

**UNIVERSIDADE FEDERAL DE JUIZ DE FORA
FACULDADE DE ENGENHARIA ELÉTRICA
PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA
ELÉTRICA (PPEE)**

Vinicius Mendes Kohl

**UMA NOVA METODOLOGIA PARA GERAÇÃO DE SÉRIES
SINTÉTICAS DE VAZÃO NATURAL AFLUENTE USANDO
MODELO PERIÓDICO AUTORREGRESSIVO COM ERRO
MULTIPLICATIVO BASEADA EM OTIMIZAÇÃO NÃO LINEAR**

Juiz de Fora

2024

Vinícius Mendes Kohl

**UMA NOVA METODOLOGIA PARA GERAÇÃO DE SÉRIES
SINTÉTICAS DE VAZÃO NATURAL AFLUENTE USANDO
MODELO PERIÓDICO AUTORREGRESSIVO COM ERRO
MULTIPLICATIVO BASEADA EM OTIMIZAÇÃO NÃO LINEAR**

Dissertação apresentada ao Programa de Pós Graduação em Engenharia Elétrica (PPEE) da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Engenharia Elétrica. Área de concentração: Sistemas Energéticos

Orientador: D.Sc. André Luís Marques Marcato

Juiz de Fora

2024

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

Kohl, Vinícius.

UMA NOVA METODOLOGIA PARA GERAÇÃO DE SÉRIES SINTÉTICAS DE VAZÃO NATURAL AFLUENTE USANDO MODELO PERIÓDICO AUTORREGRESSIVO COM ERRO MULTIPLICATIVO BASEADA EM OTIMIZAÇÃO NÃO LINEAR / Vinícius Mendes Kohl. – 2024.

85 f. : il.

Orientador: André Luís Marques Marcato

Dissertação de (Mestrado) – Universidade Federal de Juiz de Fora, Faculdade de Engenharia Elétrica. Programa de Pós Graduação em Engenharia Elétrica (PPEE), 2024.

1. Séries Sintéticas e Vazões. 2. Séries Temporais. 3. Modelo Periódico Autorregressivo. 4. Modelo Periódico Autorregressivo Multiplicativo. 5. Programação Não Linear. 6. Planejamento da Operação de Médio Prazo. I. Marcato, André, orient. II. Título.

Vinicius Mendes Kohl

Uma nova metodologia para geração de séries sintéticas de vazão natural afluyente usando modelo periódico autorregressivo com erro multiplicativo baseada em otimização não linear

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Engenharia Elétrica. Área de concentração: Sistemas de Energia Elétrica

Aprovada em 17 de setembro de 2024.

BANCA EXAMINADORA

Prof. Dr. André Luis Marques Marcato - Orientador

Universidade Federal de Juiz de Fora

Prof. Dr. Afonso Henriques Moreira Santos

Universidade Federal de Itajubá

Prof. Dr. João Alberto Passos Filho

Universidade Federal de Juiz de Fora

Juiz de Fora, 28/08/2024.



Documento assinado eletronicamente por **Andre Luis Marques Marcato, Professor(a)**, em 17/09/2024, às 10:33, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Joao Alberto Passos Filho, Professor(a)**, em 17/09/2024, às 12:55, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **afonso henriques moreira santos, Usuário Externo**, em 26/11/2024, às 08:18, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **1950953** e o código CRC **F0C088C3**.

AGRADECIMENTOS

Agradeço aos meus pais Andersonn e Edila, e ao meu irmão Victor, que ao longo dessa jornada compuseram meu núcleo familiar e me acompanharam em todos os passos enquanto me apoiavam e tornavam meus dias mais alegres.

Agradeço aos meus tios, tias e primos que sempre se preocuparam comigo e me incentivam a ir mais longe e aceitar novos desafios.

Agradeço especialmente a minha avó, Elisete, que nesse último ano cuidou de mim e eu dela sempre com o carinho de mãe no coração.

Agradeço ao meu professor orientador, André Marcato, por ter me aceitado como aluno e expandido meus conhecimentos de forma brilhante.

Agradeço também aos meus empregadores na Consultoria RegE Barros Correia por financiar meu projeto de mestrado. Tem sido um prazer e uma satisfação imensa trabalhar ao lado de vocês, Tiago, Paulo, Rafael, Natália e Renata.

“Um livro é a prova de que os homens são capazes de fazer magia”.

Carl Sagan

RESUMO

O Planejamento da Operação Energética de Médio Prazo do Sistema Interligado Nacional (SIN) é um problema complexo de decisão sob incerteza, envolvendo interações espaciais e temporais intrincadas. A política ótima atualmente é obtida através do algoritmo de Programação Dinâmica Dual Estocástica (PDDE), que modela a incerteza hidrológica por meio de um processo Autorregressivo Periódico (PAR) utilizando a metodologia Box e Jenkins. No entanto, o modelo PAR estimado via Box e Jenkins apresenta uma limitação significativa: a possibilidade de gerar cenários sintéticos com valores negativos de afluência, o que pode levar à inviabilidades na solução do problema de otimização. Para contornar essa limitação, a metodologia oficial utiliza uma transformação Lognormal de três parâmetros, que, embora evite resultados negativos, introduz uma não linearidade indesejada no modelo, afetando a eficácia da PDDE. Este trabalho propõe uma metodologia alternativa, simples e flexível, para a estimação dos coeficientes do modelo PAR. A abordagem é baseada na minimização direta dos erros multiplicativos, combinada com a restrição da média nos resíduos gerados e ajustes de escala subsequentes, visando à geração de cenários sintéticos que atendam às premissas essenciais para o funcionamento ótimo da PDDE.

Palavras-chave: Séries Sintéticas e Vazões; Séries Temporais; Modelo Periódico Autorregressivo; Modelo Periódico Autorregressivo Multiplicativo; Programação Não Linear; Planejamento da Operação de Médio Prazo

ABSTRACT

The Mid-Term Energy Operation Planning of the Brazilian National Inter-connected System (SIN) is a complex decision-making problem under uncertainty, involving intricate spatial and temporal interactions. The optimal policy is currently determined through the Stochastic Dual Dynamic Programming (SDDP) algorithm, which models hydrological uncertainty through a Periodic Autoregressive (PAR) process using the Box and Jenkins method. However, the PAR model estimated by Box and Jenkins has a significant limitation: the possibility of generating synthetic series with negative inflow values, which can lead to unfeasibilities in the optimization problem. To address this issue, the official methodology employs a three-parameter Lognormal transformation, which, while avoiding negative results, introduces an undesirable nonlinearity into the model, affecting the effectiveness of the SDDP. This work proposes an alternative, simple, and flexible methodology for estimating the coefficients of the PAR model. The approach is based on the direct minimization of multiplicative errors, combined with a mean constraint on the generated residuals and subsequent scale adjustments, aiming to generate synthetic scenarios that meet the essential assumptions for the optimal functioning of the SDDP.

Palavras-chave: Synthetic Series and Inflows; Time Series; Periodic Autoregressive Model; Periodic Autoregressive Multiplicative Model; Nonlinear Programming; Mid-term Operational Planning

LISTA DE ILUSTRAÇÕES

Figura 1.1 – Composição da Matriz Elétrica Brasileira	16
Figura 2.1 – Representação Gráfica dos Pontos Chaves das Sequências Negativas	39
Figura 3.1 – Atenuação introduzida pelo canal	45
Figura 3.2 – Atenuação introduzida pelo canal compensada	45
Figura 4.1 – Influência do Fator de Ajuste nas Séries Produzidas para Usina Euclides da Cunha	55
Figura 4.2 – Influência do Fator de Ajuste nas Séries Produzidas para Usina Estreito Tocantins	56
Figura 4.3 – Influência do Fator de Ajuste nas Séries Produzidas para Usina Cana Brava	57
Figura 4.4 – Henry Borden	58
Figura 4.5 – Influência do Fator de Ajuste nas Séries Produzidas para Usina Serra da Mesa	59
Figura 4.6 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Jirau	66
Figura 4.7 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Baguari	66
Figura 4.8 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Balbina	67
Figura 4.9 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Belo Monte	67
Figura 4.10–Séries Sintéticas Produzidas pelo Modelo Proposto para as Usinas do Complexo Paulo Afonso - Moxotó	68
Figura 4.11–Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Furnas	68
Figura 4.12–Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Itaipu	69
Figura 4.13–Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Itaparica	69
Figura 4.14–Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Porto Primavera	70
Figura 4.15–Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Santo Antônio	70

Figura 4.16–Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Três Marias	71
Figura 4.17–Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Tucuruí	71
Figura 4.18–Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Xingó	72
Figura 4.19–Erros Médios Relativos entre as Médias das Séries Sintéticas e Históricas	73
Figura 4.20–Erros Médios Relativos entre os Desvios Padrão das Séries Sintéticas e Históricas	74
Figura 4.21–Distribuição dos Erros Relativos entre as Médias das Séries Sintéticas e Históricas	75
Figura 4.22–Distribuição dos Erros Relativos entre os Desvios Padrão das Séries Sintéticas e Históricas	75
Figura 4.23–Ampliação da Distribuição dos Erros Relativos entre as Médias das Séries Sintéticas e Históricas	76
Figura 4.24–Ampliação da Distribuição dos Erros Relativos entre os Desvios Padrão das Séries Sintéticas e Históricas	77
Figura 4.25–Séries Sintéticas e Históricas Geradas pelo PAR(p)M Tradicional para Usina Jordão	78
Figura 4.26–Séries Sintéticas e Históricas Geradas pelo PAR(p)M Proposto para Usina Jordão	78
Figura 4.27–Séries Sintéticas e Históricas Geradas por Box e Jenkins para Usina Jordão	79

LISTA DE TABELAS

Tabela 2.1 – Descrição e procedimento de cálculo das sequências negativas	40
Tabela 4.1 – Fatores de Ajuste Aplicados	60
Tabela 4.2 – Taxa de Sucesso nos Testes Estatísticos das Séries	64
Tabela 4.3 – Taxa de Sucesso nos Testes Estatísticos das Sequências	65

LISTA DE ABREVIATURAS E SIGLAS

Add	Mínimos Quadrados Ordinários
ANEEL	Agência Nacional de Energia Elétrica
AR	Modelo autorregressivo
Box	Box e Jenkins
CCEE	Câmara de Comercialização de Energia Elétrica
CEPEL	Centro de Pesquisas de Energia Elétrica
DECOMP	Modelo oficial de formação de preços de energia elétrica de curto prazo desenvolvido pelo CEPEL
EPE	Empresa de Pesquisa Energética
GEVAZP	Modelo de Geração de Cenários de Energias e Vazões e <i>Gated Recurrent Unit</i>
GRU	<i>Gated Recurrent Unit</i>
ENA	Energia natural afluenta
FAC	Função de autocorrelação
FACP	Função de autocorrelação parcial
NEWAVE	Modelo oficial de formação de preços de energia elétrica de médio e longo prazo desenvolvido pelo CEPEL
DESSEM	Modelo oficial de formação de preços de energia elétrica de curtíssimo prazo desenvolvido pelo CEPEL
IC	Intervalo de confiança
i.i.d.	Independente e indenticamente distriuído
IPOPT	<i>Interior Point Optimizer</i>
LSTM	<i>Long-Short Term Memory</i>
Mult	Mínimos Quadrados Ordinários com restrição nos coeficientes e erro multiplicativo
Mult New	Modelo proposto de otimização não linear com erro multiplicativo
ONS	Operador Nacional do Sistema Elétrico
PAR	Modelo periódico autorregressivo
PAR(p)	Modelo periódico autorregressivo de ordem p
PAR(p)M	Modelo periódico autorregressivo de ordem p com erro multiplicativo
PARMA	Modelo periódico autorregressivo de média variável
PDDE	Programação Dinâmica Dual Estocástica
PMO	Planejamento Mensal da Operação

PVARm	Modelo periódico vetorial autorregressivo com erro multiplicativo
PySDDP	<i>Toolbox</i> para Python especializada na leitura e processamento de dados advindos do NEWAVE
SIN	Sistema Interligado Nacional
UHE	Usina Hidrelétrica

LISTA DE SÍMBOLOS

\approx	Aproximadamente
γ	Função de autocorrelação;
δ	Limite inferior do erro aditivo;
ϵ	Erro aditivo do modelo autorregressivo;
ε	Erro relativo;
η	Erro multiplicativo do modelo autorregressivo;
θ	Parâmetro de nível do modelo autorregressivo;
μ	Média;
ρ	Função de autocorrelação parcial;
σ	Desvio padrão;
σ^2	Variância;
ϕ	Coefficiente de defasagem do modelo autorregressivo;
χ	Fator de ajuste do modelo multiplicativo com otimização não linear
C	Conjunto contendo os erros relativos
<i>C</i>	Comprimento de sequência
<i>E</i>	Esperança;
H_0	Hipótese nula
H_1	Hipótese alternativa
<i>I</i>	Intensidade de sequência
<i>n</i>	Número de amostras no histórico;
<i>p</i>	Ordem do modelo autorregressivo;
<i>S</i>	Soma de sequência
<i>t</i>	Período
<i>W</i>	Vetor multivariado que representa a dependência temporal e espacial dos resíduos;
<i>y</i>	Índice de ano do histórico
z_t	Vazão natural afluyente normalizada;
<i>Z</i>	Vazão natural afluyente;

SUMÁRIO

1	INTRODUÇÃO	16
1.1	Considerações iniciais e relevância do tema	16
1.2	Motivação	19
1.3	Objetivo do Trabalho	19
1.4	Produção científica resultante desta pesquisa	19
1.5	Organização do trabalho	20
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	Revisão Bibliográfica	21
2.1.1	Planejamento da Operação	21
2.1.2	Modelo Autorregressivo	22
2.1.3	Modelo Autorregressivo Multiplicativo	24
2.2	PAR(p) com erro aditivo	25
2.2.1	Método de Box e Jenkins	26
<i>2.2.1.1</i>	Identificação dos Coeficientes do Modelo	26
<i>2.2.1.2</i>	Identificação da Ordem do Modelo	27
2.2.2	Método dos Mínimos Quadrados Ordinários	28
2.2.3	Resíduos	29
2.3	PAR(p) com erro multiplicativo	32
2.3.1	Estimação dos Coeficientes	33
2.3.2	Identificação da Ordem do Modelo Multiplicativo	34
2.3.3	Estimação dos Resíduos	34
2.4	Avaliação do Modelo com Testes Estatísticos	35
2.4.1	Teste de Igualdade das Médias	36
2.4.2	Teste de Igualdade das Variâncias	37
2.4.3	Testes de Aderência	38
2.4.4	Análise da Sequências Negativa e Positiva	39
2.5	Resumo do Capítulo	41
3	MODELO PROPOSTO E FERRAMENTAS	42
3.1	Definição da Função Objetivo	42
3.1.1	Restrição de média	43
3.2	Fator de Ajuste	44

3.2.1	Formulação do Modelo proposto	46
3.3	Ferramentas Utilizadas	47
3.3.1	Linguagem de Programação	47
3.3.2	Leitura de Dados do NEWAVE	48
3.3.3	<i>Solver</i>	49
3.4	Conclusão Parcial	51
4	RESULTADOS	52
4.1	Considerações Iniciais	52
4.2	Influência do Fator de Ajuste	52
4.3	Resultados dos Testes Estatísticos	63
4.4	Média e Desvio Padrão das Séries Sintéticas	65
4.5	Limitações do Modelo	77
5	CONCLUSÃO E TRABALHOS FUTUROS	80
	REFERÊNCIAS	82

1 INTRODUÇÃO

1.1 Considerações iniciais e relevância do tema

A geração de energia hidrelétrica desempenha um papel fundamental na matriz elétrica do Brasil, sendo responsável por mais de 60% da eletricidade gerada no país [EPE, 2023]. A composição completa pode ser vista na Figura 1.1. Essa predominância deve-se à abundância de recursos hídricos em território nacional, que possibilita a operação de grandes usinas hidrelétricas com alta eficiência. Entretanto, a dependência dessa fonte de energia impõe desafios significativos, sobretudo no que tange à gestão dos riscos associados à estocasticidade das séries hidrológicas. A variabilidade natural das vazões afluentes, influenciada por fatores climáticos e ambientais, impacta diretamente a capacidade de geração de energia, tornando crucial a quantificação e o gerenciamento desses riscos para garantir a segurança e a sustentabilidade do sistema elétrico.

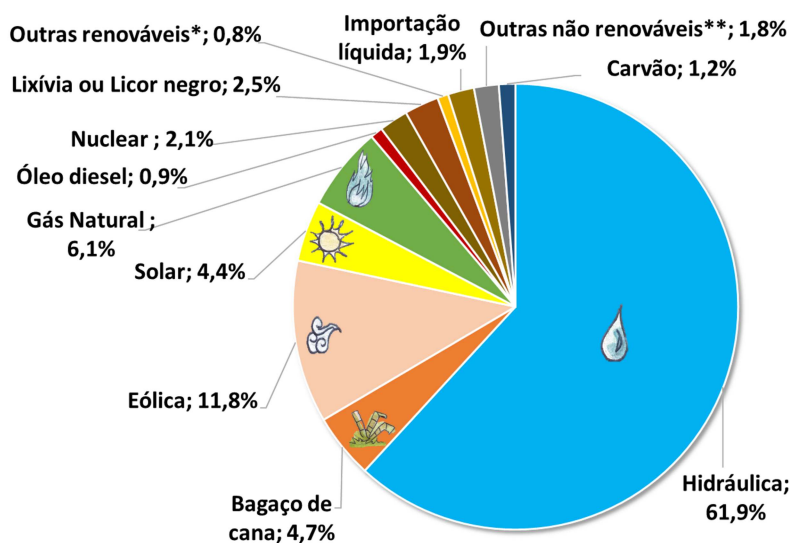


Figura 1.1 – Composição da Matriz Elétrica Brasileira

Fonte: EPE [2023]

O setor elétrico brasileiro é organizado de forma a assegurar a coordenação eficiente da operação e do planejamento da geração e transmissão de energia elétrica.

Nesse contexto, o Operador Nacional do Sistema Elétrico (ONS) é a entidade responsável pela coordenação e controle da operação das instalações de geração e transmissão, garantindo a continuidade e a qualidade do suprimento de energia em todo o território nacional. A Agência Nacional de Energia Elétrica (ANEEL), por sua vez, atua como órgão regulador, fiscalizando e regulamentando as atividades do setor, assegurando que o mercado opere de maneira justa e equilibrada. A Empresa de Pesquisa Energética (EPE) é responsável pelo planejamento da expansão do sistema, conduzindo estudos e pesquisas para apoiar a formulação de políticas públicas e subsidiar as decisões de investimento no setor.

A formação do preço da energia elétrica, conduzida pela Câmara de Comercialização de Energia Elétrica (CCEE), e o planejamento da operação, conduzida pelo ONS, no mercado brasileiro é quantificada por uma cadeia de modelos computacionais amplamente utilizados e desenvolvidos pelo Centro de Pesquisas de Energia Elétrica (CEPEL). O modelo NEWAVE é utilizado para a otimização do planejamento da operação a longo prazo, levando em consideração a incerteza das afluições e o uso eficiente dos recursos hídricos. O modelo DECOMP realiza a otimização da operação a médio prazo, ajustando a estratégia determinada pelo NEWAVE às condições hidrológicas de curto prazo. Finalmente, o modelo DESSEM é empregado para o despacho de geração a curtíssimo prazo, buscando a operação otimizada no horizonte diário. Esses modelos são fundamentais para a definição dos preços de curto e longo prazo da energia elétrica no Brasil, refletindo as condições hidrológicas e a disponibilidade dos recursos energéticos [Maceira et al., 2001].

A quantificação dos riscos hidrológicos é um processo essencial para a operação e o planejamento do setor elétrico brasileiro. Dada a imprevisibilidade das séries de vazões, é necessário dispor de ferramentas que permitam prever, com certa acurácia, os cenários futuros de afluição. Esses cenários são a base para a tomada de decisões estratégicas, como o despacho das usinas hidrelétricas, a gestão dos reservatórios e a definição de políticas de operação que minimizem os custos e garantam a oferta de energia. Nesse contexto, a modelagem estocástica surge como uma ferramenta vital, proporcionando uma forma estruturada de representar a incerteza inerente às séries temporais hidrológicas.

A metodologia oficial empregada para a quantificação desses riscos é a

Programação Dinâmica Dual Estocástica (PDDE), amplamente utilizada no planejamento da operação de sistemas hidrotérmicos. A PDDE permite a consideração de múltiplos cenários de afluências futuras, incorporando a incerteza hidrológica na tomada de decisão de forma explícita. Para que o método seja eficaz, é imprescindível a geração de cenários sintéticos de vazão que reflitam as características estatísticas das séries históricas, incluindo sua periodicidade, tendência e variabilidade.

Os modelos periódicos autorregressivos (PAR) têm sido amplamente utilizados para a geração de cenários sintéticos de vazão afluente devido à sua capacidade de capturar a sazonalidade das séries hidrológicas. Esses modelos permitem representar a dependência temporal das vazões, ajustando coeficientes que variam ao longo dos períodos, o que é essencial para a modelagem da estacionalidade presente nos dados hidrológicos. Os modelos PAR também permitem que haja correlação espacial entre as usinas, onde a vazão de outras usinas podem ser usadas para melhorar a geração das séries artificiais.

Os modelos formadores de preço utilizam os resultados de diversos modelos satélites como insumos para a montagem dos seus problemas de otimização. O modelo satélite responsável por gerar as vazões sintéticas que alimentam o NEWAVE se chama GEVAZP. Há a possibilidade de se gerar cenários de modo não condicionado ou levando em consideração a tendência hidrológica. Na geração com tendência hidrológica, utiliza-se os últimos valores reais registrados para a criação dos cenários artificiais em conjunto com os coeficientes e resíduos aleatórios representando as incertezas hidrológicas. Já a geração não condicionada, baseia-se nas médias mensais de longo termo (MLT) para criação de cenários iniciais e, a partir deles, novos cenários subsequentes são gerados [CEPEL, 2012].

Entretanto, o modelo PAR oficial assume erros aditivos na modelagem das séries temporais, o que pode não ser adequado em todas as situações, especialmente quando as séries apresentam a necessidade de valores positivos. A solução adotada, pelo CEPEL, para a produção de valores não negativos pelo modelo aditivo, que será melhor descrita na subseção **2.2.3**, afeta a linearidade do modelo além de introduzir uma dependência temporal dos resíduos assim, violando duas premissa para o funcionamento ótimo da PDDE [CEPEL, 2012].

1.2 Motivação

Diante desse cenário, os modelos com erro multiplicativo se apresentam como uma alternativa viável à geração de cenários positivos e lineares. O presente trabalho propõe uma nova abordagem para a modelagem das séries hidrológicas, baseada em um modelo PAR(p) com erro multiplicativo com os coeficientes estimados por otimização não linear. A principal motivação para a proposta é a necessidade de melhor representar a variabilidade das vazões afluentes, especialmente em cenários onde a magnitude das vazões impacta diretamente a variabilidade dos erros de previsão. O modelo proposto visa proporcionar uma maior flexibilidade na captura das características estocásticas das séries hidrológicas, garantindo que os cenários sintéticos gerados sejam mais aderentes às condições reais, o que, por sua vez, reflete em uma avaliação mais precisa dos riscos associados à operação do sistema elétrico.

Ao longo deste trabalho, serão discutidos os fundamentos teóricos do modelo PAR(p) com erro multiplicativo, sua implementação e os resultados obtidos a partir da aplicação prática no contexto da geração de cenários sintéticos de vazão afluente.

1.3 Objetivo do Trabalho

O principal objetivo deste trabalho é propor um modelo PAR univariado com erro multiplicativo e coeficientes derivados por otimização não linear como uma alternativa apta na geração de cenários sintéticos de vazão natural afluente no planejamento de médio prazo, bem como documentar suas particularidades e estabelecer uma base sólida como ponto de partida para futuros trabalhos derivados.

1.4 Produção científica resultante desta pesquisa

Este trabalho resultou em uma nova metodologia para geração de cenários sintéticos de vazão natural afluente de forma individualizada para o planejamento da operação de médio prazo. As conclusões obtidas com este trabalho levaram na publicação dos resultados parciais no trabalho. ¹

¹ Este trabalho e toda produção científica derivada dele foi financiada pela RegE Barros Correia Consultoria e deixo aqui meus mais sinceros agradecimentos

V. M. Kohl, A. L. M. Marcato, and P. de B. Correia. Using nonlinear optimization solvers to improve PAR(p) coefficients estimation in synthetic in-flow scenarios generation. In *58th International Universities Power Engineering Conference*, 2023.

1.5 Organização do trabalho

O Capítulo 2 apresenta uma revisão da literatura, juntamente com a fundamentação teórica que sustenta os modelos analisados neste trabalho. Nele, são definidos os conceitos essenciais que embasam o método proposto, proporcionando uma base sólida para a compreensão das abordagens adotadas;

O Capítulo 3 apresenta o modelo proposto, detalha suas principais premissas e discute as ferramentas utilizadas para a realização das simulações e algumas alternativas;

O Capítulo 4 apresenta os resultados das simulações computacionais, realizando uma comparação detalhada entre eles, destacando as principais tendências, padrões e discrepâncias observadas, e interpretando os achados à luz dos objetivos do estudo.

O Capítulo 5 traz as conclusões inferidas a partir dos resultados obtidos, sintetizando os principais achados e refletindo sobre sua importância e implicações. Além disso, analisa as limitações do estudo e explora possibilidades de expansão do tema para futuros trabalhos, oferecendo recomendações e sugerindo direções para novas pesquisas.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Revisão Bibliográfica

2.1.1 Planejamento da Operação

O GEVAZP é o modelo computacional desenvolvido pelo CEPEL utilizado para a geração de séries sintéticas de vazões ou energia natural afluyente, aplicado em problemas estocásticos como os modelos NEWAVE e DECOMP utilizando um modelo PAR para representar séries hidrológicas mensais. O modelo PAR utilizado oficialmente utiliza uma combinação linear de coeficientes e observações históricas de vazões naturais afluentes para produzir as séries sintéticas necessárias ao estudo. A metodologia de Box e Jenkins (1970) é empregada na obtenção da ordem, estimação dos coeficientes e cálculo dos resíduos do modelo PAR. Este algoritmo utiliza-se das equações de Yule-Walker nessa determinação para cada mês do horizonte de estudo. Após a definição do modelo autorregressivo, as séries sintéticas são geradas selecionando valores aleatórios de vazão natural afluyente presentes no histórico e adicionando-se um ruído à previsão, de forma a obter valores inéditos [CEPEL, 2012].

No entanto, um problema surge com a combinação de coeficientes e ruídos assumindo valores negativos. Torna-se possível a geração de cenários de vazões negativas, sendo um equivalente físico da inversão do fluxo do rio. Tal situação não é desejada e a alternativa utilizada oficialmente no planejamento da operação pelo CEPEL é a aplicação de uma transformação Lognormal com três parâmetros aos resíduos.

A solução adotada pelo CEPEL não é ótima, já que a incerteza incorporada no algoritmo da PDDE deve ser modelada de uma forma que a representação seja linear e a componente aleatória seja independente no tempo, condições violadas pela transformação Lognormal. Essa violação não garante que as variações estocásticas não introduzam dependências temporais indesejadas que podem tornar mais complexas as análises e a otimização do sistema [Penna, 2009].

Tais subotimalidades levam à busca por novas soluções tanto dentro do ramo dos modelos autorregressivos quanto fora dele. Em [Adriano de Melo et al., 2019],

os autores se distanciam da abordagem autorregressiva e utilizam duas técnicas de redes neurais recorrentes: *Long-Short Term Memory* (LSTM) e *Gated Recurrent Unit* (GRU) para previsão de vazões nas usinas do Rio Grande.

2.1.2 Modelo Autorregressivo

Os modelos PAR podem ser compreendidos como extensões periódicas dos modelos autorregressivos (AR), ou ainda, como tipos específicos de modelos AR multivariados. Essencialmente, os modelos PAR estendem a estrutura dos modelos AR ao associar um modelo AR específico para cada período m considerado em um ano. Isso significa que, ao invés de utilizar um único modelo AR para toda a série temporal, os modelos PAR permitem a definição de diferentes modelos para cada mês, capturando assim as variações sazonais e periódicas das séries temporais hidrológicas.

Os modelos PAR são projetados para lidar com erros independentes, o que os torna adequados para a modelagem da incerteza no contexto da PDDE. Além disso, esses modelos são amplamente reconhecidos na literatura especializada por sua eficácia e sua robustez na reprodução dos processos de Energia Natural Afluente (ENA) e na captura das características essenciais do processo estocástico envolvido na gestão de recursos hídricos.

As pesquisas sobre o comportamento estocástico de séries hidrológicas começaram no início da década de 1960 [Maass et al., 1962]. Este estudo pioneiro introduziu o modelo autorregressivo periódico (PAR), que se mostrou promissor para a análise e previsão de séries temporais hidrológicas. O modelo PAR foi projetado para capturar a periodicidade e a natureza estocástica das séries hidrológicas, oferecendo uma ferramenta valiosa para a gestão de recursos hídricos.

Avançando para 1984, o modelo PAR foi aplicado ao planejamento energético de médio prazo do Sistema Interligado Nacional (SIN) [Pereira et al., 1984]. Este trabalho foi significativo, pois demonstrou a aplicabilidade prática do modelo PAR diretamente nas séries hidrológicas, sem a necessidade de transformações específicas, como a transformação Box-Cox, que é frequentemente utilizada para estabilizar variâncias e tornar a série mais aderente aos pressupostos de normalidade. A simplicidade e a eficiência do modelo PAR tornaram-no uma escolha ideal para

essa aplicação.

Em Mendes et al. [2007] os autores descrevem a metodologia aplicada para desenvolver um modelo computacional orientado a objeto para geração mensal de vazões sintéticas multivariadas através de um Modelo PARMA multiplicativo. O modelo assume que a vazão de um determinado mês depende explicitamente da vazão do mês anterior, da vazão do mesmo mês do ano anterior e do ano retrasado, bem como de um ruído aleatório. O modelo estocástico tenta garantir que as correlações históricas de defasagem unitária das vazões agregadas anuais são devidamente reproduzidas, bem como os demais parâmetros anuais, periódicos e estatísticos.

Em Oliveira [2010], o autor tem como primeiro foco o uso da técnica Bootstrap para estimar com maior precisão a identificação das ordens p dos modelos, determinando a significância desses coeficientes. Tradicionalmente, essa identificação é realizada com base na significância dos coeficientes da função de autocorrelação parcial (FACP) usando a aproximação assintótica de Quenouille.

Ainda em Oliveira [2010], o trabalho teve como segundo objetivo aplicar o Bootstrap na geração de cenários, substituindo a tradicional distribuição Lognormal de três parâmetros, que tende a introduzir não-linearidades indesejáveis no modelo. Em vez disso, os resíduos do modelo PAR(p) aplicados às séries históricas foram usados diretamente para a geração de cenários. Os resultados indicaram que o Bootstrap não apenas levou à identificação de ordens p inferiores na maioria dos casos, mas também preservou satisfatoriamente as propriedades estatísticas das séries originais, resultado este que foi explorado também em de Castro [2012] e em Oliveira [2013].

Avanços na metodologia envolvendo modelos PAR geralmente assumem a Gaussianidade nos resíduos, permitindo a transformação dos dados em uma distribuição paramétrica. No entanto, foi observado que, na maioria dos casos reais do Setor Elétrico Brasileiro, o ruído não pode ser tratado dessa forma, devido aos comportamentos intrinsecamente enviesados das caudas, que são desafiadores para a reprodução no planejamento operacional do SIN. Em Baldiotti and Souza [2018], em resposta a essa limitação, os autores propõem uma abordagem não paramétrica para simular e amostrar os resíduos das séries de ENA utilizando a técnica de

Monte Carlo via Cadeias de Markov e a Estimação de Densidade por Kernel.

2.1.3 Modelo Autorregressivo Multiplicativo

Para processos que sempre produzem valores positivos, recomenda-se a utilização de modelos periódicos autorregressivos com erros multiplicativos, que serão doravante denominados PAR(p)M. Esses são descritos como o produto entre um fator de escala, que pode ser visto como um valor determinístico condicionado pelos valores anteriores da série, e um ruído que tem suporte positivo e valor esperado igual a um. Esses modelos são especialmente úteis em contextos onde os dados não podem assumir valores negativos, como em séries temporais financeiras, hidrológicas ou de demanda de energia, onde as quantidades observadas são inerentemente positivas [Box et al., 2008].

A abordagem multiplicativa permite que o modelo capture de forma mais precisa a variabilidade relativa dos dados, já que o erro é proporcional ao valor esperado. Isso contrasta com os modelos aditivos, onde o erro é independente do valor esperado e pode ser menos adequado para séries com grande variabilidade relativa [Harvey, 1990].

Apesar de mais recentes, produções científicas abordando esses tipos de modelo vem ascendendo rapidamente à medida que pesquisadores reconhecem suas vantagens e buscam aplicá-los em uma variedade de campos. Estudos recentes têm explorado a implementação e a eficácia desses modelos em diferentes contextos, destacando sua capacidade de fornecer previsões mais precisas e confiáveis quando comparados a métodos tradicionais. A crescente aceitação e desenvolvimento dos modelos com erros multiplicativos indicam uma tendência promissora para o futuro da análise de séries temporais, especialmente em áreas onde a positividade dos dados é um requisito indispensável.

No contexto da modelagem de séries temporais positivas, diversos estudos têm explorado modelos com erros multiplicativos. Por exemplo, trabalhos como o de Lanne [2006], Hautsch et al. [2014], Cipollini et al. [2006], e Bodnar and Hautsch [2016] abordam propostas de aplicação desses modelos no mercado financeiro. Em contrapartida, no setor elétrico brasileiro, teses como as de Braga [2011] e Ferreira [2013] introduziram modelos periódicos de distribuição gama para representar erros

multiplicativos em séries de Energia Natural Afluente (ENA). Além disso, Ribeiro et al. [2017] aplicaram um modelo PAR univariado com erro multiplicativo para gerar cenários no Planejamento Energético de Médio Prazo. Em outro estudo relevante, Cabral [2016] propôs o modelo PAR multivariado com erro multiplicativo para simular afluências positivas, superando limitações do PAR tradicional nesse aspecto específico.

Em Cabral [2016], o autor desenvolve um modelo Periódico Vetorial Autor-regressivo com erro multiplicativo (PVARm), proposto como uma alternativa mais adequada, pois, além de manter as premissas da PDDE, incorpora a correlação espacial das afluências diretamente na formulação do modelo. Este trabalho aplica o PVARm na geração de cenários para otimização do SIN, considerando mudanças de topologia das usinas ao longo do tempo e avaliando dois critérios de identificação: ordem fixa unitária e ordem determinada pela menor soma de erros de ajuste. O desempenho do modelo é avaliado de forma mais detalhada em [Cyrillo, 2018], mostrando a eficácia da abordagem.

2.2 PAR(p) com erro aditivo

Nesta seção será definida a fundamentação teórica do modelo PAR(p) com erro aditivo. Considere a vazão natural afluente Z_t como um processo periódico. Um modelo PAR de ordem (p) pode ser definido para um período t , com $t = 1, 2, \dots, 12$, da seguinte forma:

$$\left(\frac{Z_t - \mu_t}{\sigma_t} \right) = \sum_{i=1}^p \phi_i \left(\frac{Z_{t-i} - \mu_{t-i}}{\sigma_{t-i}} \right) + \epsilon_t \quad (2.1)$$

onde:

- Z_t : Vazão natural afluente no mês t ;
- μ_t : Média sazonal da série Z_t para período t ;
- σ_t : Desvio padrão sazonal da série Z_t para período t
- ϕ_i : Coeficiente AR de defasagem i , para o período t ;

- ϵ_t : é o erro associado à previsão da combinação linear dos coeficientes ϕ_i com as vazões nos períodos $t - i$

Simplificando a expressão 2.1, fazendo $\left(\frac{Z_t - \mu_t}{\sigma_t}\right) = z_t$, temos:

$$z_t = \sum_{i=1}^p \phi_i \cdot z_{t-i} + \epsilon_t \quad (2.2)$$

onde z_t é o valor normalizado da vazão natural afluyente no mês t .

Definido o modelo PAR(p) com erro aditivo, serão apresentados dois métodos para obter os coeficientes do modelo: o método de Box e Jenkins e o método dos mínimos quadrados ordinários.

2.2.1 Método de Box e Jenkins

O método de Box e Jenkins é um método clássico de identificação e estimação do modelo PAR, conforme indicado por [Hipel and McLeod, 1994]. A estimação é realizada utilizando as equações de Yule-Walker e a função de autocorrelação parcial (FACP) da série em análise.

2.2.1.1 Identificação dos Coeficientes do Modelo

Primeiramente, para o período m , a função de autocovariância periódica teórica ($\gamma_k^{(m)}$), de defasagem k , é definida para Z_t como sendo

$$\gamma_k^{(m)} = E[(Z_t - \mu_m)(Z_{m-k} - \mu_{m-k})] \quad (2.3)$$

Quando $k = 0$, a autocovariância periódica é simplesmente a variância, $\gamma_k^{(m)}$, de uma variável aleatória representante das observações do período m .

A variável normalizada, mais conveniente do que $\gamma_k^{(m)}$ para análise, é a função de autocorrelação (FAC) periódica teórica, que é definida para cada período m como sendo:

$$\rho_k^{(m)} = \frac{\gamma_k^{(m)}}{\sqrt{\gamma_0^{(m)} \gamma_0^{(m-k)}}} \quad (2.4)$$

A equação 2.4 permite analisar a dependência linear periódica da série em análise. Multiplicando-se ambos os membros da equação 2.1, que descreve a formulação do modelo PAR, pelo termo $(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}})$ e aplicando o valor esperado, o desenvolvimento da equação resultante levará à expressão:

$$\rho_k^{(m)} = \phi_1 \rho_{k-1}^{(m-1)} + \phi_2 \rho_{k-2}^{(m-2)} + \dots + \phi_p \rho_{k-p}^{(m-p)} \quad (2.5)$$

para $k > 0$ e $m = 1, 2, \dots, 12$.

Aplicando-se $k = 1, 2, \dots, p$ na equação da FAC, equação 2.5, obtém-se as equações periódicas de Yule-Walker para o período m , como apresentado abaixo:

$$\begin{aligned} \rho_k^{(m)} &= \phi_1 \rho_0^{(m-1)} + \phi_2 \rho_1^{(m-2)} + \dots + \phi_p \rho_{k-p-1}^{(m-p)} \\ \rho_k^{(m)} &= \phi_1 \rho_1^{(m-1)} + \phi_2 \rho_0^{(m-2)} + \dots + \phi_p \rho_{k-p-2}^{(m-p)} \\ &\vdots \\ \rho_p^{(m)} &= \phi_1 \rho_{p-1}^{(m-1)} + \phi_2 \rho_{p-2}^{(m-2)} + \dots + \phi_p \rho_0^{(m-p)} \end{aligned} \quad (2.6)$$

A relação que expressa os parâmetros AR para o período m é:

$$\begin{bmatrix} \rho_0^{(m-1)} & \rho_1^{(m-2)} & \dots & \rho_{k-p-1}^{(m-p)} \\ \rho_1^{(m-1)} & \rho_0^{(m-2)} & \dots & \rho_{k-p-2}^{(m-p)} \\ \vdots & \vdots & \dots & \vdots \\ \rho_{p-1}^{(m-1)} & \rho_{p-2}^{(m-2)} & \dots & \rho_0^{(m-p)} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \rho_1^{(m)} \\ \rho_2^{(m)} \\ \vdots \\ \rho_p^{(m)} \end{bmatrix} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_p \end{bmatrix} \quad (2.7)$$

2.2.1.2 Identificação da Ordem do Modelo

A função de autocorrelação (FAC) de um modelo PAR para um período m tende a atenuar sem truncar em uma defasagem específica, o que dificulta identificar diretamente a ordem do modelo AR pela FAC. Para superar essa dificuldade, utiliza-se a função de autocorrelação parcial periódica (FACP), que exibe um comportamento de corte claro. A FACP periódica no período m é definida pelo último parâmetro AR do modelo de ordem p .

A FACP periódica para um período m deve ser zero para defasagens maiores que p . Isso significa que após a defasagem p , a FACP periódica não apresenta valores significativos, indicando que a ordem do modelo AR para esse período é p .

A FACP periódica é estimada utilizando as equações de Yule-Walker, substituindo as autocorrelações teóricas pelas autocorrelações amostrais. Este processo é repetido para diferentes ordens p e diferentes períodos m .

O processo de estimativa da FACP é repetido para cada um dos períodos. A FACP é calculada para defasagens $1, 2, \dots, p$ em cada período. Se a FACP periódica amostral não é significativamente diferente de zero após a defasagem p , isso indica que a ordem do modelo AR para aquele período é p . A significância da FACP periódica é determinada comparando seus valores com o intervalo de confiança (IC) definido de 95%, dado por $\frac{1.96}{\sqrt{n}}$, com n sendo o número de amostras no histórico. Se a FACP estiver acima desse valor, ela é considerada significativa. Se nenhum valor da FACP é significativamente diferente de zero, ou seja, acima de IC, a série pode ser modelada como ruído branco, implicando em $p = 0$. Além disso, a ordem máxima permitida deve ser $p \leq 11$ para evitar que a vazão estimada seja correlacionada com a observação do mesmo mês no ano anterior.

2.2.2 Método dos Mínimos Quadrados Ordinários

O método dos mínimos quadrados ordinários é uma técnica estatística amplamente utilizada para a estimativa dos coeficientes ϕ_i em um modelo. Este método trabalha minimizando a soma dos quadrados dos erros, conhecidos como resíduos ϵ_t . Essencialmente, o objetivo principal do método dos mínimos quadrados é determinar os valores dos coeficientes de tal forma que o ajuste do modelo aos dados observados seja otimizado. Em outras palavras, o método busca reduzir ao máximo a diferença entre os valores que foram realmente observados e aqueles previstos pelo modelo, garantindo assim a melhor representação possível dos dados observacionais. A minimização da soma dos quadrados dos resíduos assegura que as discrepâncias individuais, ao serem elevadas ao quadrado, sejam somadas de maneira que qualquer grande erro seja penalizado mais severamente, resultando em uma solução que maximiza a precisão global do modelo.

Partindo da equação geral do modelo PAR com erro aditivo (2.2), e sabendo que o modelo PAR necessita de um ruído branco ϵ_t para funcionamento, conforme será abordado na subseção **2.2.3**. Uma maneira de se produzir esta série consiste na obtenção dos resíduos resultantes da previsão do modelo PAR, para o período

t , a partir do histórico de vazões, de modo que a expressão do modelo PAR que descreve uma única observação passe a ser:

$$z_t^y = \sum_{i=1}^p \phi_i \cdot z_{t-i}^y + \epsilon_t^y \quad (2.8)$$

onde:

- z_t^y é a vazão natural afluyente padronizada observada no período t no ano y ;
- ϵ_t^y é uma amostra resíduo, ou seja, a diferença entre a previsão do modelo PAR e uma vazão natural afluyente observada no período t no ano y ;

z_t segue uma distribuição normal, de modo que e_t também apresenta o mesmo comportamento, sendo este uma condição necessária. Isolando o resíduo na equação 2.8, tem-se:

$$\epsilon_t^y = z_t^y - \sum_{i=1}^p \phi_i \cdot z_{t-i}^y \quad (2.9)$$

Com o intuito de manter o modelo PAR o mais aderente possível ao histórico, define-se um problema de otimização cuja função objetivo consiste em minimizar a soma quadrática dos resíduos:

$$\min \left(\sum_{y=2}^{anos} (\epsilon_t^y)^2 \right) = \min \left(\sum_{y=2}^{anos} \left(z_t^y - \sum_{i=1}^p \phi_i \cdot z_{t-i}^y \right)^2 \right) \quad (2.10)$$

2.2.3 Resíduos

A verificação dos resíduos ϵ_t do modelo é uma etapa necessária que se baseia na realização de diversas análises estatísticas dos valores dos resíduos obtidos. Para assegurar que o modelo está bem ajustado aos dados, é fundamental que a série de resíduos satisfaça os pressupostos de um ruído branco para que não haja viés sistemático nos erros e não haja correlação significativa entre as observações dos resíduos, ou seja:

$$E[\epsilon_t^y] = 0 \quad (2.11)$$

$$\text{cov}(\epsilon_t^y, \epsilon_{t-i}^y) = \begin{cases} \sigma_t^2, & i = 0, \\ 0, & i = 1, 2, \dots, 12 \end{cases} \quad (2.12)$$

Além disso, a variância dos resíduos deve ser pequena e constante ao longo do tempo, refletindo a homocedasticidade, ou seja, a variabilidade dos erros é uniforme em todas as observações. Assim, ϵ_t^y é periodicamente independente e identicamente distribuído (i.i.d.), com valor esperado zero e variância constante σ_t^2 .

No entanto, tanto o método de Box e Jenkins quanto o método dos mínimos quadrados apresentam um fator complicador adicional por ambos permitirem que valores de afliências negativas sejam produzidos pelo PAR(p). Assim, para que sejam produzidas somente vazões positivas, a seguinte condição deve ser estabelecida a partir da manipulação da equação 2.1:

$$\left(\frac{Z_t - \mu_t}{\sigma_t} \right) = \sum_{i=1}^p \phi_i \left(\frac{Z_{t-i} - \mu_{t-i}}{\sigma_{t-i}} \right) + \epsilon_t \quad (2.13)$$

$$Z_t = \mu_t + \sum_{i=1}^p \sigma_m \cdot \phi_i \left(\frac{Z_{t-i} - \mu_{t-i}}{\sigma_{t-i}} \right) + \sigma_t \cdot \epsilon_t > 0 \quad (2.14)$$

isolando o termo ϵ_t :

$$\epsilon_t^y > -\frac{\mu_t}{\sigma_t} - \sum_{i=1}^p \phi_i \left(\frac{Z_{t-i} - \mu_{t-i}}{\sigma_{t-i}} \right) + \sigma_t. \quad (2.15)$$

Assim, temos um limite inferior para o valor de ϵ_t que será chamado de δ , então:

$$\delta = -\frac{\mu_t}{\sigma_t} - \sum_{i=1}^p \phi_i \left(\frac{Z_{t-i} - \mu_{t-i}}{\sigma_{t-i}} \right) + \sigma_t. \quad (2.16)$$

logo:

$$\epsilon_t > \delta \quad (2.17)$$

Na geração de séries sintéticas, é fundamental que os resíduos resultantes apresentem um coeficiente de assimetria próximo de zero. Esse coeficiente de assimetria reduzido é um indicativo de que os resíduos seguem uma distribuição normal, um requisito essencial para a sua utilização na PDDE [de Castro, 2012]. Para assegurar a normalidade dos resíduos, suas distribuições têm sido atualmente ajustadas para se conformarem a uma distribuição lognormal com três parâmetros [Marco et al., 1993]. Esse ajuste específico na distribuição dos resíduos é obrigatório, pois garante que eles permaneçam normais mesmo após a aplicação de uma transformação de normalização na série original. Em muitos casos, a normalização de uma série pode resultar em uma transformada inversa que não mantém a normalidade. Utilizando o logaritmo de uma série, a transformada inversa, que é o antilogaritmo, preserva a normalidade dos resíduos. Contudo, é importante notar que essa transformação, embora eficaz em garantir a normalidade, não assegura que os valores transformados sejam sempre não negativos.

Dada a equação:

$$\epsilon_t = e^{\xi_t} + \delta \quad (2.18)$$

onde: $\xi_t \sim N(\mu_\xi, \sigma_\xi^2)$. E sendo a_t uma distribuição lognormal [CEPEL, 2012], tem-se:

$$a_t \sim LNormal(\mu_\xi, \sigma_\xi^2, \delta) \quad (2.19)$$

Em séries temporais multivariadas, é essencial que os procedimentos adotados mantenham a dependência temporal. Portanto, considerando que os resíduos exibem dependência temporal, define-se o vetor multivariado W_t , sendo ($W_t \sim N_4(0, 1)$):

$$W_t = \frac{\xi_t - \mu_\xi}{\sigma_\xi} \quad (2.20)$$

isolando ξ_t :

$$\xi_t = \mu_\xi + W_t \sigma_\xi \quad (2.21)$$

Substituindo a equação 2.21 em 2.18:

$$\epsilon_t = e^{\mu_{\epsilon_t} + W_t \sigma_{\epsilon}} + \delta \quad (2.22)$$

Então a variável de vazão Z_t assume a seguinte formulação (OLIVEIRA, 2010):

$$Z_t = (e^{\mu_{\epsilon_t} + W_t \sigma_{\epsilon}}) \sigma^m \quad (2.23)$$

A utilização de transformações lognormal é primordialmente adotada para evitar valores negativos nas vazões geradas pelo PAR(p). Este método é crucial porque valores negativos de vazões naturais afluentes não são fisicamente plausíveis. No entanto, como evidenciado pelas equações 2.22 e 2.23, observa-se que ao garantir que as afluentes sejam não negativas, resulta-se em uma estrutura não linear para o cálculo desses valores [Finardi et al., 2009]. Esta não linearidade impacta na fase de otimização destinada ao cálculo do custo mínimo, visto que a otimização pela PDDE é conduzida sob a suposição de linearidade do modelo.

2.3 PAR(p) com erro multiplicativo

Considere Z_t como um processo periódico. Um modelo PAR(p)M de ordem p é definido para um período t , com $t = 1, 2, \dots, 12$, como segue:

$$Z_t = \left(\theta + \sum_{i=1}^p \phi_i Z_{t-i} \right) \cdot \eta_t \quad (2.24)$$

Sendo:

- Z_t : vazão natural afluente no período t ;
- ϕ_i : Coeficiente AR de defasagem i para o período t ;
- θ_m : parâmetro de nível;
- η_t : erro multiplicativo associado ao período t

2.3.1 Estimação dos Coeficientes

Tanto o método dos mínimos quadrados quanto o método de Yule-Walker são apropriados para aplicações do modelo PAR(p)M onde não é estritamente necessária a geração de cenários positivos, no entanto, como este é o principal objetivo por trás da utilização do modelo PAR(p)M para produção de cenários sintéticos, se fazem necessárias algumas alterações. Para isso, o método dos mínimos quadrados com restrições nos coeficientes será utilizado neste trabalho, cabendo, porém, ressaltar que é possível utilizar o método de Yule-Walker adicionando a hipótese da esperança condicional do erro multiplicativo condicionado as afluências até o período $t - 1$ ser igual a um.

Com a previsão dos valores de afluência dada pela equação 2.24, e sabendo que $\eta \geq 0$, conforme será mostrado na próxima seção, é essencial garantir que os coeficientes do modelo satisfaçam certas condições para assegurar que as vazões previstas sejam sempre positivas. Em particular, para que o modelo PAR(p)M seja coerente com a realidade física das vazões, é necessário e suficiente que os coeficientes θ e ϕ_i sejam não negativos, ou seja, $\theta, \phi_i \geq 0$.

Assim, estabelecendo essas condições, é definido o problema de otimização para o modelo PAR(p) com erro multiplicativo. A função objetivo é formulada como a minimização do erro quadrático entre as vazões observadas e as vazões previstas pelo modelo ($\eta = 1$). Isso é expresso da seguinte forma:

$$\min \left(\sum_{y=2}^{anos} (\epsilon_t^y)^2 \right) = \min \left(\sum_{y=2}^{anos} \left(Z_t^y - \left(\theta + \sum_{i=1}^p \phi_i \cdot Z_{t-i}^y \right) \right)^2 \right) \quad (2.25)$$

Aqui, ϵ_t^y representa o erro entre o valor observado da vazão Z_t^y e o valor previsto pelo modelo. A minimização da soma dos quadrados desses erros ao longo dos anos considerados visa ajustar os coeficientes θ e ϕ_i de forma a proporcionar o melhor ajuste possível entre os dados observados e o modelo.

No entanto, essa minimização deve respeitar as condições de não negatividade já mencionadas, que são formalmente expressas como:

$$\theta, \phi_i \geq 0 \tag{2.26}$$

2.3.2 Identificação da Ordem do Modelo Multiplicativo

Dada a limitação de ordem máxima $p \leq 11$, o critério de determinação da ordem dos modelos escolhido é o da menor soma dos erros quadrados da previsão [Cyrillo, 2018]. Este critério efetivamente se traduz em definir o modelo para ordem $p = 11$ e, dada a limitação nos coeficientes $\theta, \phi_i \geq 0$, o otimizador poderá livremente anular a contribuição da vazão correspondente a um mês se esta contribuir negativamente com o valor da função objetivo.

2.3.3 Estimação dos Resíduos

O modelo PAR(p) multiplicativo pode ser interpretado como um caso particular dentro da ampla classe de modelagens abordadas por Cabral [2016], onde o autor apresenta uma formulação mais geral e robusta, denominada modelo periódico multivariado autorregressivo multiplicativo, ou simplesmente PVARm. Esse modelo generaliza a ideia de periodicidade e autorregressividade, ao mesmo tempo em que incorpora uma estrutura de interconexões.

Considerando a formulação do resíduo descrita na forma aditiva, tem-se, neste caso, uma estimação baseada na diferença entre o valor realizado e o valor estimado pelo modelo 2.9. Ao desenvolver a proposta do modelo com resíduos multiplicativos, conforme ilustrado na equação 2.24, a estimativa dos resíduos adota uma formulação distinta. Essa formulação específica foi apresentada por Ribeiro et al. [2017], sendo expressa matematicamente da seguinte forma:

$$\eta_t^y = \frac{Z_t^y}{\left(\theta + \sum_{i=1}^p \phi_i Z_{t-i}^y\right)} \tag{2.27}$$

Essa expressão destaca que o resíduo multiplicativo η_t^y é calculado como a razão entre a vazão observada Z_t^y e a vazão estimada pelo modelo, que é uma combinação linear dos valores passados, ponderados pelos coeficientes θ e ϕ_i .

Para que as características da série histórica original sejam preservadas, é necessário que o ruído η_t^y siga um comportamento i.i.d., o que é uma suposição

comum em modelagens autorregressivas. Além disso, é imposta a condição $E[\eta_t^y] = 1$ sobre a expectativa do ruído, conforme apontado por Ribeiro et al. [2017]:

Essa condição assegura que, em média, o erro multiplicativo não introduza um viés sistemático nas previsões do modelo, mantendo a neutralidade do erro ao longo do tempo. Isso é essencial para garantir que o modelo preserve a tendência central dos dados, sem distorções causadas pelo erro.

Adicionalmente, dado que por definição $\theta, \phi_i \geq 0$ e considerando que Z_{t-1}^y são grandezas físicas que são inerentemente positivas, conclui-se que η_t^y também será não negativo, ou seja, $\eta_t^y \geq 0$. Isso garante a consistência física do modelo, evitando a geração de resíduos que possam conduzir a previsões de vazões negativas, o que seria fisicamente incoerente.

Uma vez que o modelo foi devidamente estimado, os resíduos multiplicativos são calculados conforme indicado na equação 2.24. Com o modelo ajustado, é possível simular cenários sintéticos de vazão natural afluyente utilizando as amostras dos resíduos obtidos.

2.4 Avaliação do Modelo com Testes Estatísticos

Após a construção de um modelo PAR(p), é necessário realizar uma avaliação do seu desempenho para assegurar que ele reproduza de maneira adequada as características das séries históricas observadas. Esse processo de validação envolve a aplicação de uma série de testes que analisam diversas propriedades estatísticas das séries geradas. Entre os principais aspectos verificados estão as estatísticas descritivas, como média, variância e assimetria, que são comparadas com as correspondentes medidas das séries históricas para garantir que as séries sintéticas possuam características similares.

Além disso, a correlação cruzada entre os dados gerados e os dados históricos é examinada para avaliar a capacidade do modelo em capturar as dependências temporais presentes nos dados originais. A correlação cruzada é uma medida importante, pois reflete como os valores de uma variável em diferentes momentos no tempo estão relacionados, o que é essencial em séries temporais que exibem comportamento periódico ou sazonal.

Outra dimensão da avaliação do modelo envolve a análise do comportamento das sequências geradas, onde se busca identificar se as séries sintéticas reproduzem adequadamente os padrões observados nas séries históricas, como ciclos, tendências e persistência.

2.4.1 Teste de Igualdade das Médias

O principal objetivo desse teste é verificar se as médias das séries sintéticas geradas pelo modelo são estatisticamente semelhantes à média da série original observada. Essa avaliação é realizada por meio de um teste de hipóteses. Nesse contexto, o teste de hipóteses envolve a formulação de duas hipóteses distintas: a hipótese nula, que será indicada como H_0 , postula que não há diferença significativa entre as médias das duas distribuições em questão, ou seja, as médias das séries sintéticas e da série original são estatisticamente semelhantes. Por outro lado, a hipótese alternativa, indicada como H_1 , sugere que existe uma diferença estatisticamente significativa entre as médias [Gujarati, 2006]. Abaixo, são detalhadas as formulações das hipóteses H_0 e H_1 :

$$\begin{aligned} H_0 : \mu_{hist} &= \mu_{gerada} \\ H_1 : \mu_{hist} &\neq \mu_{gerada} \end{aligned} \tag{2.28}$$

Dessa forma, aplicando um teste de hipótese bilateral com um nível de significância de 5% ($\alpha = 0,05$), o teste t permite determinar se a média das séries geradas é estatisticamente equivalente à média da série histórica. Para cada cenário, calcula-se um p-valor e, se o p-valor for superior ao nível de significância de 5%, a hipótese nula não é rejeitada, implicando que as médias dos cenários e da série histórica são estatisticamente idênticas.

$$\begin{aligned} \text{p-valor} \geq 0,05 &\Rightarrow \mu_{hist} = \mu_{gerada} \\ \text{p-valor} < 0,05 &\Rightarrow \mu_{hist} \neq \mu_{gerada} \end{aligned} \tag{2.29}$$

Neste trabalho, foi realizada a média de todos os cenários gerados para cada um dos 60 períodos analisados. A comparação entre as médias foi feita

verificando-se a média de todos os cenários em cada um dos 60 períodos (5 anos) de planejamento em relação à média da série histórica replicada do período em questão. Também foi analisada a média global, onde a média global da série histórica foi comparada com as médias globais de cada um dos cenários gerados.

2.4.2 Teste de Igualdade das Variâncias

O teste utilizado para verificar a igualdade das variâncias entre diferentes séries é o Teste de Levene. Este teste é conhecido por sua robustez, sendo aplicável tanto a distribuições normais quanto a não normais [Levene, 1960]. A principal função do Teste de Levene é comparar a variância de uma distribuição específica com as variâncias de outras séries, permitindo avaliar se há homogeneidade entre elas. As hipóteses consideradas no teste são formalizadas da seguinte maneira:

$$\begin{aligned} H_0 : \sigma_1^2 &= \dots = \sigma_k^2 \\ H_1 : \sigma_i^2 &\neq \sigma_j^2 \end{aligned} \tag{2.30}$$

onde $i = 1, 2, \dots, k$ e $j = 1, 2, \dots, k$, com $i \neq j$.

A hipótese nula (H_0) postula que as variâncias das diversas séries analisadas são iguais, ou seja, não há diferença significativa entre elas. Em contrapartida, a hipótese alternativa (H_1) sugere que pelo menos uma das variâncias difere das demais, indicando a presença de variabilidade nas séries comparadas. As séries em análise são amostras compostas por n variáveis aleatórias i.i.d, com médias e variâncias desconhecidas [Almeida et al., 2008].

Ao adotar um nível de significância de 5% ($\alpha = 0,05$), o critério para não rejeitar a hipótese nula é que os p-valores obtidos nos testes sejam superiores a esse nível de significância. Em outras palavras, p-valores maiores que 0,05 indicam que não há evidência suficiente para rejeitar H_0 , sugerindo que as variâncias das séries sintéticas e da série histórica são estatisticamente equivalentes.

$$\begin{aligned} \text{p-valor} \geq 0,05 &\Rightarrow \sigma_1^2 = \dots = \sigma_k^2 \\ \text{p-valor} < 0,05 &\Rightarrow \sigma_i^2 \neq \sigma_j^2 \end{aligned} \tag{2.31}$$

Após a geração das séries sintéticas, as variâncias são calculadas para cada um dos 60 períodos avaliados. Essas variâncias são então comparadas de acordo com o procedimento descrito anteriormente. Além disso, é realizada uma análise das variâncias globais, onde os valores das variâncias globais das séries históricas são replicados e comparados com as variâncias globais das séries sintéticas geradas. Esse processo assegura que o comportamento das variâncias ao longo do tempo seja consistentemente avaliado, garantindo a confiabilidade das séries sintéticas em termos de suas propriedades estatísticas.

2.4.3 Testes de Aderência

Os testes de aderência são métodos estatísticos utilizados para verificar se uma distribuição de frequência observada em um conjunto de dados é estatisticamente similar a uma distribuição de referência, ou esperada. Estes testes são classificados como não paramétricos, o que significa que eles não pressupõem que os dados sigam uma distribuição específica, oferecendo, assim, maior flexibilidade na análise de dados com características variadas.

O Teste de Kolmogorov-Smirnov é adequado para variáveis contínuas e se baseia na comparação entre a distribuição de probabilidade acumulada da amostra observada e a distribuição acumulada esperada [Conover, 1971]. Este teste é útil para avaliar a similaridade entre duas distribuições contínuas e suas respectivas características. As hipóteses do Teste de Kolmogorov-Smirnov são expressas como:

$$\begin{aligned} H_0 : F(x) &= F(x_0) \\ H_1 : F(x) &\neq F(x_0) \end{aligned} \tag{2.32}$$

O nível de significância adotado para este teste foi de 5%. Os p-valores obtidos acima desse nível de significância indicam que a hipótese nula não deve ser rejeitada, sugerindo que as distribuições comparadas são estatisticamente semelhantes. No presente trabalho, o Teste de Kolmogorov-Smirnov foi aplicado com o objetivo de comparar a distribuição das médias globais dos períodos dos cenários gerados com a distribuição das médias históricas correspondentes e comparar as

distribuições acumuladas dos períodos dos cenários gerados com as distribuições acumuladas dos períodos históricos correspondentes.

Além disso, o teste também foi utilizado na análise da aderência das séries sintéticas geradas em relação às características da série histórica original, garantindo a validade do modelo em representar os padrões observados nos dados históricos.

2.4.4 Análise da Sequências Negativa e Positiva

O propósito da análise da sequência negativa é determinar se os períodos críticos do histórico, tais como as secas ocorridas ao longo dos anos, são igualmente representados nas séries geradas. A sequência negativa corresponde aos períodos em que as vazões afluentes estão abaixo de certos valores de interesse e a positiva é o oposto. Por simplicidade, esta subseção tratará apenas do procedimento para a sequência negativa, deixando subentendido o da positiva. Nesse contexto, é avaliado o comportamento dos valores que ficam abaixo das médias mensais obtidas a partir da série histórica. A Figura 2.1 oferece uma ilustração mais detalhada desse conceito [CEPEL, 2012].

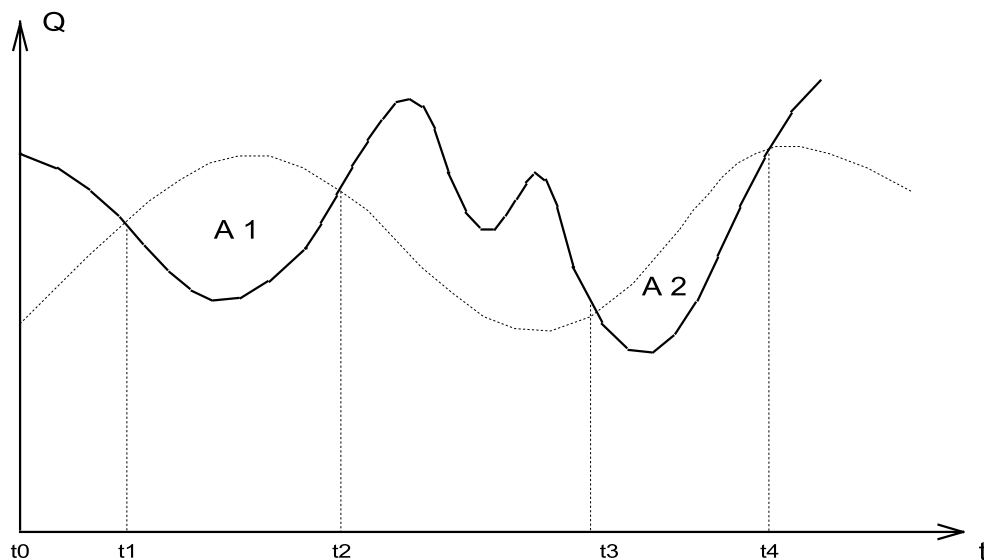


Figura 2.1 – Representação Gráfica dos Pontos Chaves das Sequências Negativas

Fonte: CEPEL [2012]

Na Figura 2.1, a linha contínua ilustra os valores das afluições da série, enquanto a linha pontilhada mostra a média histórica, que serve como a referência escolhida. O eixo horizontal representa o período de tempo analisado, e o eixo vertical exibe os valores das vazões afluentes. A sequência negativa é identificada nos períodos em que os valores das afluições ficam abaixo das médias mensais históricas. Na Figura 2.1, esses períodos correspondem aos intervalos de tempo (t_1, t_2) e (t_3, t_4) .

Seguindo o exposto em Pereira et al. [1984], serão avaliadas duas variáveis fundamentais para a análise referente às sequências: a soma da sequência, ilustrada pelas áreas A_1 e A_2 na Figura 2.1; e a intensidade da sequência, que corresponde ao valor médio das observações que estão abaixo do limite pré estabelecido. O método para calcular essas variáveis é descrito na Tabela 2.1 [CEPEL, 2012].

Tabela 2.1 – Descrição e procedimento de cálculo das sequências negativas

Variável	Descrição	Cálculo
Comprimento de sequência	Corresponde ao comprimento dos intervalos $(t_1 - t_2)$, $(t_3 - t_4)$	$C = (t_1 - t_2)$
Soma de sequência	Corresponde à área abaixo do limite durante a sequência. Na figura 2.1, equivale às áreas A_1 e A_2	$S = \sum_{i=t_1}^{t_2} (Z_i - \mu_i)$
Intensidade de sequência	Corresponde ao valor médio abaixo do limite, isto é a soma de sequência dividido pelo respectivo comprimento de sequência	$I = \frac{S}{C} = \frac{\sum_{i=t_1}^{t_2} (Z_i - \mu_i)}{(t_1 - t_2)}$

Fonte: CEPEL [2012]

Após a geração das séries sintéticas, é possível fazer uma comparação com os dados observados no histórico. Amostras são extraídas tanto do histórico quanto dos cenários, e então são conduzidos os testes de aderência descritos na subseção anterior. O objetivo é verificar se os cenários replicam as distribuições de probabilidade de cada variável, quando comparadas com a distribuição da amostra histórica.

2.5 Resumo do Capítulo

Neste Capítulo foi feita uma breve revisão bibliográfica a respeito do tema de geração de séries sintéticas, com ênfase em trabalhos relacionados ao uso do modelo $PAR(p)$ para esta tarefa no contexto do planejamento energético. Foi abordada também a formulação do modelo $PAR(p)$ com erro aditivo e multiplicativo, passando pelas etapas de identificação da ordem, métodos para identificação dos coeficientes e estimação dos resíduos. Por fim, foram estabelecidos e explanados os testes estatísticos a serem utilizados como métricas para avaliação das séries sintéticas geradas.

3 MODELO PROPOSTO E FERRAMENTAS

Neste capítulo, apresenta-se uma nova maneira de estimar os coeficientes de um modelo PAR(p)M utilizando diretamente a formulação do erro multiplicativo para minimizar seu valor, bem como as devidas considerações, para aplicação deste. Ao final deste capítulo, a última subseção detalha as ferramentas utilizadas bem como alternativas a elas junto a observações que facilitem a compreensão das escolhas feitas na implementação das simulações.

3.1 Definição da Função Objetivo

O modelo proposto parte da equação 2.27 para o erro multiplicativo e , sabendo que o valor esperado de $\eta_t = 1$, temos:

$$\eta_t = \frac{Z_t}{(\theta + \sum_{i=1}^p \phi_i Z_{t-i})} \approx 1 \quad (3.1)$$

ou então:

$$\eta_t - 1 = \frac{Z_t}{(\theta + \sum_{i=1}^p \phi_i Z_{t-i})} - 1 \approx 0 \quad (3.2)$$

Assim, utilizando-se o método dos mínimos quadrados para estimar os coeficientes que melhor se ajustam às observações presentes no histórico de vazões naturais afluentes, pode-se formular a função objetivo do problema de otimização que minimiza diretamente a soma quadrática do erro multiplicativo:

$$\min \left(\sum_{y=2}^{anos} (\eta_t^y - 1)^2 \right) = \min \left(\sum_{y=2}^{anos} \left(\frac{Z_t^y}{(\theta + \sum_{i=1}^p \phi_i \cdot Z_{t-i}^y)} - 1 \right)^2 \right) \quad (3.3)$$

Como a função objetivo não está definida em $\theta = \phi_i = 0$ e em conjunto com a restrição de que $\theta, \phi_i \geq 0$, o domínio da função objetivo só é definido para:

$$\theta, \phi_i > 0 \quad (3.4)$$

Em teoria, a modelagem apresentada até agora seria suficiente para que o problema de otimização pudesse ser resolvido. No entanto, durante as simulações, algumas dificuldades foram encontradas, levando a revisões no modelo para que ele produzisse resultados satisfatórios, o que, em um primeiro momento, significa reproduzir adequadamente os momentos de primeira e segunda ordem das séries históricas, ou seja, média e desvio padrão respectivamente.

3.1.1 Restrição de média

Um dos problemas enfrentados pelo modelo PAR(p)M foi a $E[\eta_t] \neq 1$, resultado do processo de otimização. O conjunto de coeficientes que otimiza o problema não necessariamente garante que a média dos resíduos η_t seja 1, devido a fatores como a restrição aos valores dos coeficientes $\theta, \phi_i > 0$ limitar a capacidade do modelo de ajustar corretamente $E[\eta_t]$ ou a função objetivo de minimizar os resíduos quadráticos $\left(\sum_{y=2}^{anos} (\eta_t^y - 1)^2\right)$ possuir discrepâncias significativas com relação a apenas minimizar o módulo da diferença $\left(\sum_{y=2}^{anos} |\eta_t^y - 1|\right) \Rightarrow E[\eta_t - 1] = 0$.

Utilizar um conjunto de coeficientes com $E[\eta_t] \neq 1$ ainda produz corretamente as séries sintéticas, no entanto, elas só podem ser corretamente geradas a partir dos resíduos de observações do histórico ou com um conjunto sintético de resíduos que possuam a mesma média do histórico. Como consequência, visto que cada mês possui um valor diferente para $E[\eta_t]$, o modelo adotaria mais uma premissa diretamente ligada ao histórico do respectivo período.

Assim, para se contornar esta situação e manter a premissa original, se faz necessária a inserção de uma restrição no problema para que a média dos resíduos multiplicativos η_t seja obrigatoriamente igual a 1. Tal restrição foi definida da seguinte forma:

$$media \left(\sum_{y=2}^{anos} (\eta_t^y - 1)^2 \right) = \frac{\sum_{y=2}^{anos} \left(\frac{Z_t^y}{(\theta + \sum_{i=1}^p \phi_i \cdot Z_{t-i}^y)} - 1 \right)}{anos - 1} = 0 \quad (3.5)$$

Mudando o necessário, a equação 3.5 pode ser reescrita como:

$$\sum_{y=2}^{anos} \left(\theta + \sum_{i=1}^p \phi_i \cdot Z_{t-i}^y \right) = \sum_{y=2}^{anos} Z_t^y \quad (3.6)$$

A equação 3.6 limita a região de solução a um hiperplano. Hiperplanos dividem conjuntos em dois semi-espacos convexos, mas não afins. Como os semi-espacos definidos pelo hiperplano da restrição são convexos, a propriedade da interseção de conjuntos garante que a interseção de dois conjuntos convexos também é convexa [Boyd and Vandenberghe, 2004]. Isto garante que a adição da restrição não introduz uma não convexidade no problema de otimização.

3.2 Fator de Ajuste

O modelo demonstrou uma resistência notável à presença de *outliers* no histórico de vazões. Como resultado dessa robustez, nos períodos históricos marcados por uma maior ocorrência desses valores extremos, a distribuição das séries sintéticas tende a apresentar uma dispersão reduzida, um efeito que se torna particularmente evidente na estatística do desvio padrão. Esse comportamento também impacta, embora em menor grau, o valor médio dos cenários sintéticos. Em casos de *outliers* elevados, observa-se um desvio das médias para valores menores; da mesma forma, a presença de *outliers* inferiores pode resultar em médias sintéticas mais altas.

Empiricamente, foi observado que, em períodos com *outliers* menos extremos, esses valores não distorcem de forma significativa a distribuição histórica, mas ainda geram um impacto notável na média das séries sintéticas. Tratar os *outliers* no processo de estimação dos coeficientes se mostrou ineficaz então, para mitigar os desvios entre as médias sintética e histórica, foi adotada uma solução inspirada na sondagem de canal no domínio da frequência, que a partir deste ponto será referida apenas como sondagem. Esta técnica, comumente utilizada em telecomunicações, é empregada para compensar a distorção de frequência e fase que um canal pode introduzir em um sinal transmitido.

A sondagem opera da seguinte forma: amostras em uma faixa estreita de frequências, com amplitudes conhecidas, são transmitidas através de um canal cujas características são inicialmente desconhecidas. Ao medir os dados recebidos,

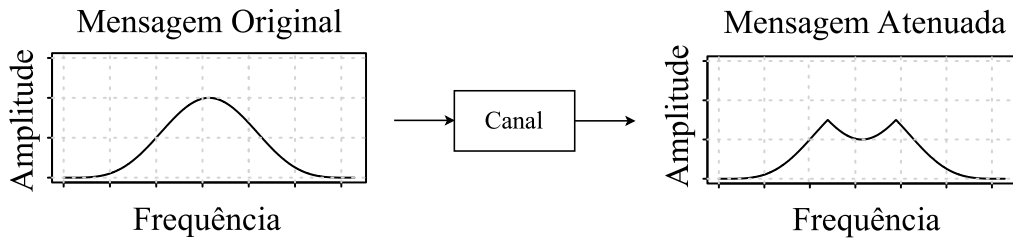


Figura 3.1 – Atenuação introduzida pelo canal

Fonte: Elaboração própria.

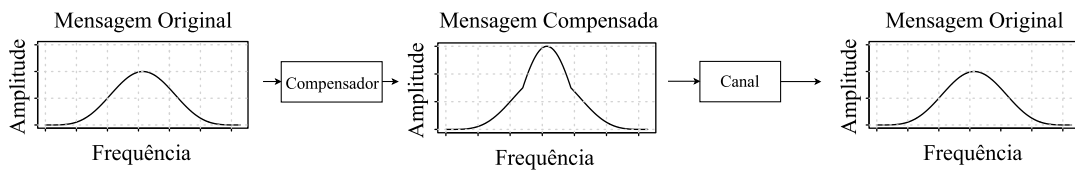


Figura 3.2 – Atenuação introduzida pelo canal compensada

Fonte: Elaboração própria.

torna-se possível calcular os desvios de frequência e a atenuação introduzidos pelo canal nessas amostras. Com base nesses cálculos, é possível determinar a função de transferência do canal para todo o espectro de frequências abrangido pelas amostras, permitindo a introdução de uma compensação nas mensagens transmitidas. O objetivo dessa compensação é garantir que o receptor consiga reconstruir o sinal desejado sem distorções significativas Filho [2020]. As Figuras 3.1 e 3.2 ilustram esse processo de maneira simplificada: a primeira figura demonstra o impacto do canal na mensagem original, enquanto a segunda figura mostra o efeito da compensação aplicada, restaurando a integridade da mensagem.

Um modelo PAR(p) pode ser interpretado de maneira análoga a um canal de comunicações, onde os coeficientes do modelo atuam como uma função de transferência. Nesse contexto, as entradas correspondem a uma sequência de vazões Z_{t-1}, \dots, Z_{t-p} e um resíduo multiplicativo η_t , enquanto a saída é a vazão

sintética Z_t . Quando uma discrepância significativa entre as médias é detectada, essa diferença pode ser entendida como a introdução de um viés na saída do modelo. Esse viés, por sua vez, pode ser corrigido ao ajustar as entradas, de modo a garantir que a vazão sintética gerada seja mais alinhada com a realidade histórica.

Como requisito fundamental, as séries geradas pelo modelo devem reproduzir a média histórica das vazões. Portanto, qualquer viés pode ser quantificado como um fator de escala, que é dado pela razão entre o valor histórico da vazão natural afluente do mês em análise e a previsão fornecida pelo modelo PAR(p)M, quando este é alimentado pelas tendências médias dos períodos anteriores e com $\eta_t = 1$. Sob essas condições, a previsão do modelo deve, teoricamente, aproximar-se da média histórica. Assim, define-se um fator de ajuste χ_t da seguinte forma:

$$\chi_t = \frac{\mu_{Z_t}}{\left(\theta + \sum_{i=1}^p \phi_i \cdot \mu_{Z_{t-i}}\right)} \quad (3.7)$$

3.2.1 Formulação do Modelo proposto

Recapitulando e formalizando as alterações discutidas ao longo deste capítulo, podemos, por fim, definir o modelo de geração de cenários sintéticos de vazões naturais afluentes, cujos coeficientes são estimados por meio de otimização não linear, da seguinte forma:

- Função Objetivo:

$$\min \left(\sum_{y=2}^{anos} (\eta_t^y - 1)^2 \right) = \min \left(\sum_{y=2}^{anos} \left(\frac{Z_t^y}{\left(\theta + \sum_{i=1}^p \phi_i \cdot Z_{t-i}^y\right)} - 1 \right)^2 \right) \quad (3.8)$$

- Sujeito à:

$$\theta, \phi_i > 0 \quad (3.9)$$

e

$$\sum_{y=2}^{anos} \left(\theta + \sum_{i=1}^p \phi_i \cdot Z_{t-i}^y \right) = \sum_{y=2}^{anos} Z_t^y \quad (3.10)$$

- Com cenários sintéticos gerados por:

$$Z_t = \left(\theta + \sum_{i=1}^p \phi_i \cdot Z_{t-i}^y \right) \cdot \chi_t \cdot \eta_t^y \quad (3.11)$$

com

$$\chi_t = \frac{\mu_{Z_t}}{\left(\theta + \sum_{i=1}^t \phi_i \cdot \mu_{Z_{t-i}}^y \right)} \quad (3.12)$$

3.3 Ferramentas Utilizadas

Conforme mencionado no início deste capítulo, e com o objetivo de garantir a transparência e a reprodutibilidade dos resultados por outros acadêmicos que venham a se deparar com este trabalho, esta subseção será dedicada a detalhar as ferramentas utilizadas, bem como a apresentar algumas alternativas a elas. Além disso, serão incluídas observações que visam facilitar a compreensão das escolhas feitas na implementação das simulações.

3.3.1 Linguagem de Programação

A linguagem de programação escolhida para o desenvolvimento do trabalho foi Python por diversas características, dentre as quais se destacam:

- Ser uma linguagem de programação de código aberto;
- Possuir uma sintaxe relativamente simples;
- Apresentar ampla disponibilidade de *toolboxes* de modelagem e resolução de problemas de otimização de forma declarativa, facilitando trabalhar com problemas mais complexos;
- Disponibilizar a *toolbox* PySDDP especializada na leitura e processamento de dados advindos do NEWAVE que teve sua estrutura detalhada em Silva et al. [2023].

Outras linguagens foram avaliadas mas, como justificado anteriormente e como será visto nas subseções **3.3.2** e **3.3.3**, as melhores soluções estão disponíveis em Python, sendo inconveniente mudar trabalhar em outra linguagem para este fim.

3.3.2 Leitura de Dados do NEWAVE

A fonte dos dados para as simulações é um *deck* de NEWAVE para o Planejamento Mensal da Operação (PMO) disponibilizado pelo ONS . A leitura do *deck* de NEWAVE costuma ser o ponto mais problemático pois diversas informações, geralmente as associadas às UHEs, estão contidas em arquivos de formato binário, incluindo o arquivo VAZOES.DAT, que contém as informações do histórico de vazões naturais afluentes em cada posto de medição de vazão desde 1932. Existem ferramentas como alternativas para se obter esses dados de forma mais palatável. Segue um pequeno comparativo dentre as três principais alternativas disponíveis, mantendo-se apenas no âmbito do necessário para condução deste trabalho:

- **VazEdit:** O VazEdit é um executável de Windows (.exe) que permite a conversão dos arquivos VAZOES.DAT e MLT.DAT livremente entre os formatos binário e texto tabular. Ele foi criado pelo ONS e é distribuído pelo ONS e pela CCEE.
 - **Prós:** Utilização intuitiva e direta sem necessidade de conhecimento prévio de linguagens de programação. Formato tabular permite utilização com praticamente qualquer ferramenta.
 - **Contras:** Disponível apenas para Windows, faz somente a leitura dos arquivos, necessitando de intervenção extra do usuário para organizar as informações e estimar os coeficientes do PAR(p) utilizando algum método já estabelecido na literatura como *benchmark*.
- **Leitorrmppe:** O Leitorrmppe é uma biblioteca de código feita pela EPE na linguagem de programação R e disponível para todas as plataformas.
 - **Prós:** Possui a capacidade de realizar a leitura de todos os dados contidos no deck de NEWAVE e organizá-los de maneira adequada em estruturas tabulares, não exigindo nenhum conhecimento prático de R para que sejam exportados e utilizados em outras ferramentas. Além disso, os dados vem discriminados com valores para todos os estágios no horizonte de estudo, com a inclusão de informações que exigiriam conhecimento

prévio do usuário como, por exemplo, a potência disponível no mês, que demanda aplicação de uma fórmula para levar em consideração a potência máxima, taxas de indisponibilidade programada e forçada, fator de capacidade e cronograma de manutenção individual das máquinas que compõem a usina.

- **Contras:** Não faz parte do escopo da ferramenta estimar os coeficientes do PAR(p), necessitando uma modelagem a parte para isto.
- **PySDDP:** Ainda está em desenvolvimento mas já consegue ler e organizar perfeitamente todos os dados do NEWAVE relacionados às UHEs e consegue fazer o mesmo para a maioria dos outros dados.
 - **Prós:** Disponível para todas as plataformas suportadas pelo Python, também possui discretização temporal dos parâmetros de forma apropriada além de realizar mais cálculos com os dados disponíveis do que o Leitorrmppe, como por exemplo a vazão incremental nos postos e estimação do modelo PAR(p).
 - **Contras:** O PySDDP coloca todos os dados do deck em um objeto com diversos atributos, métodos e funções, necessitando conhecimento básico de Python para extração de informações.

O PySDDP foi a ferramenta escolhida para obtenção dos dados provenientes do deck do NEWAVE, visto que é a mais completa para o problema proposto considerando que ele já calcula o PAR(p) utilizando o método de Box-Jenkins, detectando sua ordem máxima possível, estimando seus coeficientes e matriz de resíduos.

3.3.3 *Solver*

O problema de otimização usado no trabalho é classificado como problema não linear, mais especificamente convexo. Então, há a necessidade de um solver não linear com capacidade Boyd and Vandenberghe [2004]. O Python possui duas *toolboxes* principais adequadas ao problema em estudo, sendo elas o Pyomo e SciPy.

O SciPy é uma *toolbox* de computação científica que possui uma suíte completa de otimização embutida com diversos métodos disponíveis para utilização. A *toolbox* SciPy, com o método *Sequential Least Squares Programming* (SLSQP) Kraft [1988]. No entanto este apresentou dificuldades de convergência com as configurações padrão e por isso, não foi utilizado nas simulações.

O Pyomo é uma *toolbox* exclusivamente de modelagem de problemas de otimização, servindo apenas de interface para outros *solvers* existentes no mercado sem exigir do usuário conhecimento das suas sintaxes próprias desses, permitindo uma fácil troca de ferramentas sem nenhum esforço. A alternativa escolhida foi o IPOPT, um *solver* não linear de código aberto, parte da iniciativa COIN-OR, que implementa o método de pontos interiores primal-dual como estratégia de resolução IPOPT [2024]. O IPOPT possui como principal ponto positivo ser gratuito para uso comercial e acadêmico e, por consequência, possuir maior facilidade na distribuição de softwares oriundos de trabalhos futuros que incorporem o código utilizado neste trabalho, considerando aspectos tanto técnicos quanto legais, além de manter o espírito do código aberto neste trabalho por uma questão de preferência pessoal.

Existem também *solvers* de problemas não lineares não convexos que adotam estratégias como *branch and cut* para explorar de forma mais ampla o domínio do problema de otimização. No entanto essa categoria se mostrou extremamente exigente em relação ao poder computacional necessário, demandando um tempo de execução demasiadamente grande para resolver problemas cujos "*solvers* de ótimo local" foram bastante rápidos e precisos em resolver, dada a natureza convexa do problema, como será mostrado nos resultados no próximo capítulo. Alguns *solvers* desta categoria são Baron, Couenne e Gurobi, sendo que, este último, apesar de ser um conhecido *solver* linear, apresenta capacidade de solução de problemas não lineares e, mesmo não sendo explicitamente citado como destinado a problemas não convexos, apresenta a mesma descrição de algoritmo de solução e comportamento que os outros dois citados.

Outras linguagens de programação diferentes foram cogitadas para esta tarefa, como R e MATLAB, mas ambas não possuem suporte para modelagem declarativa, somente matricial o que seria contra contraproducente caso fossem usadas diante dos motivos já expostos em favor do Python.

3.4 Conclusão Parcial

Neste capítulo, apresentou-se a formulação do modelo proposto, detalhando sua função objetivo, a restrição que impõe média nula aos resíduos multiplicativos e a introdução de uma nova componente no modelo, representada pelo fator de ajuste. Foram também discutidas as ferramentas computacionais disponíveis para a implementação do problema e as razões que motivaram as escolhas das ferramentas adotadas, destacando sua adequação ao contexto e aos objetivos da pesquisa.

4 RESULTADOS

4.1 Considerações Iniciais

Como o objetivo principal deste trabalho é avaliar o desempenho global do modelo proposto, e dado o grande número de simulações realizadas, os resultados dos testes estatísticos e das séries sintéticas geradas serão abordados de forma agrupada para melhor visualização do panorama geral.

O conjunto de dados do NEWAVE empregado neste trabalho data do mês de julho de 2024, período em que estavam em operação 165 usinas hidrelétricas no sistema. Esse conjunto de dados foi utilizado como base para a realização das simulações de Monte Carlo, onde foram gerados 2000 cenários sintéticos. Cada cenário simula as condições hidrológicas e operacionais ao longo de um período de 5 anos, abrangendo, portanto, 60 meses consecutivos. Cabe destacar também que todos os testes foram realizados nas usinas de forma individualizada utilizando as vazões naturais afluentes sem nenhuma forma de acoplamento espacial entre vazões de diferentes usinas.

Ao longo deste capítulo, para melhor exibição dos resultados, serão usadas abreviações nas figuras e tabelas para denotar as informações provenientes das simulações dos seus respectivos métodos, sendo essas:

- **Box**: Box e Jenkins (erro aditivo);
- **Add**: Mínimos Quadrados Ordinários (erro aditivo);
- **Mult**: Mínimos Quadrados Ordinários com restrição nos coeficientes (erro multiplicativo) e;
- **Mult New**: Modelo proposto de otimização não linear (erro multiplicativo).

4.2 Influência do Fator de Ajuste

A introdução do fator de ajuste na equação de geração de cenários sintéticos de vazão natural afluente é uma estratégia fundamental para capturar corretamente variações das séries históricas. Esse fator permite ajustar a amplitude e a tendência

média dos cenários gerados e demonstrar os efeitos desse fator é essencial para validar a eficácia da modelagem proposta e evidenciar como ela contribui para reduzir discrepâncias entre as séries sintéticas e os dados observados.

Nas Figuras 4.1, 4.2, 4.3, 4.4 e 4.5 são apresentados 5 usinas onde a presença do fator de ajuste mais impactou na distribuição das séries produzidas, fazendo uma comparação dos resultados obtidos com e sem a sua aplicação. Essas representações visuais comparam as médias e os desvios padrão das séries históricas com as séries sintéticas, evidenciando as diferenças de desempenho entre os dois cenários. Optou-se por ocultar os valores das séries sintéticas para facilitar a visualização, mantendo visíveis apenas suas estatísticas. Cabe destacar que só foi julgada necessária a aplicação do fator em meses cujos valores médios das séries sintéticas apresentaram desvios maiores que 2,5% da média histórica.

Para facilitar o entendimento dos gráficos que serão apresentados a seguir, será fornecida uma explicação breve, mas detalhada, sobre o significado de suas legendas. Estas podem ser interpretadas como:

- **Séries Históricas:** Linhas contínuas sobrepostas com a trajetória de todas as séries do histórico;
- **Séries Sintéticas:** Linhas contínuas sobrepostas com a trajetória de todas as séries sintéticas geradas;
- **μ Séries Históricas:** Círculos indicando a trajetória do valor médio das séries históricas em cada mês;
- **μ Séries Sintéticas:** Linhas contínuas indicando a trajetória do valor médio das séries sintéticas em cada mês;
- **σ Séries Históricas:** Círculos indicando a trajetória do valor da média \pm o desvio padrão das séries históricas em cada mês;
- **σ Séries Sintéticas:** Linhas tracejadas indicando a trajetória do valor da média \pm o desvio padrão das séries sintéticas em cada mês;

- **Limite de *Outlier*:** Par de linhas tracejadas que delimitam valores considerados normais na distribuição histórica mensal, com *outliers* estando acima ou abaixo de ambas.

A análise das Figuras permite identificar que a inclusão do fator de ajuste leva a uma aproximação mais fiel da distribuição dos valores de média e desvio padrão. Também é possível observar que o histórico apresenta uma quantidade expressiva de *outliers* e que justamente nos meses mais afetados com a distribuição histórica mais distorcida o fator de ajuste foi mais efetivo, evidenciando assim a já mencionada resistência a *outliers* do modelo. As figuras também evidenciam que os *outliers* contribuem para a distribuição das séries históricas, e ajustar o modelo sem a presença dessas amostras tornaria o modelo aderente à uma distribuição consideravelmente diferente, mostrando que o tratamento desses valores é ineficaz.

Analisando a Tabela 4.2, observa-se que o fator de ajuste foi necessário em 28,33% dos casos, ou seja, em mais de um quarto dos casos o desvio apresentado pelo modelo foi maior do que 2,5% do valor médio do respectivo mês. Além disso, a amplitude mínima e máxima do fator de ajuste foi 0,654 e 1,826 respectivamente. O número elevado de ocorrências assim como com o alcance alto da compensação, se deve ao valor limite de ativação do ajuste ser relativo ao mês analisado. Então, pequenas distorções na média das séries sintéticas em meses de menor vazão histórica podem resultar em uma detecção para distorções às vezes insignificantes. O estabelecimento de um nível de significância global para os desvios das médias poderia resultar em menos ocasiões onde houve a necessidade de inclusão do fator de ajuste com pouco ou nenhum prejuízo à qualidade dos cenários gerados.

E. DA CUNHA

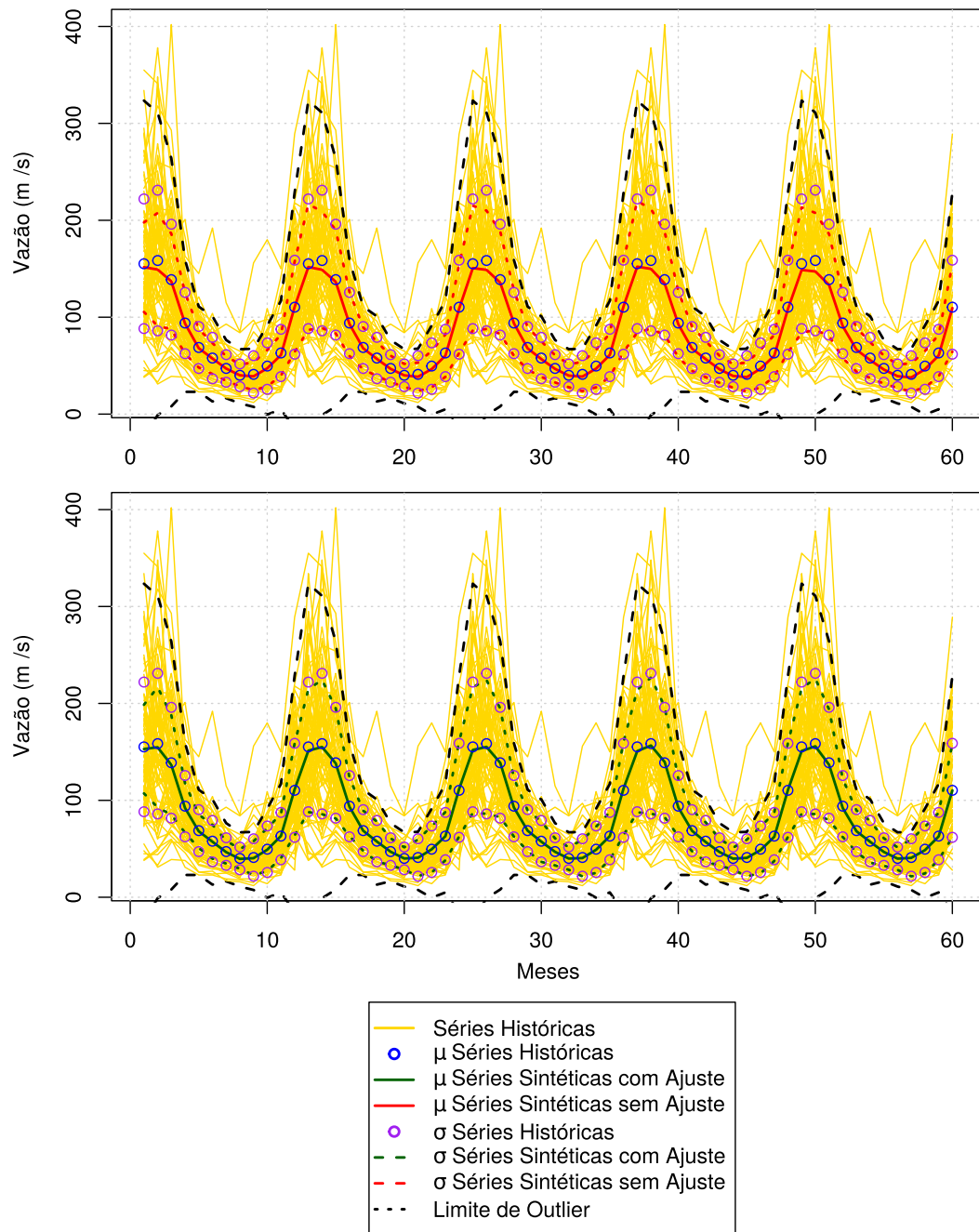


Figura 4.1 – Influência do Fator de Ajuste nas Séries Produzidas para Usina Euclides da Cunha

Fonte: Elaboração Própria.

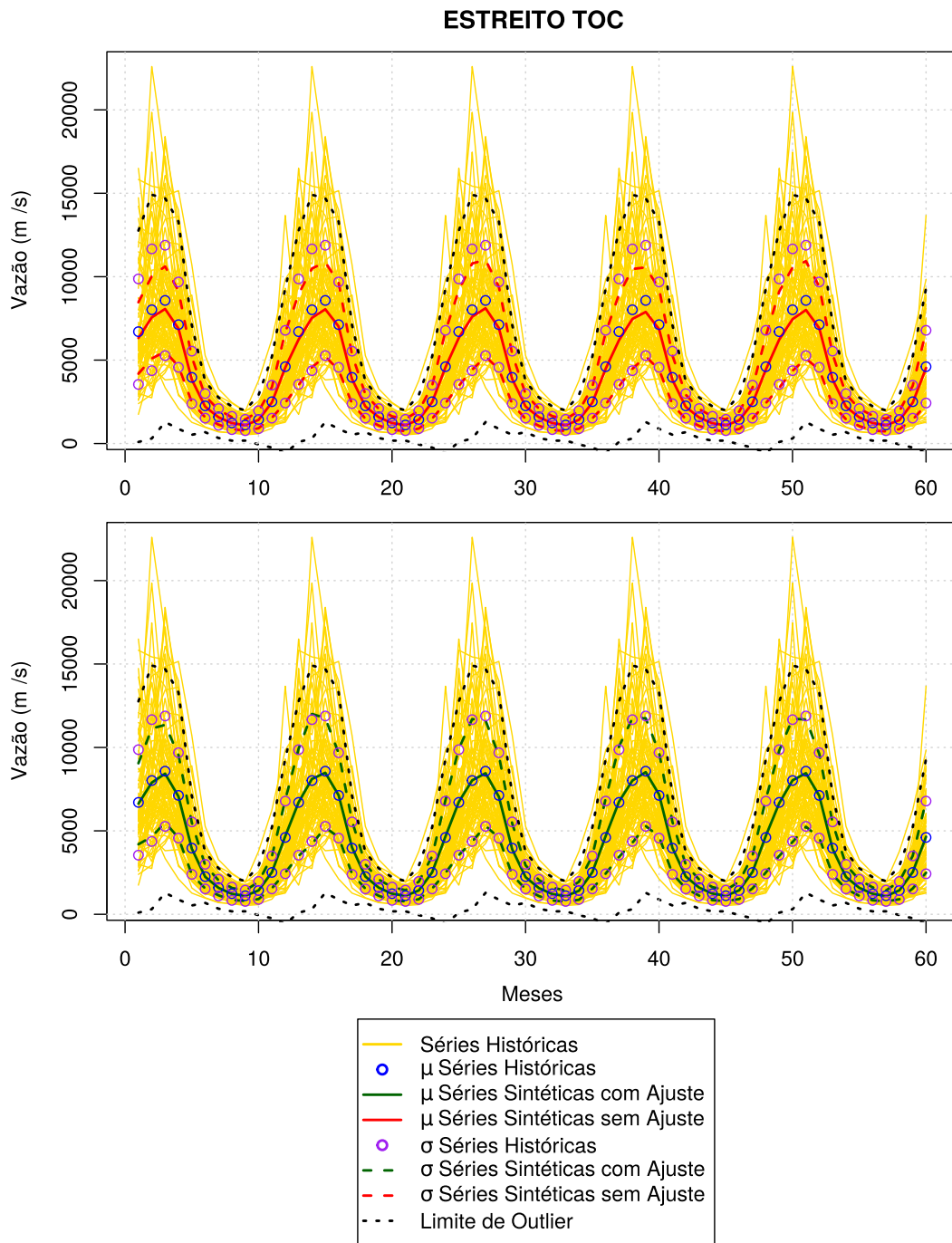


Figura 4.2 – Influência do Fator de Ajuste nas Séries Produzidas para Usina Estreito Tocantins

Fonte: Elaboração Própria.

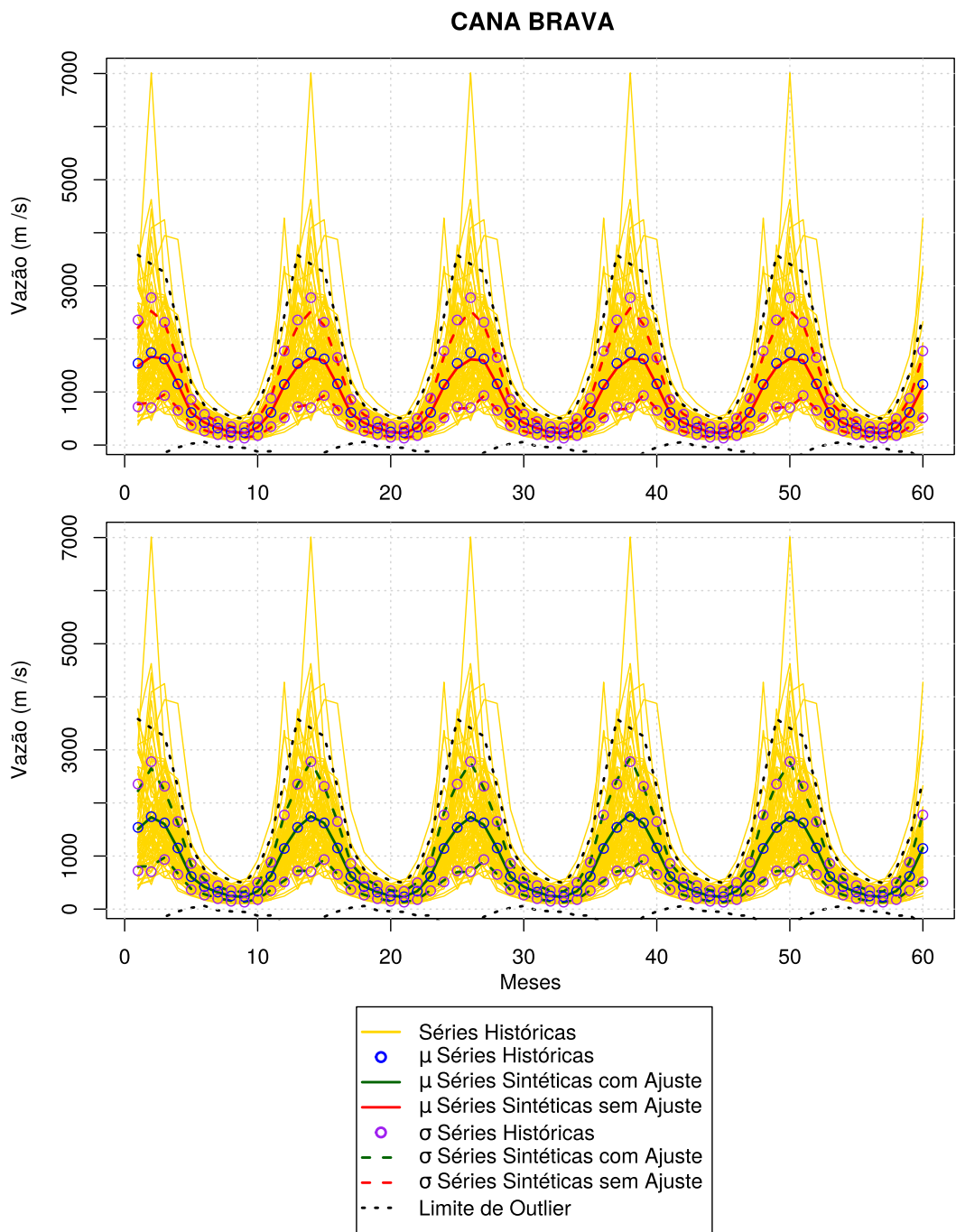


Figura 4.3 – Influência do Fator de Ajuste nas Séries Produzidas para Usina Cana Brava

Fonte: Elaboração Própria.

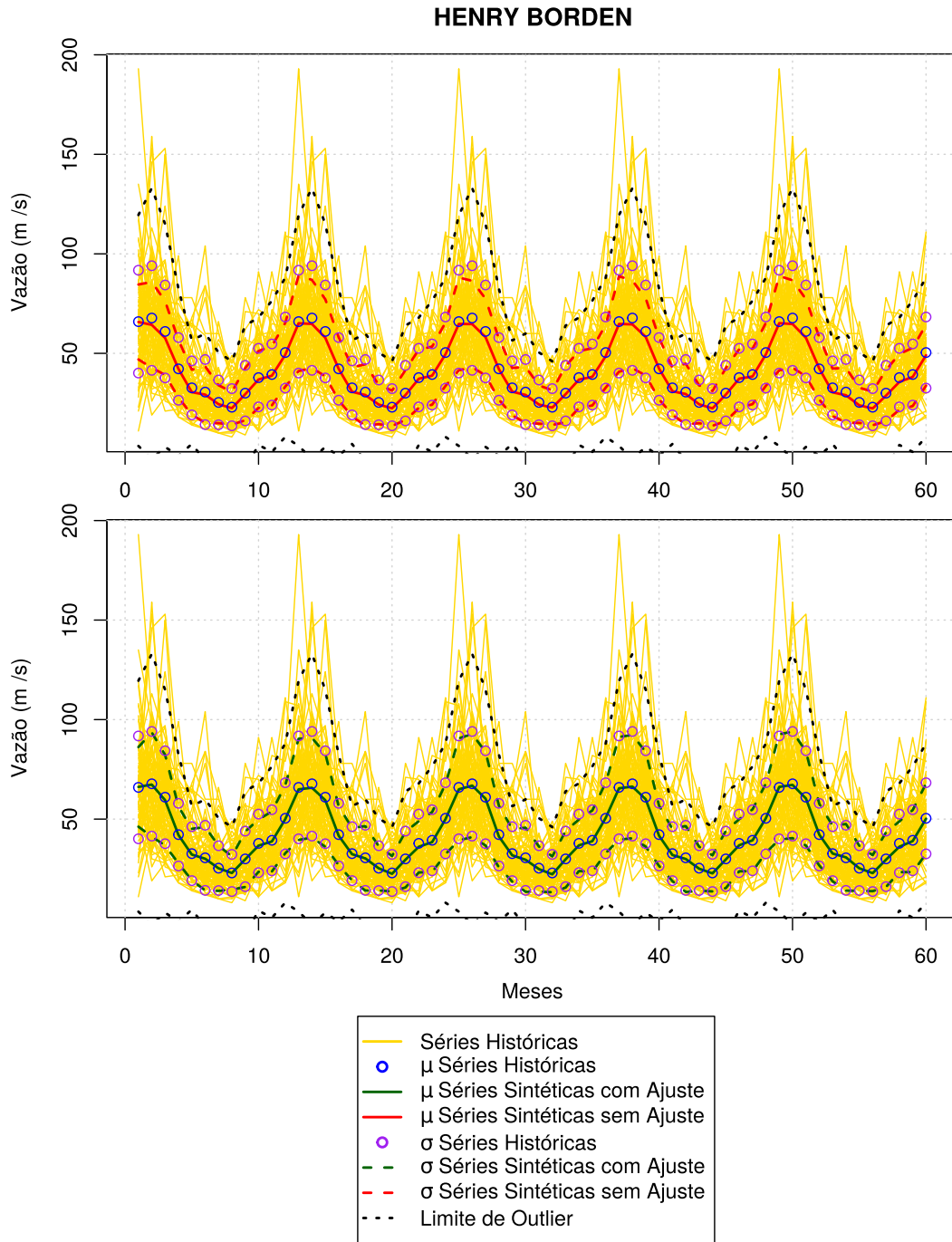


Figura 4.4 – Henry Borden

Fonte: Elaboração Própria.

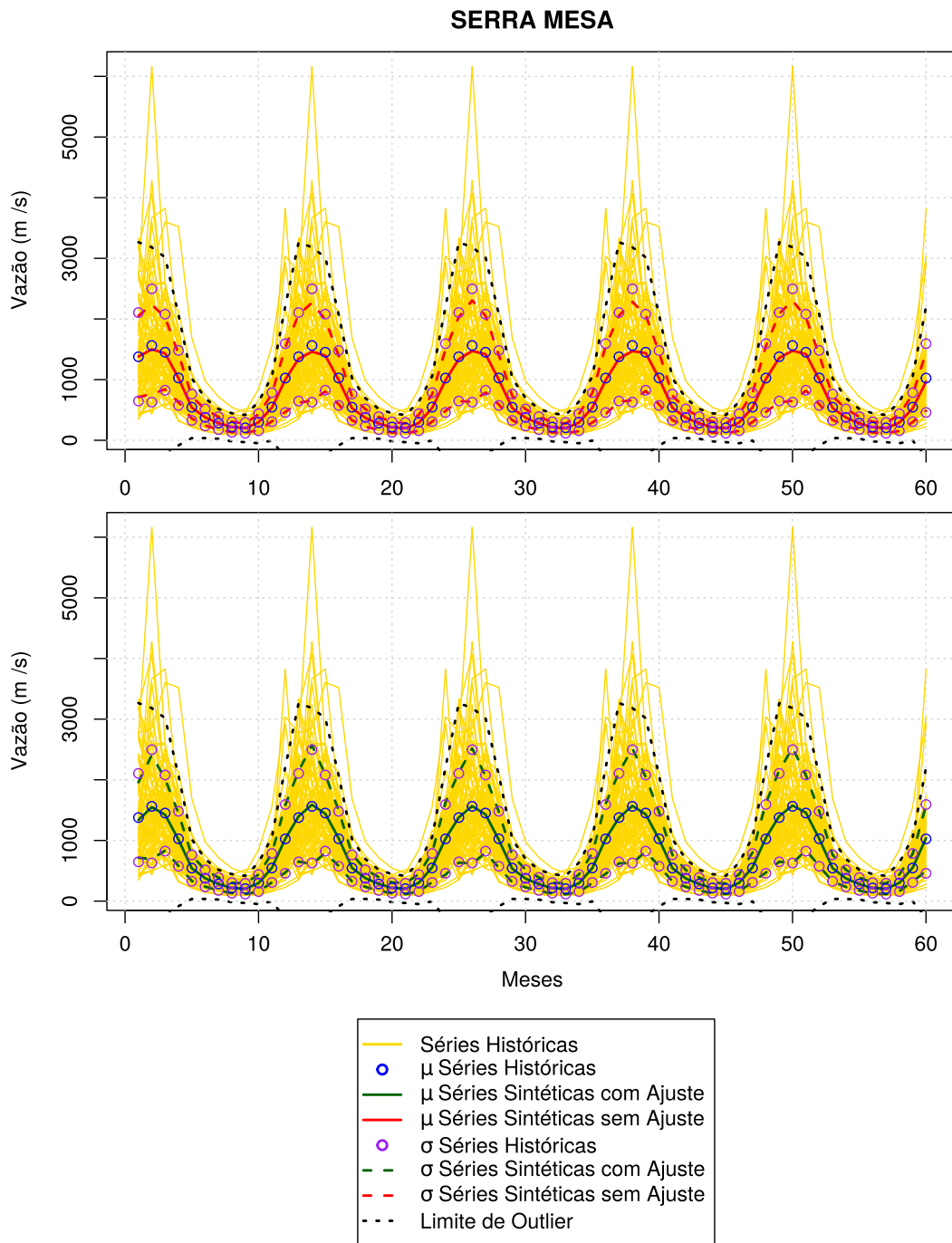


Figura 4.5 – Influência do Fator de Ajuste nas Séries Produzidas para Usina Serra da Mesa

Fonte: Elaboração Própria.

Tabela 4.1 – Fatores de Ajuste Aplicados

	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
14 DE JULHO	1	1.084	1.121	0.929	1	1.028	1	0.974	1.07	1.083	1	1
A. VERMELHA	1	1	1	1	1	1	1	1	1.038	1.026	1	1
A.A. LAYDNER	1.041	1	1	1	1	1.071	0.965	1	1	1	1	1.028
A.S. LIMA	1	1	1	1	1.027	1.073	0.955	1	1.025	1	1	1
A.S.OLIVEIRA	1	1.047	1	0.971	1	1	1	1	1.045	1	1.027	1
AIMORES	1	1.073	0.967	1	1	1	1	1	0.966	1	1	1
B. COQUEIROS	1	1	1	1	1	1	1	1	1	1	1	1
B. ESPERANCA	1	1.027	1	1	1	1	1	1	1	1	1	1
BAGUARI	1	1.076	0.962	1	1	1	1	1	1	1	1	1
BAIXO IGUACU	1	1	1	0.975	1.089	1	1	0.834	1.112	0.959	1	1.09
BALBINA	0.958	1	1.032	1	1	1.029	0.973	1	1	1.027	1	1
BARRA BONITA	1	1.03	1	1	1	1.087	0.956	1	1.028	1	1	1
BARRA BRAUNA	1	1.026	0.974	0.973	1	1	1	0.974	1	1	1.046	1
BARRA GRANDE	0.969	1.078	1	1.025	0.908	1.051	1	1.027	0.975	1.07	1.056	0.939
BATALHA	1	1.027	1	1	1	1	1	1	1	1	1	1
BELO MONTE	1	1	1	1	1	1.103	1	1.028	1.091	1.043	1.036	1.048
BILLINGS	1	1.026	1	1	1.044	1	1	1	1	1	1	1
CACH.CALDEIR	1	0.968	1	1	1	1	1	1	1	1	1	1.108
CACH.DOURADA	1	1	1	1	1	1	1	1	1	1	1.026	1
CACONDE	1	1.038	1	1	1	1	1	1	1.039	1	1.03	1
CACU	1	1	1	1	1	1	1	1	1	1	1	1
CAMARGOS	1	1	1	1	1	1	1	1	1.032	1	1	1
CAMPOS NOVOS	0.951	1.048	0.953	1	0.882	1	0.955	0.954	1.035	1	1.084	0.975
CANA BRAVA	1.032	1.026	1	1	1	1	1	1	1	1	1	1.035
CANDONGA	1.039	1	1	1	0.968	1	1	1	1	1	1	1
CANOAS I	1.036	1	0.966	1	1.037	1.044	0.962	1	1	1	1	1
CANOAS II	1	1	0.975	1	1.038	1.045	0.965	1	1	1	1	1
CAPIM BRANC1	1	1	1	1	1	1	1	1	1	1.036	1	1
CAPIM BRANC2	1	1	1	1	1	1	1	1	1	1	1.029	1
CAPIVARA	1	1	1	1	1.11	1.029	0.951	1	1	0.968	1	1
CASTRO ALVES	1	1.077	1	0.909	1	1.038	1	1	1.1	1.047	1.049	0.919
CHAVANTES	1	1	1	1	1	1.058	0.966	1	1	0.974	1	1.028
COARACY NUNE	1	1	1	1	1	1	1	1	1	1	1	1.103
COLIDER	1	1	1	1	1	1	1	1	1	1	1	1
COMP PAF-MOX	1	1.053	1	0.95	1.04	1	1	0.97	1	1	1.031	0.965
CORUMBA I	1	1	1	1	1	1	1	1	1	1	1	1
CORUMBA III	1	1.042	1	1	1	1	0.972	1	1	1.034	1	1
CORUMBA IV	1	1.031	1	1	1	1	0.972	1	1	1.034	1	1
CURUA-UNA	1	1	1	1	1	1	0.974	1	1	1	1	1.026
D. FRANCISCA	0.936	1	1	1	1	1	0.971	1	1.066	1	1.129	1.084
DARDANELOS	1	1	1	1	1	1	1	1	1	1	1	1
E. DA CUNHA	1	1.047	1	0.971	1	1	1	1	1.047	1	1	1
EMBORCACAO	1	1.038	1	1	0.969	1	1	1	1	1	1.049	1
ERNESTINA	1.074	1.044	0.967	0.943	1.047	1.054	0.938	1.103	0.968	1.098	1.06	1
ESPORA	1	1	1	1	1	1	1	1	1	1	1	1

Usina	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
ESTREITO TOC	1.053	1.029	1	1	1	1	1	1	1	1	1	1
ESTREITO	1	1	1.033	0.971	1	1	1	1	1	1.054	1	1
FERREIRA GOM	1	0.968	1	1	1	1	1	1	1	1	1	1.109
FICT.CANA BR	1.032	1.026	1	1	1	1	1	1	1	1	1	1.035
FICT.IRAPE	1.036	1.056	1	1.065	1	1	1	1	0.936	0.975	1	1.031
FICT.LAJEADO	1.054	1	1	1	1	1	1	1	1	1	1	1
FICT.MAUA	1.131	0.966	1	1	1.049	1.039	1.089	0.958	1	1	1	0.965
FICT.PEIXE A	1.049	1	0.975	1	1	1	1	1	1	1	1	1.034
FICT.QUEIMAD	1	1	1	1	0.96	1	1	1	1	1	1	1
FICT.RETIRO	1	1	1	1	0.974	1	1	1	1	1	1	1
FICT.SAO SAL	1.035	1	1	1	1	1	1	1	1	1	1	1.039
FICT.SERRA M	1	1.054	1	1	1	1	1	0.973	1	1	1	1.034
FICT.TRES MA	1.05	1	1	1	0.974	1	1	1	1	1	1.029	1
FONTES	1	1	1	1	1	1.196	0.931	0.946	1.048	1.134	0.92	1
FOZ CHAPECO	1	1	1.06	1	0.934	1	0.962	1	0.963	1	1.07	1
FOZ R. CLARO	1	1	1	1	1	1	1	1	1	1	1	1
FUNDAO	1	0.958	1	1.061	1	0.947	1	0.83	1	1	0.924	1.093
FUNIL-GRANDE	1	1	1	1	1	1	1	1	1	1	1	1
FUNIL	1	1	1	1	1	1	1	1	1	1	1	1
FURNAS	1	1	1	1	1	1	1	1	1.034	1.065	1	0.969
G.B. MUNHOZ	0.952	1	0.967	1.039	1.111	0.936	1	0.955	1.09	1	1	1.055
G.P. SOUZA	1	1.04	1.025	0.972	1.038	1	1	1	1.028	0.971	1	1
GARIBALDI	0.951	1.05	0.944	1	0.904	1	0.94	0.958	1.035	1	1.098	0.962
GUAPORE	1	1	1	1	1	1	1	1	1	1	1	1
GUARAPIRANGA	1	1	1	1	1	1.043	1	1.036	0.922	0.964	1.078	0.951
GUILMAN-AMOR	1	1.039	1	1	1	1	1	0.968	1	1	1	1
HENRY BORDEN	1	1.029	1	1	1.046	1	1	1	1	1	1	1
I. SOLTEIRA	1	1	1	1	1	1	1	1	1	1	1.032	1
IBITINGA	1	1	1	1	1	1.095	0.959	1	1	1	1	1
IGARAPAVA	1	1	1.032	1	1	1	1	1	1	1.045	1	1
ILHA POMBOS	1	1	1	1	1	1	1	1	1.027	1	1	1
IRAPE	1.036	1.056	1	1.065	1	1	1	1	0.936	0.975	1	1.031
ITA	0.948	1.049	1.058	1	0.894	1	1	1	0.955	1	1.056	1.082
ITAIPU	1	1	1	1	1	1.033	1	1	1	1	1	1
ITAPARICA	1	1.043	1	0.954	1.041	1	0.97	1	1	0.969	1.041	0.96
ITAPEBI	1.111	1.039	1	1.051	1	1	0.941	1	1	1	1.059	1.026
ITAUBA	0.954	1	1	1	1	1	1	1	1.033	1	1.109	1.073
ITIQUIRA I	1	1	1	1	1	1	1	1	1	1	1	1
ITIQUIRA II	1	1	1	1	1	1	1	1	1	1	1	1
ITUMBIARA	1	1	1	1	1	1	1	1	1	1	1	1.032
ITUTINGA	1	1	1	1	1	1	1	1	1.032	1	1	1
JACUI	0.941	1.035	1	1.1	1	0.968	1	1	1.028	1	1.092	1.077
JAGUARA	1	1	1.033	0.971	1	1	1	1	1	1.054	1	1
JAGUARI	1	1	1	1	1.036	1	1	1	1.044	1	1	1
JAURU	1	1	1	1	1	1	1	1	1	1	1	1
JIRAU	1	1	1	1	1	1	1.039	1	1	1	0.964	1

Usina	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
JORDAO	1.826	1.778	0.654	1.274	1.615	1.707	1.326	0.689	1.468	0.813	1.184	1.151
JUPIA	1	1	1	1	1	1	1	1	1.027	1	1	1
L.N. GARCEZ	1	1	0.974	1	1.037	1.044	0.967	1	1	1	1	1
LAJEADO	1.054	1	1	1	1	1	1	1	1	1	1	1
LAJES	1	1	1	1	1	1	1	1	1	1	1	1
M. DE MORAES	1	1	1.034	0.971	1	1	1	1	1.025	1.047	1	1
MACHADINHO	0.957	1.048	0.969	1.048	0.896	1.048	0.97	1.029	1	1.026	1.05	1
MANSO	1	1	1	1	1	1	1	1	1	1.036	1	1
MARIMBONDO	1	1	1	1	1	1	1	1	1.031	1	1.027	1
MASCARENHAS	1	1.069	1	1	1	0.972	1	1	1	1	1	1
MAUA	1.131	0.966	1	1	1.049	1.039	1.089	0.958	1	1	1	0.965
MIRANDA	1	1	1	1	1	1	1	1	1	1.038	1	1
MONJOLINHO	1.026	1	1	1.08	1.029	1.029	0.91	1	0.937	1.078	1.068	1.109
MONTE CLARO	1.026	1.076	1.117	0.923	1	1.033	0.973	1	1.07	1.083	1	1
NAVANHANDA	1	1.025	1	1	1	1.068	0.969	1	1	1	1	1
NILO PECANHA	1	1	1	1	1	1	1	1	1	1	1	1
NOVA PONTE	1	1	1	1	1	1	1	1	1	1.034	1	1
OURINHOS	1	1	0.972	1	1.037	1.041	0.965	1	1	0.974	1	1.028
P. CAVALO	1.033	0.749	0.952	1.142	0.86	0.97	1	0.94	1	1.214	1.045	1.209
P. COLOMBIA	1	1	1.038	1	1	1	1	1	1	1	1	1
P. ESTRELA	1	1.085	1	1	1	1	1	1	0.958	1.03	1	1
P. PASSOS	1	1	1	1	1	1	1	1	1	1	1	1
P. PRIMAVERA	1	1	1	1	1	1	1	1	1	1	1	1
PARAIBUNA	1	1	1	1	1	1	1	1	1	1	1	1
PASSO FUNDO	1	1.04	1	1.06	1.076	1	0.938	1	0.967	1	1.064	0.964
PASSO REAL	0.942	1.036	1	1.1	1	0.967	1	1	1.028	1	1.092	1.075
PASSO S JOAO	0.938	0.938	1	1.083	1.034	1.128	1	1	1.059	1.131	1.071	1.045
PEIXE ANGIC	1.049	1	0.975	1	1	1	1	1	1	1	1	1.034
PICADA	1.033	0.971	1	1	1	1	1	1	1	1	1	1
PIMENTAL	1.11	1.206	0.97	1.2	1	1	1	1	1	1	1	1
PIRAJU	1.039	1	1	1	1	1.071	0.973	1	1	1	1	1.026
PONTE PEDRA	1	1	1	1	1	1	1	1	1	1	1	1
PROMISSAO	1	1	1	1	1	1.085	0.972	1	1	1	1	1
QUEBRA QUEIX	1	1	0.945	1	1.066	0.942	1.044	1	0.966	1.026	1	1.164
QUEIMADO	1	1	1	1	0.96	1	1	1	1	1	1	1
RETIRO BAIXO	1	1	1	1	0.974	1	1	1	1	1	1	1
RONDON II	1	1	1	1	1	1	1	1	1	1	1	1
ROSAL	1.035	1	1	1	1	1	1	0.975	1	1	1	1.031
ROSANA	0.973	1	1	1	1.12	1	1	0.951	1	1	1	1
SA CARVALHO	1	1.04	1	1	1	1	1	0.966	1	1	1	1
SALTO CAXIAS	1	1	1	1	1.075	1	1	0.835	1.113	0.96	1	1.09
SALTO GRANDE	1	1.086	1	1	1	1	1	1	0.959	1.03	1	1
SALTO OSORIO	1.052	1	1	1	1.129	0.954	1	0.86	1.096	0.964	1	1.061
SALTO PILAO	1.063	1.042	1	1.031	0.912	1.027	0.967	1.054	1	1	1.073	0.957
SALTO	1	1	1	1	1	1	1	1	1	1	1	1
SAMUEL	1	1	1	1	1	1	1.031	1.032	1.029	1.036	1.077	0.974

Usina	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
SANTA BRANCA	1	1	1	1	1	1	1	1	1	1	1	1
SAO DOMINGOS	1	1	1	1	1	1	1	1	1	1	1	1
SAO JOSE	0.945	0.939	0.97	1.091	1	1.156	0.972	1.028	1.053	1.13	1.074	1.029
SAO MANOEL	1	1	1	1	1	1	1	1	1	1.055	1.036	1
SAO ROQUE	0.974	1.037	1	0.962	0.916	1	0.88	1	1.043	1.056	1.09	0.9
SAO SALVADOR	1.035	1	1	1	1	1	1	1	1	1	1	1.039
SAO SIMAO	1	1	1	1	1	1	1	1	1	1	1	1.049
SEGREDO	1.04	1	1	1	1.052	1	1	0.924	1.093	1	1	1.057
SERRA FACAO	1	1.036	1	1	1	0.974	1	1	1	1	1.031	1
SERRA MESA	1	1.054	1	1	1	1	1	0.973	1	1	1	1.034
SIMPLICIO	1	1	1	1	1	1.028	1.027	1	1.035	1	1	1
SINOP	1	1	1	1	1	1	1	1	1	1	1	1
SLT VERDINHO	1	1	1	1	1	1	1	1	1	1	1	1
SLT.SANTIAGO	0.948	1	1	1	1.094	1	1	0.863	1.098	0.951	1	1.061
SOBRADINHO	1	1.055	1	0.955	1.04	1	0.972	1	1	0.97	1.026	0.971
SOBRAGI	1.037	1	1	0.975	1	1	1	1	1	1	1	1
STA CLARA MG	1.134	1	1	1	0.973	1	1	1	1	1.049	1.064	1.026
STA CLARA PR	1	0.957	1	1.063	1	0.948	1	0.829	1	1	0.925	1.094
STO ANT JARI	1	1.042	1	1	1	1	1	1	1	1	1.077	0.961
STO ANTONIO	1	1	1	1	1	1	1.038	1	1	1	0.963	1
SUICA	1	1.061	0.963	0.969	1	1	1	1	1	1.027	1	1
TAQUARUCU	1	1	0.975	1	1.123	1.03	0.958	0.957	1	1	1	1
TELES PIRES	1	1	1	1	1	1	1	1	1	1.055	1.036	1
TRES IRMAOS	0.973	1.027	1	1	1	1.082	0.969	1	1	1	1	1
TRES MARIAS	1.05	1	1	1	0.974	1	1	1	1	1	1.029	1
TUCURUI	1.038	1	1	1	1	1	1	1	1	1	1	1
VOLTA GRANDE	1	1	1.037	1	1	1	1	1	1	1.027	1	1
XINGO	1	1.053	1	0.95	1.04	1	1	0.971	1	1	1.032	0.964

Fonte: Elaboração Própria.

4.3 Resultados dos Testes Estatísticos

Nesta subseção, os resultados serão apresentados em termos percentuais, indicando a proporção de simulações em que os P-valores excederam o nível de significância proposto de 5%. Nas Tabelas 4.2 e 4.3, estão descritas as taxas de sucesso alcançadas por cada modelo nos testes estatísticos, em relação ao número total de simulações realizadas.

Para a Tabela 4.2, foram realizados testes em 165 usinas, considerando individualmente os 12 meses do ano, além de um teste adicional para as estatísticas globais de cada série, totalizando 2145 simulações por teste. No teste de igualdade

das médias, todos os modelos apresentaram desempenho perfeito, indicando sucesso do fator de ajuste χ_t no ajuste das médias. Nos testes de igualdade das variâncias e de aderência, as taxas de sucesso foram elevadas para todos os modelos, com uma leve vantagem para os modelos multiplicativos, sendo que o modelo proposto ficou ligeiramente atrás do modelo multiplicativo tradicional.

Tabela 4.2 – Taxa de Sucesso nos Testes Estatísticos das Séries

	Médias (%)	Variâncias (%)	Aderência (%)
Box	100	93.94	92.21
Add	100	94.22	92.68
Mult	100	98.14	95.34
Mult New	100	96.41	93.99

Fonte: Elaboração Própria.

Na Tabela 4.2, são apresentados os resultados dos testes de sequências negativas e positivas. Ao analisar o teste de sequência negativa, na variável soma, os modelos com erro aditivo superaram os modelos com erro multiplicativo. Em contrapartida, na variável intensidade, os modelos multiplicativos, especialmente o proposto, apresentaram desempenho significativamente superior, sugerindo uma tendência da modelagem em capturar melhor essa característica. No entanto, nas sequências positivas, observou-se o oposto, com o modelo proposto exibindo a pior taxa de aderência entre todos os modelos para ambas as variáveis, mesmo que ainda sejam números altos. Normalmente, a robustez contra *outliers* é uma característica desejada no entanto, como os *outliers* afetam as estatísticas da série histórica, isto afeta diretamente a capacidade de reprodução da soma e duração das sequências positivas das vazões sintéticas. Esse desempenho inferior aos concorrentes não se repetiu para a sequência negativa devido à presença de valores extremos superiores ser mais comum que os inferiores.

Tabela 4.3 – Taxa de Sucesso nos Testes Estatísticos das Sequências

	Sequência Negativa		Sequência Positiva	
	Soma (%)	Intensidade (%)	Duração (%)	Soma (%)
Box	95.15	64.85	96.97	91.52
Add	95.76	67.27	96.36	92.73
Mult	80.61	79.39	85.45	93.33
Mult New	90.3	90.91	73.94	82.42

Fonte: Elaboração Própria.

4.4 Média e Desvio Padrão das Séries Sintéticas

Conforme discutido anteriormente, a produção de valores exclusivamente não negativos juntamente com a proximidade subjetiva entre os valores de média e desvio padrão das séries sintéticas e das séries históricas foram adotados como critérios fundamentais para a aprovação dos cenários gerados. Esses critérios visam garantir que os cenários sintéticos sejam representativos das séries históricas e proporcionem uma análise mais precisa e realista. Com isso em mente, o modelo proposto atingiu seu objetivo de maneira satisfatória, conforme evidenciado pelas Figuras 4.6, 4.7, 4.8, 4.9, 4.10, 4.11, 4.12, 4.13, 4.14, 4.15, 4.16, 4.17, 4.18, que ilustram a geração de cenários bem sucedida.

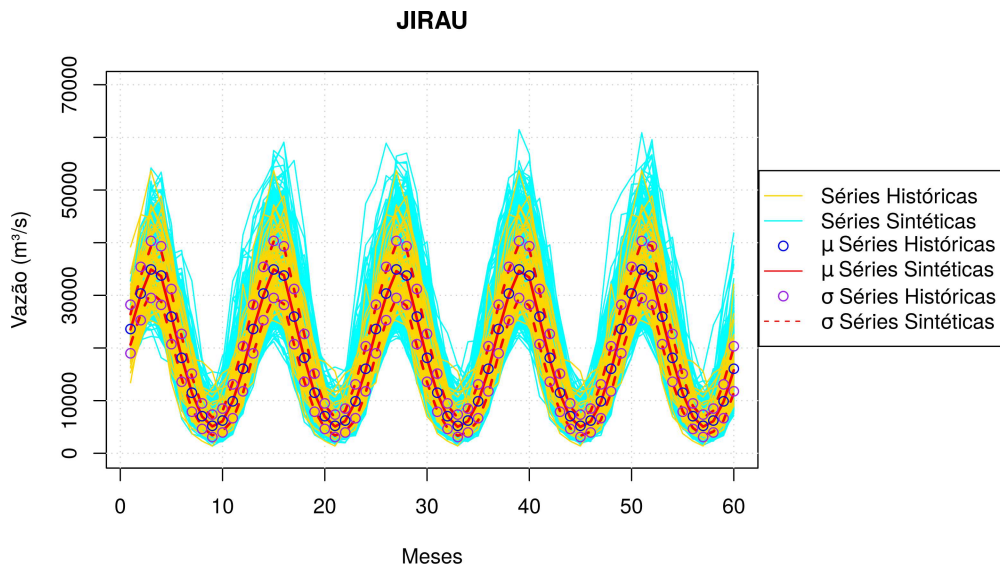


Figura 4.6 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Jirau

Fonte: Elaboração Própria.

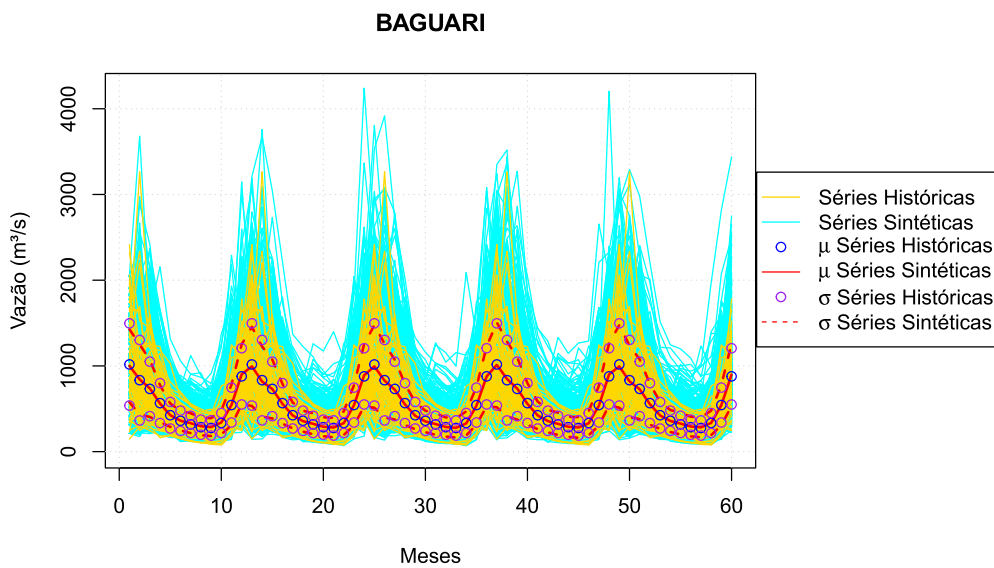


Figura 4.7 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Baguari

Fonte: Elaboração Própria.

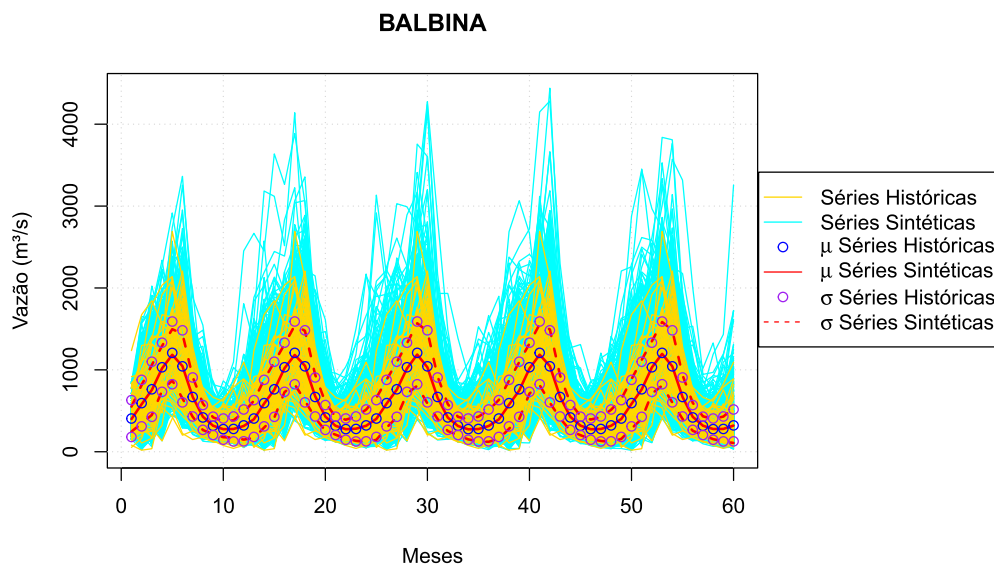


Figura 4.8 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Balbina

Fonte: Elaboração Própria.

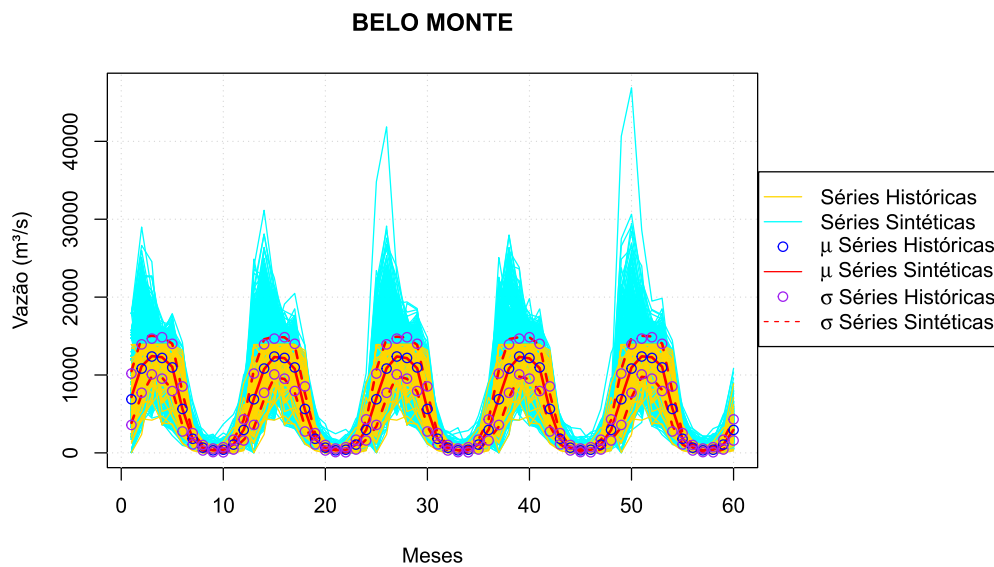


Figura 4.9 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Belo Monte

Fonte: Elaboração Própria.

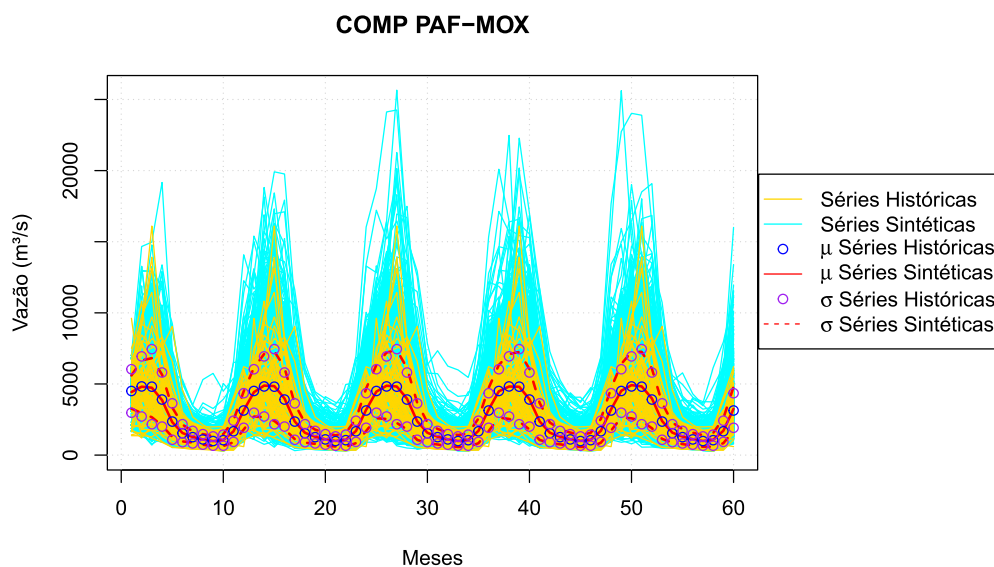


Figura 4.10 – Séries Sintéticas Produzidas pelo Modelo Proposto para as Usinas do Complexo Paulo Afonso - Moxotó

Fonte: Elaboração Própria.

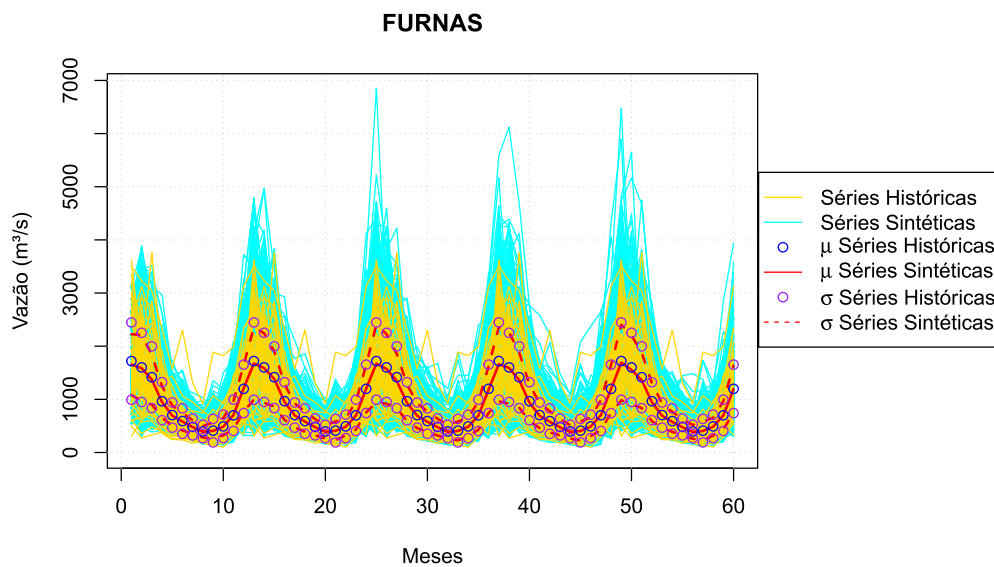


Figura 4.11 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Furnas

Fonte: Elaboração Própria.

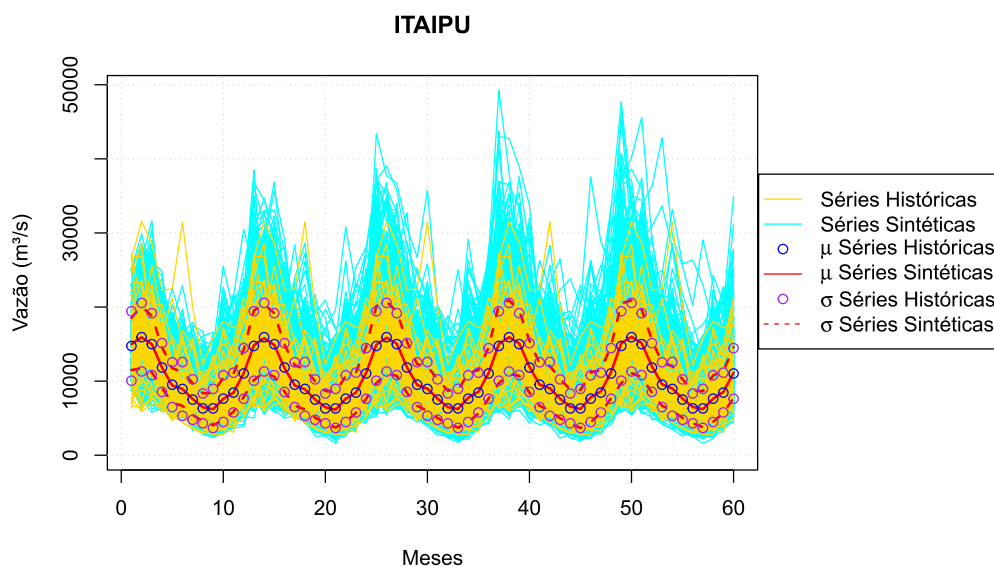


Figura 4.12 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Itaipu

Fonte: Elaboração Própria.

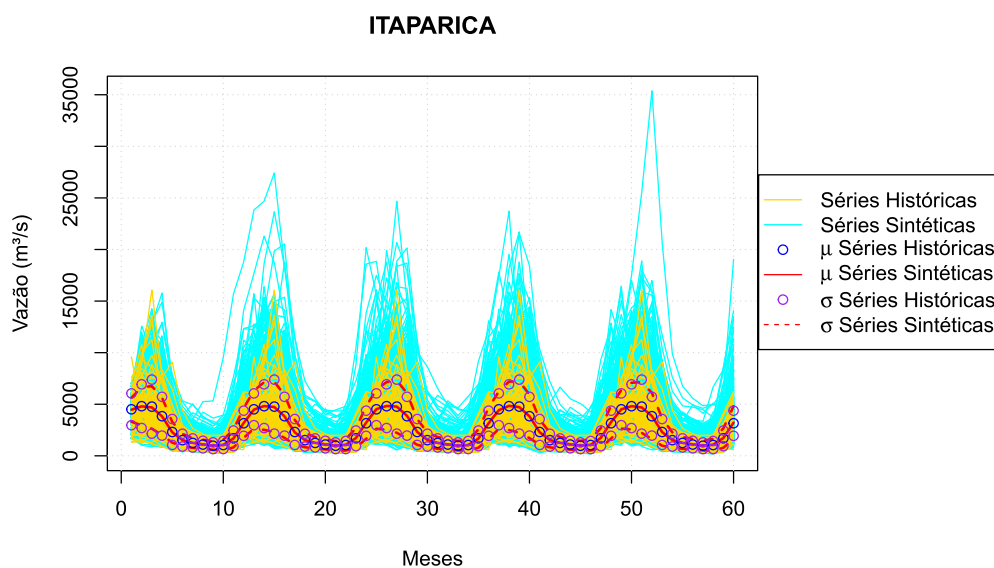


Figura 4.13 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Itaparica

Fonte: Elaboração Própria.

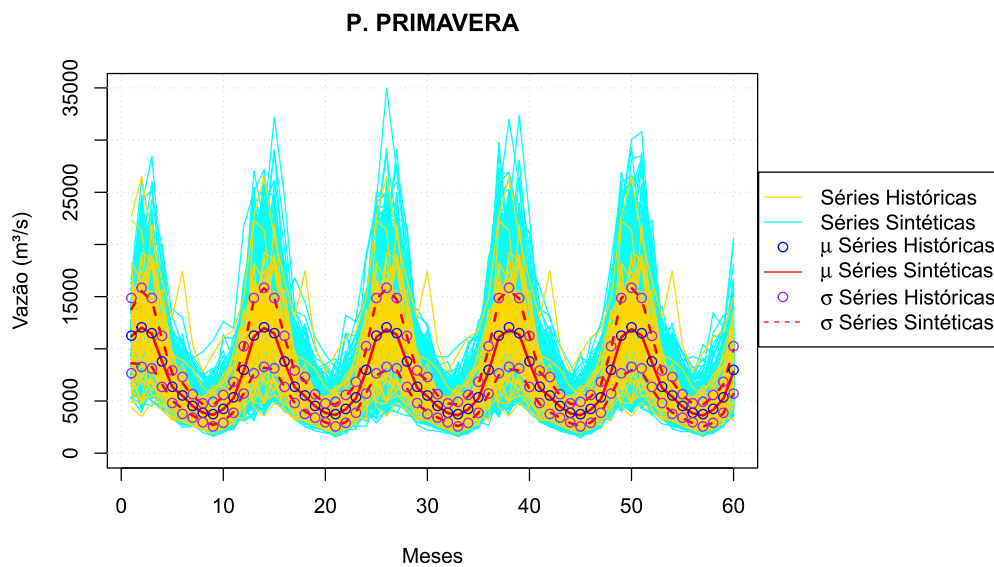


Figura 4.14 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Porto Primavera

Fonte: Elaboração Própria.

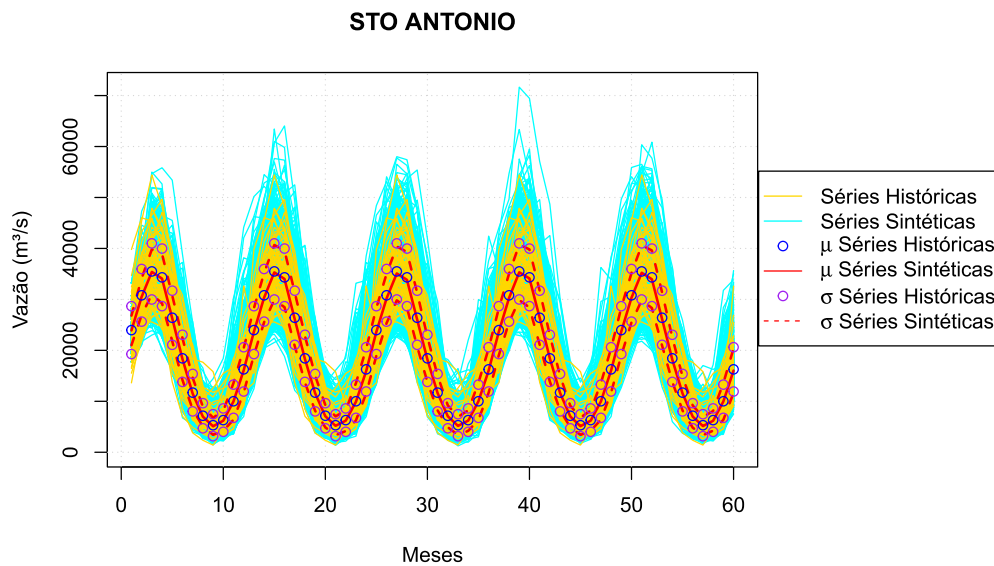


Figura 4.15 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Santo Antônio

Fonte: Elaboração Própria.

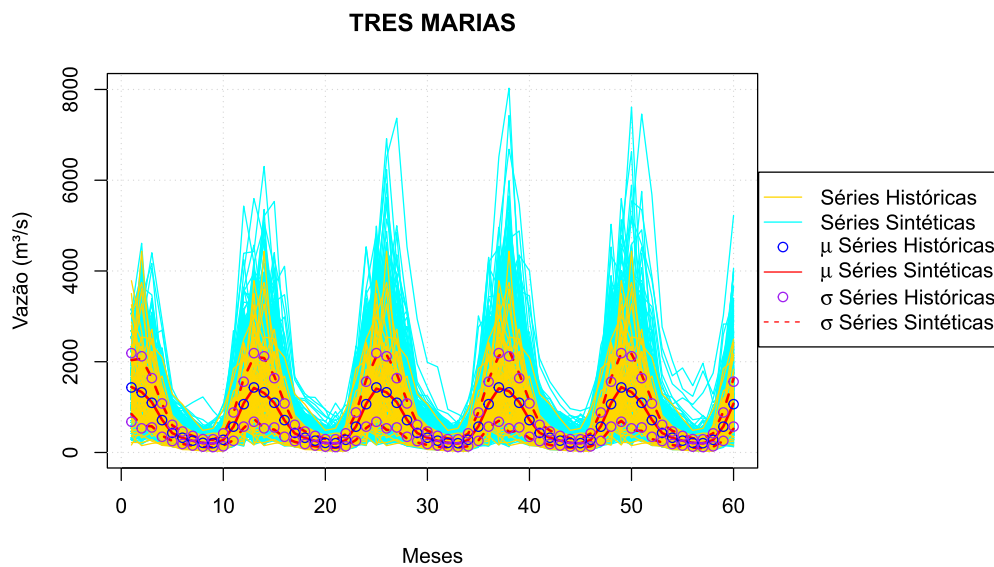


Figura 4.16 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Três Marias

Fonte: Elaboração Própria.

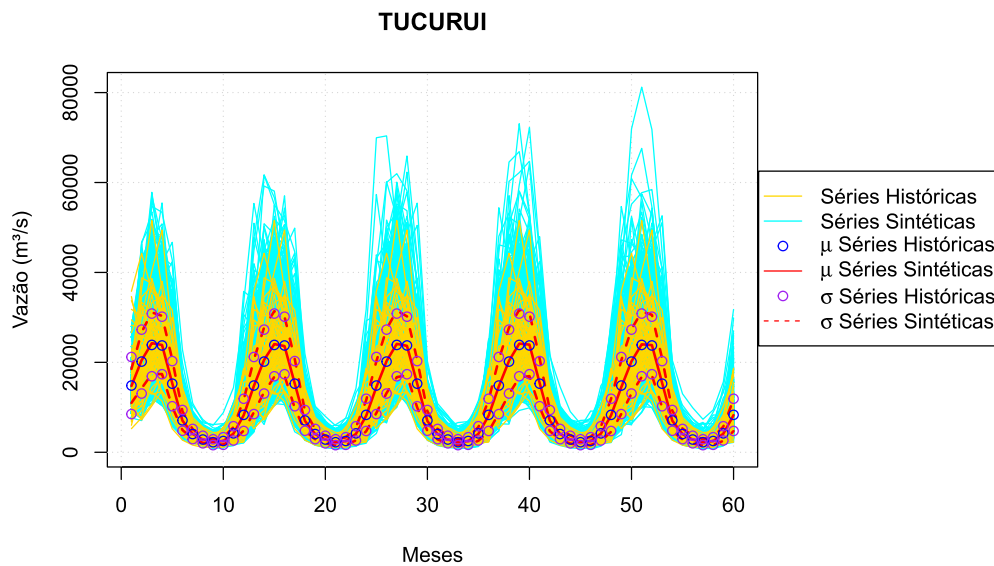


Figura 4.17 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Tucuruí

Fonte: Elaboração Própria.

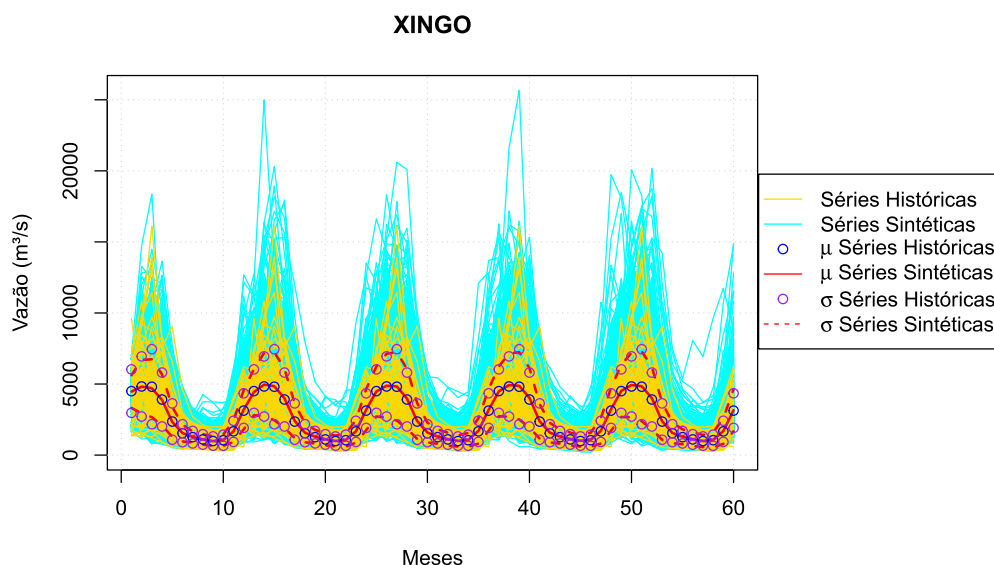


Figura 4.18 – Séries Sintéticas Produzidas pelo Modelo Proposto para Usina Xingó

Fonte: Elaboração Própria.

Devido ao elevado número de usinas, a análise individual de gráficos seria inviável então, para facilitar a interpretação dos resultados gerais e tornar a comparação entre os diferentes modelos mais clara e intuitiva, os dados serão apresentados por gráficos que agrupam os comportamentos das séries geradas. Essa abordagem gráfica facilita a visualização de tendências e padrões, permitindo uma análise comparativa mais eficaz entre os modelos..

O primeiro gráfico, exibido na Figura 4.19, é um gráfico de barras ilustrando a tendência média de erro relativo que cada modelo apresentou em cada mês e globalmente. Para sua elaboração, foram calculadas as médias dos cenários sintéticos e históricos para uma usina específica, em um determinado mês e para um modelo particular (caso). Em seguida, foi determinado o erro relativo entre essas médias da seguinte forma:

$$\varepsilon_{Rel} = \frac{\mu_{sint} - \mu_{hist}}{\mu_{hist}} \quad (4.1)$$

O processo foi então repetido para as demais usinas dentro do mesmo caso e mês, resultando em um conjunto de erros relativos de cada usina, denominado

$\mathbf{C} = \{\varepsilon_{Rel\ Usina\ 1}, \varepsilon_{Rel\ Usina\ 2}, \dots\}$. A média desse conjunto \mathbf{C} representa uma das observações apresentadas no gráfico da Figura 4.19. Outras observações foram obtidas aplicando o mesmo processo a diferentes períodos e modelos. Adicionalmente, os resultados referentes ao desvio padrão, obtidos pelo mesmo método, estão ilustrados na Figura 4.20.

Pode-se observar que todos os modelos exibem desvios insignificantes nas médias, o que ratifica o resultado perfeito obtido no teste de igualdade das médias. No entanto, ao analisar o desvio padrão, nota-se que o modelo multiplicativo tradicional tende a se desviar mais significativamente do valor histórico em comparação com os outros modelos. Apesar disso, a tendência geral é que os desvios permaneçam baixos.

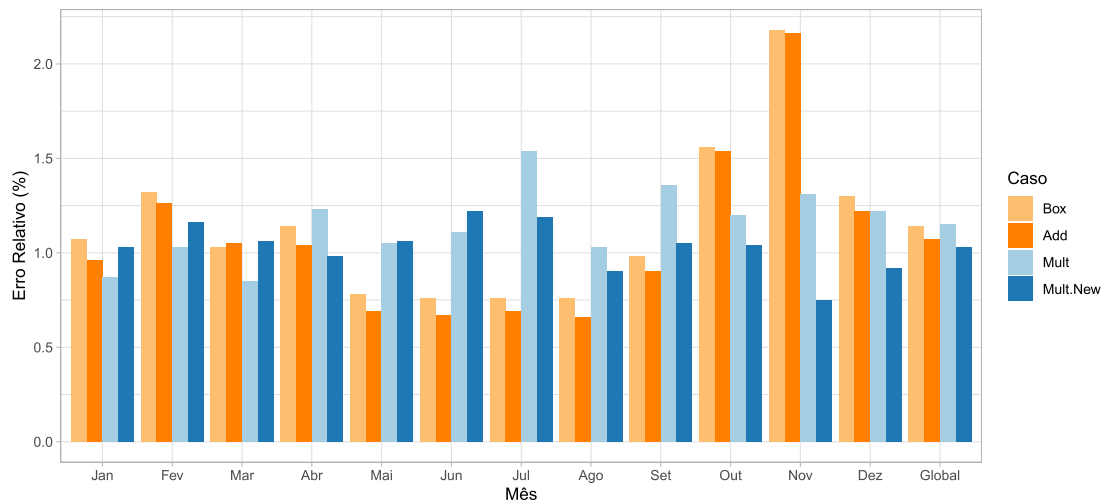


Figura 4.19 – Erros Médios Relativos entre as Médias das Séries Sintéticas e Históricas

Fonte: Elaboração Própria.

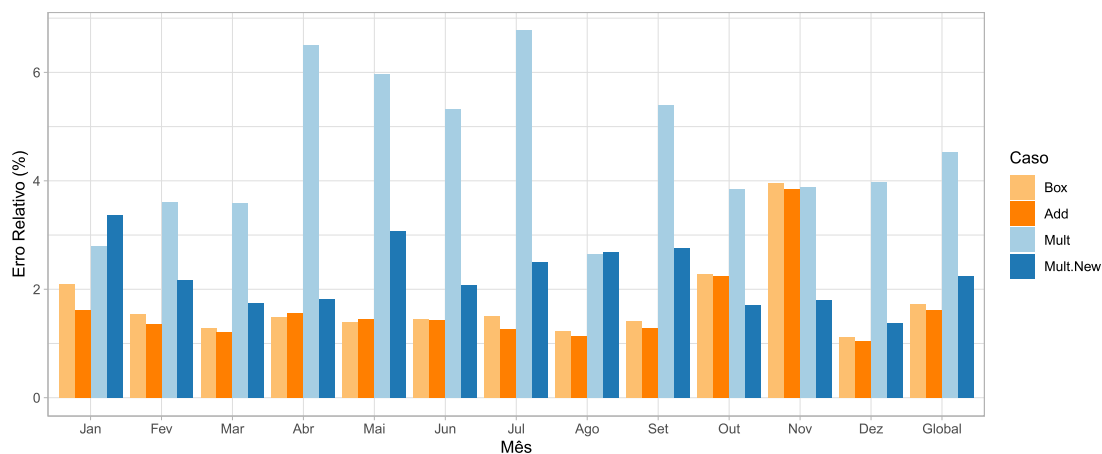


Figura 4.20 – Erros Médios Relativos entre os Desvios Padrão das Séries Sintéticas e Históricas

Fonte: Elaboração Própria.

No entanto, os conjuntos \mathbf{C} podem ser explorados de maneira mais aprofundada para entender melhor o comportamento individual das usinas. Com esse objetivo, foram elaborados gráficos do tipo *Boxplot*, detalhados nas Figuras 4.21 e 4.22, com cada caixa representando um conjunto \mathbf{C} . A primeira observação relevante é a ausência de *outliers* extremos no modelo proposto tanto para a média quanto para o desvio padrão. Os modelos aditivos, ainda que apresentassem menor quantidade de ocorrência de valores extremos em geral, ainda exibiram esses, especialmente nos dois últimos meses do ano.

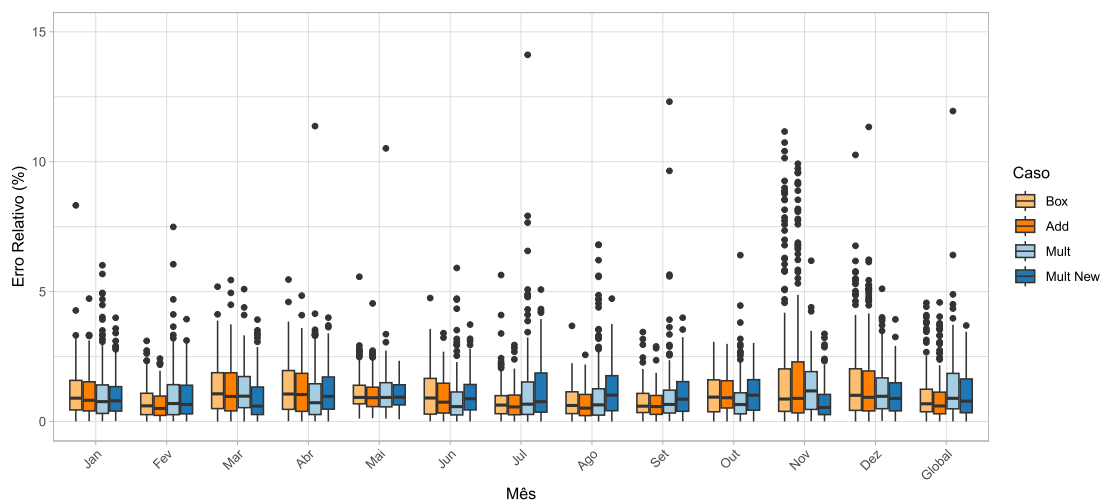


Figura 4.21 – Distribuição dos Erros Relativos entre as Médias das Séries Sintéticas e Históricas

Fonte: Elaboração Própria.

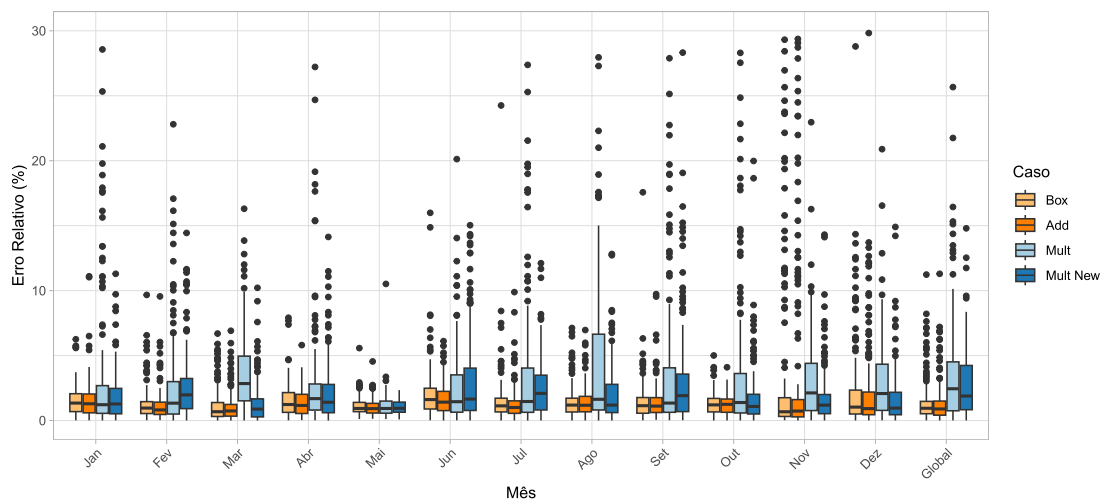


Figura 4.22 – Distribuição dos Erros Relativos entre os Desvios Padrão das Séries Sintéticas e Históricas

Fonte: Elaboração Própria.

Os gráficos 4.21 e 4.22 foram ampliados para uma visualização mais deta-

lhada, conforme mostrado nas Figuras 4.23 e 4.24. Ao analisar os gráficos ampliados, não foram identificadas peculiaridades significativas, além da tendência previamente constatada de que os modelos PAR(p)M costumam apresentar valores ligeiramente superiores de erros no desvio padrão.

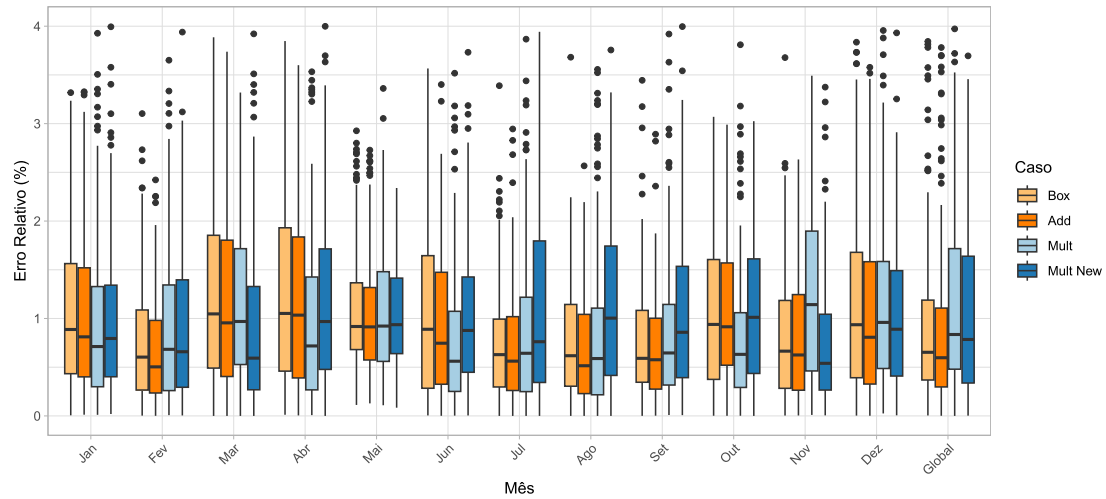


Figura 4.23 – Ampliação da Distribuição dos Erros Relativos entre as Médias das Séries Sintéticas e Históricas

Fonte: Elaboração Própria.

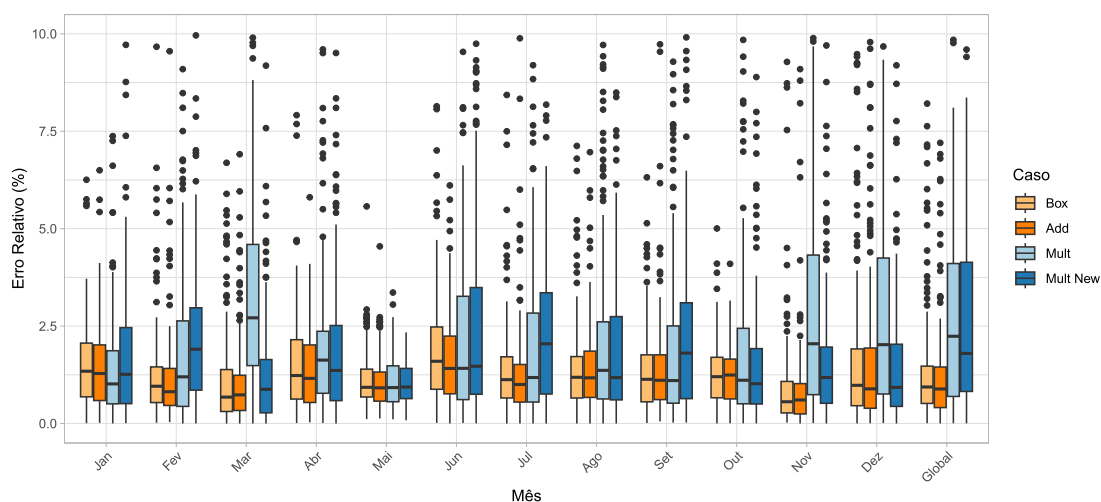


Figura 4.24 – Ampliação da Distribuição dos Erros Relativos entre os Desvios Padrão das Séries Sintéticas e Históricas

Fonte: Elaboração Própria.

4.5 Limitações do Modelo

Embora o modelo proposto tenha demonstrado bons resultados, especialmente após a aplicação do fator de ajuste, a resistência intrínseca aos *outliers* não pôde ser completamente superada em todos os cenários. Verificou-se que, em meses com vazões extremamente elevadas, o desvio padrão das séries sintéticas tende a ser subestimado em comparação com as séries históricas, mesmo quando a média é devidamente ajustada. Esse comportamento é evidente nas Figuras 4.25, 4.26, 4.27, onde, nos meses de junho, as usinas Piraju, Canoas I e Jurumirim (Armando Avellanal Laydner) apresentam vazões naturais afluentes excepcionalmente altas, e a distribuição das séries sintéticas não conseguiu replicar com precisão essas anomalias.

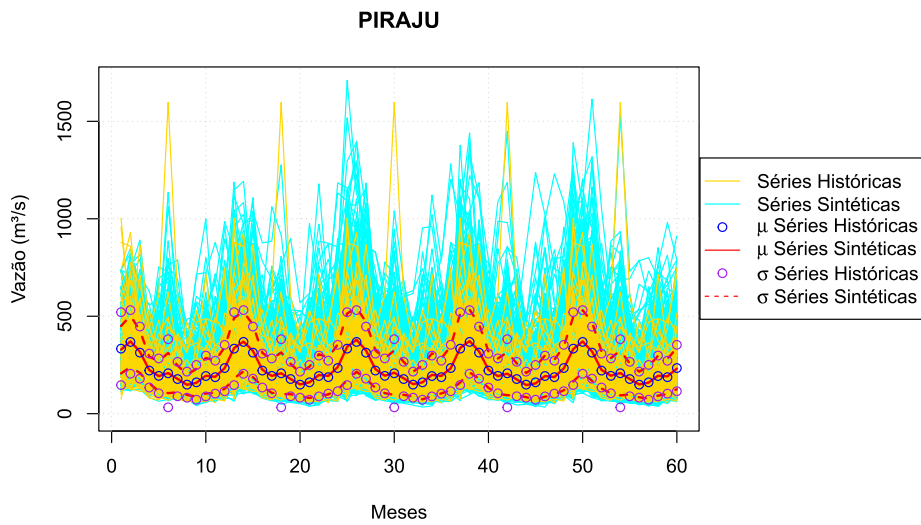


Figura 4.25 – Séries Sintéticas e Históricas Geradas pelo PAR(p)M Tradicional para Usina Jordão

Fonte: Elaboração Própria.

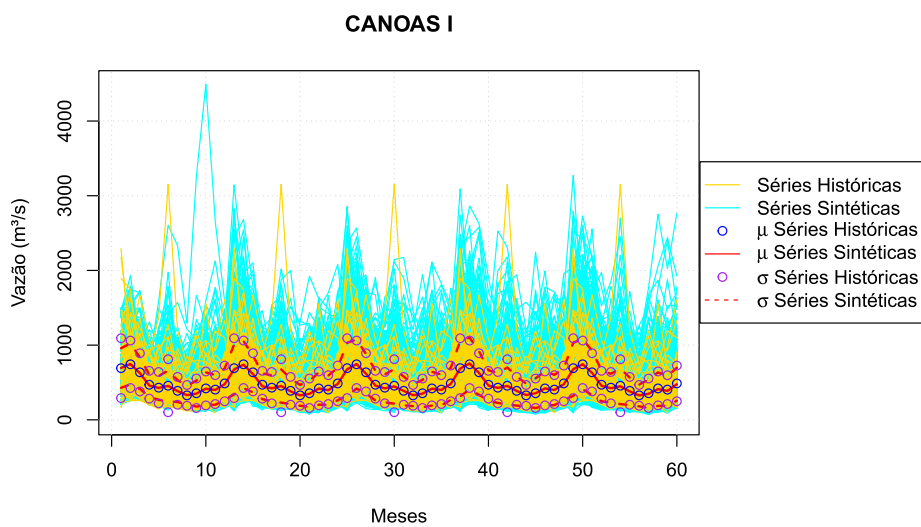


Figura 4.26 – Séries Sintéticas e Históricas Geradas pelo PAR(p)M Proposto para Usina Jordão

Fonte: Elaboração Própria.

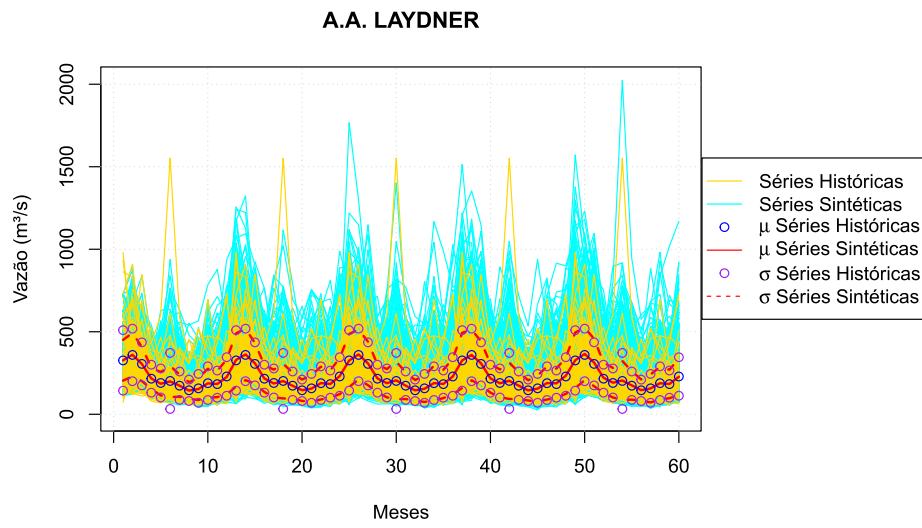


Figura 4.27 – Séries Sintéticas e Históricas Geradas por Box e Jenkins para Usina Jordão

Fonte: Elaboração Própria.

5 CONCLUSÃO E TRABALHOS FUTUROS

Conforme pode ser observado, o modelo PAR(p) com erro multiplicativo proposto demonstrou-se altamente eficaz em alcançar o objetivo de replicar de forma precisa e satisfatória os valores de média e desvio padrão do histórico de vazões naturais afluentes, atingindo desempenho similar aos concorrentes no geral enquanto evita a produção de valores negativos.

O modelo proposto ainda se destacou na replicação da intensidade das sequências negativas em comparação às positivas, considerada estrategicamente mais relevante para o planejamento energético, uma vez que a previsão precisa dos riscos de esvaziamento dos reservatórios é crucial para evitar prejuízos significativos. As implicações de uma limitação nesse aspecto são substanciais, pois, ao subestimar a intensidade de eventos abaixo da média, corre-se o risco de uma percepção inadequada dos riscos envolvidos, o que é particularmente crítico no setor de energia, onde tais eventos podem comprometer tanto a produção quanto a estabilidade do fornecimento.

Ao final, foi possível definir um modelo PAR(p)M com coeficientes estimados por meio de otimização não linear, algo que não teria sido alcançado sem as devidas adaptações no processo. Por um lado, o modelo exigiu ajustes significativos para atingir os resultados desejados, destacando a complexidade inerente ao processo de modelagem; por outro, demonstrou uma flexibilidade notável, permitindo que ele fosse adaptado de maneira eficiente às necessidades específicas do problema em questão. Essa combinação de adaptabilidade e precisão ressalta o potencial do modelo PAR(p)M como uma ferramenta poderosa na análise e previsão de séries temporais, especialmente em contextos onde é necessário equilibrar rigor estatístico com a capacidade de resposta a condições variáveis e desafiadoras.

A flexibilidade demonstrada pelo modelo serve como um ponto de partida significativo para aprimoramentos em trabalhos futuros dentre os quais os seguintes se destacam:

- Implementar e validar o modelo proposto em um cenário multivariado;
- Verificar a convexidade da função objetivo para ordens até $p = 11$;

- Avaliar outras se funções objetivos sem encaixam melhor com o problema como a minimização da dispersão dos erros ou a função de Huber-Loss;
- Avaliar o desempenho na reprodução da assimetria e curtose do modelo;
- Avaliar técnicas de modelagem adaptáveis para garantir o sucesso dos testes estatísticos para todas as usinas.
- Validar a metodologia desenvolvida e explorar seus efeitos em comparação com o uso da distribuição log-normal na resolução do problema de despacho hidrotérmico usando a PDDE.
- Identificar possíveis reconstruções imprecisas das afluências disponíveis para geração de energia elétrica no histórico de vazões naturais afluentes, resultado de um processo de medição com métricas diferentes das atuais.

REFERÊNCIAS

Gabriel Adriano de Melo, Dylan Nakandakari Sugimoto, Paulo Marcelo Tasinaffo, Afonso Henriques Moreira Santos, Adilson Marques Cunha, and Luiz Alberto Vieira Dias. A new approach to river flow forecasting: Lstm and gru multivariate models. *IEEE Latin America Transactions*, 17(12):1978–1986, 2019. doi: 10.1109/TLA.2019.9011542.

Alexandre Almeida, Sabrina Ellian, and José Nobre. Modificações e alternativas aos testes de levene e de brown e forsythe para igualdade de variâncias e médias. *Revista Colombiana de Estatística*, 31:241–260, 2008.

Hugo Ribeiro Baldioti and Reinaldo Castro Souza. Using a markov chain monte carlo technique to simulate synthetic natural inflow energy scenarios. In *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 1–6, 2018. doi: 10.1109/PMAPS.2018.8440277.

T. Bodnar and N. Hautsch. Dynamic conditional correlation multiplicative error processes. *Journal of Empirical Finance*, 36:41–67, 2016.

George E. P. Box, Gwilym M. Jenkins, and Gregory C. Reinsel. *Time Series Analysis: Forecasting and Control*. Wiley, Hoboken, NJ, 4th edition, 2008.

Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004. ISBN 978-0521833783.

D. B. M. Braga. Processos periódicos autorregressivos gama: Uma contribuição para modelagem da energia natural afluyente com possíveis usos na determinação da função de custo futuro. Dissertação de mestrado, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, 2011.

F. G. Cabral. Uma proposta de um modelo periódico multivariado autorregressivo multiplicativo para geração de cenários de afluência aplicável ao modelo de planejamento do setor elétrico brasileiro. Dissertação de mestrado, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil, 2016.

CEPEL. *Manual de Referência do Modelo NEWAVE*. Rio de Janeiro, 2012.

F. Cipollini, R. F. Engle, and G. M. Gallo. Vector multiplicative error models: Representation and inference. *Statistics*, pages 1–53, 2006.

W. U. Conover. *Practical Nonparametric Statistics*. John Wiley & Sons, New York, 1971.

Yasmin Monteiro Cyrillo. Avaliação do modelo pvarm interconfigurações para geração de cenários de ena no planejamento da operação de médio prazo. Dissertação de mestrado, Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2018. URL <https://www.lattes.cnpq.br>. Banca: Reinaldo Castro Souza, Pedro Guilherme Costa Ferreira, Fernando Luiz Cyrino Oliveira.

Cristina Márcia Barros de Castro. *Planejamento energético da operação de médio prazo conjugando as técnicas de PDDE, PAR(p) e Bootstrap*. Tese de doutorado, Universidade Federal de Itajubá, 2012.

EPE. Balanço energético nacional 2023. Technical report, Empresa de Pesquisa Energética, Rio de Janeiro, 2023. URL <https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/balanco-energetico-nacional-2023>. Ano base 2022.

P. G. C. Ferreira. *A Estocasticidade Associada ao Setor Elétrico Brasileiro e uma Nova Abordagem para a Geração de Afluências Via Modelos Periódicos Gama*. Tese de doutorado, Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil, 2013.

José Marcos Leal Barbosa Filho. Sondagem do canal banda larga sem fio: técnicas, simulação e pós-processamento de dados. Dissertação (mestrado em engenharia elétrica e de computação), Centro de Tecnologia, Natal, Brasil, 2020.

E. C. Finardi et al. Investigações de propostas metodológicas nos modelos de precificação para minimizar a volatilidade do preço de liquidação de diferenças: Possíveis problemas nos cortes de benders. Technical report, Tractebel Energia, Labplan (UFSC), Florianópolis, Brasil, 2009.

Damodar N. Gujarati. *Econometria Básica*. Campus, [S.l.], 2006.

Andrew C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, UK, 1990.

N. Hautsch, P. Malec, and M. Schienle. Capturing the zero: A new class of zero-augmented distributions and multiplicative error processes. *Journal of Financial Econometrics*, 12(1):89–121, jan 2014.

K. W. Hipel and A. I. McLeod. *Time Series Modelling of Water Resources and Environmental Systems*. Elsevier, Amsterdam, 1994.

IPOPT. Ipopt documentation, 2024. <https://coin-or.github.io/Ipopt/>.

D. Kraft. A software package for sequential quadratic programming. Technical Report DFVLR-FB 88-28, DLR German Aerospace Center — Institute for Flight Mechanics, Koln, Germany, 1988.

M. Lanne. A mixture multiplicative error model for realized volatility. *Journal of Financial Econometrics*, 4(4):594–616, 2006.

Howard Levene. Robust tests for equality of variances. In I. Olkin et al., editors, *Contributions to Probability and Statistics*, pages 278–292. Stanford University Press, Stanford, CA, 1960.

A. Maass, M. M. Hufschmidt, R. Dorfman, H. A. Thomas Jr., S. A. Marglin, and G. M. Fair. *Design of Water-Resource Systems: New Techniques for Relating Economic Objectives, Engineering Analysis, and Governmental Planning*. Harvard University Press, 1962.

M. E. P. Maceira, M. A. M. Rodrigues, R. P. D. Ross, A. Kopiler, F. L. Tito, L. A. C. Henriques, A. Castro, A. C. P. Araújo, P. D. Zarur, J. C. O. Mello, and D. B. Andrades. Sistema de encadeamento de modelos energéticos. In *XVI Seminário Nacional de Produção e Transmissão de Energia Elétrica*, Campinas, 2001. XVI Seminário Nacional de Produção e Transmissão de Energia Elétrica, 2001.

J.B. Marco, R. Harboe, and J.D. Salas, editors. *Stochastic Hydrology and its Use in Water Resources Systems Simulation and Optimization*. Peñíscola, Spain, 1993.

Antonio Geraldo Mendes, A.L.M. Marcato, Thiago Correa Cesar, Ivo Chaves S., Paulo Augusto Nepomuceno Garcia, Anderson M. Iung, Tales Pulinho Ramos, and Rafael da Silva Brandi. Generation of multivariate monthly synthetic water streamflows through multiplicative parma model. In *2007 42nd International Universities Power Engineering Conference*, pages 539–544, 2007. doi: 10.1109/UPEC.2007.4469006.

F. L. C. Oliveira. Nova abordagem para geração de cenários de afluências no planejamento da operação energética de médio prazo. Dissertação de mestrado, PUC-Rio, 2010.

F. L. C. Oliveira. *Modelo de Séries Temporais para Construção de Árvores de Cenários Aplicadas à Otimização Estocástica*. Tese de doutorado, PUC-Rio, 2013.

D. D. J. Penna. *Definição da Árvore de Cenários de Afluências para o Planejamento da Operação Energética de Médio Prazo*. Tese de doutorado, PUC-Rio, 2009.

M. V. F. Pereira et al. Stochastic streamflow models for hydroelectric systems. *Water Resources Research*, 20(3):379–390, 1984.

B. A. Ribeiro, H. R. Baldioti, and R. C. Souza. Modelo par(p) com resíduo multiplicativo aplicado na simulação de cenários de energia natural afluyente. In *XLIX Simpósio Brasileiro de Pesquisa Operacional*, 2017.

Amanda P. Silva, André L. M. Marcato, Alessandro L. Castro, Diogo B. Pinto, and Camila T. S. Cunha. Pysddp: An open-source python tool applied to the operation planning problem in the age of energy transition. In *2023 58th International Universities Power Engineering Conference (UPEC)*, pages 1–6, 2023. doi: 10.1109/UPEC57427.2023.10294355.