansão 3-3 é obtida, o procedimento particiona o conjunto de índices $\{1, \dots, d\}$ em m conjuntos disjuntos I_1, \dots, I_m . Defina $\mathbf{X}_I = \sum_{i \in I} \mathbf{X}_i$. A expansão 3-3 leva à decomposição:

$$\mathbf{X} = \mathbf{X}_{I_1} + \dots + \mathbf{X}_{I_m}. \quad (3-4)$$

O procedimento de escolha dos conjuntos I_1, \dots, I_m é chamado de agrupamento das autotriplas. Se $m = d$ e $I_j = \{j\}, j = 1, \dots, d$, o agrupamento correspondente é chamado de elementar. A escolha das autotriplas principais corresponde a aproximação da série temporal em vista da propriedade de otimização do SVD, portanto m é o único parâmetro a ser definido neste estágio, isto é, quais os grupos que representarão o sinal da série.

No último passo, cada matriz \mathbf{X}_{I_j} de 3-4 é transformada em uma nova série de tamanho N . Seja \mathbf{Y} uma matriz $L \times K$ com elementos $y_{i,j}$, $1 \leq i \leq L$, $1 \leq j \leq K$, e seja $L \leq K$. Ao aplicar a Média Diagonal a matriz \mathbf{Y} é transformada na série $(\tilde{y}_N, \dots, \tilde{y}_1)$ usando a fórmula

$$\tilde{y}_s = \frac{\sum_{(l,k) \in A_s} y_{lk}}{|A_s|}, \quad (3-5)$$

onde $A_s = \{(l, k) : l + k = s + 1\}, 1 \leq l \leq L, 1 \leq k \leq K$ e $|A_s|$ denota o número de elementos no conjunto A_s . Este procedimento corresponde em calcular a média das “antidiagonais” da matriz elementar.

A média diagonal aplicada a matriz resultante \mathbf{X}_{I_j} produz uma série reconstruída $\tilde{X}^{(k)} = (\tilde{x}_1^{(k)}, \dots, \tilde{x}_N^{(k)})$. Assim, a série inicial (x_1, \dots, x_N) é decomposta na soma de m séries reconstruídas, como na Equação 3-6.

$$x_n = \sum_{k=1}^m \tilde{x}_n^{(k)}, n = 1, \dots, N. \quad (3-6)$$

3.1.2

Multichannel Singular Spectrum Analysis (MSSA)

Como exposto anteriormente, a técnica SSA lida com a decomposição de uma única série temporal, porém como um dos objetivos deste trabalho é verificar a influência de variáveis climáticas nas séries de vazão, a extensão multivariada da técnica SSA, chamada *Multichannel Singular Spectrum Analysis*, ou MSSA, será aplicada permitindo o estudo desta influência já na fase de filtragem da modelagem.

Segundo Golyandina & Stepanov [79], a MSSA é útil para lidar com a análise e previsão e/ou simulação de múltiplas séries temporais de uma só vez, seguindo a mesma estrutura do algoritmo da técnica SSA com a diferença de fazê-lo usando um sistema de séries temporais em face de uma única série.

Considere um sistema $F^{(k)} = \left(f_j^{(k)}\right)_{j=1}^N$, $k = 1, \dots, s$ com s séries temporais de tamanho N . Escolhendo um único comprimento de janela L para todo o sistema $F^{(k)}$, onde $1 < L < N$, na fase de incorporação são obtidos $K = N - L + 1$ vetores defasados $X_j^{(k)} = (f_j^{(k)}, \dots, f_{j+L-1}^{(k)})^T$, $1 \leq j \leq K$. A matriz de trajetória do sistema multidimensional $(F^{(1)}, \dots, F^{(s)})$ é uma matriz $L \times Ks$ e tem a forma

$$\mathbf{X} = [X_1^{(1)} : \dots : X_K^{(1)} : \dots : X_1^{(s)} : \dots : X_K^{(s)}] = [\mathbf{X}^{(1)} : \dots : \mathbf{X}^{(s)}]$$

Segundo Golyandina [80], o espaço trajetória é definido por um espaço linear gerado pelos vetores defasados (colunas da matriz trajetória \mathbf{X}). Como \mathbf{X} tem dimensão $L \times Ks$, $S = \mathbf{X}\mathbf{X}^T$ tem dimensão $L \times L$ e o SVD da matriz \mathbf{X} pode ser obtido de forma análoga ao SSA.

De forma análoga, a fase de agrupamento particiona o conjunto de índices $\{1, \dots, d\}$ em m subconjuntos disjuntos I_1, \dots, I_m de modo que a matriz trajetória \mathbf{X} possa ser reescrita no que é chamado de decomposição agrupada, assim como em 3-4.

Por fim a média diagonal é aplicada a cada uma das séries decomposta e então o grupo de sinais reconstruídos, dado por $\tilde{F}^{(k)} = \left(\tilde{f}_j^{(k)}\right)_{j=1}^N$ é obtido.

3.1.3

Escolha dos parâmetros L e m

O principal objetivo de aplicação da técnica SSA e sua extensão MSSA nesse estudo é a filtragem das séries de vazão de forma que a utilização da série filtrada nos modelos de séries temporais forneçam melhores medidas de erro quando comparados com os modelos aplicados às séries sem o filtro.

Em Zhang et al. [23] e Du et al. [81], os autores identificaram que diversos estudos comprovaram o desempenho superior de previsão de modelos híbridos associados a técnicas de pré-processamento de dados. No entanto, muitos destes estudos primeiro decompõem toda a série de dados em componentes e depois dividem cada componente em conjuntos de dados de calibração e validação para estabelecer os modelos, o que envia alguma quantidade de informações futuras para os processos de decomposição e reconstrução. Como consequência, os componentes resultantes usados para prever o valor de determinado momento são calculados usando informações de valores futuros, que não estão disponíveis naquele momento específico em um exercício de previsão. Tal procedimento é chamado de *hindcast* por Zhang et al. [23].

Portanto, para evitar tal problema e garantir que em nenhum momento é transferido a série filtrada qualquer conhecimento do futuro, a escolha dos parâmetros tamanho de janela (L) e grupos para reconstrução (m) será definida a partir da combinação de parâmetros que obtiver melhor performance na fase de treinamento dos modelos. Baseado em Hassani et al. [77], o procedimento para escolha segue o passo a passo a seguir:

1. Considere uma série temporal $Y = (y_{(1,1)}, \dots, y_{(1,S)}, \dots, y_{(N,S)})$ com NS observações e periodicidade S .
2. Divida a série temporal em duas partes, digamos treinamento: $Y = (y_{(1,1)}, \dots, y_{(1,S)}, \dots, y_{(K,S)})$ com tamanho $T = KS$; e calibração: $Y = (y_{(K+1,1)}, \dots, y_{(K+1,S)}, \dots, y_{(N,S)})$ com $h = (N - K - 2)S$ observações.
3. Para $L = S, 2S, \dots, \frac{KS}{2}$
 - (a) Construa a matriz de trajetória \mathbf{X}
 - (b) Obtenha o SVD: $S = \mathbf{X}\mathbf{X}^T$
 - (c) Para $m = 1, \dots, L$
 - i. Agrupe as m primeiras componentes $\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_m$
 - ii. Aplique a média diagonal em \mathbf{X} e obtenha a série filtrada \tilde{Y}
 - iii. Aplique o modelo de série temporal escolhido em \tilde{Y} e obtenha h passos a frente $\tilde{Y} = (\tilde{y}_{(K+1,1)}, \dots, \tilde{y}_{(K+1,S)}, \dots, \tilde{y}_{(N,S)})$
 - iv. Compute o erro obtido ao comparar a previsão \tilde{Y} com o conjunto de validação.
4. O vetor de parâmetros L e m escolhido será o que obtiver menor erro.

Vale ressaltar que este procedimento automatiza a escolha da dupla L e m , porém pode custar a interpretabilidade das técnicas SSA e MSSA, dado que dessa maneira as componentes são escolhidas a partir da melhor performance em uma dada janela ao invés de componentes interpretáveis.

3.2

Modelo periódico autorregressivo (PAR)

O modelo periódico autorregressivo (PAR) é o modelo adotado pelo SEB para geração de cenários da variável Energia Natural Afluente para cada subsistema brasileiro. De acordo com Hipel & McLeod [7], ajustar um modelo PAR a uma série sazonal é similar a ajustar um modelo autorregressivo (AR) para cada período do ano, isto é, uma série com periodicidade mensal será modelada de forma que cada mês possua seu próprio modelo AR. Por isso, o PAR também pode ser chamado de PAR(p), onde p corresponde a ordem do modelo, em outras palavras, o número de termos autorregressivos que identificam o modelo. No caso da série ser mensal, p é um vetor com doze elementos, $p = [p_1, p_2, \dots, p_{12}]$, onde cada elemento fornece a ordem do modelo autorregressivo de cada mês.

Seguindo a notação comumente utilizada ao se referir ao modelo PAR, seja Z uma série com S períodos e N número de anos, então $Z = [z_{(1,1)}, z_{(1,2)}, \dots, z_{(1,S)}, \dots, z_{(N,S)}]$. O modelo PAR da série Z no período m está matematicamente descrito na Equação 3-7.

$$\left(\frac{z_{(t,m)} - \mu_m}{\sigma_m} \right) = \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}} \right) + a_{t,m}, \quad (3-7)$$

onde μ_m é a média do período m , σ_m é o desvio padrão do período m , $\varphi_i^{(m)}$ é o i -ésimo coeficiente autorregressivo do período m , p_m é a ordem do operador autorregressivo do período m e $a_{t,m}$ é a série de ruídos independentes com média 0 e desvio padrão σ_m^a .

Existem, basicamente, duas maneiras de determinar a ordem autorregressiva de cada mês: a partir da análise da função de autocorrelação periódica (PeACF) e da função de autocorrelação parcial periódica (PePACF) [82]; ou através dos critérios de informação, como o *Akaike's Information Criterion* (AIC) e o *Bayesian Information Criterion* (BIC), que são funções penalizadoras para a seleção de modelos e medidas relativas da qualidade do ajuste, com base no número de parâmetros e na função de verossimilhança [83].

A PeACF do período m e lag k pode ser calculada como na Equação 3-8.

$$\rho_k^{(m)} = \frac{\gamma_k^{(m)}}{\sigma_m \sigma_{m-k}}, \quad (3-8)$$

onde

$$\gamma_k^{(m)} = \frac{1}{N} \sum_{i=1}^N (z_{(i,m)} - \mu_m)(z_{(i,m-k)} - \mu_{m-k}). \quad (3-9)$$

Na Equação 3-10 são apresentadas as equações periódicas de Yule-Walker que servem como base para o cálculo da PePACF. A PePACF pode ser traduzida como a correlação entre $z_{(t,m)}$ e $z_{(t,m-k)}$ quando a influência de

$z_{(t,m-1)}, \dots, z_{(t,m-k+1)}$ é removida e é calculada através da resolução do sistema de Yule-Walker para diferentes valores de k . O conjunto de valores $\varphi_{kk}^{(m)}$ para $k = 1, \dots$ é considerado a função de autocorrelação parcial periódica do período m .

$$\begin{bmatrix} \rho_1^{(m)} \\ \rho_2^{(m)} \\ \rho_3^{(m)} \\ \vdots \\ \rho_k^{(m)} \end{bmatrix} = \begin{bmatrix} 1 & \rho_1^{(m-1)} & \rho_2^{(m-1)} & \dots & \rho_{k-1}^{(m-1)} \\ \rho_1^{(m-1)} & 1 & \rho_1^{(m-2)} & \dots & \rho_{k-2}^{(m-2)} \\ \rho_2^{(m-1)} & \rho_1^{(m-2)} & 1 & \dots & \rho_{k-3}^{(m-2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1}^{(m-1)} & \rho_{k-2}^{(m-2)} & \rho_{k-3}^{(m-3)} & \ddots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k1}^{(m)} \\ \varphi_{k2}^{(m)} \\ \varphi_{k3}^{(m)} \\ \vdots \\ \varphi_{kk}^{(m)} \end{bmatrix} \quad (3-10)$$

A determinação da ordem do modelo é realizada ao selecionar os p_m lags significantes, isto é, se os valores da PePACF não forem significativamente diferente de zero após o *lag* p_m , então a ordem do modelo ajustado para a série Z no período m será p_m . Para determinar se um *lag* é signficante ou não, verifica-se se o valor $\varphi_{kk}^{(m)}$ não está contido no intervalo de confiança calculado como $\pm \frac{1,96}{\sqrt{N}}$ para cada período.

Como dito, uma outra maneira de determinar a ordem autorregressiva de cada período é através dos critérios de informação, sendo os mais utilizados o AIC e o BIC. Ambos os critérios realizam um *trade-off* entre a qualidade do ajuste do modelo testado e a sua complexidade, traduzida pelo número de parâmetros. Estes critérios não fornecem um teste de hipótese e sim um estimador da qualidade relativa dos modelos. Sendo k o número de parâmetros estimados pelo modelo e \hat{L} o valor máximo da função de verossimilhança, o AIC e o BIC são calculados como segue nas Equações 3-11 e 3-12, respectivamente.

$$AIC = 2k - 2\ln(\hat{L}) \quad (3-11)$$

$$BIC = \ln(n)k - 2\ln(\hat{L}) \quad (3-12)$$

A maior diferença entre esses critérios se dá na penalização dos parâmetros, sendo mais forte no BIC do que no AIC e por isso sendo indicado o BIC no caso de seleção de um modelo dentre uma classe de modelos paramétricos, enquanto o AIC pode ser denominado como uma medida de bondade de ajuste de qualquer modelo [84]. Mais detalhes de cada critério podem ser encontrados em Akaike [85], Schwarz [83] e Aho et al. [86].

Por conta do BIC ser melhor na seleção de um modelo dentre um grupo este será o utilizado para selecionar as ordens de cada período das séries de vazão do modelo PAR, ou seja, a ordem escolhida de cada período será aquela que apresentar menor BIC.

Após definir a ordem do modelo é necessário realizar a estimação dos

parâmetros $\varphi_i^{(m)}$. Seja $\beta_m = (\varphi_1^{(m)}, \dots, \varphi_{p_m}^{(m)})$ o vetor de parâmetros autorregressivos do período m . Um estimador assintoticamente eficiente, $\hat{\beta}_m$, pode ser obtido ao resolver as Equações 3-13,

$$\gamma_l^{(m)} = \sum_{i=1}^{p_m} \hat{\varphi}_i^{(m)} \gamma_{l-i}^{(m-i)}, l = 1, \dots, p_m. \quad (3-13)$$

3.3

Modelo periódico autorregressivo com uma variável exógena (PARX)

Seja Z a série periódica já definida anteriormente e X a variável exógena na modelagem de Z , com o mesmo número de observações ($N \times S$) e com a mesma periodicidade (S) de Z . Segundo Ursu & Pereau [87] e Silveira et al. [16], baseados em Andel [88], o modelo periódico autorregressivo com uma variável exógena (PARX) da variável dependente Z e da variável exógena X pode ser matematicamente descrito como:

$$\begin{aligned} \left(\frac{z_{(t,m)} - \mu_m}{\sigma_m} \right) &= \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}} \right) \\ &+ \sum_{j=0}^{v_m} \theta_j^{(m)} \left(\frac{x_{(t,m-j)} - \mu_{m-j}^{(x)}}{\sigma_{m-j}^{(x)}} \right) + a_{t,m} \end{aligned} \quad (3-14)$$

onde, μ_m é a média da variável dependente Z do período m , σ_m é o desvio padrão da variável dependente Z do período m , $\varphi_i^{(m)}$ é o i -ésimo coeficiente autorregressivo da variável dependente Z do período m , p_m é a ordem do operador autorregressivo da variável dependente Z do período m . $\mu_m^{(x)}$ é a média da variável independente X do período m , $\sigma_m^{(x)}$ é o desvio padrão de X do período m , $\theta_j^{(m)}$ é o j -ésimo coeficiente autorregressivo da variável exógena X do período m , v_m é a ordem do operador autorregressivo da variável exógena X do período m , e $a_{t,m}$ é a série de ruídos independentes com média 0 e desvio padrão σ_m^a .

A definição das ordens autorregressivas da variável dependente e da variável exógena para cada período (p_m, v_m) é uma questão delicada no modelo PARX, pois é difícil de ser especificado. Algumas soluções já foram apresentadas, como por exemplo, o uso de programação genética por Ursu & Pereau [87]. Porém, como o foco deste trabalho não é a pesquisa por método de seleção de ordem será utilizado novamente o critério BIC. No caso do modelo PARX, para cada período, o BIC obtido será referente ao conjunto (p_m, v_m) , isto é, o conjunto de parâmetros que obtiver menor BIC será o escolhido.

A estimação dos parâmetros do modelo, assim como para o modelo PAR, é realizada via Mínimos Quadrados Ordinários [87]. Considere $Y_{ns+m} =$

$\left(\frac{z_{(t,m)} - \mu_m}{\sigma_m}\right)$ e $X_{ns+m} = \left(\frac{x_{(t,m-j)} - \mu_{m-j}^{(x)}}{\sigma_{m-j}^{(x)}}\right)$, onde $n = 0, \dots, N-1$ e $m = 1, \dots, s$, com tamanho Ns . Seja $\mathbf{w}_m = [Y_m, Y_{m+s}, \dots, Y_{(N-1)s+m}]$ e $\mathbf{a}_m = [a_m, a_{m+s}, \dots, a_{(N-1)s+m}]$ vetores de dimensão $(N \times 1)$ com T sendo o operador de transposição e $\mathbf{W}_m = [\mathbf{Y}_m \mathbf{X}_m]$ a matriz com dimensão $N \times (p_m + 1 + v_m)$, onde \mathbf{Y}_m e \mathbf{X}_m são descritos nas Equações 3-15 e 3-16, respectivamente.

$$\mathbf{Y}_m = \begin{bmatrix} Y_{m-1} & Y_{m-2} & \dots & Y_{m-p_m} \\ Y_{s+m-1} & Y_{s+m-2} & \dots & Y_{s+m-p_m} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{(N-1)s+m-1} & Y_{(N-1)s+m-2} & \dots & Y_{(N-1)s+m-p_m} \end{bmatrix} \quad (3-15)$$

$$\mathbf{X}_m = \begin{bmatrix} X_m & X_{m-1} & \dots & X_{m-v_m} \\ X_{s+m} & X_{s+m-1} & \dots & X_{s+m-v_m} \\ \vdots & \vdots & \ddots & \vdots \\ X_{(N-1)s+m} & X_{(N-1)s+m-1} & \dots & X_{(N-1)s+m-v_m} \end{bmatrix} \quad (3-16)$$

Seja

$$\beta_m = \left(\varphi^{(m)}, \theta^{(m)}\right)^T$$

o vetor paramétrico, onde

$$\varphi^{(m)} = \left(\varphi_1^{(m)}, \dots, \varphi_{p_m}^{(m)}\right)^T \text{ e}$$

$$\theta^{(m)} = \left(\theta_1^{(m)}, \dots, \theta_{v_m}^{(m)}\right)^T.$$

Dado que a Equação 3-14 é um modelo linear, pode ser escrita na forma de modelo de regressão:

$$\mathbf{w}_m = \mathbf{W}_m \beta_m + \mathbf{a}_m, m = 1, \dots, s.$$

A matriz de variância-covariância do vetor aleatório \mathbf{a}_m é $\sigma_m^2 \mathbf{I}_N$, onde \mathbf{I}_N é a matriz identidade de tamanho N . O estimador de mínimos quadrados de β_m é obtido ao minimizar a Equação 3-17.

$$\begin{aligned} S(\beta) &= \sum_{m=1}^s \mathbf{a}_m^T \mathbf{a}_m \\ &= \sum_{n=0}^{N-1} \sum_{m=1}^s \left(Y_{ns+m} - \sum_{i=1}^{p_m} \varphi_i^{(m)} Y_{ns+m-i} - \sum_{j=1}^{v_m} \theta_j^{(m)} X_{ns+m-j} \right)^2 \end{aligned} \quad (3-17)$$

Finalmente, ao aplicar o operador de diferença na Equação 3-17 tem-se os estimadores via mínimos quadrados para $\hat{\beta}_m = (\varphi^{(m)}, \theta^{(m)})^T$.

$$\hat{\beta}_m = \{\mathbf{W}_m^T \mathbf{W}_m\}^{-1} \mathbf{W}_m^T \mathbf{w}_m.$$

3.4

Modelo Aditivo Generalizado (GAM)

Na revisão da literatura foi identificado que modelos de inteligência artificial que não têm como requisito a linearidade dos dados mostraram resultados superiores aos que requerem [62, 63, 67, 69, 60, 70, 89]. Porém, não foram encontrados trabalhos que utilizem variáveis aplicativas e apliquem modelos estatísticos que não requerem tal linearidade e relacionem tais variáveis de forma periódica. Por esse motivo, optou-se pelo Modelo Aditivo Generalizado (GAM), pois sua formulação permite a modelagem das vazões de forma não-linear e possibilita a construção de equações periódicas, como poderá ser visto posteriormente nesta seção.

Neste primeiro momento será apresentada a formulação formal e que é atualmente encontrada na literatura e em seguida, na sub-seção 3.4.1 será proposta a extensão periódica, chamada de Modelo Periódico Aditivo Generalizado (PGAM).

O modelo GAM foi originalmente proposto em um artigo por Hastie & Tibshirani [90] e posteriormente, em 1990, no livro chamado *Generalized Additive Models* também da autoria de Hastie & Tibshirani [91]. Sua formulação formal pode ser vista na Equação 3-18:

$$\begin{aligned} g[E(y_i)] &= f_1(x_{i1}) + \dots + f_p(x_{ip}) + \epsilon_i, \\ y_i &\sim \text{distribuição da família exponencial}, \end{aligned} \quad (3-18)$$

onde $i = 1, \dots, N$, $g(\cdot)$ é a função de ligação, y é a variável resposta, x_1, \dots, x_p são as variáveis independentes, f_1, \dots, f_p são funções suaves desconhecidas e ϵ é a série de ruídos independentes com média 0 e desvio padrão σ^ϵ .

A função de suavização $f(\cdot)$ é composta pela soma de funções bases $b(\cdot)$ e seus coeficientes de regressão correspondentes β , formalmente escrita como:

$$f(x) = \sum_{i=1}^q b_i(x) \beta_i, \quad (3-19)$$

onde q é a dimensão da base. Funções de suavização são mais comumente conhecidas por *splines* podendo ser traduzidas como funções reais que são definidas por partes de funções polinomiais (funções base). Uma função polinomial por partes $f(x)$ é obtida dividindo-se o domínio de x em intervalos contíguos e

representando $f(\cdot)$ por polinômios separados em cada intervalo [92]. Os lugares onde dos polinômios se conectam são chamados de nós.

Se absolutamente quaisquer funções de suavização fossem permitidas, então a estimação de tais modelos invariavelmente resultaria em estimativas complexas e no *overfitting* de $f(\cdot)$. Por esta razão, os modelos são, geralmente, ajustados por regressões de penalização de *splines* a fim de regularizar a suavização de uma *spline*.

Portanto, o modelo pode ser escrito em uma forma linear como:

$$g[E(y)] = \beta \mathbf{X} + \epsilon, \quad (3-20)$$

onde \mathbf{X} é a matriz modelo e β é o vetor de coeficientes da regressão.

Então, a função objetivo a ser minimizada é:

$$\|y - \beta \mathbf{X}\|^2 + \lambda \int_0^1 [f''(x)]^2 dx, \quad (3-21)$$

onde λ é o parâmetro de suavização e a integral dos quadrados das segundas derivadas pode ser escrita como:

$$\int_0^1 [f''(x)]^2 dx = \beta^T \mathbf{S} \beta, \quad (3-22)$$

onde \mathbf{S} é a matriz de coeficientes conhecidos.

Tudo isso implica que os coeficientes da regressão podem ser obtidos (estimados) pela Equação 3-23:

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{S})^{-1} \mathbf{X}^T y. \quad (3-23)$$

$\hat{\beta}$ é chamado de estimador de mínimos quadrados ponderados neste caso. O método para obter a estimativa de β é chamado de Mínimos Quadrados Reponderados Iterativamente (P-IRLS).

Existem diversas *splines* que são adequadas para a regressão, porém, segundo Cantoni & Hastie [92] não há uma boa razão para ir além das *splines* cúbicas já que só é possível notar visualmente descontinuidade até a segunda derivada, por esse motivo nessa tese essa foi a função escolhida.

Após a definição do tipo de base a ser utilizada, um procedimento importante é escolher (estimar) um parâmetro ótimo de suavização λ e o número de dimensões da base (ou seja, seus graus de liberdade). A abordagem mais comumente utilizada é chamada Pontuação de Validação Cruzada Generalizada (GCV), que minimiza a Equação 3-24:

$$\nu_g = \frac{n \sum_{i=1}^n (y_i - \hat{f}_i)^2}{[tr(\mathbf{I} - \mathbf{A})]^2} \quad (3-24)$$

onde \mathbf{A} é a matriz de projeção. Nota-se que quando λ está próximo de 1, então a *spline* será suavizada em excesso, por outro lado quando λ está próximo de 0 então a *spline* não é penalizada. Com o número de dimensões da

base (graus de liberdade) é o oposto. Uma dimensão maior implica que o ajuste será menos suave (*overfitting*), enquanto que dimensões inferiores implicam em um comportamento mais suave dos valores ajustados.

3.4.1

GAM Periódico (PGAM)

Por natureza, os dados hidrológicos são geralmente sazonais e estacionários [8]. Ou seja, as vazões de uma determinada estação do ano são estatisticamente semelhantes de ano para ano, mas podem variar drasticamente em todas as estações do ano. Estudiosos também descobriram que as séries temporais de vazão exibem uma estrutura de autocorrelação que depende não apenas do intervalo de tempo entre as observações, mas também da estação do ano [27]. Portanto, abordagens que busquem modelar tais séries de forma periódica, isto é, de mês a mês, podem oferecer vantagem frente as que não consideram tal variação.

Não foi possível encontrar na literatura a aplicação do GAM de forma periódica, portanto, uma das inovações proposta nesta tese encontra-se justamente na proposição de uma metodologia baseada nos Modelos Aditivos Generalizados e também no PAR que resultou no Modelo Periódico Aditivo Generalizado (PGAM), com formulação:

$$g \left[E \left(\frac{z_{t,m} - \mu_m}{\sigma_m} \right) \right] = \sum_{i=1}^{p_m} f_i \left(\frac{z_{t,m-i} - \mu_{m-i}}{\sigma_{m-i}} \right) + \sum_{j=0}^{v_m} f_j \left(\frac{x_{(t,m-j)} - \mu_{m-j}^{(x)}}{\sigma_{m-j}^{(x)}} \right) + a_{t,m}, \quad (3-25)$$

onde $g(\cdot)$ é a função de ligação, f_i e f_j são *splines*, μ_m é a média da variável dependente Z do período m , σ_m é o desvio padrão da variável dependente Z do período m , p_m é a ordem da defasagem da variável dependente Z do período m . $\mu_m^{(x)}$ é a média da variável independente X do período m , $\sigma_m^{(x)}$ é o desvio padrão da variável independente X do período m , v_m é a ordem da defasagem da variável exógena X do período m , e $a_{t,m}$ é a série de ruídos independentes com média 0 e desvio padrão σ_m^a .

A escolha do parâmetro ótimo de suavização λ e o número de dimensões da base (cúbica) será através do GCV (Equação 3-24). Assim como foi feito para o PARX, as ordens p_m e v_m serão selecionadas a partir do conjunto que produzir o menor BIC para o período de treinamento.

3.5

Geração de Cenários Sintéticos de Vazão

Após a seleção do método estatístico serão construídos cenários sintéticos de vazão. Portanto, o objetivo do procedimento a ser descrito nesta seção é dado um dos modelos PAR, PARX ou PGAM ajustado à série original, aproximar o comportamento estocástico e, sinteticamente, gerar tantas novas séries temporais quanto se queira, diferentes do histórico original, mas igualmente possíveis do ponto de vista estatístico.

Segundo o manual do NEWAVE [93], a abordagem oficial atualmente aplicada para geração das séries sintéticas envolve a utilização da distribuição Lognormal com três parâmetros ajustada aos resíduos mensais ($a_{t,m}$) do modelo PAR, por dois principais motivos [6]. O primeiro está relacionado com a necessidade de geração de valores sempre positivos, por se tratar de dados hidrológicos. O segundo motivo é referente à presença de forte assimetria nos dados (e resíduos) o que inviabiliza a utilização da distribuição Normal.

Manipulando a Equação 3-7 do modelo PAR para isolar Z_t :

$$\frac{z_{(t,m)}}{\sigma_m} = \frac{\mu_m}{\sigma_m} + \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}} \right) + a_{t,m}, \quad (3-26)$$

$$z_{(t,m)} = \frac{\mu_m}{\sigma_m} \sigma_m + \sigma_m \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}} \right) + \sigma_m a_{t,m}, \quad (3-27)$$

$$z_{(t,m)} = \mu_m + \sigma_m \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}} \right) + \sigma_m a_{t,m}, \quad (3-28)$$

Assim, para garantir que não sejam gerados valores negativos de $z_{(t,m)}$:

$$z_{(t,m)} > 0, \quad (3-29)$$

$$\mu_m + \sigma_m \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}} \right) + \sigma_m a_{t,m} > 0, \quad (3-30)$$

$$\sigma_m a_{t,m} > -\mu_m - \sigma_m \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}} \right), \quad (3-31)$$

$$a_{t,m} > -\frac{\mu_m}{\sigma_m} - \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}} \right), \quad (3-32)$$

$$a_{t,m} > -\frac{\mu_m}{\sigma_m} - \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}} \right), \quad (3-33)$$

$$a_{t,m} > \Delta. \quad (3-34)$$

Portanto, a variável Δ é função apenas dos momentos (média e variância) do período m e dos coeficientes autorregressivos e é dada por:

$$\Delta = -\frac{\mu_m}{\sigma_m} - \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}} \right). \quad (3-35)$$

Definindo μ_m^a e σ_m^a como a média e o desvio padrão, respectivamente, da série de resíduo do período m ($a_{t,m}$), temos que:

$$\xi_{t,m} \sim N(\mu_\xi, \sigma_\xi^2), \quad (3-36)$$

$$a_{t,m} = e^{\xi_{t,m}} + \Delta, \quad (3-37)$$

$$a_{t,m} \sim \text{LNormal}(\mu_\xi, \sigma_\xi^2, \Delta), \quad (3-38)$$

Por se tratar de ruídos aleatórios:

$$W \sim N(0, 1), \quad (3-39)$$

$$\frac{\xi_{t,m} - \mu_\xi}{\sigma_\xi} = W, \quad (3-40)$$

$$\xi_{t,m} = W\sigma_\xi + \mu_\xi, \quad (3-41)$$

$$a_{t,m} = e^{W\sigma_\xi + \mu_\xi} + \Delta \quad (3-42)$$

Os parâmetros μ_ξ e σ_ξ são estimados de forma a preservar os momentos dos resíduos, conforme Charbeneau [94] e reproduzido por Pereira et al. [95].

$$\theta = 1 + \frac{\sigma_\xi^2}{(\mu_\xi - \Delta)^2} \quad (3-43)$$

$$\mu_\xi = \log \left(\frac{\sqrt{\sigma_m^a}}{\sqrt{\theta(\theta - 1)}} \right), \quad (3-44)$$

$$\sigma_\xi = \sqrt{\log(\theta)} \quad (3-45)$$

O parâmetro Δ , como foi visto na manipulação das equações do PAR, é o deslocamento necessário para garantir a não geração de valores negativos. Manipulando as equações dos modelos PARX e PGAM da mesma maneira é possível obter o Δ para cada metodologia, conforme exposto nas Equações 3-46 e 3-47, respectivamente.

$$\Delta_{PARX} = -\frac{\mu_m}{\sigma_m} - \sum_{i=1}^{p_m} \varphi_i^{(m)} \left(\frac{z_{(t,m-i)} - \mu_{m-i}}{\sigma_{m-i}} \right) - \sum_{j=0}^{v_m} \theta_j^{(m)} \left(\frac{x_{(t,m-j)} - \mu_{m-j}^{(x)}}{\sigma_{m-j}^{(x)}} \right) \quad (3-46)$$

$$\Delta_{PGAM} = -\frac{\mu_m}{\sigma_m} - \sum_{i=1}^{p_m} f_i \left(\frac{z_{t,m-i} - \mu_{m-i}}{\sigma_{m-i}} \right) - \sum_{j=0}^{v_m} f_j \left(\frac{x_{(t,m-j)} - \mu_{m-j}^{(x)}}{\sigma_{m-j}^{(x)}} \right) \quad (3-47)$$

Uma limitação do método de simulação via distribuição Lognormal é a não-linearidade dos resíduos, conforme afirma Cyrino Oliveira et al. [96] que propõe a utilização da técnica de computação intensiva de *Bootstrap* para

solucionar tal problema. Porém, ressalta-se que nesta tese o objetivo principal é apresentar metodologias que possuam capacidade superior de modelagem do que a atualmente utilizada, portanto, o método de geração de cenários utilizado foi exatamente o mesmo do NEWAVE, de forma que qualquer melhoria encontrada seja proveniente dos modelos propostos e não da maneira de geração de cenários.

Recapitulando, foram apresentadas neste capítulo a técnica de decomposição de dados SSA e sua extensão multivariada MSSA. Também foram expostas as equações dos modelos PAR, PARX e GAM, assim como o desenvolvimento das equações periódicas do GAM, chamado aqui de PGAM. Por fim, foram exibidas as formulações dos parâmetros de deslocamento para cada método aqui proposto.

No próximo capítulo, a base de dados de vazões e das variáveis climáticas será apresentada em conjunto com uma análise descritiva detalhada.

4

Análise descritiva dos dados

O Setor Elétrico Brasileiro utiliza a variável Energia Natural Afluente para tomar a decisão acerca do despacho das hidrelétricas e termelétricas. Porém, nesta Tese decidiu-se utilizar a série de vazão natural, dado que a transformação de água em energia depende da configuração do sistema, como por exemplo o número de turbinas em operação. Essa transformação em ENA também pode ocultar as correlações com as variáveis climáticas.

Nas próximas seções será realizada a análise descritiva das séries temporais de vazão e também das variáveis climáticas assim como a relação entre tais variáveis. Toda a análise aqui realizada assim como os resultados foram obtidos utilizando o *software* R [97].

4.1

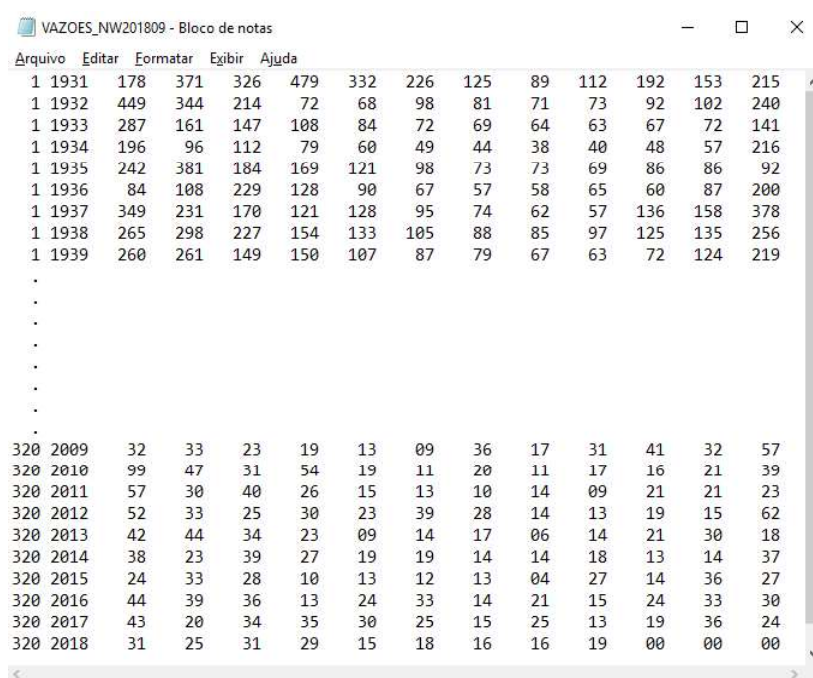
Variáveis de vazão

A implantação de reservatórios integrados a usinas hidrelétricas modifica a distribuição espacial e temporal das vazões. Esses reservatórios acumulam água nos períodos de maior disponibilidade e liberam o volume acumulado em períodos de menor disponibilidade hídrica, o que provoca um efeito de regularização dos cursos d'água onde estão implantados, de forma a garantir uma capacidade de geração, ao longo de um período, necessária para atender à demanda de energia elétrica do sistema ao qual a usina está conectada. Assim, a vazão, em um determinado instante, em um ponto de controle de um curso d'água, a jusante de um reservatório, normalmente não é mais aquela que ocorreria caso a bacia contribuinte permanecesse em suas condições naturais. Além da operação de acumulação e liberação de água, os reservatórios apresentam outras possíveis alterações no regime natural, como as alterações na quantidade de água evaporada em função da implantação do reservatório. A área do reservatório, que anteriormente apresentava uma perda de água por evaporação do solo e transpiração das plantas, passa a apresentar uma evaporação de superfície líquida que, dependendo do tamanho do reservatório, das condições climáticas regionais e da cobertura vegetal original, pode representar uma alteração significativa [98].

O setor elétrico tem adotado o termo vazão natural para mensurar a

quantidade média de água, por unidade de tempo, que cada reservatório recebe dos rios afluentes, lagos e sua própria área de drenagem [14]. As séries de vazão natural são disponibilizadas pelo ONS, mensalmente, para as usinas hidrelétricas controladas por ele. Para setembro de 2018 consideram-se 163 unidades, sendo 70 com reservatório, 89 a fio d'água e 4 com bombeamento, totalizando 101.86 MW de capacidade de geração, pois considera a entrada em operação de cinco usinas até dezembro de 2022.

Os dados históricos de vazão são medidos em m^3/s e estão disponíveis em base mensal para cada usina desde janeiro de 1931. Para usinas criadas após 1931, o ONS faz um estudo de *backforecasting* para que todas as usinas possuam a mesma quantidade de informações. Um fragmento do arquivo com as séries de vazão pode ser visto na Figura 4.1, onde a primeira coluna corresponde ao número de identificação da estação fluviométrica, a segunda identifica o ano da medição e as demais doze colunas são a média mensal da vazão natural de janeiro a dezembro do ano.



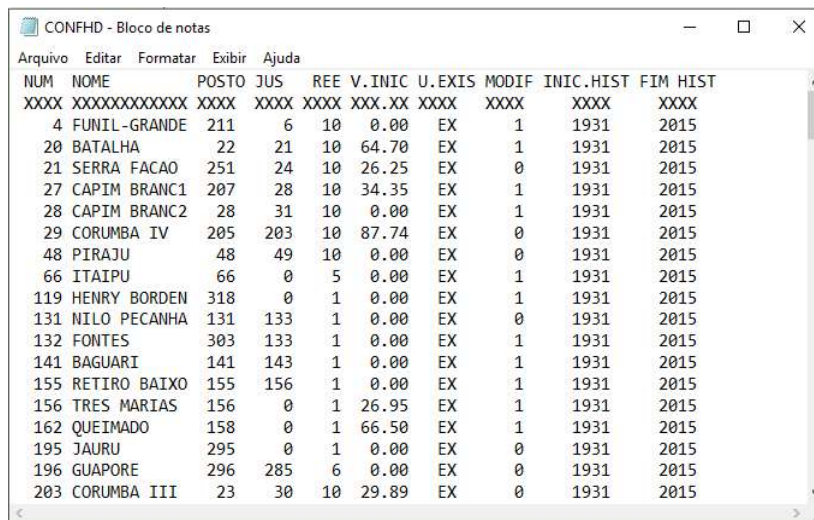
Arquivo	Editar	Formatar	Exibir	Ajuda											
1	1931	178	371	326	479	332	226	125	89	112	192	153	215		
1	1932	449	344	214	72	68	98	81	71	73	92	102	240		
1	1933	287	161	147	108	84	72	69	64	63	67	72	141		
1	1934	196	96	112	79	60	49	44	38	40	48	57	216		
1	1935	242	381	184	169	121	98	73	73	69	86	86	92		
1	1936	84	108	229	128	90	67	57	58	65	60	87	200		
1	1937	349	231	170	121	128	95	74	62	57	136	158	378		
1	1938	265	298	227	154	133	105	88	85	97	125	135	256		
1	1939	260	261	149	150	107	87	79	67	63	72	124	219		
.		
.		
.		
.		
.		
320	2009	32	33	23	19	13	09	36	17	31	41	32	57		
320	2010	99	47	31	54	19	11	20	11	17	16	21	39		
320	2011	57	30	40	26	15	13	10	14	09	21	21	23		
320	2012	52	33	25	30	23	39	28	14	13	19	15	62		
320	2013	42	44	34	23	09	14	17	06	14	21	30	18		
320	2014	38	23	39	27	19	19	14	14	18	13	14	37		
320	2015	24	33	28	10	13	12	13	04	27	14	36	27		
320	2016	44	39	36	13	24	33	14	21	15	24	33	30		
320	2017	43	20	34	35	30	25	15	25	13	19	36	24		
320	2018	31	25	31	29	15	18	16	16	19	00	00	00		

Figura 4.1: Fragmento do arquivo VAZOES.DAT após usar o VazEdit.exe.
Fonte: [99].

Esse arquivo, chamado VAZOES.DAT, é disponibilizado mensalmente dentro do *deck* de preços do NEWAVE, que é o programa utilizado pelo ONS para o planejamento da expansão e operação do setor energético brasileiro, em um formato que somente o NEWAVE compreende. Para transformar o VAZOES.DAT em um formato que seja legível é preciso utilizar o *software*

disponibilizado pela Câmara de Comercialização de Energia Elétrica (CCEE) chamado VazEdit.exe.

Para associar as vazões do arquivo VAZOES.DAT, que são por posto fluviométrico, é necessário utilizar outro arquivo do *deck* de preços, chamado CONFHD.DAT, que contém cada usina controlada pelo ONS e o número do posto fluviométrico correspondente. Veja na Figura 4.2 um fragmento deste arquivo que contém na primeira coluna o número de identificação da usina, na segunda coluna o nome da usina, na terceira coluna o posto da estação fluviométrica que o representa, na quarta coluna a identificação de qual usina é jusante, se houver (0, caso não haja) e na quinta coluna está a informação de qual Reservatório Equivalente de Energia (REE) aquela usina pertence. As próximas três colunas são relevantes apenas para o problema de otimização. E as últimas duas colunas contêm o ano de início e fim do histórico de vazões.



NUM	NOME	POSTO	JUS	REE	V.INIC	U.EXIS	MODIF	INIC.HIST	FIM HIST
4	FUNIL-GRANDE	211	6	10	0.00	EX	1	1931	2015
20	BATALHA	22	21	10	64.70	EX	1	1931	2015
21	SERRA FACA0	251	24	10	26.25	EX	0	1931	2015
27	CAPIM BRANC1	207	28	10	34.35	EX	1	1931	2015
28	CAPIM BRANC2	28	31	10	0.00	EX	1	1931	2015
29	CORUMBA IV	205	203	10	87.74	EX	0	1931	2015
48	PIRAJU	48	49	10	0.00	EX	0	1931	2015
66	ITAIPU	66	0	5	0.00	EX	1	1931	2015
119	HENRY BORDEN	318	0	1	0.00	EX	1	1931	2015
131	NILO PECANHA	131	133	1	0.00	EX	0	1931	2015
132	FONTES	303	133	1	0.00	EX	1	1931	2015
141	BAGUARI	141	143	1	0.00	EX	1	1931	2015
155	RETIRO BAIXO	155	156	1	0.00	EX	1	1931	2015
156	TRES MARIAS	156	0	1	26.95	EX	1	1931	2015
162	QUEIMADO	158	0	1	66.50	EX	1	1931	2015
195	JAURU	295	0	1	0.00	EX	0	1931	2015
196	GUAPORE	296	285	6	0.00	EX	0	1931	2015
203	CORUMBA III	23	30	10	29.89	EX	0	1931	2015

Figura 4.2: Fragmento do arquivo CONFHD.DAT. Fonte: [99].

4.1.1

Principais bacias hidrográficas Brasileiras

A partir do diagrama esquemático das usinas hidrelétricas do SIN, as usinas presentes no arquivo CONFHD.DAT podem ser agregados em doze grandes bacias hidrográficas, ver Figura 4.3. A bacia do Paraná, sozinha, representa 41% do potencial hidrelétrico de todo o Brasil e por isso pode ser separada em outras seis bacias: Paranaíba, Grande, Tietê, Paraná, Paranapanema e Iguaçu.



Figura 4.3: Grandes bacias hidrográficas. Fonte: [100].

Na Figura 4.4 é representado o esquema das usinas hidrelétricas em cada bacia considerada neste trabalho além da capacidade de geração de cada um. Em que as linhas azuis representam os rios, os triângulos representam as usinas com reservatório e os círculos as usinas a fio d'água, sendo que quando os símbolos estão completamente preenchidos significa que a usina já está em operação, enquanto que símbolos vazios representam usinas que entrarão em operação após o horizonte de estudo (2022, nesse caso). Os símbolos parcialmente coloridos equivalem a usinas que estão em construção, porém entrarão em operação até o final do período vigente. Os quadrados preenchidos são unidades de bombeamento de água e os quadrados com triângulos dentro são apenas reservatórios de água, às vezes seguidos por linhas azuis pontilhadas que correspondem a desvios da água.

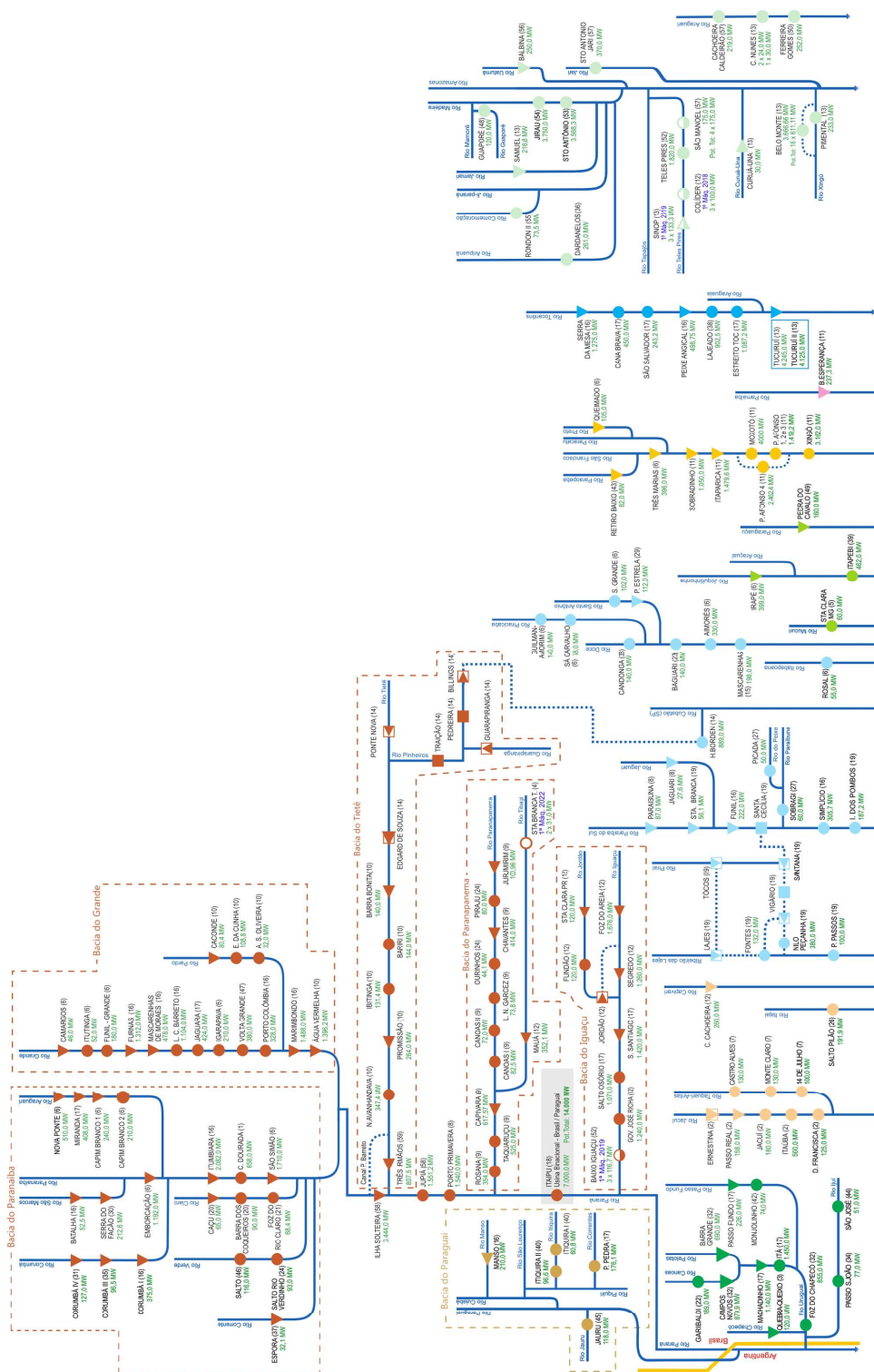


Figura 4.4: Esquema das usinas hidrelétricas por bacia. Fonte: [100].

Na Tabela 4.1 são apresentados os totais de capacidade de geração para cada bacia, considerando a bacia do Paraná como 6 sub-bacias, levando a um total de 15 bacias, onde a bacia com maior capacidade de geração passa a ser a do Amazonas, seguida pela do Paraná.

Tabela 4.1: Capacidade instalada por bacia hidrográfica.

Número	Bacia	Capacidade Instalada [MW]	% do Total
1	Paranaíba	8,34	8,07%
2	Grande	6,52	6,30%
3	Tietê	5,88	5,69%
4	Paraná	14,00	13,54%
5	Paranapanema	4,22	4,08%
6	Iguaçu	4,35	4,21%
7	Paraguai	0,66	0,64%
8	Uruguai	5,75	5,56%
9	Atlântico Sul	1,51	1,47%
10	Atlântico Sudeste	3,79	3,67%
11	Atlântico Leste	1,08	1,05%
12	São Francisco	10,56	10,21%
13	Parnaíba	0,24	0,23%
14	Tocantins	12,83	12,41%
15	Amazonas	23,64	22,87%
Total		103,36	100%

4.1.2

Séries temporais das vazões por bacia

Com inspiração no trabalho de Lima et al. [14], ao invés de realizar a modelagem e previsão e/ou simulação de cada um dos 163 usinas, as vazões serão agrupadas por bacia hidrográfica criando assim usinas equivalentes. Para calcular a vazão de cada usina equivalente é necessário usar a vazão incremental, pois a soma das vazões naturais consideraria mais de uma vez uma mesma vazão já que lida-se com um esquema de cascata.

A vazão incremental é aquela correspondente a toda vazão lateral captada entre um posto de medição e todos os postos de medição imediatamente a montante deste. A Figura 4.5 mostra um trecho da Bacia do Paranaíba onde podem ser observados os reservatórios de Corumbá IV, Corumbá III, Corumbá I, Batalha, Serra do Facão, Emborcação, Nova Ponte, Miranda, Capim Branco 1, Capim Branco 2 e Itumbiara. As vazões naturais e incrementais aos mesmos para o mês de janeiro de 1931 estão representadas na Tabela 4.2. A vazão incremental à usina de Miranda é oriunda da afluência lateral ao Rio Araguari entre os postos de medição das usinas de Nova Ponte e Miranda e é igual a $88 \text{ m}^3/\text{s}$. Já a vazão incremental da usina de Itumbiara é toda afluência lateral que ocorre entre a usina de Itumbiara e todos os seus aproveitamentos imediatamente à montante que são Corumbá I, Emborcação e Capim Branco 2. Todas as usinas de cabeceira têm suas vazões incrementais dadas diretamente

pela suas vazões naturais.

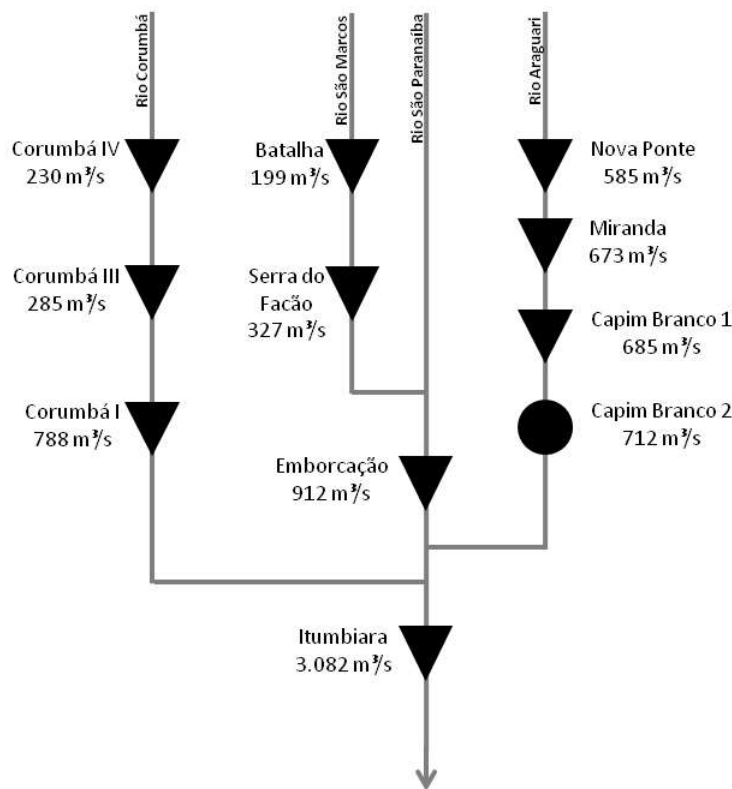


Figura 4.5: Vazões naturais em um trecho da Bacia do Paranaíba (janeiro de 1931). Fonte: Elaboração própria.

Tabela 4.2: Exemplo de procedimento de cálculo das vazões incrementais

Usina	Vazão Natural (m³/s)	Vazão Incremental (m³/s)	Procedimento de cálculo da vazão incremental
Corumbá IV	230	230	Cabeceira
Corumbá III	285	55	=285-230
Corumbá I	788	503	=788-285
Batalha	199	199	Cabeceira
Serra do Falcão	327	128	=327-199
Emborcação	912	585	=912-327
Nova Ponte	585	585	Cabeceira
Miranda	673	88	=673-585
Capim Branco 1	685	12	=685-673
Capim Branco 2	712	27	=712-685
Itumbiara	3.082	670	=3082-788-912-712

O *deck* de preços utilizado nesta Tese é referente a setembro de 2018, portanto as séries históricas de vazão estão disponíveis até o mês de dezembro

de dois anos antes, isto é, até dezembro de 2016. Portanto, após o cálculo das vazões incrementais é possível obter a vazão histórica de cada bacia hidrográfica de janeiro de 1931 até dezembro de 2016, ver Figura 4.6.

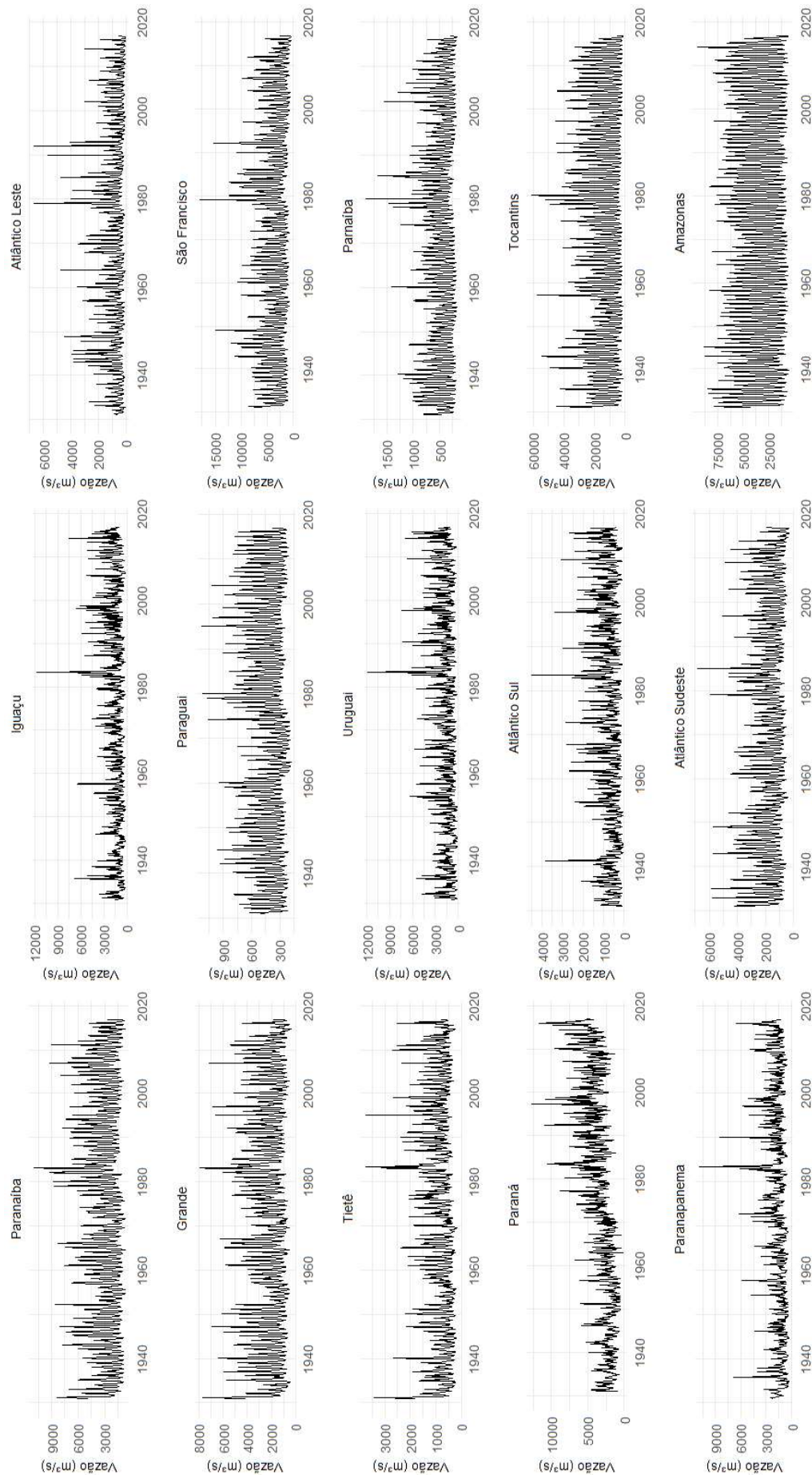


Figura 4.6: Séries temporais de vazão das bacias. Fonte: [99].

Observando os gráficos percebe-se que no geral as bacias apresentam um comportamento sazonal bem marcado com a presença de alguns *outliers*. Nota-se também que a bacia do Paraná apresenta uma tendência crescente de vazão média mensal, o que não pode ser observado nas demais séries.

Uma forma de confirmar o que foi observado através dos gráficos de séries temporais é o cálculo das medidas de dispersão que são mostrados na Tabela 4.3 como desvio padrão e coeficiente de variação. O coeficiente de variação é uma medida padronizada e que pode ser usada para comparação entre dois ou mais conjuntos de dados. A bacia do Paraguai foi a que revelou menor valor de variação (0,43) enquanto a do Atlântico Leste apresentou o maior valor (1,08). A bacia do Paraná apresenta coeficiente de variação de apenas 0,71, mostrando que a presença de uma suposta tendência nos anos finais da série ainda não afetaram tal coeficiente.

As medidas de assimetria e curtose são medidas de formato das séries e indicam o padrão da distribuição dos valores ao longo do intervalo dos dados. As séries hidrológicas tem como característica serem assimétricas positivas, ou seja, uma grande concentração de dados em valores menores do que a média, sendo traduzidos por valores de assimetria maiores do que zero. Concordando com essa característica, todas as bacias apresentaram assimetria positiva, sendo a mais assimétrica a bacia do Atlântico Leste (que também apresentou maior coeficiente de variação) e a menos assimétrica a bacia do Amazonas. Já a medida de curtose verifica o grau de achatamento da distribuição, sendo três o valor de referência de uma distribuição Normal padrão. Note que, com exceção do Amazonas, todas as bacias apresentaram curtose acima de três, chegando ao valor de 17,6 (Paranapanema), mostrando que tais distribuições possuem uma curva de frequência mais fechada ou mais aguda em sua parte superior.

As medidas de centralidade, média e mediana, são as estatísticas descritivas mais famosas e podem representar a grandeza dos dados. Na base de estudo, verifica-se através destas medidas que, a bacia do Amazonas é a que tem maior contribuição na média de vazões total (49%), seguida pelo Tocantins (20%), enquanto as bacias do Paraguai e do Parnaíba são as de menor representatividade (0,6% e 0,7%, respectivamente).

Tabela 4.3: Estatísticas descritivas das séries de vazão por bacia.

Bacia	Média	Mediana	Desvio padrão	Coefficiente de variação	Assimetria	Curtose
Paranaíba	2870	2298	1722	0,60	1,2	4,3
Grande	2077	1620	1296	0,62	1,3	4,6
Tietê	1474	1243	789	0,53	1,6	6,3
Paraná	1776	1520	1258	0,71	1,5	6,7
Paranapanema	2418	2156	1350	0,56	2,5	17,6
Iguaçu	1489	1187	1141	0,77	2,3	12,6
Paraguai	406	342	173	0,43	1,2	4,0
Uruguai	1622	1286	1287	0,79	2,0	9,3
Atlântico Sul	737	612	561	0,76	1,8	7,6
Atlântico Sudeste	1613	1283	988	0,61	1,5	5,4
Atlântico Leste	718	447	774	1,08	3,0	15,4
São Francisco	3431	2432	2549	0,74	1,4	5,4
Parnaíba	458	388	242	0,53	1,4	5,7
Tocantins	13340	9232	11129	0,83	1,1	3,7
Amazonas	33003	28076	21220	0,64	0,5	2,0

Na Figura 4.7 as características de estacionariedade e sazonalidade são verificadas através da função de autocorrelação (FAC). Nota-se a presença de picos nos *lags* multiplicativos de 12 indicando que as séries são extremamente correlacionadas com o passado da própria série e portanto sazonais.

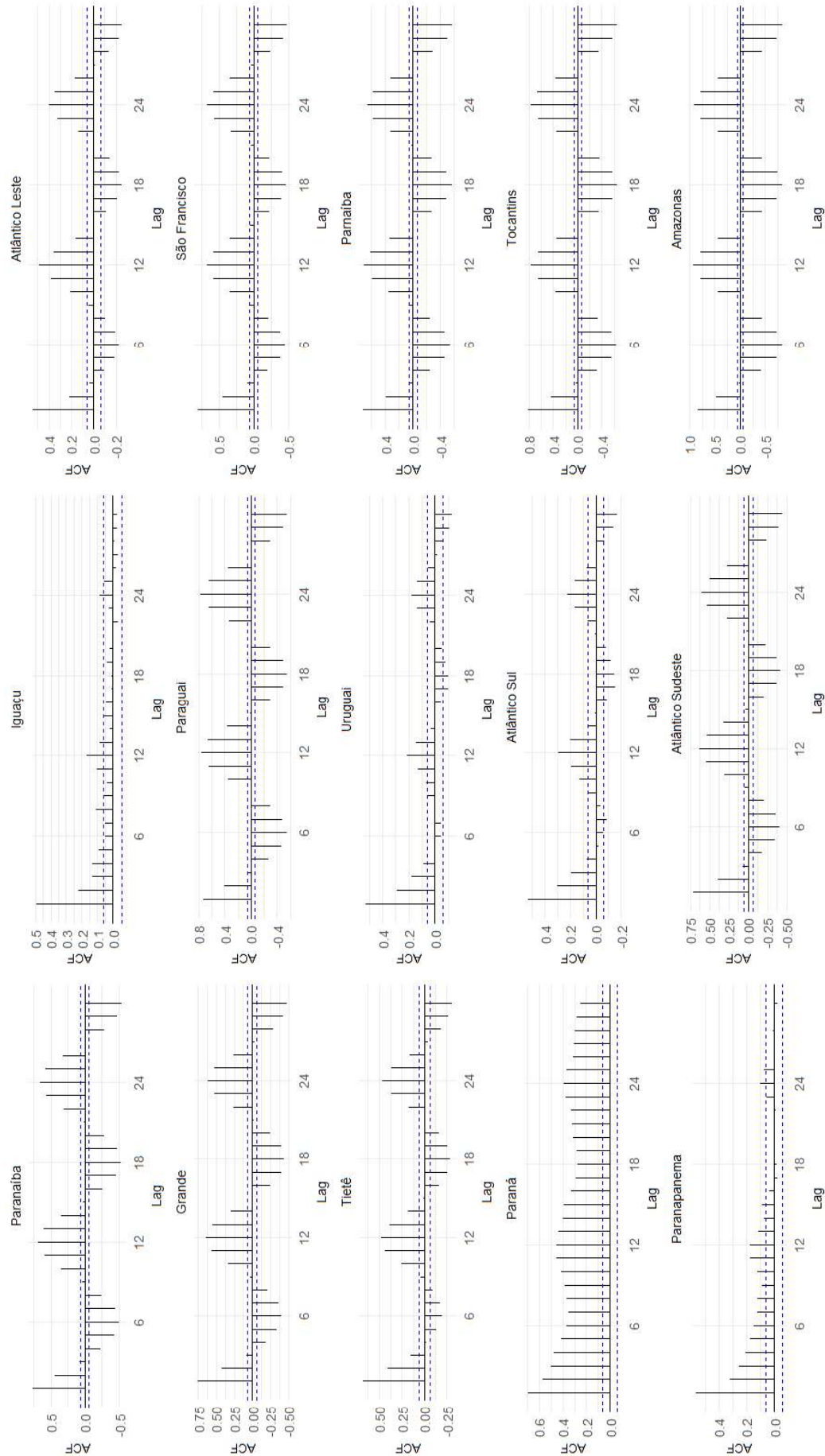


Figura 4.7: Função de autocorrelação das séries temporais de vazão das bacias.
Fonte: Elaboração própria.

Para comprovar a estacionariedade ou não da série, pode-se utilizar o teste de Dickey-Fuller Aumentado (ADF) [101] e o teste de Phillips-Perron [102]. Na Tabela 4.4 estão expostos os resultados obtidos com ambos os testes para cada uma das bacias. Contrariando o que foi observado visualmente, a série da bacia do Paraná pode ser considerada estacionária tanto pelo teste de Dickey-Fuller quanto pelo teste de Phillips-Perron que tem como hipótese alternativa a estacionariedade da série. Sugere-se portanto a utilização do modelo ARFIMA para pesquisar a ordem fracionária de diferenciação.

Tabela 4.4: P-valor dos testes de estacionariedade.

Bacia	ADF	PP	Estacionária?
Paranaíba	0,01	0,01	Sim
Grande	0,01	0,01	Sim
Tietê	0,01	0,01	Sim
Paraná	0,01	0,01	Sim
Paranapanema	0,01	0,01	Sim
Iguaçu	0,01	0,01	Sim
Paraguai	0,01	0,01	Sim
Uruguai	0,01	0,01	Sim
Atlântico Sul	0,01	0,01	Sim
Atlântico Sudeste	0,01	0,01	Sim
Atlântico Leste	0,01	0,01	Sim
São Francisco	0,01	0,01	Sim
Parnaíba	0,01	0,01	Sim
Tocantins	0,01	0,01	Sim
Amazonas	0,01	0,01	Sim

4.2

Variáveis climáticas

Nesta seção será apresentada a análise descritiva das variáveis climáticas. Algumas variáveis são locais e diferem para cada bacia, como a precipitação. Outras, como os índices oceânicos, são iguais para todas as bacias.

4.2.1

Precipitação

A Administração Nacional do Espaço e da Aeronáutica (NASA) disponibiliza, através do programa Giovanni [103], a taxa de precipitação total média de uma área em intervalos de 0,25 graus para o período de janeiro de 1979 até o mês atual. Tal base de dados foi utilizada para obter a média mensal de precipitação na área de cada bacia hidrográfica, seguindo as mesmas coordenadas utilizadas por Lima et al. 2014 [14] e expostas na Tabela 4.5. Quando uma

mesma bacia é dividida em duas áreas significa que a média entre os valores determinou a precipitação daquela bacia.

Tabela 4.5: Coordenadas das bacias para obtenção das precipitações.

Número	Bacia	Área	Latitude		Longitude	
			Sul	Norte	Oeste	Leste
1	Paranaíba	1	-19,9	-17,5	-53,7	-45,8
2	Grande	1	-24,2	-19,9	-55,3	-44,6
3	Tietê	1	-24,2	-19,9	-55,3	-44,6
4	Paraná	1	-26,3	-24,2	-54,6	-47,9
5	Paranapanema	1	-24,2	-19,9	-55,3	-44,6
6	Iguaçu	1	-26,3	-24,2	-54,6	-47,9
7	Paraguai	1	-20,3	-13,7	-59,2	-56,0
8	Uruguai	1	-28,3	-26,2	-54,0	-49,6
9	Atlântico Sul	1	-31,3	-28,3	-54,3	-50,5
10	Atlântico Sudeste	1	-23,3	-18,6	-43,5	-39,8
		2	-23,3	-22,3	-47,0	-43,5
11	Atlântico Leste	1	-18,0	-15,0	-42,0	-39,0
12	São Francisco	1	-10,6	-7,1	-44,2	-36,4
		2	-20,0	-10,6	-46,7	-44,2
13	Parnaíba	1	-8,9	-5,3	-46,1	-40,4
14	Tocantins	1	-15,8	-1,8	-51,7	-46,8
15	Amazonas	1	-12,5	-1,8	-69,9	-50,3

As sub-bacias do Paraná são tão próximas que uma mesma coordenada foi considerada para as bacias do Paraná e do Iguaçu, e também uma mesma área foi usada para representar as bacias do Grande, Tietê e Paranapanema. Os gráficos de série temporal dos dados de precipitação são mostrados na Figura 4.8 e apresentam as mesmas características das séries de vazão, isto é, componente sazonal e observações extremas, além de estacionariedade.

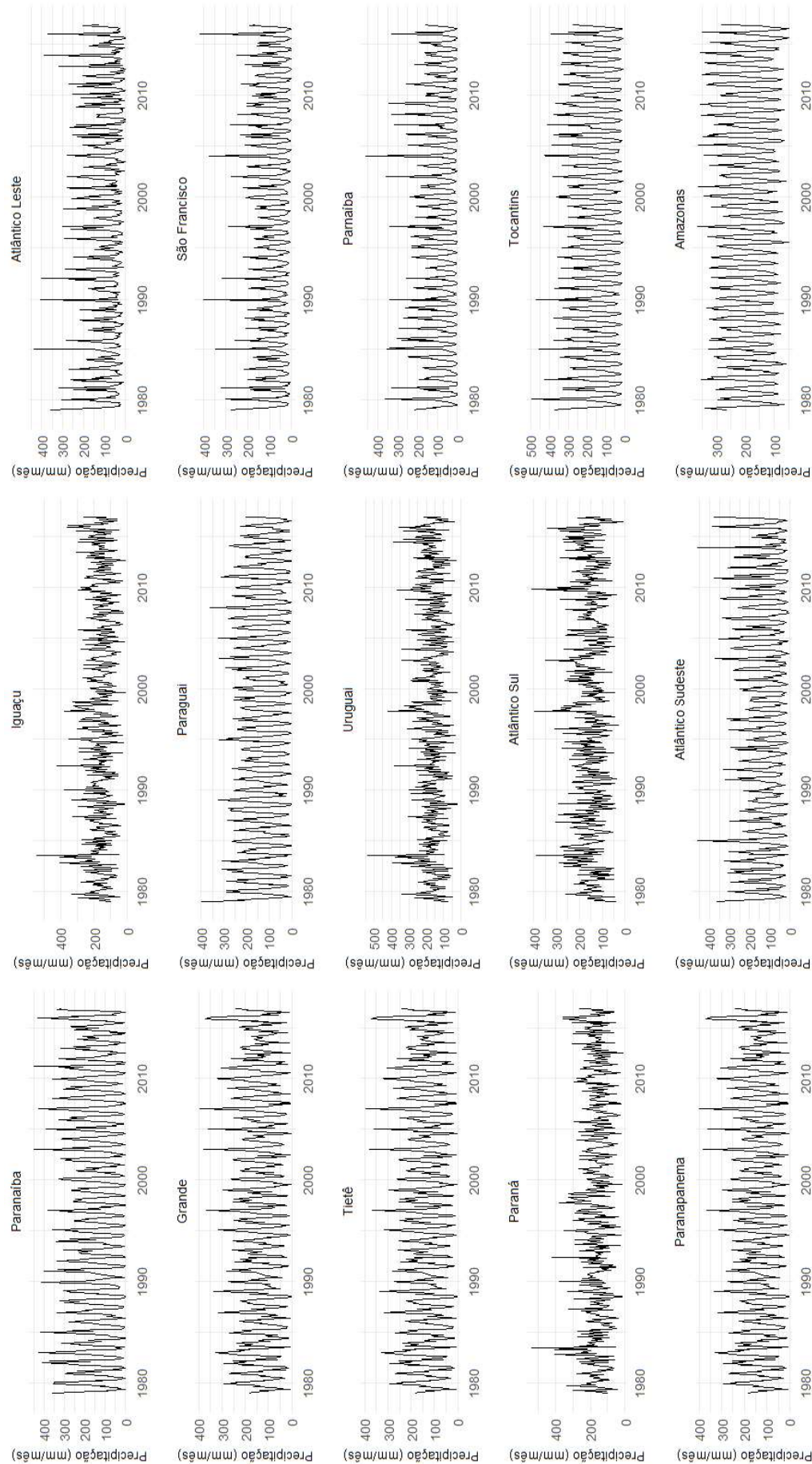


Figura 4.8: Séries temporais de precipitação. Fonte: [103].

As estatísticas descritivas de cada série de precipitação podem ser vistas na Tabela 4.6. Não surpreendentemente, por se tratar de séries hidrológicas, o formato das séries de precipitação apresentam assimetria positiva e são leptocúrticas. A série com maior variação é a pertencente a bacia do Parnaíba e a de maior média e mediana pertencente ao Amazonas que também é a com menor assimetria.

Tabela 4.6: Estatística descritivas das séries de precipitação por bacia.

Bacia	Média	Mediana	Desvio padrão	Coefficiente de variação	Assimetria	Curtose
Paranaíba	132	105	110	0,84	0,66	2,5
Grande	132	128	81	0,62	0,48	2,8
Tietê	132	128	81	0,62	0,48	2,8
Paraná	162	158	76	0,47	0,7	4,2
Paranapanema	132	128	81	0,62	0,48	2,8
Iguaçu	162	158	76	0,47	0,7	4,2
Paraguai	122	115	89	0,73	0,38	2,1
Uruguai	167	162	73	0,44	0,75	4,5
Atlântico Sul	156	152	65	0,42	0,62	3,6
Atlântico Sudeste	118	87	93	0,79	0,96	3,2
Atlântico Leste	88	56	79	0,9	1,51	5,2
São Francisco	82	62	74	0,89	1,14	4,2
Parnaíba	85	62	85	1	1,2	4,2
Tocantins	156	146	115	0,74	0,46	2,2
Amazonas	196	196	84	0,43	0,11	1,8

4.2.2

O Fenômeno ENSO

A ocorrência do fenômeno El Niño Oscilação Sul (ENSO) provoca alterações na circulação atmosférica, afetando em grande escala mudanças climáticas em diferentes regiões do mapa terrestre. A nomenclatura ENSO é usada para se referir aos fenômenos El Niño e La Niña, uma vez que há influência direta entre a atmosfera (representada pela Oscilação Sul) e o oceano (referente ao El Niño/La Niña). A Oscilação Sul, descoberta e definida pelo matemático britânico Sir Gilbert Walker em 1928, consiste na diferença de pressão atmosférica entre as regiões do Pacífico Sul, no Oriente do Tahiti, e do Oceano Índico, no Ocidente de Darwin, Austrália. A pressão atmosférica quando alta na primeira, encontra-se baixa na segunda e vice-versa.

No final dos anos 60, Bjerknes 1969 [104] constatou que a interação entre o mar e a atmosfera podia ter um grande impacto na circulação dos ventos, nas chuvas e no clima, assim, dando continuidade aos estudos de Walker, ele descreveu um padrão de circulação do ar sobre todo o Pacífico, a Circulação de Walker. Sob condições normais, ou seja, em períodos sem El Niño ou La Niña, existe um grande fluxo de massa de ar ascendente na Oceania e fluxos menores no leste da África e no norte da América do Sul (Figura 4.9). Em períodos

de El Niño e La Niña, esses fluxos são deslocados e intensificados (Figura 4.10 e Figura 4.11). As cores laranja e azul nos oceanos representam águas mais quentes e mais frias, respectivamente.

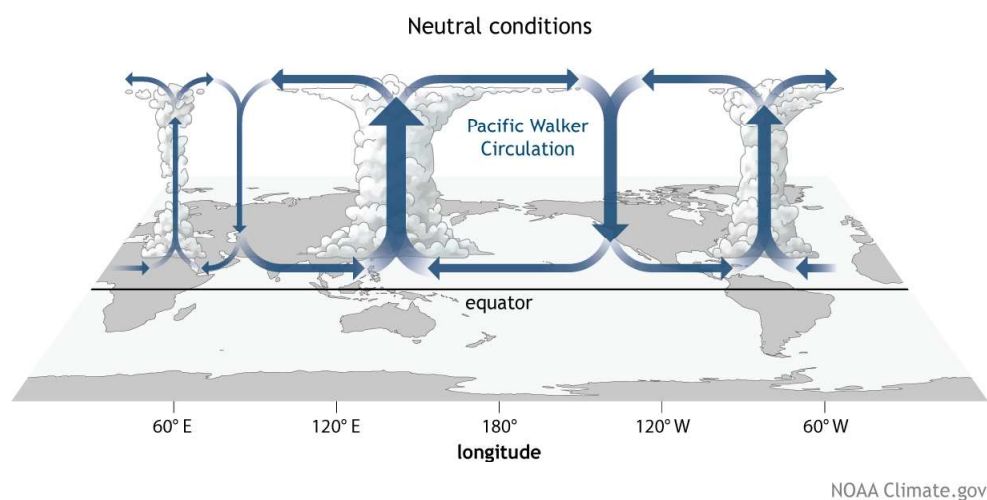


Figura 4.9: Circulação de Walker sob condições normais. Fonte: [105].

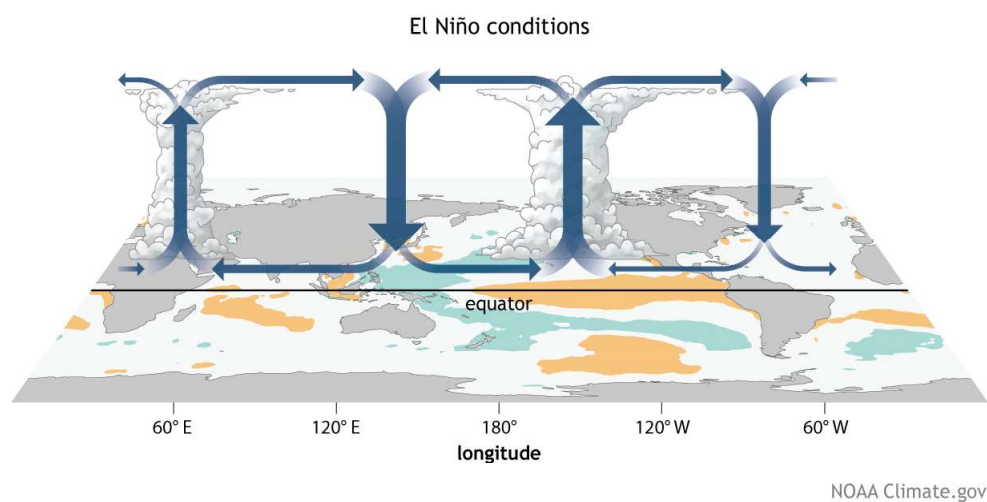


Figura 4.10: Circulação de Walker durante El Niño. Fonte: [105].

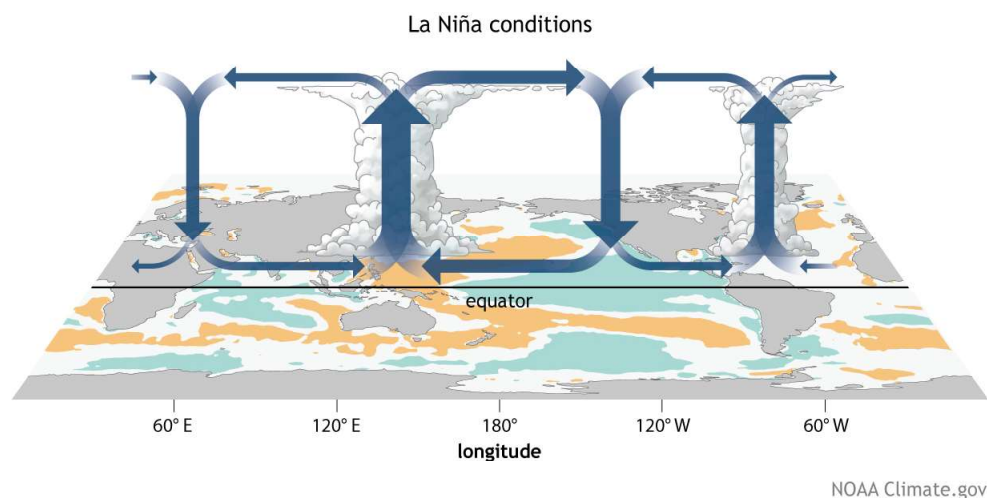


Figura 4.11: Circulação de Walker durante La Niña. Fonte: [105].

Segundo Oliveira [106], o fenômeno El Niño consiste no aquecimento anormal da temperatura das águas do Oceano Pacífico Equatorial, enquanto o La Niña representa o resfriamento anormal dessas águas na mesma região. Assim, durante esses dois estados, os fluxos de ar ascendente da circulação de Walker se intensificam, devido à anormalidade na temperatura das águas, e se deslocam para as regiões de águas mais aquecidas. Como consequência, pode haver um aumento na quantidade de chuvas nessas regiões e seca em outras, influenciando diretamente a variação do volume de águas nos reservatórios das hidrelétricas.

Nas últimas décadas, têm sido observados acontecimentos de El Niño fora do comum, o chamado El Niño Modoki. Nesse fenômeno, as águas aquecem mais no Pacífico Central, área conhecida como Niño 3.4, enquanto a área Niño 1+2 não é afetada, apesar de encontrar-se no Pacífico Oriental. Como consequência, atualmente, diversos modelos da literatura utilizam o índice Niño 3.4 para realizar previsões dos fenômenos El Niño e La Niña. Porém, muitos outros índices podem ser usados para caracterizar esses fenômenos, segundo o National Oceanic and Atmospheric Administration (NOAA), que é uma agência americana e referência quando se fala de ENSO.

4.2.2.1

Southern Oscillation Index (SOI) e Equatorial SOI

O indicador do ENSO mais antigo é o Southern Oscillation Index (SOI) e tem seu cálculo baseado na diferença de pressão atmosférica ao nível do mar entre as regiões do Tahiti, no Pacífico Ocidental, e Darwin, na Austrália (Pacífico Oriental), ver Figura 4.12.

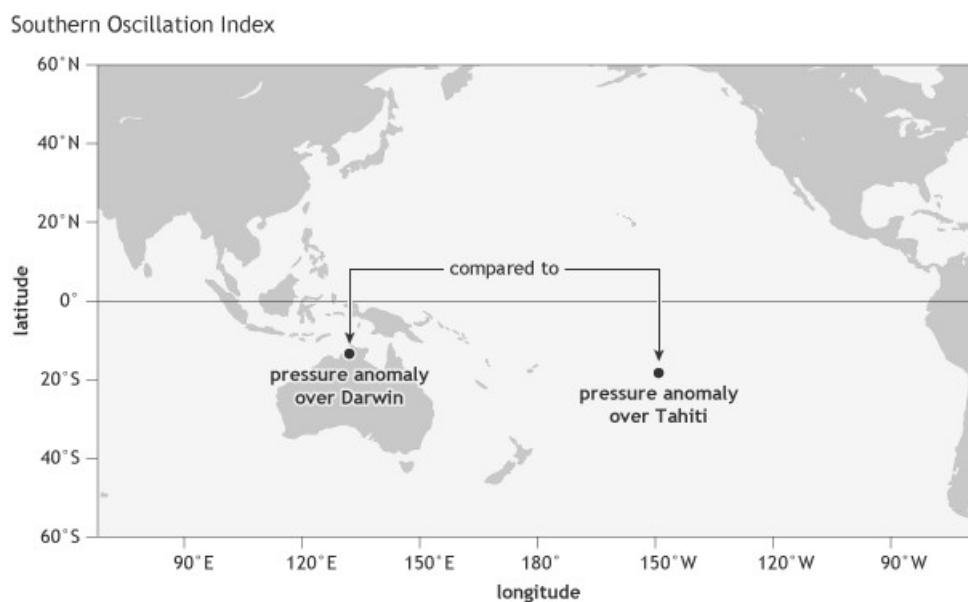


Figura 4.12: Localização do Southern Oscillation Index (SOI). Fonte: [107].

O fato do SOI se basear na pressão do nível do mar em apenas dois pontos mostra que pode ser afetado por flutuações de curto prazo, dia a dia ou semana a semana, não relacionadas ao ENSO. Mas a média dos valores do índice em meses ou estações ajuda a isolar desvios mais sustentados da média, como os associados ao ENSO.

Outra limitação do SOI é que tanto o Tahiti quanto Darwin estão localizados ao sul da linha do Equador, enquanto o fenômeno ENSO se concentra mais perto da mesma linha. Tendo essa situação em vista, o Equatorial SOI supera tal limitação uma vez que calcula a diferença da média de pressão atmosférica ao nível do mar entre duas regiões centradas na linha do Equador: a Indonésia e o Pacífico Oriental, ver Figura 4.13.

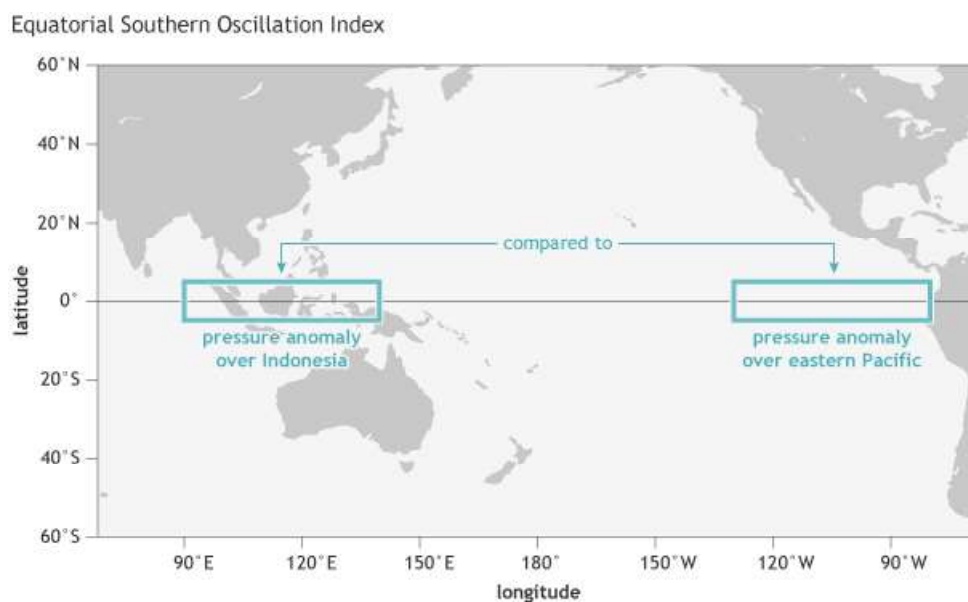


Figura 4.13: Localização do Equatorial SOI. Fonte: [107].

Para constatar a ocorrência de um fenômeno ENSO são utilizadas as séries de anomalia dos índices, isto é, a diferença de valores entre o período anterior e o período atual. A faixa de valores para indicar a ocorrência ou não dos fenômenos El Niño ou La Niña é a mesma tanto para o índice SOI quanto para o Equatorial SOI. Períodos consecutivos de índices negativos evidenciam a ocorrência do fenômeno El Niño, períodos consecutivos positivos evidenciam a presença do La Niña e por fim, valores próximos a zero indicam a situação de normalidade, ou seja, não há ocorrência de nenhum dos dois fenômenos.

Considerando que a existência do fenômeno ENSO seja confirmada após cinco valores, consecutivos, acima de 0.5° (La Niña) e cinco valores, consecutivos, abaixo de -0.5° (El Niño), veja na Figura 4.14 e 4.15 o histórico das variáveis SOI e Equatorial SOI, respectivamente. Os pontos em azul indicam a ocorrência de eventos La Niña e os pontos em vermelho de eventos El Niño, seguindo as diretrizes acima.

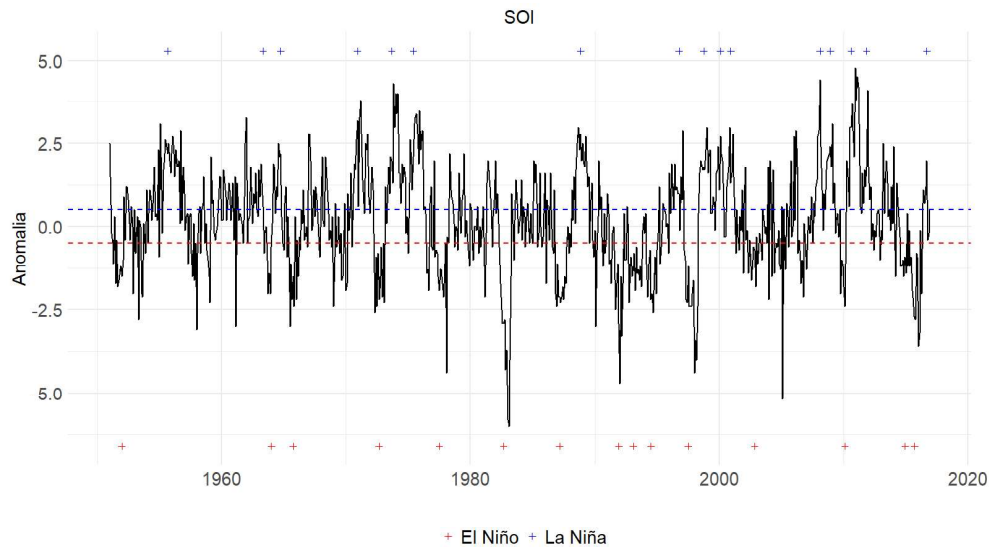


Figura 4.14: Série temporal do SOI. Fonte: [108].

El Niño: jan/52, jan/64, out/65, set/72, jul/77, set/1982, mar/87, dez/91, fev/93, jul/94, jul/97, nov/02, fev/10, dez/14, set/2015

La Niña: set/55, mai/63, out/64, dez/70, set/93, junho/75, nov/88, out/96, out/98, fev/00, dez/00, fev/08, dez/08, ago/10, nov/11, set/16

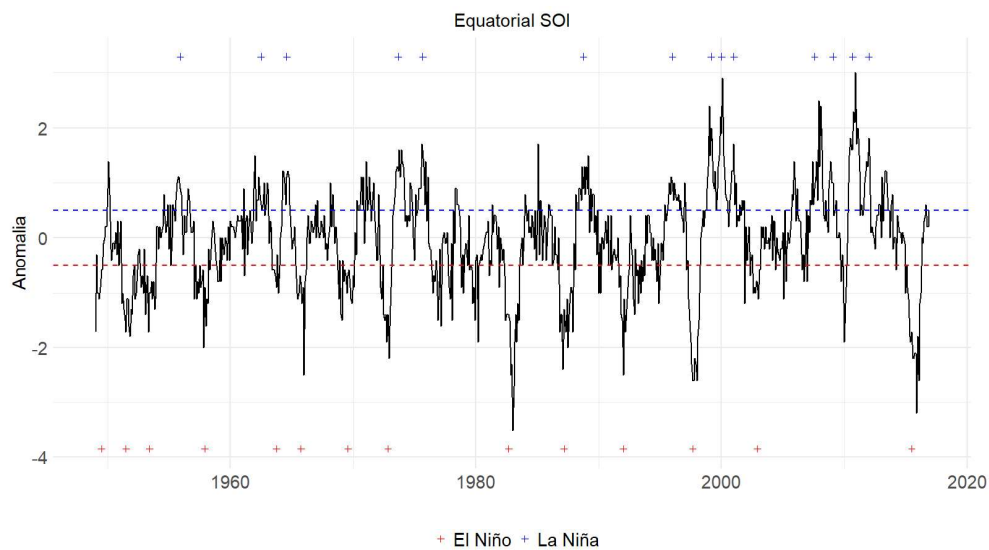


Figura 4.15: Série temporal do Equatorial SOI. Fonte: [108].

El Niño: jul/49, jul/51, jun/53, dez/57, out/63, out/65, ago/69, nov/72, set/82, mar/87, jan/92, set/97, dez/02, jul/15

La Niña: dez/55, jul/62, ago/64, set/73, set/75, out/88, jan/96, mar/99, jan/00, jan/01, ago/07, fev/09, set/10, jan/12

Como foi dito, os valores até então apresentados correspondem a anomalia das séries, porém é interessante investigar também se as séries acumuladas (originais) apresentam algum tipo de tendência, o que pode indicar um aumento ou não da temperatura do mar.

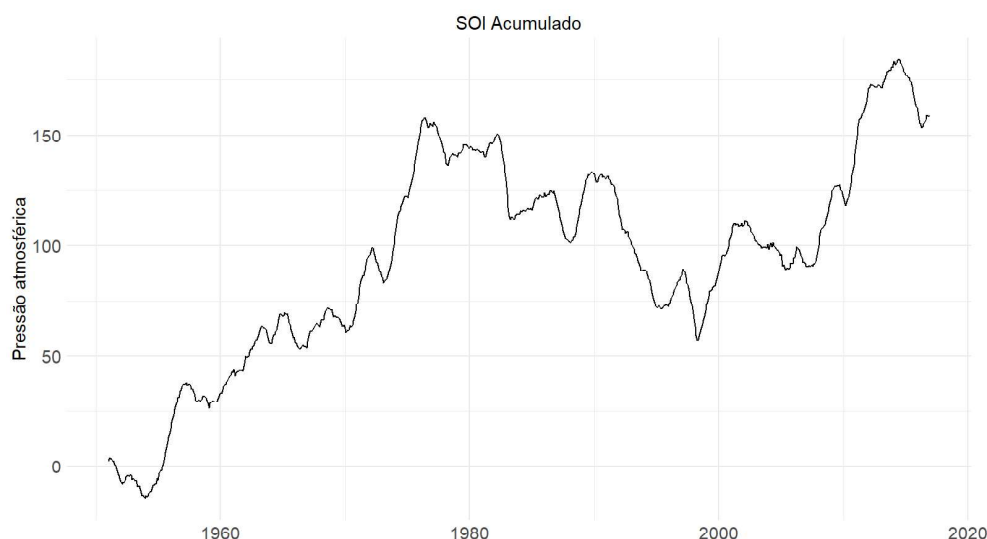


Figura 4.16: Série temporal do SOI acumulado. Fonte: [108].

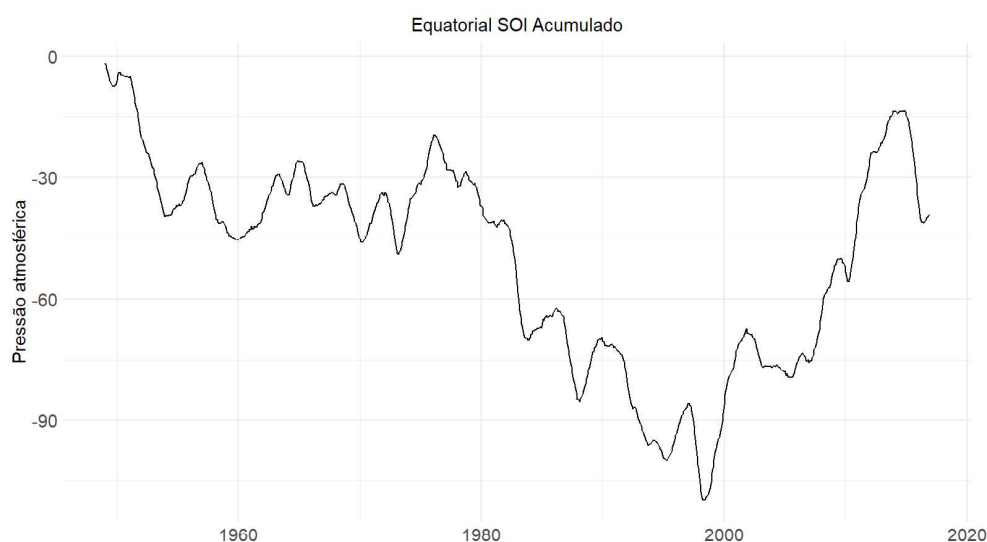


Figura 4.17: Série temporal do Equatorial SOI acumulado. Fonte: [108].

Observe na Figura 4.16 que segundo o índice SOI há uma tendência de crescimento da pressão atmosférica na superfície do mar entre as regiões do Taiti e Darwin, usando como base janeiro de 1951. Porém a Figura 4.17 mostra que segundo o índice Equatorial SOI esta tendência não existe entre a Indonésia e o Pacífico Oriental (ambas centradas na linha do Equador), também usando como referência janeiro de 1951.

4.2.2.2

Sea Surface Temperature (SST) e Oceanic Niño Index (ONI)

Em 1969, Bjerknes [104] descobriu que a temperatura da superfície do mar (SST) era peça chave para determinar a ocorrência do fenômeno ENSO. Inicialmente, certas regiões foram definidas para as medições: Niño 1, Niño 2 (combinado em Niño 1+2), Niño 3 e Niño 4 - por causa de dados consistentemente disponíveis provenientes de navios que passam por essas áreas. Mais tarde, uma área chamada Niño 3.4 foi identificada como sendo a mais representativa do ENSO [109]. Localizado entre (e sobreposto com) Niño 3 e Niño 4 (ver Figura 4.18), esta é a região cuja anomalia de temperatura é refletida pelo Oceanic Niño Index (ONI), com a diferença de que o ONI tenta remover a tendência de aquecimento existente na região.

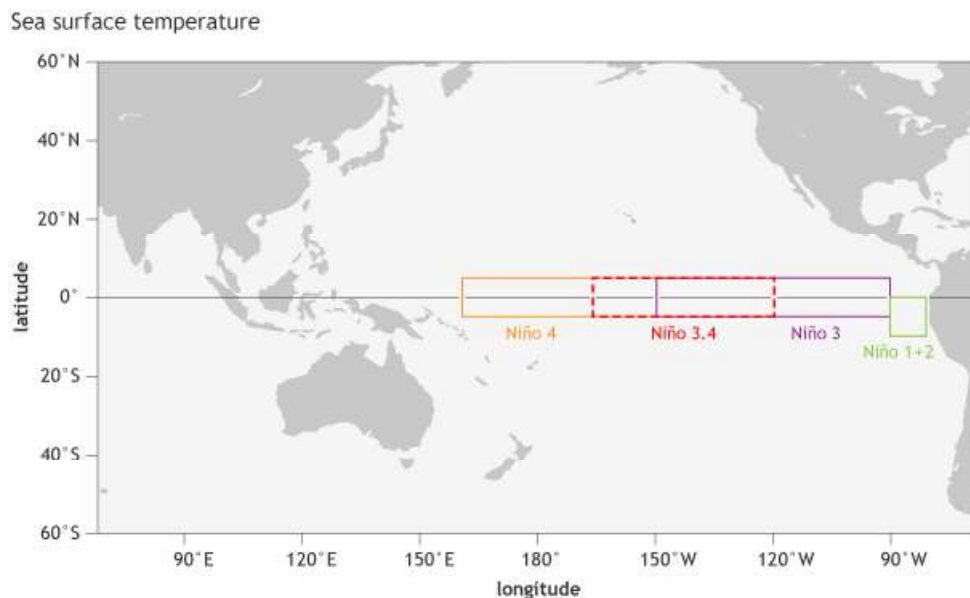


Figura 4.18: Localização do SST e ONI. Fonte: [107].

Assim como para o SOI e para o Equatorial SOI, o limiar para o estado de normalidade desses índices está entre -0.5° e $+0.5^\circ$ e é aplicado na diferença da temperatura do mar de um período para o outro. O critério comumente utilizado para classificar estados de El Niño consiste na constatação de cinco médias consecutivas de anomalias de SST acima de $+0.5^\circ$. Similarmente, para o La Niña, esse critério permanece, mas agora a anomalia deve estar abaixo de -0.5° .

Veja nas Figuras 4.19, 4.20, 4.21, 4.22 e 4.23 o histórico das variáveis Niño 1+2, Niño 3, Niño 4, Niño 3.4 e ONI, respectivamente. Assim como para os índices SOI, os pontos vermelhos representam os eventos El Niño, enquanto os pontos azuis representam os eventos La Niña, porém valores positivos marcam

o evento El Niño e negativos o La Niña, exatamente o contrário do SOI e do Equatorial SOI.

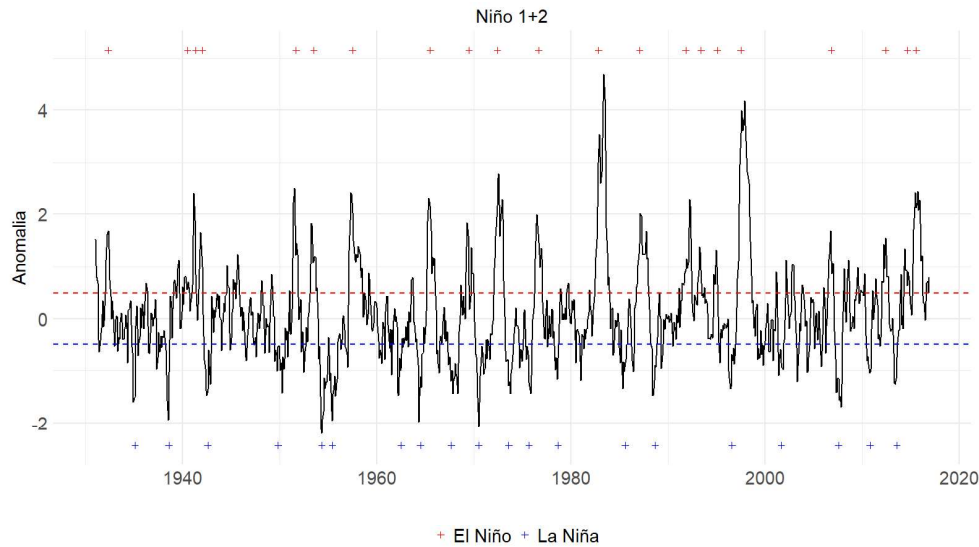


Figura 4.19: Série temporal do Niño 1+2. Fonte: [108].

El Niño: mai/32, jul/40, mai/41, jan/42, set/51, jul/53, jul/57, jul/65, jul/69, jun/72, set/76, nov/82, fev/87, nov/91, jun/93, fev/95, jul/97, nov/06, jun/12, set/14, ago/15
La Niña: fev/35, ago/38, ago/42m nov/49, mai/54, jun/55, jul/62, jul/64, set/67, jul/70, ago/73, set/75, set/78, ago/85, set/88, ago/96, set/01, ago/07, nov/10, ago/13

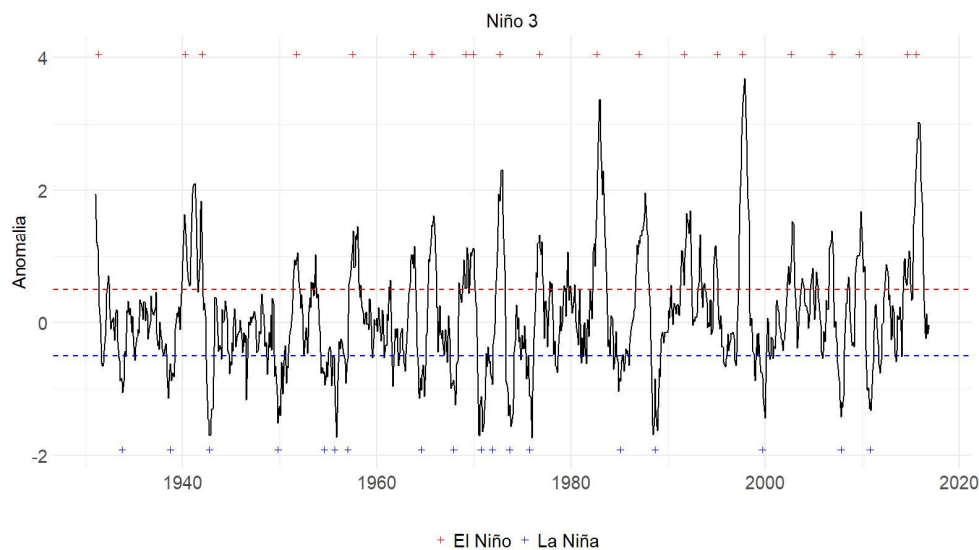


Figura 4.20: Série temporal do Niño 3. Fonte: [108].

El Niño: mai/31, abr/40, jan/42, out/51, jul/57, out/63, set/65, mar/69, dez/69, set/72, out/76, set/82, jan/87, set/91, fev/95, set/97, set/02, dez/06, set/09, set/14, ago/15
La Niña: out/33, out/38, out/42, nov/49, ago/54, set/55, jan/57, ago/64, dez/67, out/70, dez/71, set/73, out/75, fev/85, set/88, out/99, nov/07, nov/10

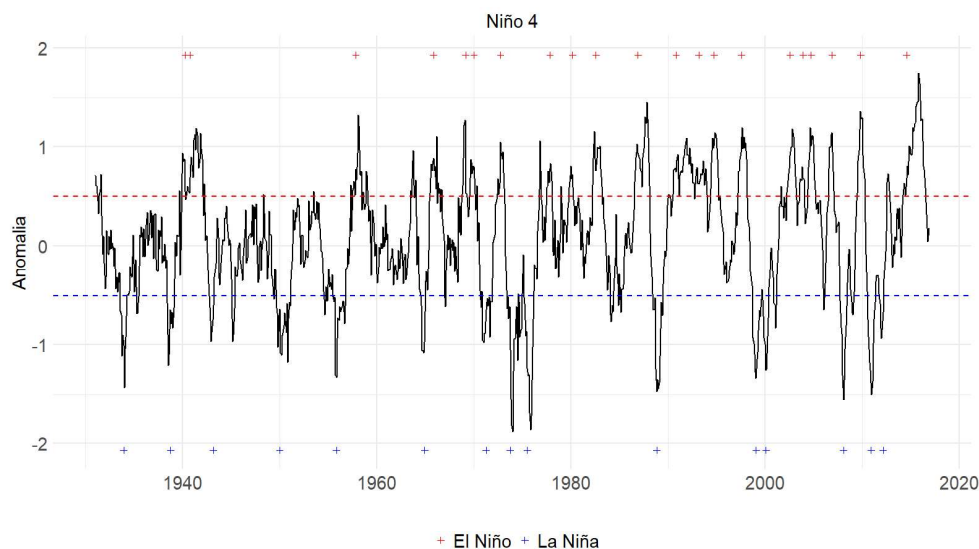


Figura 4.21: Série temporal do Niño 4. Fonte: [108].

El Niño: abr/40, out/40, nov/57, nov/65, mar/69, jan/70, out/72, nov/77, mar/80, ago/82, dez/86, nov/90, mar/93, out/94, ago/97, ago/02, dez/03, out/04, dez/06, nov/09, ago/14

La Niña: dez/33, out/38, mar/43, jan/50, nov/55, dez/64, abr/71, out/73, jul/75, nov/88, fev/99, fev/00, fev/08, dez/10, mar/12

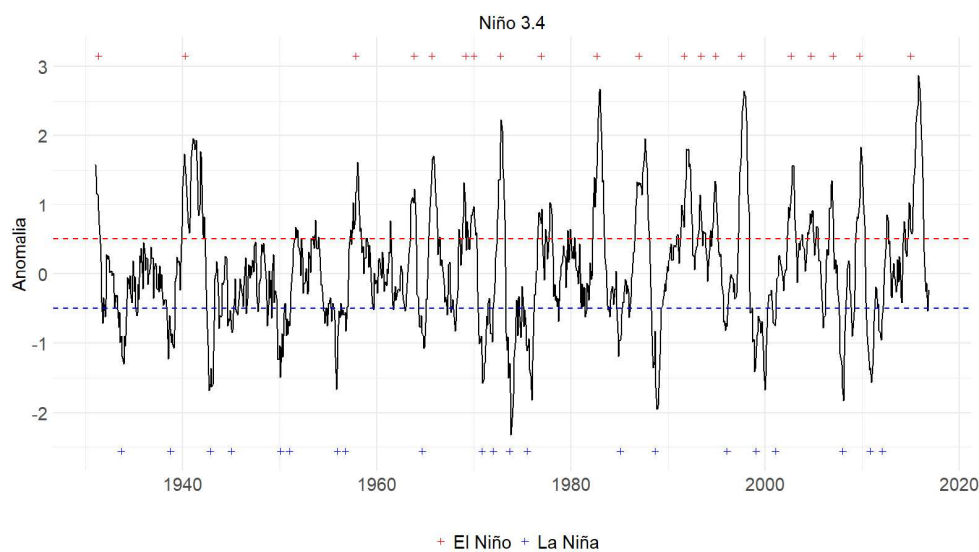


Figura 4.22: Série temporal do Niño 3.4. Fonte: [108].

El Niño: mai/31, abr/40, nov/57, nov/63, set/65, mar/69, jan/70, out/72, dez/76, set/82, jan/87, set/91, jun/93, dez/94, ago/97, set/02, out/04, jan/01, out/09, jan/15

La Niña: set/33, out/38, nov/42, jan/45, fev/50, jan/51, dez/55, out/56, set/64, nov/70, jan/72, set/73, jul/75, fev/85, set/88, fev/96, fev/99, fev/01, jan/08, nov/11, fev/12

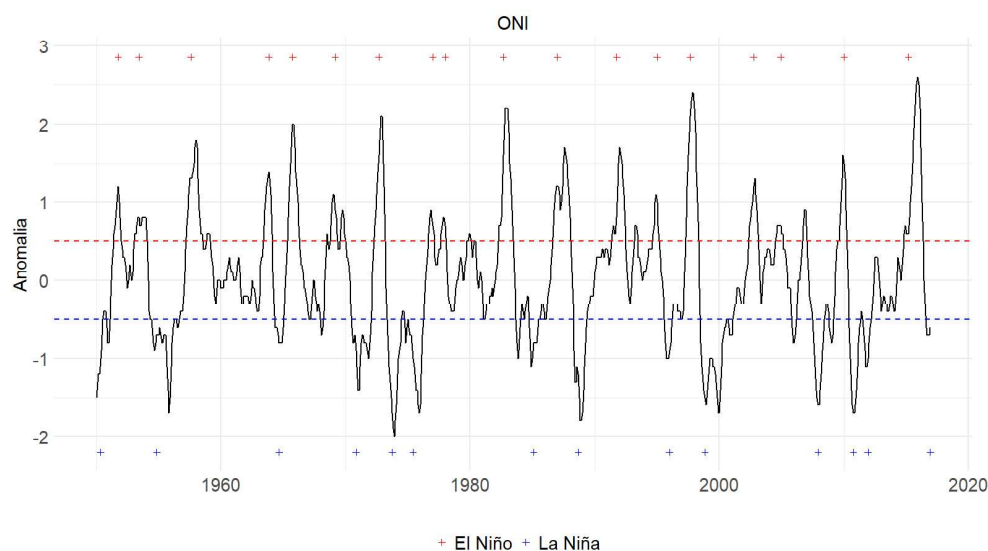


Figura 4.23: Série temporal do ONI. Fonte: [108].

El Niño: out/51, jun/53, ago/57, nov/63, out/65, mar/69, set/72, jan/77, jan/78, set/82, jan/87, out/91, jan/95, set/97, out/02, dez/04, jan/10, mar/15

La Niña: mai/50, nov/54, set/64, nov/70, out/73, jun/75, fev/85, set/88, jan/96, nov/98, dez/07, out/10, dez/11, dez/16

Assim como foi apresentado para os índices SOI e Equatorial SOI, na Figura 4.24 são apresentados os índices SST e também o ONI em sua escala original, isto é, temperatura. Note que após 1980, todos os índices, com exceção do ONI, apresentam tendência de crescimento. Ressalta-se também que o índice ONI é exatamente o Niño 3.4 com remoção de tendência, portanto não seria esperado tal índice apresentar crescimento ou decréscimo.

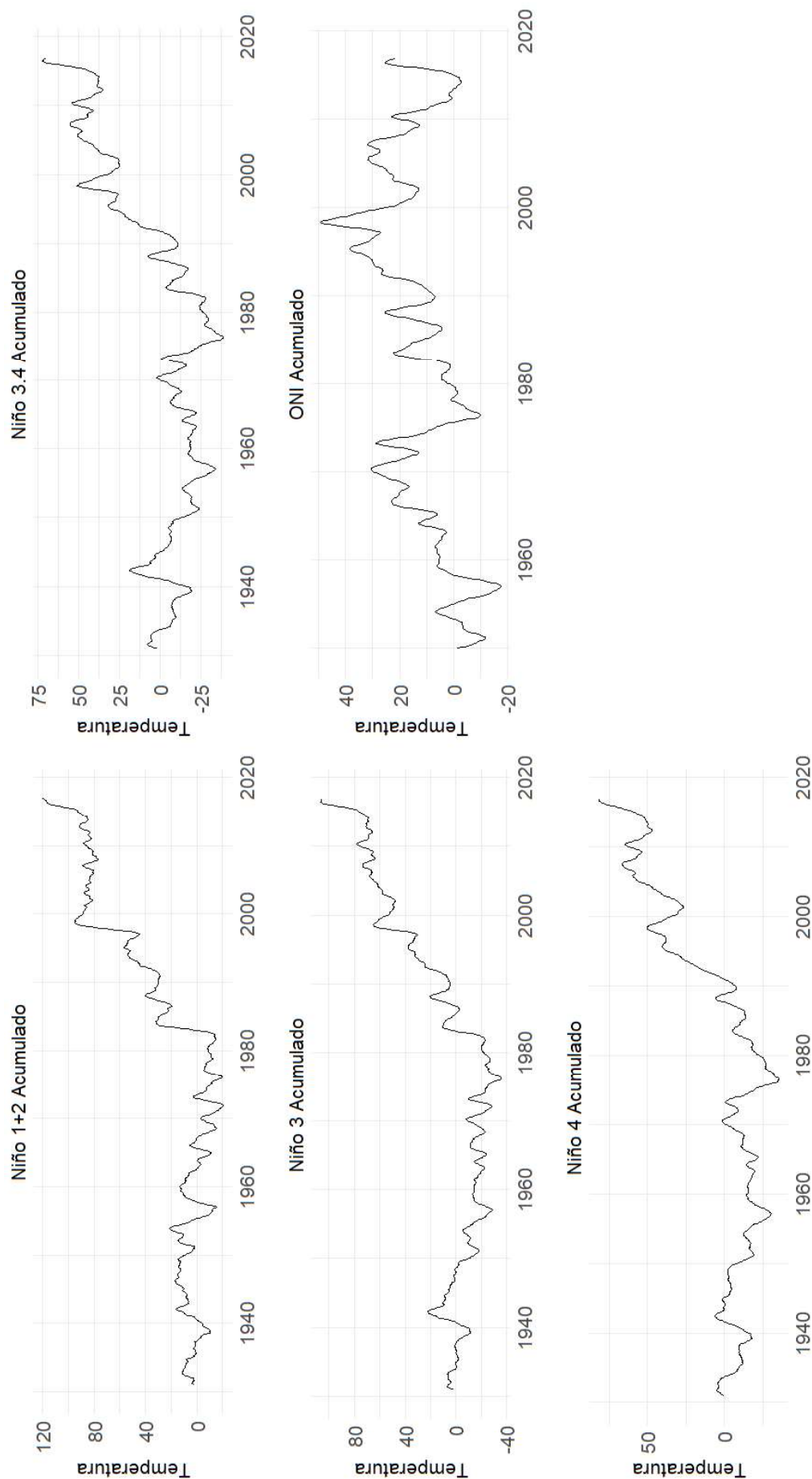


Figura 4.24: Série temporal dos índices SST e ONI acumulados. Fonte: [108].

Após apresentação de tantas medidas diferentes para o ENSO e a diferença entre elas, é possível concluir que a vantagem em testar a influência de cada uma delas nas séries de vazão se dá pelo fato do ENSO ser multifacetado, envolvendo diferentes aspectos do oceano e da atmosfera sobre o Pacífico tropical. Além disso, não é possível medir perfeitamente um evento que envolve todo o Pacífico tropical com apenas uma medida, uma imagem melhor do evento é obtida quando se usa algumas medidas relacionadas.

4.3

Relação entre as vazões e as variáveis climáticas

A primeira abordagem adotada para apresentar a relação entre as séries de vazão e as variáveis climáticas será através de dois coeficientes de correlação, de Pearson [110] e de Spearman [111]. Enquanto o coeficiente de Pearson mede a correlação linear entre duas variáveis, a correlação de Spearman avalia relações monotônicas (lineares ou não). Os valores de classificação de ambas as técnicas são iguais, variando entre -1 e 1 , onde 1 representa a correlação positiva perfeita, -1 a correlação negativa perfeita e 0 ausência de correlação.

Na Figura 4.25 é possível verificar os resultados das correlações de Pearson e de Spearman entre as vazões e as variáveis de precipitação e do fenômeno ENSO. A escala de cores empregada obedece a seguinte regra: quanto mais verde maior a correlação positiva observada e quanto mais vermelho maior a correlação negativa observada.

Pearson	Precipitação	SOI	SOI Acumulado	Equatorial SOI	Equatorial SOI Acumulado	Niño 1+2	Niño 3	Niño 3.4	Niño 4	Niño 1+2 Acumulado	Niño 3 Acumulado	Niño 3.4 Acumulado	Niño 4 Acumulado	ONI	ONI Acumulado
Paraná	0.68	-0.05	0.07	-0.08	-0.09	0.13	0.09	0.04	0.01	0.01	0.01	0.01	0.01	0.08	0.05
Grande	0.62	-0.04	0.08	-0.08	-0.06	0.16	0.10	0.05	-0.01	-0.05	-0.04	-0.05	-0.05	0.05	0.03
Tietê	0.62	-0.05	0.15	-0.09	-0.15	0.23	0.17	0.10	0.03	0.09	0.08	0.06	0.07	0.07	0.10
Paraná	0.50	-0.08	0.46	-0.10	-0.33	0.29	0.30	0.22	0.18	0.47	0.45	0.42	0.44	0.15	0.30
Paranapanema	0.45	-0.10	0.19	-0.17	-0.15	0.30	0.27	0.20	0.14	0.17	0.16	0.16	0.17	0.18	0.15
Iguaçu	0.64	-0.08	0.11	-0.13	-0.16	0.23	0.23	0.19	0.13	0.18	0.17	0.17	0.17	0.15	0.16
Paraguai	0.78	0.03	0.06	-0.02	-0.10	0.08	0.02	-0.03	-0.05	0.05	0.03	0.00	0.02	-0.02	-0.03
Uruguai	0.68	-0.13	0.08	-0.17	-0.13	0.22	0.26	0.24	0.20	0.16	0.15	0.14	0.14	0.20	0.14
Atlântico Sul	0.70	-0.13	0.07	-0.15	-0.16	0.19	0.25	0.26	0.21	0.15	0.13	0.13	0.12	0.19	0.15
Atlântico Sudeste	0.73	-0.02	-0.03	-0.07	0.01	0.06	-0.02	-0.06	-0.07	-0.11	-0.10	-0.10	-0.11	0.00	-0.02
Atlântico Leste	0.73	-0.05	-0.02	-0.08	0.04	0.00	-0.03	-0.03	0.00	-0.16	-0.16	-0.16	-0.15	0.03	-0.06
São Francisco	0.57	-0.03	-0.01	-0.10	0.03	0.08	0.02	-0.02	-0.05	-0.17	-0.16	-0.16	-0.16	0.04	-0.05
Parnaíba	0.81	0.06	0.08	0.04	0.02	-0.02	-0.04	-0.08	-0.10	-0.15	-0.15	-0.16	-0.16	-0.05	-0.08
Tocantins	0.67	0.03	0.08	-0.01	-0.02	0.04	0.02	-0.02	-0.05	-0.07	-0.06	-0.07	-0.07	0.00	-0.05
Amazonas	0.73	0.06	0.05	0.01	0.01	0.03	0.04	-0.01	-0.08	-0.01	-0.01	-0.02	-0.02	-0.04	-0.06

Spearman	Precipitação	SOI	SOI Acumulado	Equatorial SOI	Equatorial SOI Acumulado	Niño 1+2	Niño 3	Niño 3.4	Niño 4	Niño 1+2 Acumulado	Niño 3 Acumulado	Niño 3.4 Acumulado	Niño 4 Acumulado	ONI	ONI Acumulado
Paraná	0.64	-0.04	0.07	-0.10	-0.12	0.15	0.10	0.04	0.01	0.05	0.04	0.03	0.04	0.07	0.07
Grande	0.60	-0.03	0.10	-0.11	-0.09	0.17	0.11	0.05	0.01	-0.02	-0.03	-0.04	-0.04	0.05	0.04
Tietê	0.65	-0.05	0.20	-0.09	-0.19	0.21	0.14	0.08	0.06	0.13	0.10	0.07	0.09	0.05	0.11
Paraná	0.45	-0.05	0.49	-0.02	-0.34	0.19	0.24	0.18	0.18	0.40	0.38	0.36	0.38	0.07	0.29
Paranapanema	0.49	0.06	0.21	0.08	0.18	0.18	0.18	0.14	0.11	0.16	0.16	0.16	0.18	0.08	0.17
Iguaçu	0.60	-0.05	0.13	-0.07	-0.14	0.11	0.16	0.16	0.15	0.16	0.16	0.17	0.17	0.08	0.17
Paraguai	0.80	0.00	0.07	-0.08	-0.12	0.14	0.05	-0.02	-0.04	0.09	0.04	-0.02	0.00	0.00	-0.06
Uruguai	0.63	-0.10	0.11	-0.11	-0.15	0.08	0.18	0.22	0.22	0.16	0.15	0.14	0.14	0.13	0.15
Atlântico Sul	0.68	-0.11	0.10	-0.09	-0.19	0.07	0.17	0.23	0.22	0.15	0.14	0.13	0.12	0.13	0.16
Atlântico Sudeste	0.71	-0.02	-0.03	-0.13	-0.03	0.12	0.01	-0.05	-0.07	-0.08	-0.07	-0.09	-0.09	0.03	-0.02
Atlântico Leste	0.71	-0.02	-0.07	-0.11	0.02	0.02	-0.04	-0.05	-0.04	-0.21	-0.22	-0.20	-0.20	0.04	-0.08
São Francisco	0.68	0.00	-0.05	-0.12	0.00	0.10	0.02	-0.03	-0.05	-0.16	-0.15	-0.15	-0.15	0.04	-0.04
Parnaíba	0.81	0.04	0.10	-0.02	0.02	0.02	-0.01	-0.07	-0.10	-0.14	-0.13	-0.14	-0.13	-0.03	-0.07
Tocantins	0.69	0.05	0.06	-0.03	-0.03	0.06	0.06	0.00	-0.06	-0.07	-0.05	-0.06	-0.05	0.01	-0.03
Amazonas	0.72	0.05	0.05	-0.03	0.00	0.06	0.09	0.02	-0.07	0.01	0.01	-0.01	0.00	0.00	-0.06

Figura 4.25: Correlação de Pearson e Spearman. Fonte: Elaboração própria.

Note que a precipitação é a que melhor se relaciona com as vazões, de forma positiva, como esperado. Segundo o coeficiente de Pearson, o índice Niño 1+2 Acumulado foi o que apresentou maior correlação entre as variáveis ENSO, se relacionando de forma positiva com a bacia do Paraná. Já a correlação de Spearman destacou o índice SOI Acumulado, se relacionando também no mesmo sentido com a bacia do Paraná.

Vale destacar que os cálculos foram baseados no tempo corrente das variáveis, isto é, y_t e x_t , porém espera-se que a relação entre elas aconteça de forma mais forte com defasagens. Para verificar tal efeito será utilizado a Função de Correlação Cruzada, ou FCC, que calcula justamente a correlação entre uma série univariada y_t e os valores passados de uma outra série univariada x_{t+h} . Portanto para analisar a influência dos valores passados de x em y deve-se olhar os *lags* negativos.

Na Figura 4.26 estão expostos os gráficos da FCC para cada uma das bacias e a precipitação associada. Observe que as bacias do Iguaçu, Uruguai, Paraná, Atlântico Sul e Paranapanema apresentam correlação significativa com as séries de precipitação somente no tempo corrente (t) e até dois períodos anteriores ($t-1$ e $t-2$). No entanto, as demais bacias apresentam correlação significativa para todos os períodos múltiplos de 12 ($t, 2t, 3t$ etc.) e meses ao redor ($t-1, t-2, 2t-1, 2t-2, 3t-1, 3t-2$ etc.).

Com relação as variáveis climáticas, para evitar repetição, optou-se por mostrar somente as que apresentaram mais correlação significativa com as bacias. Note na Figura 4.27 que para todas as bacias foi possível encontrar pelo menos um índice ENSO que mostrasse influência nas vazões. Os índices acumulados não foram considerados nesta análise, pois para calcular a FCC é preciso aplicar o operador diferença na série acumulada, por ser não estacionária, assim tornando-a exatamente igual a série como anomalia.

Uma última análise realizada foi o cálculo da média da vazão das bacias quando é caracterizado o El Niño ou La Niña e também em períodos considerados neutros, isto é, quando não há nenhum evento. Observe na Figura 4.28 que para as bacias do Paranaíba, Grande, Tietê, Paraná, Paranapanema, Iguaçu, Uruguai, Atlântico Sul, Sudeste e Leste e São Francisco na ocorrência do fenômeno El Niño as médias das vazões são superiores quando comparadas com o período neutro, enquanto durante o La Niña influencia de forma a diminuir as médias das vazões.

Neste capítulo foram apresentadas as séries temporais de vazão, de precipitação e dos índices ENSO, assim como a relação existente entre elas. No próximo capítulo serão abordados os resultados obtidos ao aplicar os modelos propostos no capítulo Metodologia com as variáveis analisadas neste capítulo.

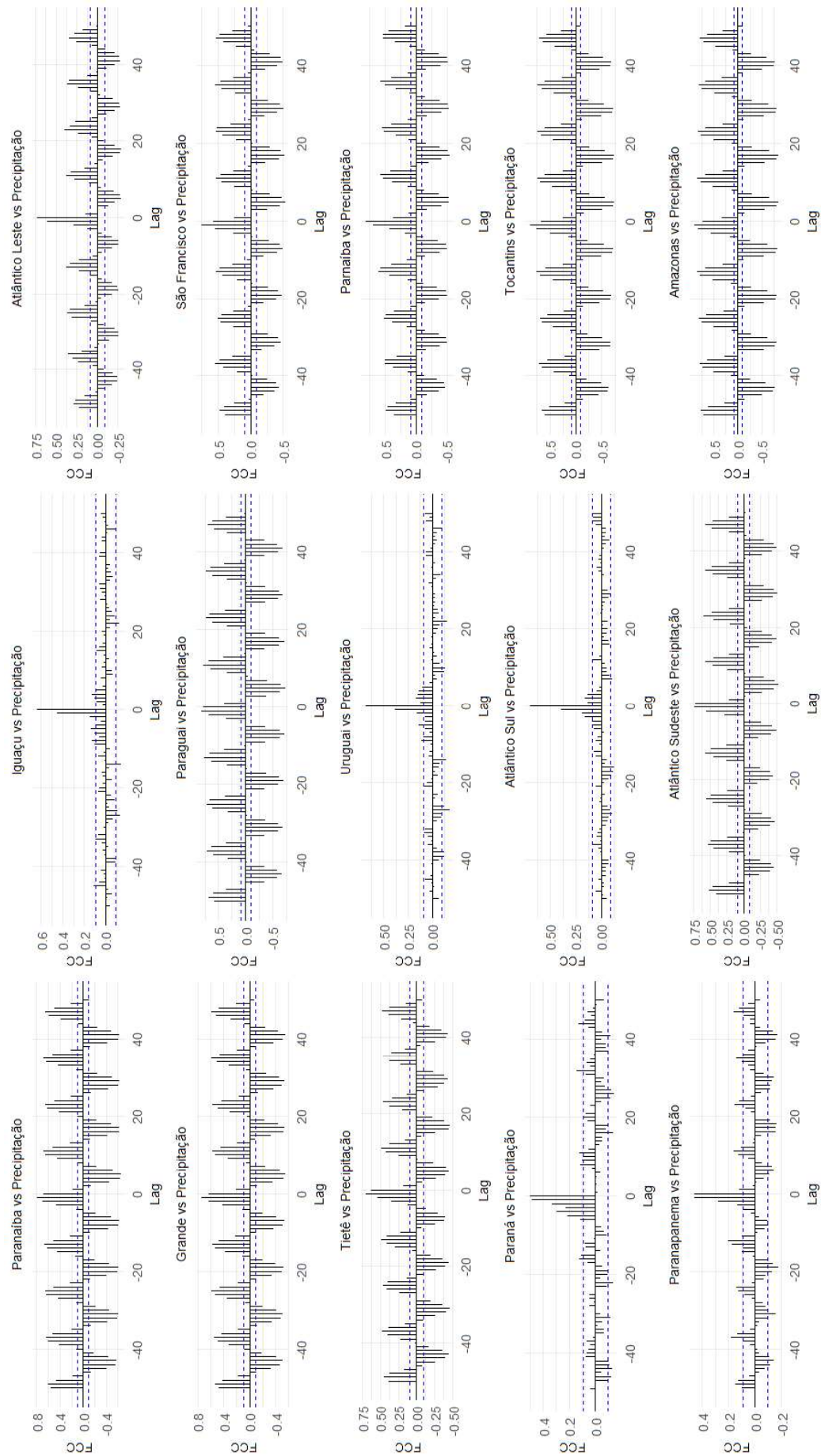


Figura 4.26: Função de correlação cruzada vazões *versus* precipitação. Fonte: Elaboração própria.

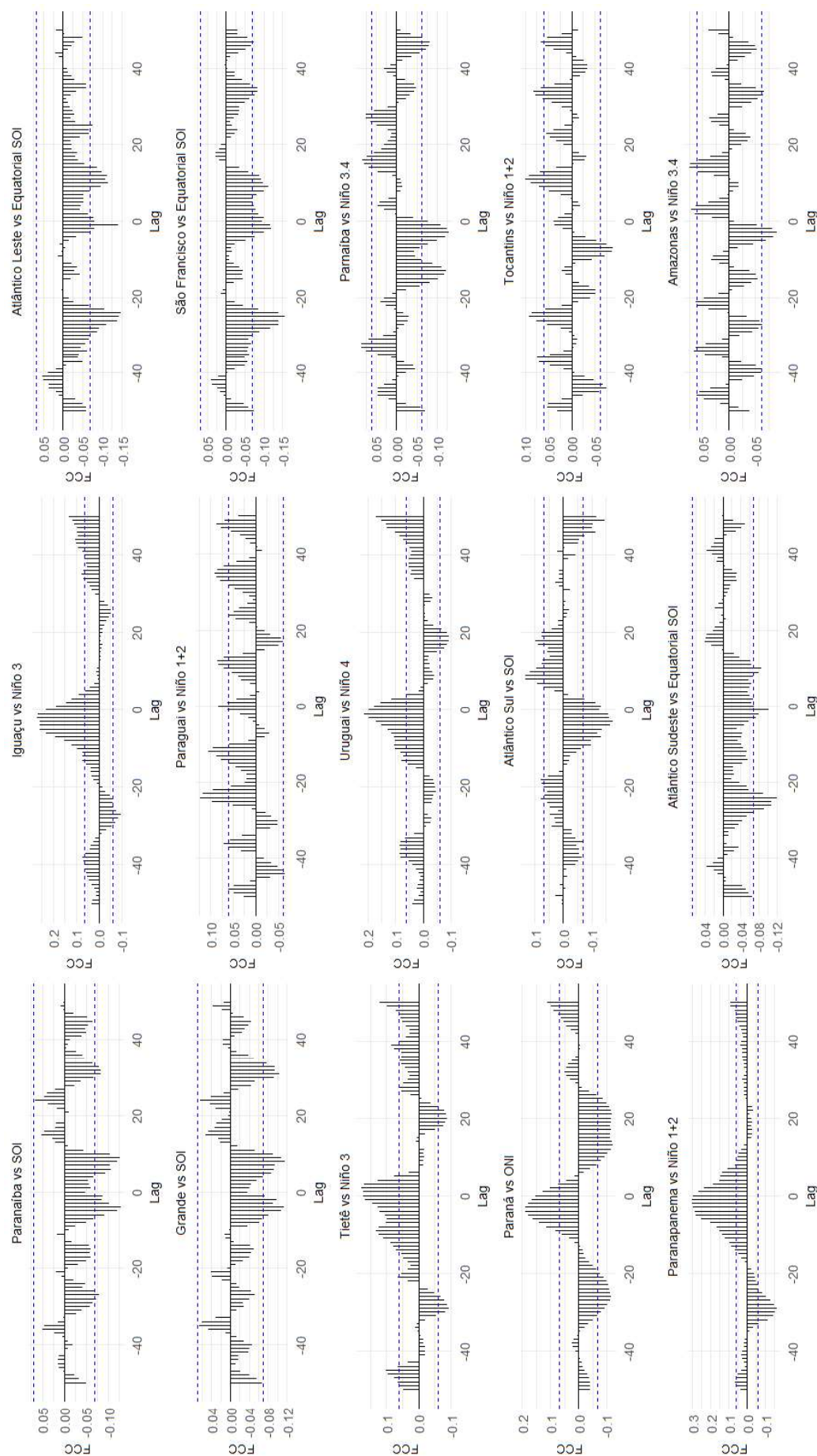


Figura 4.27: Função de correlação cruzada vazões *versus* ENSO. Fonte: Elaboração própria.

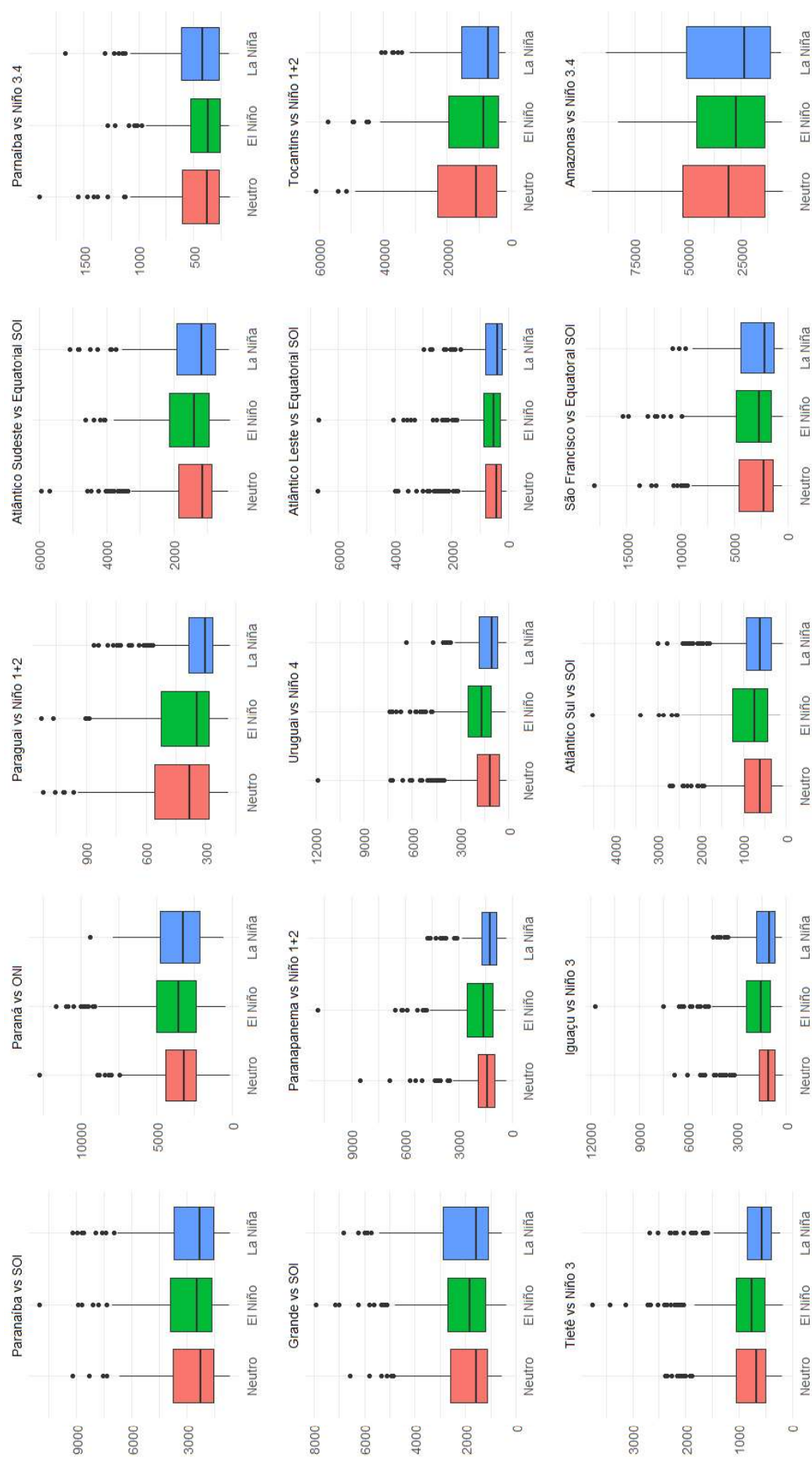


Figura 4.28: Média de vazões por evento ENSO. Fonte: Elaboração própria.

5

Resultados alcançados

No capítulo Metodologia foi explicado que, primeiramente os modelos propostos serão testados a partir do seu poder de previsão e os cenários de vazão serão gerados apenas com o modelo selecionado.

Para verificar a performance de cada modelo, a base de dados apresentada no capítulo Análise descritiva dos dados será separada em seis janelas de ajuste e previsão, descritas na Tabela 5.1.

Tabela 5.1: Janelas de ajuste e previsão.

Janela	<i>In-sample</i>	<i>Out-of-sample</i>
1	jan/1931-dez/2006	jan/2007-dez/2011
2	jan/1931-dez/2007	jan/2008-dez/2012
3	jan/1931-dez/2008	jan/2009-dez/2013
4	jan/1931-dez/2009	jan/2010-dez/2014
5	jan/1931-dez/2010	jan/2011-dez/2015
6	jan/1931-dez/2011	jan/2012-dez/2016

A janela 1 será utilizada para definir a escolha dos parâmetros L e m das técnicas SSA e MSSA e também a dupla $p(v)$ e $m(v)$ das metodologias PARX e PGAM. Isto é, para cada bacia e para cada combinação dos parâmetros deve-se realizar o passo-a-passo a seguir, considerando somente a primeira janela.

1. Ajustar a série de vazão para o período *in-sample*;
2. Simular o período *out-of-sample*;
3. Comparar os valores previstos (média dos cenários) com o observado;
4. Guardar os erros obtidos.

O parâmetro escolhido para cada bacia e modelo será aquele que obtiver menor erro. Esses parâmetros serão utilizados nas janelas 2 a 6 para o cálculo das métricas de erro repetindo os passos da janela 1, porém somente para a melhor combinação de parâmetros na primeira janela.

Os valores de erro apresentados nesta seção são obtidos ao realizar a média dos erros em cada uma das cinco janelas (2 a 6) com o objetivo de verificar de forma mais robusta o poder de previsão dos modelos. Serão

utilizadas três métricas de avaliação, a primeira é a *Mean Square Relative Error* (MSRE), dada por

$$\text{MSRE} = \frac{1}{T} \sum_{t=1}^T \left(\frac{f_t - y_t}{y_t} \right)^2 \quad (5-1)$$

a segunda métrica é chamada de *Mean Absolute Relative Error* (MARE), dada por

$$\text{MARE} = \frac{1}{T} \sum_{t=1}^T \left| \frac{f_t - y_t}{y_t} \right| \quad (5-2)$$

e a terceira será o Nash-Sutcliffe *Criterion* (NSC), com a seguinte formulação

$$\text{NSC} = 1 - \frac{\sum_{t=1}^T (f_t - y_t)^2}{\sum_{t=1}^T (y_t - \bar{y})^2} \quad (5-3)$$

onde y_t é a vazão observada no momento t , f_t é a vazão prevista pelo modelo no tempo t , \bar{y} é a média observada e T é o tamanho do horizonte de previsão.

O terceiro critério foi desenvolvido por Nash e Sutcliffe [112] e desde então tem sido amplamente utilizado para estimar o erro de modelos hidrológicos. Esse critério basicamente verifica quão bom é o modelo desenvolvido ao compará-lo com um modelo que usaria como previsão a média observada. O NSC varia de $-\infty$ a 1, sendo 1 o modelo perfeito. Se $\text{NSC} \geq 0$, então o modelo proposto é melhor do que a média observada, já se $\text{NSC} < 0$ então o modelo proposto não é funcional.

Para facilitar o entendimento do leitor será explicado a seguir quais os modelos que efetivamente serão aplicados:

- **Modelo 1) PAR:** Pode ser considerado o *benchmark*, pois é o modelo PAR aplicado diretamente às séries de vazão. É o modelo a ser melhorado com as técnicas aqui propostas;
- **Modelo 2) SSA + PAR:** Esta metodologia aplica o SSA nas séries de vazão a fim de remover possíveis *outliers* antes de utilizar o PAR para simulação;
- **Modelo 3) MSSA + PAR:**
 - **3a:** O MSSA é utilizado para remover possíveis *outliers* utilizando a informação de todas as vazões conjuntamente para posterior aplicação do PAR em cada bacia;
 - **3b:** Utiliza-se uma informação climática (Precipitação, SOI, Equatorial SOI etc.) por vez para filtrar as séries de vazão e em seguida aplicar o PAR em cada bacia.

- **Modelo 4) PARX:** Aplica o PARX a cada série de vazão utilizando as séries de precipitação e do fenômeno ENSO como explicativas, uma de cada vez.
- **Modelo 5) PGAM:** Similar ao Modelo 4, porém, utiliza o PGAM para realizar a modelagem e simulação futura.

Os modelos PARX e PGAM necessitam da variável climática especificada para o período *out-of-sample*, porém, por conta da complexidade de previsão das séries de precipitação e dos índices ENSO, foram utilizados os próprios valores realizados no período *out-of-sample* para essas variáveis. Este ponto é uma limitação da Tese, porém o foco não é a previsão de variáveis climáticas e sim a inserção de tais variáveis nos modelos de previsão e/ou simulação, para tanto, além da utilização dos próprios valores observados sugere-se a utilização da previsão de órgãos oficiais como o próprio NOAA.

Uma tabela de abreviatura dos modelos será provida a fim de facilitar a interpretação das siglas a serem utilizadas nas próximas seções.

Tabela 5.2: Resumo das variáveis a serem utilizadas.

Abreviação	Variável
PRECI	Precipitação
ANOM	Anomalia
ACUM	Acumulado
SOI	Southern Oscillation Index
EQ SOI	Equatorial SOI
NINO1+2	Sea Surface Temperature - Niño 1+2
NINO3	Sea Surface Temperature - Niño 3
NINO3.4	Sea Surface Temperature - Niño 3.4
NINO4	Sea Surface Temperature - Niño 4
ONI	Oceanic Niño Index

5.1

Melhores modelos

Na Figura 5.1 é apresentada a performance, segundo o critério MSRE, de todos os 48 modelos testados para cada bacia (em cinza) assim como a métrica para o modelo atualmente utilizado pelo SEB (em vermelho), o modelo que obteve a menor métrica utilizando a variável precipitação (em azul) e utilizando o fenômeno ENSO (em verde). A bacia do Atlântico Leste foi a que apresentou maiores erros e por isso foi necessário adicionar um eixo secundário (a direita) para representa-la, de forma que o gráfico ficasse legível. Segundo a Figura 5.1

é possível concluir que para todas as bacias foi possível encontrar um modelo que apresente erro menor que a metodologia atualmente utilizada pelo SEB ao utilizar um índice que represente o El Niño e o La Niña, e ao utilizar a variável precipitação com exceção da bacia do Paranaíba e do Paraguai.

A Tabela 5.3 apresenta o modelo que obteve menor erro segundo a métrica MSRE e também a melhoria obtida em relação ao modelo PAR. Note que a bacia do Atlântico Leste foi a que alcançou maior melhoria, porém continua sendo a que possui maior MSRE, mostrando que ainda existe espaço para melhoria em sua modelagem e previsão e/ou simulação. A bacia do Paraguai foi a que aferiu menor melhoria, ainda assim maior do que 14% de ganho.

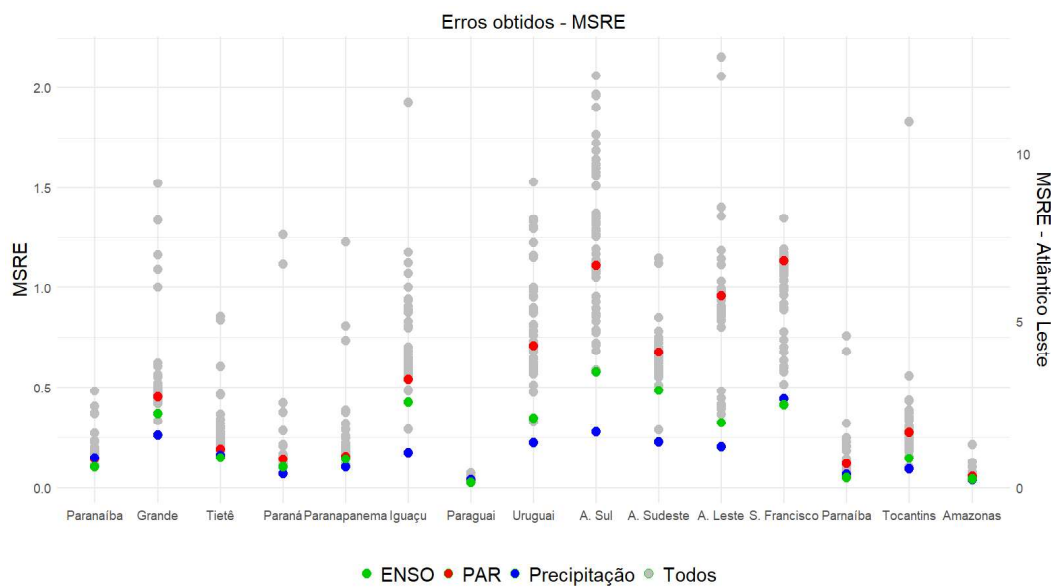


Figura 5.1: MSRE obtido para todos os 48 modelos (cinza), com o melhor modelo que usa precipitação (azul), melhor modelo que usa um índice ENSO (verde) e o PAR (vermelho) para cada bacia.

Tabela 5.3: Melhores modelos segundo a métrica MSRE.

Bacia	MSRE (PAR)	Melhor modelo	MSRE (Melhor)	Melhoria (%)
Paranaíba	0,14	PARX+EQ SOI ACUM	0,10	-29,14
Grande	0,45	PARX+PRECI	0,26	-42,39
Tietê	0,19	PGAM+ONI ACUM	0,15	-22,20
Paraná	0,14	PGAM+PRECI	0,07	-50,64
Paranapanema	0,16	PARX+PRECI	0,10	-34,06
Iguaçu	0,54	PGAM+PRECI	0,17	-67,75
Paraguai	0,03	PGAM+EQ SOI ACUM	0,03	-14,19
Uruguai	0,71	PGAM+PRECI	0,22	-68,71
A. Sul	1,11	PARX+PRECI	0,28	-75,14
A. Sudeste	0,67	PARX+PRECI	0,23	-66,56
A. Leste	5,75	PGAM+PRECI	1,23	-78,62
S. Francisco	1,13	PGAM+NINO1+2 ACUM	0,41	-63,53
Parnaíba	0,12	PGAM+NINO3.4 ACUM	0,05	-58,92
Tocantins	0,27	PARX+PRECI	0,09	-65,88
Amazonas	0,06	PGAM+PRECI	0,04	-31,45

Na Figura 5.2 assim como foi feito para o MSRE, os erros obtidos com a métrica MARE são apresentados para cada um dos 48 modelos testados (em cinza), o menor erro obtido utilizando a precipitação (em azul) e um índice ENSO (em verde), utilizando um eixo secundário para o Atlântico Leste. Também similarmente ao MSRE, foi possível encontrar erros menores do que o atualmente empregado para todas as bacias com o índice ENSO, porém para as bacias do Paranaíba e do Paraguai a precipitação não foi bem sucedida quando comparada com o PAR. A partir da Tabela 5.4 percebe-se novamente que a maior melhoria foi conquistada pela bacia do Atlântico Leste, enquanto a menor foi para a do Paraguai.

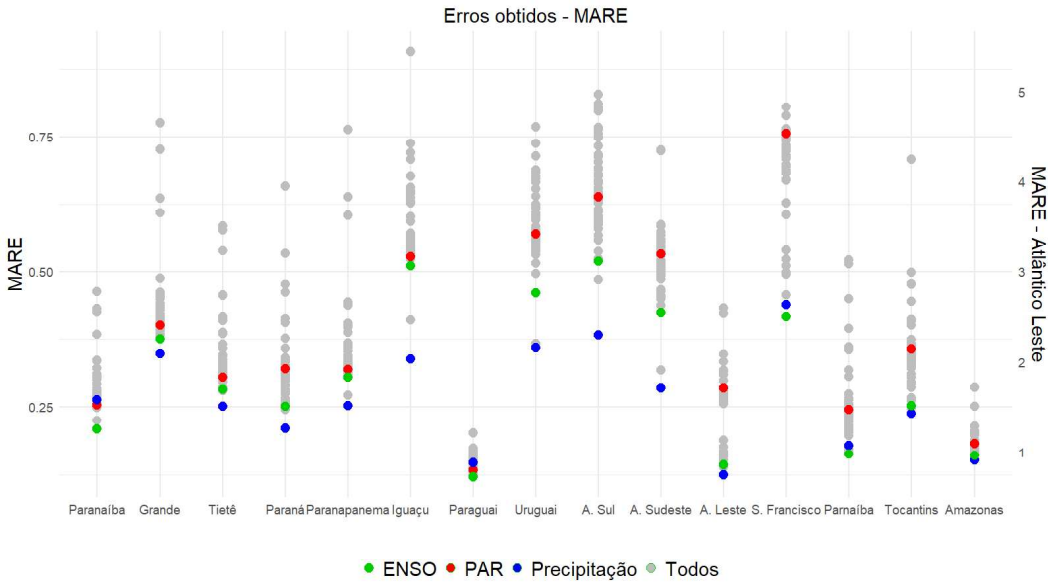


Figura 5.2: MARE obtido para todos os 48 modelos (cinza), com o melhor modelo que usa precipitação (azul), melhor modelo que usa um índice ENSO (verde) e o PAR (vermelho) para cada bacia.

Tabela 5.4: Melhores modelos segundo a métrica MARE.

Bacia	MARE (PAR)	Melhor modelo	MARE (Melhor)	Melhoria (%)
Paranaíba	0,25	PARX+EQ SOI ACUM	0,21	-17,52
Grande	0,40	PARX+PRECI	0,35	-13,05
Tietê	0,30	PGAM+PRECI	0,25	-17,60
Paraná	0,32	PGAM+PRECI	0,21	-34,79
Paranapanema	0,32	PARX+PRECI	0,25	-21,09
Iguaçu	0,53	PGAM+PRECI	0,34	-35,67
Paraguai	0,13	PGAM+EQ SOI ACUM	0,12	-8,87
Uruguai	0,57	PGAM+PRECI	0,36	-36,89
A. Sul	0,64	PARX+PRECI	0,38	-40,19
A. Sudeste	0,53	PARX+PRECI	0,29	-46,34
A. Leste	1,71	PGAM+PRECI	0,74	-56,52
S. Francisco	0,76	PGAM+NINO1+2 ACUM	0,42	-44,77
Parnaíba	0,24	PARX+NINO4 ACUM	0,16	-33,96
Tocantins	0,36	PARX+PRECI	0,24	-33,31
Amazonas	0,18	PGAM+PRECI	0,15	-16,68

Recordando que a métrica NSC compara a acurácia do modelo testado contra a acurácia de um modelo que utiliza somente a média histórica como previsão, observe na Figura 5.3 que todos os melhores modelos que utilizaram a precipitação conquistaram NSC maior que 0, significando que são melhores

do que a média histórica, e com exceção da bacia do Iguaçu, todos os que utilizaram um índice ENSO, também auferiram valores maior que 0. Até mesmo para os casos da bacia do Paraná, Paranapanema, Iguaçu, Atlântico Leste e São Francisco, que apresentam valores de NSC menores do que 0 quando o PAR é utilizado, foi possível encontrar modelos que são melhores do que a utilização da média histórica.

Diferente das métricas estudadas anteriormente, para o NSC, a última coluna da Tabela 5.5 apresenta somente o ganho absoluto obtido ao utilizar o melhor modelo encontrado em relação ao PAR, isto porque calcular ganho percentual de números negativos para números positivos pode ocasionar distorções. O maior ganho verificado foi para a bacia do Paraná e o menor para a do Paraguai, seguida de perto pela do Amazonas, Grande, Paranaíba e Tocantins, sendo essas as que possuem maior espaço para melhorias futuras.

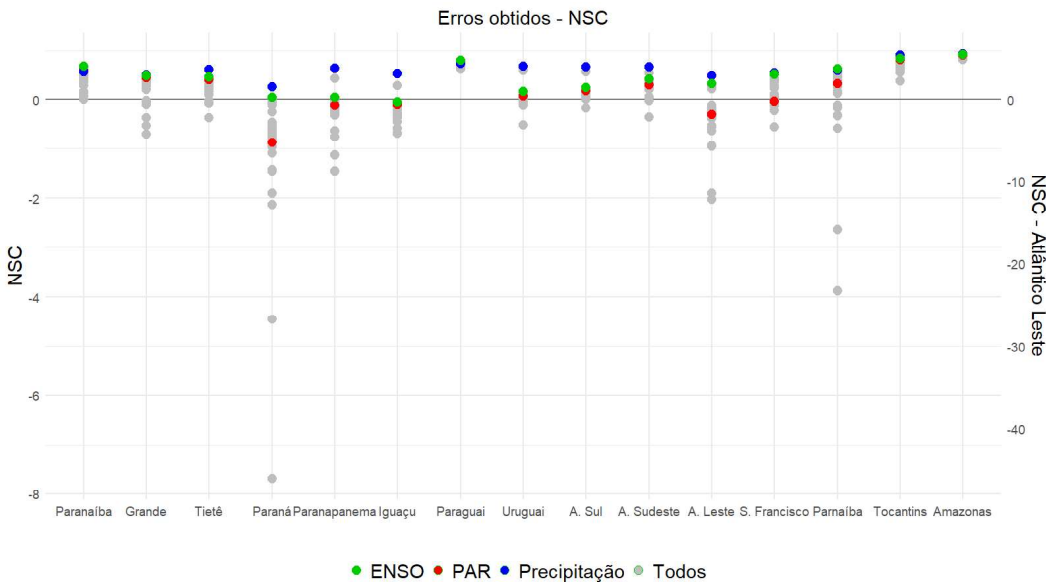


Figura 5.3: NSC obtido para todos os 48 modelos (cinza), com o melhor modelo que usa precipitação (azul), melhor modelo que usa um índice ENSO (verde) e o PAR (vermelho) para cada bacia.

Tabela 5.5: Melhores modelos segundo a métrica NSC.

Bacia	NSC (PAR)	Melhor modelo	NSC (Melhor)	Melhoria
Paranaíba	0,59	PGAM+NINO1+2 ANOM	0,68	0,09
Grande	0,45	PARX+PRECI	0,50	0,05
Tietê	0,40	PARX+PRECI	0,61	0,21
Paraná	-0,87	PARX+PRECI	0,26	1,13
Paranapanema	-0,11	PARX+PRECI	0,63	0,74
Iguaçu	-0,10	PGAM+SOI ANOM	0,53	0,64
Paraguai	0,77	PARX+SOI ACUM	0,80	0,02
Uruguai	0,07	PARX+PRECI	0,67	0,60
A. Sul	0,18	PARX+PRECI	0,66	0,49
A. Sudeste	0,30	PARX+PRECI	0,66	0,36
A. Leste	-0,30	PGAM+SOI ANOM	0,48	0,78
S. Francisco	-0,03	PGAM+SOI ANOM	0,54	0,57
Parnaíba	0,33	PARX+ONI ANOM	0,62	0,29
Tocantins	0,81	PGAM+SOI ANOM	0,90	0,09
Amazonas	0,90	PARX+PRECI	0,93	0,03

A Tabela 5.6 apresenta um resumo dos modelos que foram escolhidos como melhores de acordo com os três critérios de erro. Note que o “PARX+PRECI” e o “PGAM+PRECI” figuram nos dois primeiros lugares, mostrando que existe benefício em adicionar a precipitação como covariável. No que diz respeito as metodologias, dos modelos selecionados, 23 aplicam o PARX e 22 o PGAM.

Tabela 5.6: Resumo dos modelos mais escolhidos como melhores.

Modelo	Frequência (de 45)
PARX+PRECI	18
PGAM+PRECI	11
PGAM+SOI ANOM	4
PARX+EQ SOI ACUM	2
PGAM+EQ SOI ACUM	2
PGAM+NINO1+2 ACUM	2
PGAM+NINO1+2 ANOM	1
PARX+SOI ACUM	1
PARX+ONI ANOM	1
PARX+NINO4 ACUM	1
PGAM+ONI ACUM	1
PGAM+NINO3.4 ACUM	1

Já a Tabela 5.7 mostra a frequência com que as variáveis climáticas foram selecionadas. Como esperado, a precipitação foi a mais utilizada seguida pelo SOI (Anomalia) e Equatorial SOI (Acumulado).

Tabela 5.7: Resumo das variáveis climáticas mais escolhidas como melhores.

Variável	Frequência (de 45)
PRECI	29
SOI ANOM	4
EQ SOI ACUM	4
NINO1+2 ACUM	2
NINO1+2 ANOM	1
SOI ACUM	1
SOI ANOM	1
ONI ANOM	1
NINO4 ACUM	1
NINO3.4 ACUM	1

Na Figura 5.4 são apresentadas as vazões observadas durante o período de validação da janela 6 (jan/2012 a dez/2016) em preto, a previsão obtida ao utilizar o PAR em vermelho e com o melhor modelo proposto pelo MARE em azul quando a variável exógena utilizada é a precipitação e em verde quando é um índice ENSO.

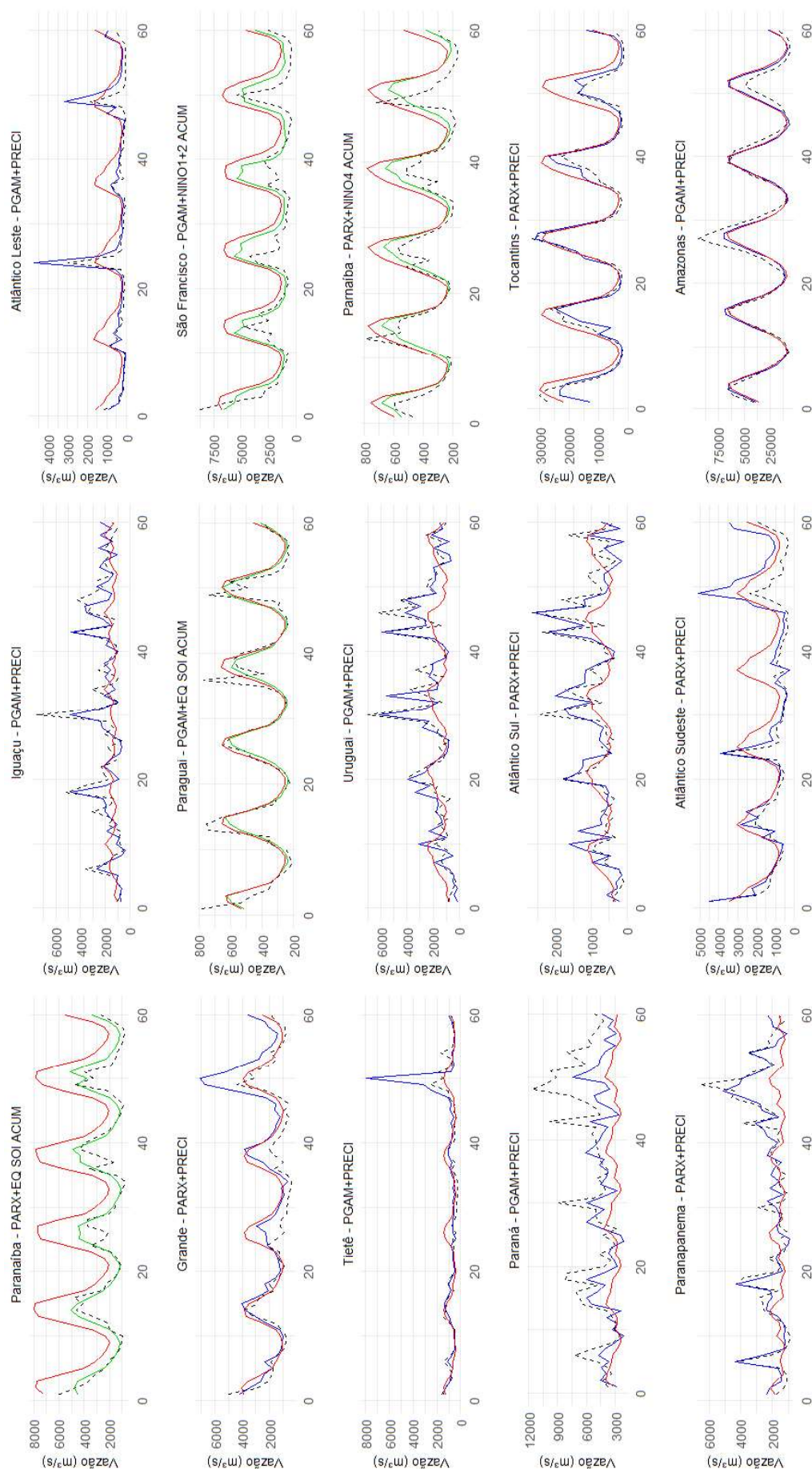


Figura 5.4: Vazão observada (preto), previsão obtida via modelo PAR (vermelho), modelo com precipitação (azul) ou com índice ENSO (verde).

5.2

Resultados comparados ao artigo Lima et al.

O tema desta Tese surgiu a partir do trabalho de Lima et al. [14] que aplicou um Modelo Dinâmico Linear às séries de vazão das principais bacias brasileiras utilizando a precipitação e o índice Niño 3.4 como variáveis exógenas. Portanto é natural que os resultados aqui obtidos sejam comparados com tal trabalho. Porém, a base de dados utilizada em Lima et al. [14] foi referente ao *deck* de dez/2010, o que impossibilita a comparação direta com os resultados apresentados na seção anterior. Assim, os métodos sugeridos nesta Tese foram aplicados na mesma base de dados do trabalho de referência para ser possível a comparação. Isto é, a janela *in-sample* data de jan/1931 até dez/2000 e a janela *out-of-sample* vai de jan/2001 até dez/2005, além disso as bacias do Paraná e do Iguaçu são tratadas como uma só. Os valores das variáveis exógenas para o período *out-of-sample* serão os observados para o período, assim como feito em Lima et al. [14].

Em um primeiro momento serão analisados os resultados utilizados como *benchmark* que referem-se ao modelo PAR. Nas Tabelas 5.8, 5.9 e 5.10 é feita a comparação entre tais resultados e os erros obtidos com os melhores modelos encontrados para cada bacia segundo as métricas MSRE, MARE e NSC, respectivamente. Observe que em todos os casos foi possível encontrar pelo menos um modelo que obtivesse resultados melhores do que o *benchmark*, chegando a 91% de melhoria para a bacia do Uruguai segundo o MARE.

Tabela 5.8: Comparação entre a métrica MSRE do PAR (*benchmark*) e do modelo proposto.

Bacia	PAR	Proposta	Melhoria (%)	Modelo
Paranaíba	0.196	0.045	-77.04	MSSA+PAR
Grande	0.330	0.132	-60.00	PGAM+PRECI
Tietê	0.119	0.020	-83.19	PARX+PRECI
Paraná/Iguaçu	0.707	0.111	-84.30	PARX+PRECI
Paranapanema	0.339	0.059	-82.60	MSSA+NINO3 ANOM+PAR
Paraguai	0.075	0.027	-64.00	PARX+PRECI
Uruguai	2.292	0.192	-91.62	PGAM+PRECI
Atlântico Sul	1.334	0.697	-47.75	PGAM+NINO3 ACUM
Atlântico Sudeste	0.384	0.044	-88.54	PARX+PRECI
Atlântico Leste	2.511	0.564	-77.54	PARX+NINO1+2 ACUM
São Francisco	1.029	0.191	-81.44	PGAM+NINO1+2 ACUM
Parnaíba	0.226	0.062	-72.57	PGAM+NINO1+2 ACUM
Tocantins	0.277	0.069	-75.09	PGAM+ONI ACUM
Amazonas	0.163	0.030	-81.60	PGAM+NINO4 ACUM

Tabela 5.9: Comparação entre a métrica MARE do PAR (*benchmark*) e do modelo proposto.

Bacia	PAR	Proposta	Melhoria (%)	Modelo
Paranaíba	0.330	0.153	-53.64	MSSA+PAR
Grande	0.422	0.268	-36.49	SSA+PAR
Tietê	0.275	0.111	-59.64	PARX+PRECI
Paraná/Iguaçu	0.611	0.260	-57.45	PARX+PRECI
Paranapanema	0.447	0.193	-56.82	MSSA+NINO3 ANOM+PAR
Paraguai	0.203	0.120	-40.89	PGAM+SOI ANOM
Uruguai	0.995	0.327	-67.14	PGAM+PRECI
Atlântico Sul	0.780	0.536	-31.28	PGAM+PRECI
Atlântico Sudeste	0.462	0.168	-63.64	PGAM+PRECI
Atlântico Leste	1.047	0.494	-52.82	PGAM+PRECI
São Francisco	0.731	0.347	-52.53	PGAM+NINO1+2 ACUM
Parnaíba	0.353	0.165	-53.26	PGAM+NINO1+2 ACUM
Tocantins	0.403	0.204	-49.38	PGAM+ONI ACUM
Amazonas	0.290	0.137	-52.76	PGAM+NINO4 ACUM

Tabela 5.10: Comparação entre a métrica NSC do PAR (*benchmark*) e do modelo proposto.

Bacia	PAR	Proposta	Melhoria	Modelo
Paranaíba	0.332	0.779	0.45	MSSA+PAR
Grande	0.261	0.777	0.52	SSA+PAR
Tietê	0.142	0.808	0.67	PARX+PRECI
Paraná/Iguaçu	-0.806	0.597	1.40	PGAM+PRECI
Paranapanema	-1.914	0.404	2.32	SSA+PAR
Paraguai	0.452	0.816	0.36	PARX+PRECI
Uruguai	-0.770	0.710	1.48	PGAM+PRECI
Atlântico Sul	-0.658	0.266	0.92	PGAM+PRECI
Atlântico Sudeste	0.259	0.859	0.60	PARX+PRECI
Atlântico Leste	-0.656	0.454	1.11	PGAM+NINO4 ANOM
São Francisco	0.023	0.635	0.61	MSSA+PRECI+PAR
Parnaíba	0.214	0.553	0.34	PGAM+EQ SOI ANOM
Tocantins	0.630	0.868	0.24	PARX+PRECI
Amazonas	0.788	0.940	0.15	PGAM+NINO4 ACUM

Em relação as modelos, metodologias e variáveis, a Tabela 5.11 resume a frequência com que apareceram. Note que os modelos “PARX+PRECI” e “PGAM+PRECI” foram os que mais obtiveram melhores resultados.

Tabela 5.11: Resumo dos melhores resultados encontrados.

Modelo	Freq.	Metodologia	Freq.	Var. Climática	Freq.
PGAM+PRECI	6	PGAM	16	PRECI	12
PARX+PRECI	6	PARX	7	NINO1+2 ACUM	5
PGAM+NINO1+2 ACUM	4	MSSA	5	NINO3 ANOM	2
MSSA+PAR	2			ONI ACUM	2
MSSA+NINO3 ANOM+PAR	2			NINO4 ACUM	2
PGAM+ONI ACUM	2			NINO3 ACUM	2
PGAM+NINO4 ACUM	2			SOI ANOM	1
PGAM+NINO3 ACUM	1				
PARX+NINO1+2 ACUM	1				
MSSA+NINO3 ACUM+PAR	1				
PGAM+SOI ANOM	1				

Finalmente, iremos comparar os resultados obtidos com os melhores modelos encontrados e também os resultados de Lima et al. Obviamente, os modelos que melhor performaram frente ao *benchmark* serão também os selecionados para serem confrontados com o modelo Dinâmico Linear. Nas Figuras 5.5, 5.6 e 5.7 estão expostas as comparações entre as métricas obtidas com cada um dos modelos. Observe que segundo as métricas MSRE e NSC o modelo proposto na Tese alcança resultados melhores do que em Lima et al. para 8 das 14 bacias, já com a métrica MARE este número é de 7 bacias.

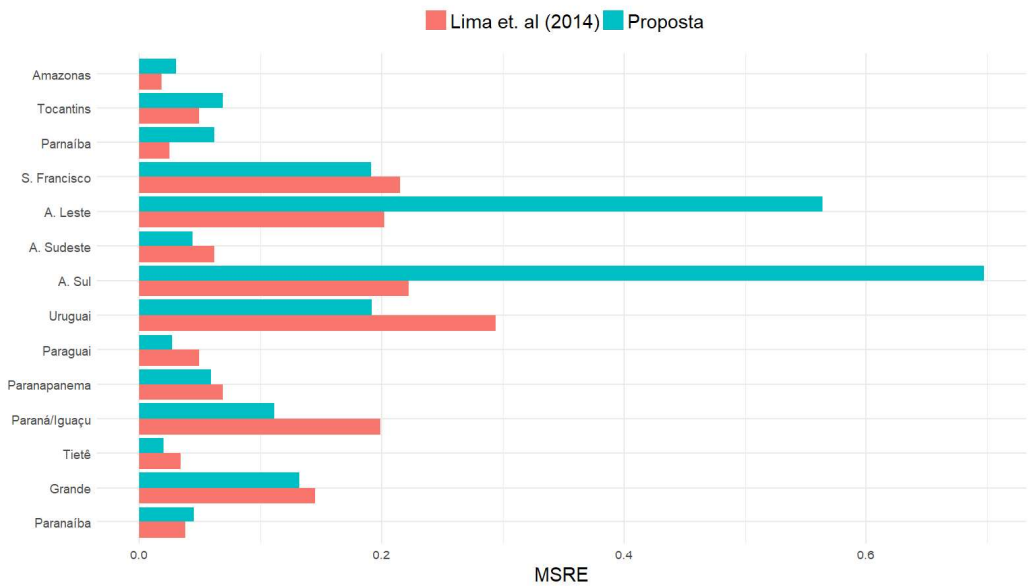


Figura 5.5: Comparação do MSRE entre os modelos de Lima et al. e da Tese.

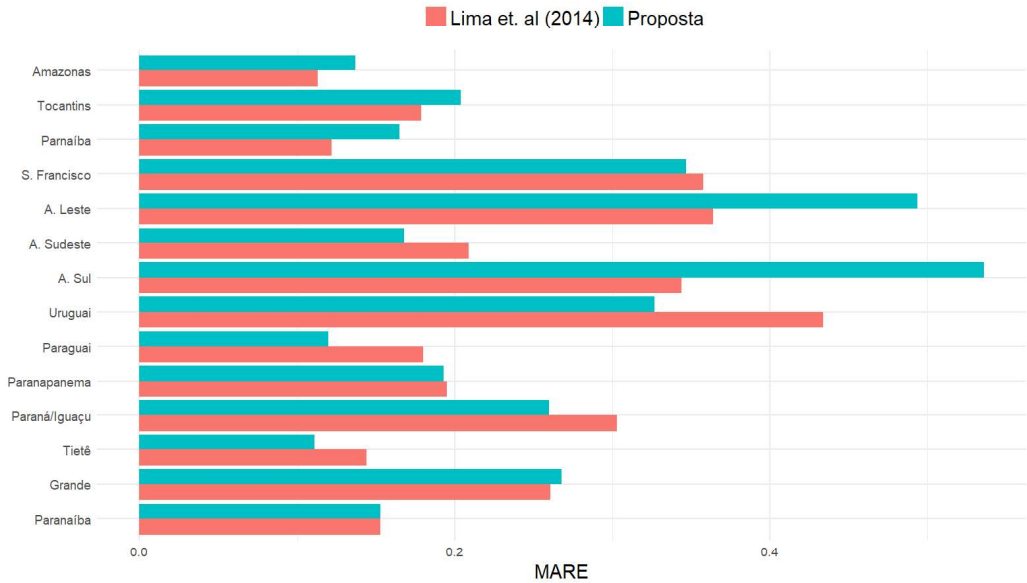


Figura 5.6: Comparação do MARE entre os modelos de Lima et al. e da Tese.

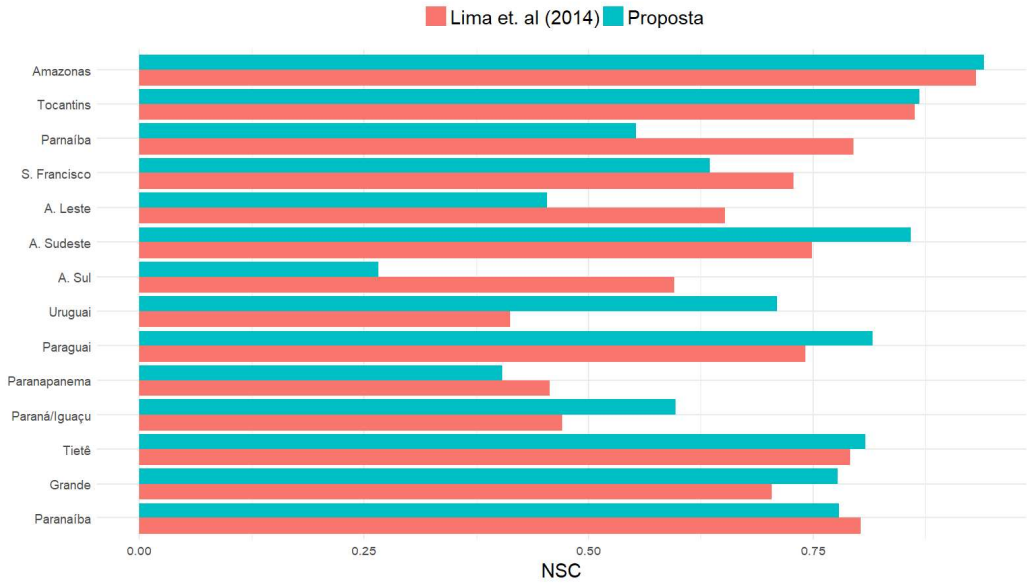


Figura 5.7: Comparação do NSC entre os modelos de Lima et al. e da Tese.

Em suma, os resultados apresentados nesta Tese validam os objetivos descritos como a serem alcançados no capítulo Introdução, isto é: incorporação de variáveis exógenas na previsão e/ou simulação das séries de vazão de forma linear e não-linear. Além disso, a metodologia proposta se mostra superior, em alguns casos, a modelos já testados e publicados na literatura. No próximo capítulo são expostas as conclusões finais e diretivas para continuação do trabalho.

A motivação deste trabalho é a importância dos modelos de previsão e simulação de séries de vazão para o setor elétrico brasileiro, que é basicamente formado por usinas hidrelétricas e consequentemente depende dos regimes hidrológicos. Definiu-se como objetivo a proposição de modelos que possam aperfeiçoar a modelagem e previsão das séries de vazão. Após identificar, na revisão da literatura, que existem vantagens em aplicar técnicas de decomposição em modelos hidrológicos e também verificar que as variáveis climáticas como precipitação e índices do evento El Niño e La Niña são correlacionadas com essas séries, definiu-se que seriam testados: a utilização da SSA e da MSSA em conjunto com o PAR, o modelo PARX e o modelo PGAM para alcançar os objetivos traçados.

A análise descritiva das séries de vazão e índices climáticos permitiu identificar previamente que todas as séries de vazão das bacias apresentavam correlação alta e significativa com as séries de precipitação e somente algumas bacias e alguns índices ENSO se relacionam no mesmo instante. Porém, tal análise também mostrou que há diferença significativa nas vazões para períodos considerados neutros, com ocorrência de El Niño ou La Niña.

Para obter os resultados finais foram utilizadas seis janelas de treinamento e validação neste trabalho. A primeira janela foi empregada com o objetivo de obter a melhor combinação dos parâmetros L e m para as técnicas SSA e MSSA e dos parâmetros $p(m)$ e $v(m)$ para o PARX e o PGAM. As demais cinco janelas foram ajustadas no período *in-sample* e verificado o erro obtido para o período *out-of-sample* e o erro final do modelo para cada bacia foi calculado como a média do erro para as janelas 2 a 6. Uma limitação deste trabalho encontra-se na ausência de modelo de previsão de variáveis explicativas sendo usado o próprio valor observado no período *out-of-sample*. Para sanar esta limitação sugere-se a utilização da previsão de órgãos oficiais ou de modelos encontrados na literatura.

Quando comparados com o modelo vigente (PAR) verifica-se que os modelos aqui propostos possuem um desempenho superior na modelagem das séries de vazão, concluindo que a inclusão de variáveis climáticas influencia significativamente na vazão das bacias brasileiras e consequentemente a geração

de energia do país. Em relação aos modelos que envolveram a utilização de variáveis climáticas, conclui-se que o Equatorial SOI deve ser inserido para modelagem das bacias do Paranaíba e do Paraguai, o ONI para a bacia do Tietê, o Niño 1+2 para a bacia do São Francisco e o Niño 3.4 ou o Niño 4 para a bacia do Parnaíba. Para todas as demais bacias sugere-se a utilização da precipitação.

A partir desses resultados conclui-se também que a inclusão de índices climáticos na modelagem das séries de vazão pode ocasionar uma mudança no nível histórico de vazão que o modelo PAR não é capaz de capturar, dado que tal modelo busca reproduzir a média histórica. Portanto, o modelo vigente pode subestimar ou sobrestimar o volume de água no caso da ocorrência de um evento climático e na mudança de regime das chuvas. A consequência de tal distorção pode ser a não-otimização do despacho hidrotérmico e um futuro racionamento de energia.

Como esta Tese foi inspirada no trabalho de Lima et al. [14] os métodos propostos foram aplicados a base de dados de tal trabalho. Quando comparados com os valores de *benchmark*, os modelos propostos superaram a performance em todas as bacias. Já quando confrontados com os resultados do modelo Dinâmico Linear, foi possível aprimorar o desempenho de pelo menos sete bacias. Concluindo assim que os modelos propostos podem e devem ser explorados para modelagem e simulação das séries de vazões brasileiras, assim como próprio modelo proposto em Lima et al.

Além de todos os pontos aqui descritos, esta Tese contribui também com o desenvolvimento do modelo Periódico Aditivo Generalizado (PGAM) e as equações de deslocamento para simulação de cenários sintéticos via distribuição Log-normal. O modelo PGAM mostrou-se eficiente para os objetivos aqui propostos dado que consegue capturar o comportamento sazonal das séries hidrológicas e relaciona-as de forma não-linear com as variáveis climáticas.

Enxerga-se como continuação deste trabalho a utilização de previsões das variáveis climáticas no lugar de valores observados, uso de diferentes funções base no modelo PGAM, consideração de mais de uma variável no PARX e aplicação de tais modelos às séries de Energia Natural Afluente.

Referências bibliográficas

- [1] AGÊNCIA NACIONAL DE ENERGIA ELÉTRICA. **BIG - Banco de Informações de Geração**, Dec. 2018.
- [2] LOHMANN, T.; HERING, A. ; REBENNACK, S.. **Spatio-temporal hydro forecasting of multireservoir inflows for hydro-thermal scheduling**. *European Journal of Operational Research*, 255(1):243–258, 2016.
- [4] OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **Histórico da Operação**, Dec. 2017.
- [6] CYRINO OLIVEIRA, F.. **Nova abordagem para geração de cenários de afluências no planejamento da operação energética de médio prazo**. Dissertação de mestrado, PUC-Rio, 2010.
- [7] HIPEL, K.; MCLEOD, A.. **Time Series Modelling of Water Resources and Environmental Systems**. Elsevier, 1994.
- [8] NOAKES, D.; MCLEOD, A. ; HIPEL, K.. **Forecasting monthly riverflow time series**. *International Journal of Forecasting*, 1(2):179–190, 1985.
- [9] AMARASEKERA, K.; LEE, R.; WILLIAMS, E. ; ELTAHIR, E.. **ENSO and the natural variability in the flow tropical rivers**. *Journal of Hydrology*, 200(1-4):24–39, 1997.
- [10] UVO, C.; GRAHAM, N.. **Seasonal runoff forecast for northern South America: A statistical model**. *Water Resources Research*, 34:3515–3524, 1998.
- [11] BERRI, G.; GHIETTO, M. ; GARCIA, N.. **The Influence of ENSO in the Flows of the Upper Paraná River of South America over the Past 100 Years**. *Journal of Hydrometeorology*, 3:57–65, 2002.
- [12] CARDOSO, A.; SILVA DIAS, P.. **The relationship between ENSO and Paraná River flow**. *Advances in Geosciences*, 6:189–193, 2006.

- [13] LIMA, C.; LALL, U.. Climate informed monthly streamflow forecasts for the Brazilian hydropower network using a periodic ridge regression model. *Journal of Hydrology*, 380(3-4):438–449, 2010.
- [14] LIMA, L.; POPOVA, E. ; DAMIEN, P.. Modeling and forecasting of Brazilian reservoir inflows via dynamic linear models. *International Journal of Forecasting*, 30(3):464–476, 2014.
- [15] MAÇAIRA, P.; CYRINO OLIVEIRA, F.; FERREIRA, P.; DE ALMEIDA, F. ; SOUZA, R.. Introducing a causal PAR(p) model to evaluate the influence of climate variables in reservoir inflows: a brazilian case. *Pesquisa Operacional*, 37(1):107–128, 2017.
- [16] SILVEIRA, C.; ALEXANDRE, A.; DE SOUZA FILHO, F.; VASCONCELOS JUNIOR, F. ; CABRAL, S.. Monthly streamflow forecast for National Interconnected System (NIS) using Periodic Autoregressive Endogenous Models (PAR) and Exogenous (PARX) with climate information. *Brazilian Journal of Water Resources*, 22, 2017.
- [17] MARQUES, C.; FERREIRA, J.; ROCHA, A.; CASTANHEIRA, J.; MELO-GONÇALVES, P.; VAZ, N. ; DIAS, J.. Singular spectrum analysis and forecasting of hydrological time series. *Physics and Chemistry of the Earth*, 31:1172–1179, 2006.
- [18] ADAMOWSKI, J.. River flow forecasting using wavelet and cross-wavelet transform models. *Hydrological Processes*, 22:4877—4891, 2008.
- [19] WU, C.; CHAU, K. ; LI, Y.. Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *Water Resources Research*, 45:1–23, 2009.
- [20] HUANG, S.; CHANG, J.; HUANG, Q. ; CHEN, Y.. Monthly streamflow prediction using modified EMD-based support vector machine. *Journal of Hydrology*, 511:764—775, 2014.
- [21] MAÇAIRA, P.; THOMÉ, A.; CYRINO OLIVEIRA, F. ; DE ALMEIDA, F.. Time series analysis with explanatory variables: A systematic literature review. *Environmental Modelling & Software*, 107:199—209, 2018.

- [22] HUANG, X.; MAÇAIRA, P.; HASSANI, H.; CYRINO OLIVEIRA, F. ; DHESI, G.. **Hydrological natural inflow and climate variables: Time and frequency causality analysis.** *Physica A*, 516:480—495, 2019.
- [23] ZHANG, X.; PENG, Y.; ZHANG, C. ; WANG, B.. **Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences.** *Journal of Hydrology*, 530:137–152, 2015.
- [24] BOX, G.; JENKINS, G.. **Time series analysis: forecasting and control.** Holden-Day, San Francisco, 1976.
- [25] BOX, G.; JENKINS, G. ; REINSEL, G.. **Time Series Analysis: Forecasting and Control.** Prentice Hall, 1994.
- [26] SALAS, J.; BOES, D. ; SMITH, R.. **Estimation of ARMA Models with seasonal parameters.** *Water Resources Research*, 18(4):1006–1010, 1982.
- [27] MOSS, M.; BRYSON, M.. **Autocorrelation structure of monthly streamflows.** *Water Resources Research*, 10:737–744, 1974.
- [28] THOMAS, H.; FIERING, M.. **Mathematical synthesis of streamflow sequences for the analysis of river basins by simulation.** *Design of Water Resource Systems*, (459-463), 1962.
- [29] PAGANO, M.. **On Periodic and Multiple Autoregressions.** *The Annals of Statistics*, 6(6):1310–1317, 1978.
- [30] TROUTMAN, B.. **Some results in periodic autoregression.** *Biometrika*, 66(2):219–228, 1979.
- [31] GLADYSHEV, E.. **Periodically and almost-periodically correlated random processes with a continuous time parameter.** *Theory of Probability and its Applications*, 8:173–177, 1963.
- [32] JONES, R.; BRELSFORD, W.. **Time series with periodic structure.** *Biometrika*, 54:403–408, 1967.
- [33] PEN-CHIH, T.. **Seasonal and Nonseasonal ARMA Models in Hydrology.** *Journal of the Hydraulics Division*, 102(10):1541–1559, 1976.
- [34] CROLEY II, T.; RAO, K.. **A manual for hydrologic time series deseasonalization and serial independence reduction.** Technical report, Iowa Institute of Hydraulic Research, Iowa City, Iowa, 1977.

- [35] TIAO, G.; GRUPPE, M.. Hidden periodic autoregressive-moving average models. *Biometrika*, 67(2):365–373, 1980.
- [36] DUNSMUIR, W.. Estimation of periodically varying means and standard deviations in time series. *Journal of Time Series Analysis*, 2(3):129–153, 1981.
- [37] CIPRA, T.. Statistical analysis of multiple moving average processes under periodicity. *Kybernetika*, 21:335–345, 1985.
- [38] CIPRA, T.; TLUSTY, P.. Estimation of multiple autoregressive-moving average models using periodicity. *Journal of Time Series Analysis*, 8:293–301, 1987.
- [39] VECCHIA, A.. Periodic autoregressive-moving average (PARMA) modelling with applications to water resources. *Water Resources Bulletin*, 21(5):721–730, 1985.
- [40] VECCHIA, A.. Maximum likelihood estimation for periodic autoregressive-moving average models. *Technometric*, 27:375–384, 1985.
- [41] THOMPSTONE, R.; HIPEL, K. ; MCLEOD, A.. Forecasting Quarter-Monthly Riverflow. *Journal of the American Water Resources Association*, 21(5):731–741, 1985.
- [42] ESHETE, Z.; VANDEWIELE, G.. Comparison of non-Gaussian multicomponent and periodic autoregressive models for river flow. *Stochastic Hydrology and Hydraulics*, 6(4):223–238, 1992.
- [43] WANG, Z.; VANDEWIELE, G.. Forecast and Monte Carlo simulation of Zaire River flow. *Hydrologie Continentale*, 9(1):69–83, 1994.
- [44] SRINIVASAN, K.; PHILIPOSE, M.. Evaluation and selection of hedging policies using stochastic reservoir simulation. *Water Resources Management*, 10(3):163–188, 1996.
- [45] MACEIRA, M.; DAMÁZIO, J.. The use of PAR(p) model in the stochastic dual dynamic programming optimization scheme used in the operation planning of the Brazilian hydropower system. *VIII International Conference on Probabilistic Methods Applied to Power Systems*, p. 397–402, 2004.

- [46] MONDAL, M.; WASIMI, S.. **Generating and forecasting monthly flows of the Ganges river with PAR model.** *Journal of Hydrology*, 323(1-4):41–56, 2006.
- [47] SHAO, Q.. **Robust Estimation For Periodic Autoregressive Time Series.** *Journal of Time Series Analysis*, 29(2):251–263, 2008.
- [48] CYRINO OLIVEIRA, F.; COSTA FERREIRA, P. ; CASTRO SOUZA, R.. **A parsimonious bootstrap method to model natural inflow energy series.** *Mathematical Problems in Engineering*, 2014, 2014.
- [49] DETTINGER, M.; DIAZ, H.. **Global Characteristics of Stream Flow Seasonality and Variability.** *Journal of Hydrometeorology*, 1(4):289–310, 2000.
- [50] KELMAN, J.; VIEIRA, A. ; RODRIGUEZ-AMAYA, J.. **Estimating the dimension of a model.** *Stochastic Environmental Research and Risk Assessment*, 14:123–138, 2000.
- [51] FOLEY, J.; BOTTA, A.; COE, M. ; COSTA, M.. **El Niño-Southern oscillation and the climate, ecosystems and rivers of Amazonia.** *Global Biogeochemical Cycles*, 16(4):79–1–79–20, 2002.
- [52] SOUZA FILHO, F.; LALL, U.. **Seasonal to interannual ensemble streamflow forecasts for Ceara, Brazil: Applications of a multivariate, semiparametric algorithm.** *Water Resources Research*, 39(11):1–13, 2003.
- [53] SOARES, J.; CARRIELLO, F.; FERREIRA, N. ; RENNÓ, C.. **Mapping the Hydrologic Response of the Brazilian hydrologic regions and their variability associated with El Niño and La Niña.** *Revista Ambiente & Água*, 1:21–36, 2006.
- [54] STUCK, J.; GUNTNER, A. ; MERZ, B.. **ENSO impact on simulated South American hydro-climatology.** *Advances in Geosciences*, 6:227–236, 2006.
- [55] FU, G.; CHARLES, S. ; CHIEW, F.. **A two-parameter climate elasticity of streamflow index to assess climate change effects on annual streamflow.** *Water Resources Research*, 43(11):1–12, 2007.
- [56] KUMAR, D.; MAITY, R.. **Bayesian dynamic modelling for nonstationary hydroclimatic time series forecasting along with uncertainty quantification.** *Hydrological Processes*, 22:3488–3499, 2008.

- [57] GARCÍA, N.; MECHOSO, C.. **Hydrological Sciences–Journal–des Sciences Hydrologiques**. Hydrological Sciences–Journal–des Sciences Hydrologiques, 50:459–478, 2009.
- [58] BOYER, C.; CHAUMONT, D.; CHARTIER, I. ; ROY, A.. **Impact of climate change on the hydrology of St. Lawrence tributaries**. Journal of Hydrology, 384(1-2):65–83, 2010.
- [59] CÓRDOBA-MACHADO, S.; PALOMINO-LEMUS, R.; GÁMIZ-FORTIS, S.; CASTRO-DÍEZ, Y. ; ESTEBAN-PARRA, M.. **Seasonal streamflow prediction in Colombia using atmospheric and oceanic patterns**. Journal of Hydrology, 538:1–12, 2016.
- [60] YANG, T.; ASANJAN, A.; WELLES, E.; GAO, X.; SOROOSHIAN, S. ; LIU, X.. **Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information**. Water Resources Research, 53(4):2786–2812, 2017.
- [61] HE, R.; CHEN, Y.; HUANG, Q.; YU, S. ; HOU, Y.. **Evaluation of ocean-atmospheric indices as predictors for summer streamflow of the yangtze river based on roc analysis**. Stochastic Environmental Research and Risk Assessment, 32:1903–1918, 2018.
- [62] WANG, W.; DING, J.. **Wavelet network model and its application to the prediction of hydrology**. Nature Science, 1:67–71, 2003.
- [63] ANCTIL, F.; TAPE, D.. **An exploration of artificial neural network rainfall– runoff forecasting combined with wavelet decomposition**. Journal of Environmental Engineering and Science, 31:21–28, 2004.
- [64] ZHOU, H.; PENG, Y. ; LIANG, G.. **The research of monthly discharge predictor-corrector model based on wavelet decomposition**. Water Resources Management, 22:217–227, 2008.
- [65] SANG, Y.. **A review on the applications of wavelet transform in hydrology time series analysis**. Atmospheric Research, 122:8–15, 2013.
- [66] NOURANI, V.; BAGHANAM, A.; ADAMOWSKI, J. ; KISI, O.. **Applications of hybrid wavelet–Artificial Intelligence models in hydrology: a review**. Journal of Hydrology, 514:358–377, 2014.
- [67] WU, C.; CHAU, K.. **Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis**. Journal of Hydrology, 399:394–409, 2011.

- [68] ZHANG, Q.; WANG, B.; HE, B.; PENG, Y. ; REN, M.. Singular spectrum analysis and ARIMA hybrid model for annual runoff forecasting. *Water Resources*, 25:2683–2703, 2011.
- [69] WANG, Y.; GUO, S.; CHEN, H. ; ZHOU, Y.. Comparative study of monthly inflow prediction methods for the Three Gorges Reservoir. *Stochastic Environmental Research and Risk Assessment*, 28:555–570, 2011.
- [70] YU, Y.; WANG, P.; WANG, C.; QIAN, J. ; HOU, J.. Combined monthly inflow forecasting and multiobjective ecological reservoir operations model: Case study of the Three Gorges Reservoir. *Journal of Water Resources Planning and Management*, 148, 2017.
- [71] GOLYANDINA, N.; KOROBEYNIKOV, A.. Basic Singular Spectrum Analysis and Forecasting with R. *Computational Statistics & Data Analysis*, 71:934–954, 2014.
- [72] GOLYANDINA, N.; NEKRUTKIN, V. ; ZHIGLJAVSKY, A.. *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman & Hall, 2001.
- [73] ELSNER, J.; TSONIS, A.. *Singular Spectrum Analysis: A New Tool in Time Series Analysis*. Springer US, 1996.
- [74] GOLYANDINA, N.; ZHIGLJAVSKY, A.. *Singular Spectrum Analysis for Time Series*. Springer-Verlag Berlin Heidelberg, 2013.
- [75] HASSANI, H.. Singular spectrum analysis: Methodology and comparison. *Journal of Data Science*, 5:239–257, 2007.
- [76] HASSANI, H.; HERAVI, S. ; ZHIGLJAVSKY, A.. Forecasting european industrial production with singular spectrum analysis. *International Journal of Forecasting*, 25:103–118, 2007.
- [77] HASSANI, H.; WEBSTER, A.; SILVA, E. ; HERAVI, S.. Forecasting u.s. tourist arrivals using optimal singular spectrum analysis. *Tourism Management*, 46:322–335, 2015.
- [78] ESQUIVEL, R.; SENNA, V. ; GOMES, G.. Modelagem de séries temporais via singular spectrum analysis: Uma otimização do window length. In: XLIII SIMPÓSIO BRASILEIRO DE PESQUISA OPERACIONAL, 2011.

- [79] GOLYANDINA, N.; STEPANOV, D.. Ssa-based approaches to analysis and forecast of multidimensional time series. In: PROCEEDINGS OF THE 5TH ST.PETERSBURG WORKSHOP ON SIMULATION, p. 293–298, 2005.
- [80] GOLYANDINA, N.. On the choice of parameters in singular spectrum analysis and related subspace-based methods. *Statistics and Its Interface*, 3:259–279, 2010.
- [81] DU, K.; YING, Z. ; LEI, J.. The incorrect usage of singular spectral analysis and discrete wavelet transform in hybrid models to predict hydrological time series. *Journal of Hydrology*, 552:44–51, 2017.
- [82] MCLEOD, A.. Diagnostic Checking Periodic Autoregression Models With Application. *The Journal of Time Series Analysis*, 15:221–233, 1995.
- [83] SCHWARZ, G.. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.
- [84] VRIEZE, S.. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, 2012.
- [85] AKAIKE, H.. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19:716—723, 1974.
- [86] AHO, K.; DERRYBERRY, D. ; PETERSON, T.. Model selection for ecologists: the worldviews of AIC and BIC. *Ecology*, 95:631–636, 2014.
- [87] URSU, E.; PEREAU, J.. Estimation and identification of periodic autoregressive models with one exogenous variable. *Journal of the Korean Statistical Society*, 46:629–640, 2017.
- [88] ANDEL, J.. Periodic autoregression with exogenous variables and periodic variances. *Aplikace Matematiky*, 34:387—395, 1989.
- [89] LUNA, I.; HIDALGO, I.; PEDRO, P.; BARBOSA, P.; FRANCATO, A. ; CORREIA, P.. Fuzzy Inference Systems for Multi-Step Ahead Daily Inflow Forecasting. *Pesquisa Operacional*, 37:129–144, 2017.

- [90] HASTIE, T.; TIBSHIRANI, R.. **Generalized additive models: some applications**. Journal of the American Statistical Association, 82:371–386, 1987.
- [91] HASTIE, T.; R, T.. **Generalized Additive Models**. Chapman & Hall/CRC, 1990.
- [92] CANTONI, E.; HASTIE, T.. **Degrees-of-freedom tests for smoothing splines**. Biometrika, 89:251–263, 2002.
- [93] CEPEL. **Manual de Referência do Modelo NEWAVE**. Technical report, Eletrobrás, Rio de Janeiro.
- [94] CHARBENEAU, R.. **Comparison of the two- and three-parameter log normal distributions used in streamflow synthesis**. Water Resources Research, 14:149–150, 1978.
- [95] PEREIRA, M.; G, O.; COSTA, C. ; KELMAN, J.. **Stochastic streamflow models for hydroelectric systems**. Water Resources Research, 20:379–390, 1984.
- [96] CYRINO OLIVEIRA, F.; CASTRO SOUZA, R. ; COSTA FERREIRA, P.. **A time series model for building scenarios trees applied to stochastic optimisation**. Mathematical Problems in Engineering, 2014, 2014.
- [97] R CORE TEAM. **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2017.
- [98] OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **Atualização de Séries Históricas de Vazões - Período 1931 a 2015**. Technical report, ONS, 2016.
- [99] CÂMARA DE COMERCIALIZAÇÃO DA ENERGIA ELÉTRICA. **Deck de Pregos**, Oct. 2018.
- [100] OPERADOR NACIONAL DO SISTEMA ELÉTRICO. **Diagrama Esquemático das Usinas Hidroelétricas do SIN - Janeiro/2018**, Oct. 2018.
- [101] SAID, S.; D, D.. **Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order**. Biometrika, 71:599–607, 1984.

- [102] PERRON, P.. Trends and random walks in macroeconomic time series. *Journal of Economic Dynamics and Control*, 12:297–332, 1988.
- [103] ACKER, J.; LEPTOUKH, G.. Online Analysis Enhances Use of NASA Earth Science Data. *Eos, Transaction American Geophysical Union*, 88:14–17, 2007.
- [104] BJERKNES, J.. Atmospheric teleconnections from the equatorial Pacific. *Journal of Physical Oceanography*, 97:163–172, 1969.
- [105] NOAA CLIMATE.GOV. The Walker Circulation: ENSO's atmospheric buddy, Oct. 2018.
- [106] OLIVEIRA, J.. O El Niño e Você - o fenômeno climático . Transtec, mar 2001.
- [107] NOAA CLIMATE.GOV. Why are there so many ENSO indexes, instead of just one?, Oct. 2018.
- [108] NATIONAL WEATHER LINKAGE: EL NIÑO SOUTHERN OSCILLATION. Climate Prediction Center, Oct. 2018.
- [109] BARNSTON, A.; CHELLIAH, M. ; GOLDENBERG, S.. Documentation of a highly ENSO-related SST region in the equatorial Pacific. *Atmosphere-Ocean*, 35:367–383, 1997.
- [110] STIGLER, S.. Francis galton's account of the invention of correlation. *Statistical Science*, 4:73–79, 1989.
- [111] SPEARMAN, C.. The proof and measurement of association between two things. *American Journal of Psychology*, 15:72–101, 1904.
- [112] NASH, J.; SUTCLIFFE, J.. River Flow Forecasting Through Conceptual Models. *Journal of Hydrology*, 10:282–290, 1970.