



Yasmin Monteiro Cyrillo

**Avaliação do Modelo PVARm Interconfigurações para
Geração de Cenários de ENA no Planejamento da
Operação de Médio Prazo**

Dissertação de Mestrado

Dissertação apresentada ao Programa de Pós Graduação em Engenharia Elétrica da PUC-Rio como requisito parcial para a obtenção do grau de Mestre em Engenharia Elétrica.

Orientador: Prof. Reinaldo Castro Souza

Co-orientador: Prof. Fernando Luiz Cyrino Oliveira

Rio de Janeiro

Abril de 2018



Yasmin Monteiro Cyrillo

**Avaliação do Modelo PVARm Interconfigurações para Geração
de Cenários de ENA no Planejamento da Operação de Médio
Prazo**

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

Prof. Reinaldo Castro Souza

Orientador

Departamento de Engenharia Industrial – PUC-Rio

Prof. Fernando Luiz Cyrino Oliveira

Co-Orientador

Departamento de Engenharia Industrial – PUC-Rio

Dr. Joari Paulo da Costa

Operador Nacional do Sistema Elétrico

Prof. André Luís Marques Marcato

UFJF

Prof. Pedro Guilherme Costa Ferreira

FGV

Prof. Márcio da Silveira Carvalho

Coordenador Setorial do Centro

Técnico Científico – PUC-Rio

Rio de Janeiro, 06 de abril de 2018

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

Yasmin Monteiro Cyrillo

Graduou-se em Engenharia Elétrica na Faculdade de Engenharia da
Universidade Federal de Juiz de Fora.

Ficha Catalográfica

Cyrillo, Yasmin Monteiro

Avaliação do modelo PVARm Interconfigurações para
Geração de Cenários de ENA no Planejamento da Operação de
Médio Prazo / Yasmin Monteiro Cyrillo ; orientador: Reinaldo Castro
Souza ; coorientador: Fernando Luiz Cyrino Oliveira. – 2018.

98 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do
Rio de Janeiro, Departamento de Engenharia Elétrica, 2018.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Planejamento energético.
3. PVAR. 4. Modelos de erro multiplicativo. 5. PAR. I. Souza,
Reinaldo Castro. II. Oliveira, Fernando Luiz Cyrino. III. Pontifícia
Universidade Católica do Rio de Janeiro. Departamento de
Engenharia Elétrica. IV. Título.

CDD: 621.3

Agradecimentos

À Deus, em primeiro lugar.

Aos meus pais e minhas irmãs, que nunca mediram esforços para contribuir com meu crescimento pessoal e profissional.

Ao Dirceu, amigo e companheiro, que sempre esteve ao meu lado, nos momentos de alegria e nos momentos que necessitei de força nessa caminhada.

Ao Reinaldo, meu orientador, e ao Fernando, meu coorientador, pela acolhida, incentivo e direcionamento.

Aos professores e amigos da PUC-Rio, pela troca de conhecimentos. Em especial aqueles da equipe MDDH e do LAMPS, que contribuíram com esta caminhada, tanto para torná-la mais leve, quanto para torná-la mais madura.

Aos amigos do ONS e da FGV-Energia, que sempre mostraram compreensão com as atividades do mestrado e enriqueceram minha bagagem.

Ao CNPq, à CAPES e à PUC-Rio, pelos auxílios financeiros concedidos.

A todos aqueles citados nas referências deste trabalho, pela valiosa contribuição literária. Em especial, aos professores André Marcato, Joari Costa e Pedro Guilherme Ferreira.

Resumo

Cyrillo, Yasmin Monteiro; Souza, Reinaldo Castro (Orientador). Oliveira, Fernando Luiz Cyrino (Co-orientador). **Avaliação do Modelo PVARm Interconfigurações para Geração de Cenários de ENA no Planejamento da Operação de Médio Prazo**. Rio de Janeiro, 2018. 98p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O Planejamento da Operação Energética de Médio Prazo do Sistema Interligado Nacional (SIN) é um problema de decisão sob incerteza, com acoplamentos espacial e temporal. A solução vigente determina a política ótima através do algoritmo Programação Dinâmica Dual Estocástica (PDDE), onde a incerteza hidrológica é representada como um processo Periódico Autorregressivo (PAR). Para a aplicação do planejamento, uma restrição do modelo PAR é a possibilidade do mesmo simular valores negativos de afluência. A abordagem usada no modelo para que este gere apenas valores positivos criou uma relação de dependência entre os resíduos do modelo e as afluências passadas, em detrimento da premissa de independência temporal assumida pela PDDE. Neste contexto, o modelo Periódico Vetorial Autorregressivo com erro multiplicativo (PVARm) torna-se interessante para representação da incerteza das afluências, visto que, junto ao método de estimação adotado, garante as premissas da PDDE, além de incorporar a correlação espacial das afluências na própria formulação do modelo. Este trabalho apresenta o modelo PVARm aplicado à geração de cenários para atendimento do processo de otimização, a partir de dados do SIN. A estimação do modelo considera as mudanças de topologia das usinas ao longo do horizonte de planejamento e avalia dois critérios de identificação – ordem fixa unitária e ordem selecionada a partir da menor soma de erros de ajuste. Os cenários gerados são comparados aos cenários gerados pelo modelo que representa aquele vigente, o PAR. A comparação se baseia nos quesitos reprodução de estatísticas mensais e anuais e na adequação dos modelos, medida por testes de sequências. Os resultados mostraram que o PVARm de ordem unitária fixa apresentou comportamento satisfatório na maioria dos testes e motivou a continuidade dos estudos para aplicação no Planejamento da Operação Energética do SIN.

Palavras-chave

Planejamento Energético; PVARm; modelos de erro multiplicativo; PAR.

Abstract

Cyrillo, Yasmin Monteiro; Souza, Reinaldo Castro (Advisor). Oliveira, Fernando Luiz Cyrino (Co-Advisor). **Evaluation of the PVARm Dynamic Model for Streamflow Scenario Generation in Medium-Term Energy Planning Context.** Rio de Janeiro, 2018. 98p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The Medium-Term Energy Planning of the National Interconnected System (SIN) is a decision-making problem under uncertainty, coupled in space and time. The official solution uses the Stochastic Dual Dynamic Programming (SDDP) algorithm, where uncertainty is represented as a Periodic Autoregressive (PAR) process. Specifically for the Planning application, the possibility of PAR simulates negative streamflow values is a limitation of the model. The approach used in PAR to solve this question creates a link between the current residual and the past streamflow in spite of the stochastic component stage-wise independency assumption of SDDP algorithm. In this context, the Periodic Vector Autoregressive model with multiplicative error (PVARm) becomes interesting for streamflow formulation, since it guarantees the PDDE assumptions, it can simulate only positive streamflow values, besides incorporating the spatial inflows correlation. This work presents the PVARm model applied to the simulation of scenarios to attend the optimization process, based on data from the SIN. The estimation of the model is considered as the topology change of the last decades for the planning and selection of two classification criteria - unit and alternative fixed order from the smallest sum of model errors of fit. The scenarios generated are compared to the scenarios generated by the model that represent the current one, the PAR. The comparison is based on the results of the monthly and annual statistics measurement and the adequacy of the models, as measured by sequence tests. The results were that the PVAR is of a type of test that was not implemented in most of the tests and motivated the continuity of the studies for the application in the Planning of the Energy Operation of the SIN.

Keywords

Energy Planning; PVARm; multiplicative error models; PAR.

Sumário

1 Introdução	13
1.1. Contribuições deste Trabalho	18
1.2. Organização da Dissertação	19
2 Planejamento Energético de Médio Prazo	20
2.1. O Modelo de Otimização	20
2.2. O Modelo da Série Temporal	24
2.2.1. Caracterização das Séries de Energia Natural Afluente (ENA)	24
2.2.2. Modelo Periódico Autorregressivo – PAR	28
2.3. Geração de Cenários	34
2.3.1. Modelagem dos Ruídos para Simulação de Cenários a partir do Modelo PAR	35
2.3.2. Técnica Monte Carlo	36
2.3.3. Construção dos Cenários para o Modelo de Otimização	37
2.4. Critérios de Avaliação de Desempenho de Modelos	39
2.4.1. Teste de Kolmogorov-Smirnov (K-S)	42
2.4.2. Teste Qui-Quadrado	42
3 O Modelo Periódico Vetorial Autorregressivo com Erro Multiplicativo – PVARm	44
3.1.1. Formulação do Modelo PVARm	45
3.1.2. Estimação dos Parâmetros	46
3.1.3. Identificação da Ordem do Modelo	52
4 Resultados	54
4.1. Análise dos Dados de Entrada	55
4.2. Verificação de Estatísticas Mensais e Anuais	58
4.2.1. Envoltória dos Cenários	58
4.2.2. Desvio Padrão	60
4.2.3. Assimetria	63

4.2.4. Correlação Espacial entre Subsistemas	65
4.3. Testes de Avaliação e Adequação dos Modelos	69
4.3.1. Teste de Kolmogorov-Smirnov (K-S)	69
4.3.2. Teste de Wilcoxon	70
4.3.3. Teste de Ansari – Bradley	71
4.3.4. Análise de Sequências	72
5 Conclusões e Trabalhos Futuros	84
6 Referências bibliográficas	87
7 Apêndice A	91
7.1. Algoritmo para Geração de Cenários com Modelo PVARm	91
7.1.1. Montagem do Histórico Dinâmico de ENAs	92
7.1.2. Identificação do Modelo e Estimação de Parâmetros	93
7.1.3. Geração de Cenários	96
7.2. Geração de Cenários de Afluências pelo Modelo PAR	98

Lista de figuras

Figura 1: Capacidade Instalada do Sistema Interligado Nacional.	13
Figura 2: Decisão sob incerteza da operação centralizada de um sistema majoritariamente hidrotérmico.	15
Figura 3: Função de custo futuro para determinado período, para a energia armazenada (EARM) em um reservatório.	22
Figura 4: Árvore de Cenários.	22
Figura 5: Histórico dinâmico	27
Figura 6: Construção da árvore de cenários e geração de cenários para as etapas <i>forward</i> e <i>backward</i> .	39
Figura 7: Ilustração do conceito de sequência.	40
Figura 8: Regiões agrupadas para formação dos REE's.	55
Figura 9: Cronograma do incremento da capacidade instalada de hidrelétricas do SIN para o horizonte de 2015 – 2019.	57
Figura 10: Envoltória dos cenários gerados pelo modelo PAR.	59
Figura 11: Envoltória dos cenários gerados pelo modelo PVAR _m ($p=1$)	59
Figura 12: Envoltória dos cenários gerados pelo modelo PVAR _m ($1 \leq p \leq 6$)	60
Figura 13: Desvio Padrão dos cenários gerados pelo modelo PAR	61
Figura 14: Desvio Padrão dos cenários gerados pelo modelo PVAR _m ($p=1$)	62
Figura 15: Desvio Padrão dos cenários gerados pelo modelo PVAR _m ($1 \leq p \leq 6$)	62
Figura 16: Assimetria dos cenários gerados pelo modelo PAR	63
Figura 18: Assimetria dos cenários gerados pelo modelo PVAR _m ($1 \leq p \leq 6$)	64
Figura 19: Correlação anual entre subsistemas dos cenários gerados pelo modelo PAR.	66

Figura 20: Correlação anual entre subsistemas dos cenários gerados pelo modelo PVARm ($p=1$)	67
Figura 21: Correlação anual entre subsistemas dos cenários gerados pelo modelo PVARm ($1 \leq p \leq 6$)	68
Figura 22: Percentual de períodos aprovados pelo teste estatístico de Kolmogorov-Smirnov	69
Figura 23: Percentual de períodos aprovados pelo teste estatístico de Wilcoxon.	70
Figura 24: Percentual de períodos aprovados pelo teste estatístico de Ansari – Bradley.	71
Figura 25: Histórico dinâmico para PMO de 2015, supondo 60 configurações diferentes.	73
Figura 26: Estatística qui-quadrado para o teste de comprimento de sequência positiva.	75
Figura 27: Estatística qui-quadrado para o teste de comprimento de sequência negativa.	75
Figura 28: Probabilidade de máximos para comprimento de sequências positivas ($C < H$)	76
Figura 29: Probabilidade de máximos para comprimento de sequências negativas ($C < H$)	76
Figura 30: P-valor para o teste de hipótese de K-S para soma de sequência positiva.	78
Figura 31: P-valor para o teste de hipótese de K-S para soma de sequência negativa.	78
Figura 32: Probabilidade de máximos para soma de sequências positivas ($C < H$)	79
Figura 33: Probabilidade de máximos para soma de sequências negativas ($C < H$)	79
Figura 34: P-valor para o teste de hipótese de K-S para intensidade de sequência positiva.	81
Figura 35: P-valor para o teste de hipótese de K-S para intensidade de sequência negativa.	81
Figura 36: Probabilidade de máximos para intensidade de sequências positivas ($C < H$)	82

Figura 37: Probabilidade de máximos para intensidade de sequências negativas ($C < H$)	82
Figura 38: Etapas do algoritmo para ajuste e geração de cenários com o modelo PVARm Interconfigurações.	91
Figura 39: Histórico dinâmico montado a partir do PMO de 12/2015	92
Figura 40: Fluxograma do código implementado para identificação e estimação do modelo PVARm Interconfigurações	95
Figura 41: Fluxograma do algoritmo da geração de cenários <i>forward</i> e <i>backward</i> .	97

Lista de tabelas

Tabela 1: Mudanças de configuração ao longo de um horizonte de 5 anos de planejamento.	28
Tabela 2: Mudanças de configuração previstas no PMO de 12/2015.	
Fonte: Deck de Preços do Newave 12/2015.	56

1

Introdução

O Sistema Interligado Nacional (SIN), composto pelos sistemas de produção e transmissão de energia elétrica, é um sistema hidro-termo-eólico de grande porte, com predominância de usinas hidrelétricas. A capacidade instalada do SIN somou 142,042 GW em dezembro de 2017, montante do qual aproximadamente 72% corresponde às hidrelétricas, 7% às eólicas e 20% às termelétricas, como apresentado na Figura 1 (ONS, 2017).

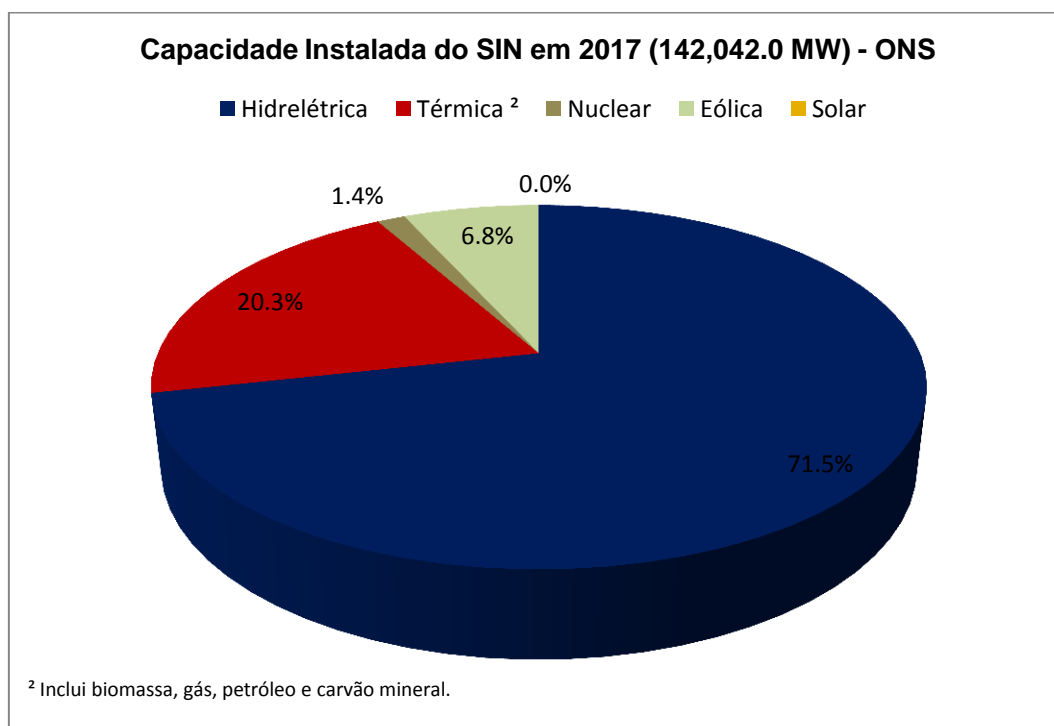


Figura 1: Capacidade Instalada do Sistema Interligado Nacional.

Fonte: Elaborada pela autora

O SIN (ONS, 2018) é dividido em quatro subsistemas – Sul, Sudeste/Centro-Oeste, Nordeste e a maior parte da região Norte. As usinas hidrelétricas estão distribuídas pelo país em doze bacias hidrográficas, compondo 69

aproveitamentos com reservatório, 85 aproveitamentos à fio d'água¹ e 4 usinas à bombeamento². Os parques eólicos, por sua vez, concentram-se nas regiões Nordeste e Sul, enquanto as termelétricas geralmente estão próximas dos principais centros de carga, estes situados nas regiões Sudeste e Sul.

Desde meados da década de 1970, o SIN tem sua operação centralizada, que visa redução dos custos globais da operação através da exploração da interligação elétrica dos subsistemas, da interdependência operativa entre as usinas e da integração dos recursos de geração e transmissão no atendimento à carga.

A operação coordenada é vantajosa por promover o intercâmbio de energia, aumentar a confiabilidade do suprimento, além de prover melhor aproveitamento dos recursos, uma vez que explora a diversidade entre os regimes hidrológicos das bacias e a complementariedade das fontes primárias de energia elétrica. Entretanto, tal coordenação traz o desafio de planejar de maneira centralizada um sistema com as especificidades do SIN.

O planejamento da operação energética de um sistema com tais características é um problema de decisão de grande porte, sujeito a incertezas como as afluências futuras, demanda, o custo futuro dos combustíveis, o cumprimento de prazo para a entrada em operação de usinas, entre outros.

O operador do sistema deve optar diariamente entre poupar água nos reservatórios ou deplecioná-los, ciente de que as condições hidrológicas no futuro indicarão se a política realizada no presente foi bem empregada ou não (Figura 2). Uma política mal empregada impacta diretamente no custo total da operação energética: se o operador decide por deplecionar os reservatórios e no futuro há escassez de água, tornar-se-á necessário o acionamento de térmicas mais caras; do contrário, se o operador decide armazenar água e no futuro há abundância deste recurso, haverá desperdício através do vertimento.

¹ Usina hidrelétrica ou pequena central hidrelétrica que utiliza reservatório com acumulação suficiente apenas para prover regularização diária ou semanal, ou ainda que utilize diretamente a vazão afluente do aproveitamento. (ANEEL, 2011)

² Usinas a bombeamento permitem o bombeamento de água desde um reservatório inferior para um reservatório superior durante períodos de carga leve, utilizando energia extra de qualquer outra fonte geradora do sistema. A água armazenada pode ser usada para movimentar as turbinas hidráulicas na geração de eletricidade nos horários de ponta.

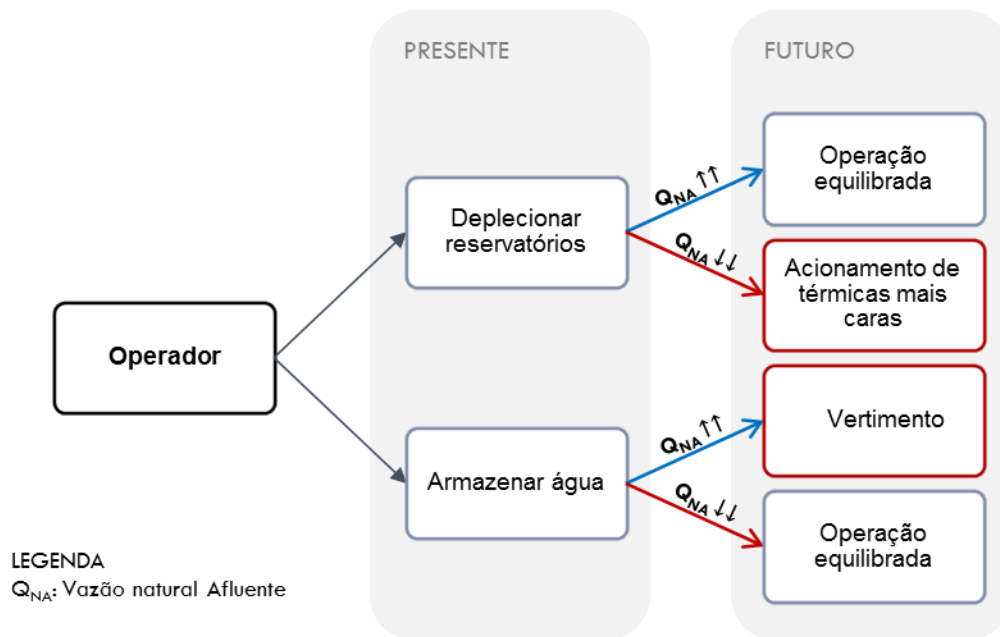


Figura 2: Decisão sob incerteza da operação centralizada de um sistema majoritariamente hidrotérmico. Fonte: Elaborada pela autora.

A escolha do operador sob a incerteza da afluência é acoplada no tempo, uma vez que a decisão tomada no presente impacta nas decisões que serão tomadas no futuro. Além do acoplamento temporal, a decisão do operador também é acoplada no espaço, uma vez que a ação de despachar uma usina afeta o estado dos demais aproveitamentos hidrelétricos da cascata.

O Planejamento da Operação Energética do SIN busca, em todas as suas fases, definir a política de operação de seus recursos energéticos ao longo do horizonte de planejamento considerado. A política é definida sob o objetivo de alcançar o menor custo esperado para todo o horizonte de planejamento, de forma a evitar o risco de se expor a cenários hidrológicos desfavoráveis, penalizando o não atendimento à demanda e respeitando restrições operativas e de segurança energética.

A solução vigente para o planejamento é dada por uma cadeia de modelos (Pereira e Pinto, 1984) que se diferem pelo horizonte e discretização de tempo observado e pela representação da incerteza das afluências futuras. Quanto maior o horizonte de planejamento, maior é a simplificação adotada na representação do sistema e maiores são as incertezas.

No horizonte de médio prazo, foco desta dissertação, está o modelo Newave, que contempla até 5 anos à frente. Neste modelo, as decisões são

tomadas em base mensal e as usinas hidrelétricas são agregadas e representadas como Reservatórios Equivalentes de Energia (REEs). A política de operação para os subsistemas equivalentes tem seu cálculo baseado no algoritmo de Programação Dinâmica Dual Estocástica (PDDE), e considera apenas a incerteza relacionada às afluições (CEPEL, 2006).

A PDDE (Pereira e Pinto, 1991) é um método baseado na decomposição de Benders, onde uma função linear por partes é construída de forma iterativa, compondo uma aproximação inferior da função de custo futuro, isto é, a função que relaciona o custo futuro da operação energética com a energia armazenada em forma de água nos reservatórios no momento presente. Neste método, o problema completo de otimização é decomposto em subproblemas de programação linear, que podem ser resolvidos de forma independente. A metodologia requer que a componente aleatória seja independente e modelada como um problema convexo.

Em geral, o desempenho da operação do sistema é afetado pela qualidade das simulações hidrológicas realizadas, que desenvolvem papel crítico nos estudos de planejamento e operação de sistemas hidrelétricos. Um modelo inadequado do fluxo dos reservatórios tende a distorcer índices importantes que auxiliam no processo de tomada de decisão.

Por conseguinte, é importante desenvolver um modelo de simulação de afluição realista, que forneça uma representação razoável da variabilidade dessa variável e atenda aos requisitos de convergência da PDDE (Shapiro, 2011), (Shapiro e Philpott, 2007). Estes requisitos são supridos pelo modelo Periódico Autorregressivo (PAR) (Hipel e McLeod, 1994), aplicado na solução vigente de médio prazo. No entanto, por se tratar de um modelo com erro aditivo, o PAR(p) pode gerar afluições negativas que inviabilizam o procedimento de otimização necessário para a realização do despacho hidrotérmico. A tentativa de solucionar esse problema gerou outro (CEPEL, 2006), que foi a dependência temporal dos resíduos, o que confronta uma das premissas na qual se baseia o algoritmo de otimização.

Recentemente surgiram propostas de modelagem na literatura que se utilizam do conceito de erro multiplicativo, definido como a razão entre o valor observado e o valor previsto da variável aleatória. Citam-se na literatura os trabalhos (Lanne, 2006), (Hautsch, Malec e Schienle, 2014), (Cipollini, Engle e Gallo, 2006) e (Bodnar e Hautsch, 2016). Em (Braga, 2011) e (Ferreira, 2013),

foram propostos modelos periódicos de distribuição gama para os erros multiplicativos. Em (Ribeiro, Baldioti e Souza, 2017) um modelo PAR(p) com erro multiplicativo é utilizado para geração de cenários para o caso do Planejamento Energético de Médio Prazo.

O trabalho descrito em (Cabral, 2016) propõe a utilização do modelo Periódico Vetorial Autorregressivo com erro multiplicativo (PVARm) para geração de cenários de afluência, com o intuito de contornar a dificuldade encontrada pelo uso do PAR (p). A proposta, sob a ótica da otimização estocástica, atende aos requisitos necessários para aplicação na PDDE para solução do problema do planejamento: a equação do modelo PVARm relaciona as afluências de maneira linear e garante a independência temporal dos ruídos multiplicativos.

Partindo do exposto, é objetivo da presente dissertação apresentar a implementação do modelo de geração de cenários sintético PVARm (Cabral, 2016) para o sistema hidrelétrico brasileiro, voltado ao atendimento do processo de otimização do planejamento da operação de médio prazo, levando em consideração a configuração do parque gerador dos períodos avaliados. Os cenários gerados serão avaliados e comparados aos cenários gerados pelo modelo que representa aquele vigente, o PAR.

A primeira etapa utilizada para avaliar o modelo PVARm foi a implementação do próprio método. Entretanto, existem questões em aberto sobre o PVARm, tais como: a identificação da ordem do modelo; a análise de consistência do estimador proposto; novas propostas de estimação dos parâmetros; e novas hipóteses sobre a distribuição do erro multiplicativo.

Nota-se que há uma série de aprimoramentos no próprio modelo que podem ser abordados. O intuito deste trabalho é adaptação do modelo à realidade do parque hidrelétrico brasileiro e posterior avaliação da geração de cenários e comparação com o modelo vigente. Portanto, objetiva-se trabalhar com os aprimoramentos primários exigidos pelo modelo, que permitam a avaliação no problema do planejamento da operação. Neste sentido, foram avaliadas duas propostas para identificação do modelo: a primeira considerando ordem unitária fixa e a segunda baseada na menor soma dos quadrados dos desvios. As propostas são comparadas junto ao modelo PAR.

Para comparação dos modelos, foram selecionados testes estatísticos não paramétricos para avaliação da reprodução das estatísticas mensais média, desvio padrão e assimetria, além da estatística anual coeficiente de correlação cruzada. Adicionalmente, foram utilizadas as variáveis comprimento e intensidade dos eventos de sequências positiva e negativa para avaliação da adequação dos modelos.

As próximas seções apresentam as maiores contribuições do trabalho desenvolvido, além da organização desta dissertação.

1.1. Contribuições deste Trabalho

Essa dissertação busca apresentar e comparar os modelos PAR e PVARm para representação das afluições no contexto do Planejamento Energético de Médio Prazo. A avaliação é feita sob a ótica de que tal modelo deve tanto reproduzir as propriedades estatísticas esperadas, quanto atender aos requisitos exigidos pelo modelo de otimização. A representação da incerteza é de grande importância para que os cenários gerados sejam fidedignos ao processo estocástico que os gerou. Entretanto, uma boa representação por si só não basta, já que o modelo de otimização define a política de operação baseado em premissas que podem não ser atendidas com tal modelo.

Este trabalho ressalta a importância de um modelo que atenda aos requisitos da otimização e que represente de maneira adequada o comportamento das afluições. Apresenta-se a comparação da metodologia vigente com o modelo proposto, apontando os pontos fortes e as oportunidades de melhorias em cada um deles. Os cenários são gerados de maneira adequada ao processo de otimização que os consome, considerando a mudança de configuração do parque de geração hidrelétrico ao longo do horizonte de planejamento.

Adicionalmente, apresenta-se revisão da literatura no que concerne aos temas supracitados e os aspectos práticos de implementação do modelo PVARm desenvolvido neste trabalho.

1.2. Organização da Dissertação

A dissertação está organizada como segue. O capítulo 2 apresenta revisão da literatura no que concerne aos modelos de otimização e de geração de cenários adotados na solução vigente para o Planejamento da Operação Energética de Médio Prazo. O modelo de geração de cenários aborda as características das séries de Energia Natural Afluente (ENA) bem como o modelo estatístico PAR para sua representação; o capítulo 3 apresenta o modelo PVARm e as modificações propostas para sua utilização na solução de otimização do médio prazo; o capítulo 4 mostra os resultados obtidos para cada um dos modelos e a comparação entre eles; o capítulo 5 apresenta as conclusões do trabalho realizado, bem como propostas de trabalhos futuros; o capítulo 6 lista as referências utilizadas e indicadas para os temas abordados nesta dissertação; e o capítulo 7 consiste de um apêndice para abordar o desenvolvimento do trabalho, com foco nos aspectos práticos da implementação do modelo.

2

Planejamento Energético de Médio Prazo

Este capítulo apresenta revisão da literatura no que concerne aos modelos de otimização e de geração de cenários adotados na solução vigente para o Planejamento da Operação Energética de Médio prazo. No modelo de otimização, apresenta-se a técnica da PDDE, além de abordar as questões relacionadas à representação da incerteza e ao processo de geração de cenários.

O modelo de geração de cenários aborda as características das séries de Energia Natural Afluente (ENA) bem como o modelo estatístico PAR para sua representação.

Além disso, apresentam-se as técnicas Monte Carlo e Bootstrap aplicadas à simulação de modelos hidrológicos e também os testes de adequação utilizados na avaliação dos modelos PAR e PVARm.

2.1.

O Modelo de Otimização

Apresentadas as características do SIN na seção anterior, nota-se que o problema do planejamento da operação desse sistema é complexo, envolvendo decisão sob incerteza para múltiplos períodos, acoplada no espaço e no tempo. Essas especificidades levam ao estudo das abordagens de Programação Dinâmica Estocástica, apontadas na literatura como solução de problemas complexos e de larga escala.

Métodos de decomposição que demandam discretização do espaço de estados e a representação completa da árvore de cenários estão suscetíveis ao “mal da dimensionalidade”. Para solução de problemas complexos de grande escala, são indicados pela literatura os métodos de decomposição baseados em amostragem do processo estocástico, como a Programação Dinâmica Dual Estocástica (PDDE) (Pereira e Pinto, 1985), *Abridged Nested Decomposition* (AND) (Donohue e Birge, 2006) e *Convergent Cutting-Plane and Partial-*

Sampling (CUPPS) (Chen e Powell, 1999), que se valem do princípio de independência temporal da componente aleatória envolvida.

A PDDE teve origem no setor elétrico (Pereira e Pinto, 1991) e permanece como o método utilizado nos estudos de planejamento de médio e longo prazos da operação de sistemas hidrotérmicos (Pereira *et al.*, 2015), (Brigatto, Street e Valladão, 2016), (Matos, De, Philpott e Finardi, 2015), (Shapiro *et al.*, 2012). Atualmente, o modelo oficial para o problema de médio prazo utiliza a técnica e considera apenas a incerteza relacionada às afluições (CEPEL, 2013).

O algoritmo PDDE aplicado ao planejamento de médio prazo busca aproximar a função do custo futuro da operação em relação à energia armazenada nos reservatórios no instante atual (Figura 3). A política de operação é definida sob o objetivo de alcançar o menor custo esperado para todo o horizonte de planejamento, de forma avessa ao risco de se expor a cenários que apresentem custos altos de operação, penalizando o não atendimento à demanda e respeitando restrições operativas e de segurança energética. A incerteza hidrológica é considerada através de cenários de energia natural afluyente (ENAs) às sub regiões.

A técnica se baseia no método de Decomposição de Benders (Benders, 1962), que particiona problemas de programação linear de larga escala em múltiplos estágios. Uma vez que o problema original é decomposto, a solução é obtida através de um processo iterativo, composto por etapas de recursão *forward* e *backward*, que continuamente refinam a função de custo futuro de cada estágio e a definem com complexidade expressivamente menor do que aqueles métodos que consideram a árvore completa (Shapiro, 2011). Em cada uma das etapas, os limites superior e inferior são calculados para verificação da convergência do processo iterativo. O resultado final é uma aproximação inferior da função de custo futuro do problema original, representada por uma função linear por partes, como apresenta a Figura 3.

A técnica exige i) convexidade do problema modelado e ii) independência temporal da componente aleatória envolvida. A seguir, descrevem-se sumariamente as etapas do algoritmo, destacando a importância da representação da estocasticidade orientada ao processo de otimização.

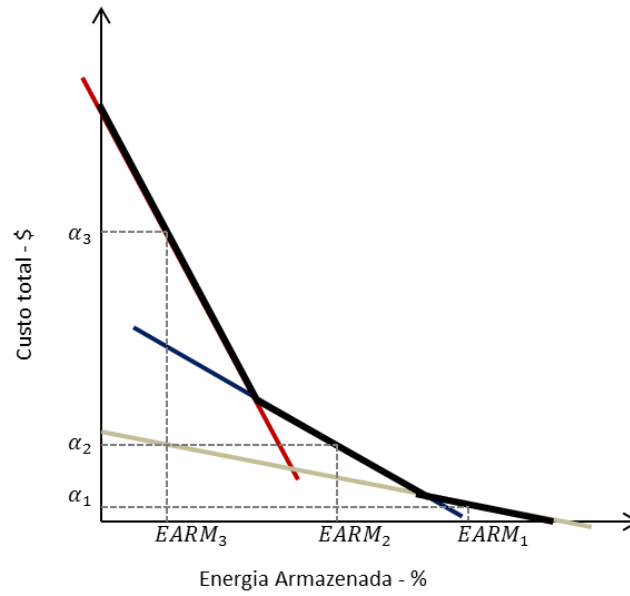


Figura 3: Função de custo futuro para determinado período, para a energia armazenada (EARM) em um reservatório. Fonte: Elaborada pela autora.

Inicialmente, uma árvore finita de cenários é gerada por amostragem aleatória a partir da distribuição original e então o problema construído é resolvido via algoritmo PDDE.

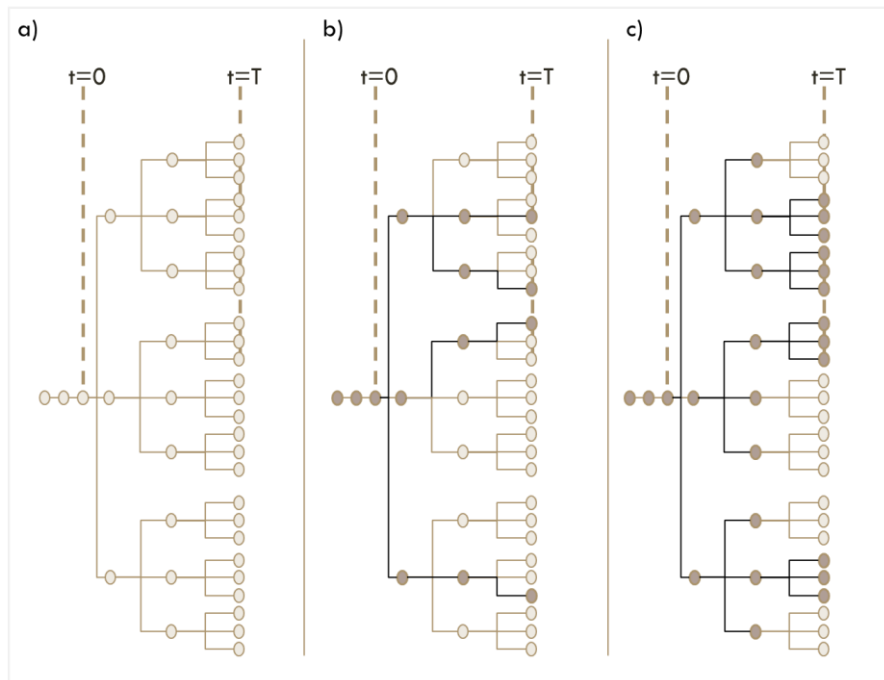


Figura 4: Árvore de Cenários: a) completa; b) cenários varridos na *forward*; c) cenários varridos na *backward*. Fonte: Elaborada pela autora.

Suponha uma árvore de cenários construída a partir do sorteio aleatório de N_t amostras para cada período tratado ($t = 0, \dots, T$). Em consequência, a distribuição da população original é substituída pela distribuição empírica construída a partir das N_t amostras, com igual probabilidade de ocorrência. A título de exemplo, considere a árvore de cenários apresentada pela Figura 4a, onde $N_t = 3, \forall t$ e $T = 3$. Para maiores detalhes sobre a estratégia, consultar (Shapiro, 2011), (Bortolossi e Pagnoncelli, 2006).

Cada nó (ou abertura) do estágio ($t - 1$) tem os mesmos ramos correspondendo às N_t aberturas selecionadas. Cada abertura da árvore representa um problema de Programação Linear a ser resolvido. O número total de cenários do problema é $N = \prod_{t=1}^T N_t$. No exemplo apresentado pela Figura 4, são 27 cenários possíveis. Note que, mesmo com um número reduzido de aberturas por estágio, o número total de cenários cresce exponencialmente e o aumento do número de estágios faz crescer ainda mais, o que sugere que uma abordagem baseada em cenários pode rapidamente se tornar intratável. Assim, a amostragem torna-se interessante e pode ser aplicada à PDDE como descrito a seguir.

A etapa *forward* do algoritmo consiste em gerar M cenários aleatórios, onde $M < N$. Na Figura 4b, foram sorteados $M = 4$ cenários aleatórios, destacados na imagem. Para cada estágio t , com $t = 1, \dots, T$ (em $t = 0$ o problema é determinístico), calcula-se o custo de operação para as M realizações. No final de cada estágio é possível calcular a média e variância amostral do custo de operação. Após o cálculo para $t = T$, a soma do custo esperado por estágio caracteriza um limite superior para o custo do problema original.

Na etapa *backward* (Figura 4c), a partir de um cenário e sua respectiva política inicialmente definida na etapa *forward* (Figura 4b), resolve-se o problema iniciando pelo último estágio, $t = T, T - 1, \dots, 1$, assumindo-se, por simplicidade, que o custo futuro neste ponto é nulo. O problema em T é resolvido para todas as N_T aberturas sorteadas e para o passado definido pelo cenário da *forward*. Com o resultado, obtém-se o custo esperado do estágio T e mais um corte da função de custo em T em relação ao armazenamento de água no início deste estágio, isto é, a função de custo futuro para o estágio $T - 1$. A regressão segue até $t = 0$, quando

o problema é determinístico. No final da etapa *backward*, é possível calcular um limite inferior para a função de custo futuro.

A convergência do processo iterativo se dá quando o limite inferior está dentro do intervalo de confiança do limite superior.

O algoritmo assume independência temporal do processo estocástico, o que implica em uma função de custo futuro única para cada estágio de decisão. O conjunto de funções de custo futuro e seus problemas associados fornecem uma política de operação, isto é, um conjunto de decisões ótimas para cada período e cenários considerados.

A partir do exposto destaca-se a importância em se desenvolver um modelo de simulação para afluências, que forneça uma representação razoável da variabilidade dessa variável, que permita a viabilidade do algoritmo e respeite os requisitos do algoritmo PDDE.

2.2.

O Modelo da Série Temporal

Nesta seção apresentam-se as características da série de Energia Natural Afluente (ENA) bem como o modelo utilizado em sua representação no Newave: o modelo Periódico Autorregressivo (PAR).

2.2.1.

Caracterização das Séries de Energia Natural Afluente (ENA)

A representação de usinas hidrelétricas agrupadas por reservatórios equivalentes de energia (REEs) reduz a complexidade e o esforço computacional do problema do planejamento da operação do SIN. No processo de agregação de usinas em REEs, as grandezas hidrológicas (volume [hm^3], vazão [m^3/s]) são convertidas para energia ($MW_{médio}$), considerando a capacidade de geração das usinas agregadas a partir da vazão ou volume de água disponível para geração de eletricidade naquele grupo. Maiores detalhes acerca da representação por sistemas equivalentes de energia podem ser encontrados em (Souza *et al.*, 2014) ou (Marcato, 2002).

A Energia Natural Afluente (ENA) é aquela que se obtém quando a vazão natural afluente a um ponto de observação é turbinada nas usinas situadas à

jusante do ponto e é dada pelo produto entre a vazão de entrada e a capacidade de produção da usina. A energia natural afluenta a uma bacia é a soma das energias naturais afluentes a todos os pontos de observação existentes na bacia.

Uma hipótese plausível para as séries de ENA e adotada neste trabalho é a invariância de suas propriedades, verificada de um mês de determinado ano para o mesmo mês do ano subsequente, isto é, a hipótese de que essa série é periodicamente correlacionada, em uma configuração estática de usinas. Adicionalmente, esse processo é considerado periodicamente ergódico, isto é, as médias temporais convergem com probabilidade 1 para as respectivas esperanças, de forma periódica (Cabral, 2016).

Uma das peculiaridades da série de ENA é que esta é função da configuração do parque de geração hidrelétrica do período em análise. Assim, a entrada de uma nova usina em operação altera a topologia do sistema e, portanto, a série histórica de energia natural afluenta.

Neste contexto, a entrada em operação de uma nova usina pode ser dividida em três fases:

1. Período de enchimento do volume morto: o reservatório da usina não está disponível para o sistema e a produção da usina é nula. Considera-se apenas a subtração do volume de água desviado para encher o reservatório;
2. Período de motorização da usina: O reservatório entrou em operação, mas o número de máquinas instaladas não é suficiente para atender a energia firme, isto é, a usina está submotorizada. Neste caso, o reservatório da usina está disponível para o sistema, mas a produção da usina é descontada diretamente do mercado.
3. Usina motorizada: O reservatório está em operação e o número de máquinas instaladas é maior ou igual ao número de máquinas de base. Neste caso, o reservatório e também a usina hidrelétrica estão totalmente disponíveis para o sistema.

Uma configuração hidroelétrica fica definida por um grupo de usinas hidráulicas. Cada uma delas deve estar em um dos estados supracitados. Uma mudança de configuração é definida em (CEPEL, 2013) quando uma hidrelétrica passa do estado 1 para o estado 2, ou do estado 2 para o estado 3.

Visto que os cenários gerados pelo modelo PAR serão utilizados no algoritmo da PDDE, que representa as afluições passadas como variáveis de estado, o modelo de geração de cenários deve ser aplicado diretamente aos dados de ENA originais, sem quaisquer transformações estatísticas, e deve ser capaz de lidar com resíduos que apresentam forte coeficiente de assimetria (CEPEL, 2006).

Uma das consequências disto é que, no momento de cálculo da política de operação, onde a afluição passada é um valor determinístico, o modelo receberá a ENA do período anterior, considerando a topologia deste período, e espera que a ENA do período corrente esteja na topologia corrente. Assim, o processo de estimação do modelo ajustado à série de ENA deve considerar as diferentes configurações do sistema.

A estratégia para trabalhar com esta série é abordada em (Ferreira, 2013) e (Souza *et al.*, 2014) e consiste em utilizar um histórico dinâmico na modelagem do processo estocástico. O modelo de um período T qualquer deve ser baseado em um histórico calculado a partir das vazões afluentes históricas e da configuração do parque gerador do próprio período T . Se o modelo exige uma regressão de ordem 1, por exemplo, e o período $T - 1$ possui configuração hidrelétrica diferente, então o histórico que representa este último período deve ser calculado com base na configuração de $T - 1$.

A título de exemplo, suponha um sistema dividido em dois subsistemas, Norte e Sul, representados pela Figura 5. Deseja-se ajustar um modelo periódico autorregressivo de ordem unitária para o período de fevereiro de 2018. Suponha ainda que, para este sistema está disponível uma base histórica de vazão natural afluente em ambos os subsistemas, de janeiro de 1931 a janeiro de 2018.

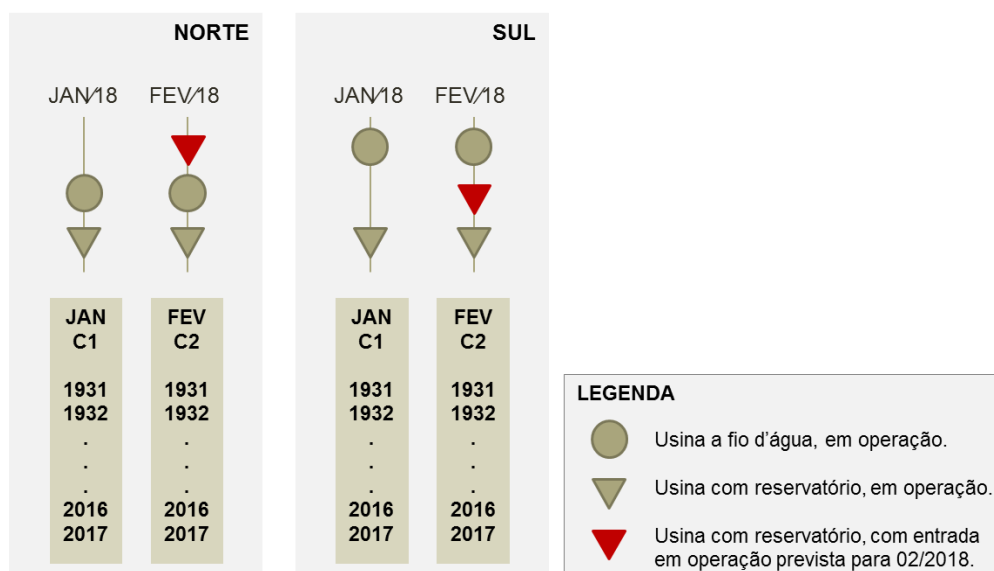


Figura 5: Histórico dinâmico - recalculado de acordo com a configuração topológica do parque de geração hidrelétrica. Fonte: Elaborada pela autora.

A topologia dos subsistemas em janeiro de 2018 é composta por duas usinas em cascata, representadas pelo triângulo e pelo círculo conectados (Figura 5). Em fevereiro de 2018, entram em operação duas novas usinas: uma em cada subsistema, destacadas em vermelho na imagem. Assim, para obtenção do modelo, são calculados dois históricos de ENA de 1931 a 2018, a partir do histórico de vazões:

- O primeiro histórico considera no cálculo a capacidade de produção das usinas com base na topologia do sistema esperada para de janeiro de 2018;
- O segundo histórico considera a capacidade de produção das usinas com base na topologia do sistema esperada para de fevereiro de 2018, quando uma nova usina entra em operação.

O modelo deve então se basear nos dados de todos os meses de janeiro, de 1931 a 2018, do histórico calculado em (i) e nos dados de todos os meses de fevereiro, de 1931 a 2018, do histórico calculado em (ii). A estimação do parâmetro de regressão abrange ambos os históricos.

O modelo Periódico Autorregressivo (PAR), aplicado em uma configuração estática, caracteriza cada um dos meses (ou períodos) de um ano. Ao adotar o

histórico dinâmico, para cada período em que há uma mudança de configuração, outro modelo deve ser ajustado para caracterizar aquele mês do ano.

Tabela 1: Mudanças de configuração ao longo de um horizonte de 5 anos de planejamento. Fonte: Elaborada pela autora

ANO/MÊS	JA N	FE V	MA R	AB R	MA I	JU N	JU L	AG O	SE T	OU T	NO V	DE Z
2015	C1	C1	C1	C1	C1	C1	C2	C2	C2	C2	C2	C2
2016	C2	C3	C3	C3	C3	C3	C3	C3	C3	C4	C4	C4
2017	C4	C4	C4	C4	C5	C5	C5	C5	C5	C5	C5	C5
2018	C5	C6	C6	C6	C6	C6	C6	C6	C6	C6	C6	C6
2019	C6	C6	C6	C7	C7	C7	C7	C7	C7	C7	C7	C7

A Tabela 1 apresenta as mudanças de configuração do parque gerador para um horizonte de planejamento de 5 anos, com início em janeiro de 2015, supondo o total de 7 topologias diferentes. Nota-se que o mês de janeiro será representado por 5 modelos de diferentes, pois nos 5 anos representados, a topologia do parque possui configurações diferentes. Já o mês de março possui os mesmos modelos para os anos de 2018 e 2019, já que nenhuma hidrelétrica entrou ou saiu de operação neste período.

Em (Ferreira, 2013), adota-se a nomenclatura ‘PAR(p) Interconfigurações’ para o modelo PAR ajustado a partir do histórico dinâmico. Esta dissertação adota a mesma nomenclatura e a estende ao modelo PVARm, para designar que o mesmo foi ajustado a partir do histórico dinâmico.

2.2.2. Modelo Periódico Autorregressivo – PAR

A seção 2.1 mostrou que a incerteza incorporada no algoritmo da PDDE deve ser modelada de tal forma que representação seja linear e a componente aleatória seja independente no tempo. Além disso, é importante que o modelo incorpore as principais características do processo estocástico representado e que seja adequado às alterações topológicas do parque gerador.

Dados os pré-requisitos, candidatos naturais à formulação da série são os modelos PAR, por trabalharem com a proposta de erro independente, além de ser reconhecido na literatura como um modelo que reproduz satisfatoriamente o

processo de ENA (CEPEL, 2006), (Oliveira, 2010), (Oliveira, 2013), (Penna, 2009).

As pesquisas acerca do comportamento estocástico de séries hidrológicas surgiram no início da década de 60, no campo de recursos hídricos (Maass *et al.*, 1962), com o modelo PAR. Este modelo foi delineado para aplicação no planejamento energético de médio prazo do SIN em (Pereira *et al.*, 1984), construído para aplicação direta na série, sem a necessidade de transformações específicas, como a Box-Cox. Uma estratégia para geração de cenários apenas positivos foi traçada e desde então o modelo vem sendo aperfeiçoado no âmbito do planejamento energético (Maceira e Damázio, 2006), (Penna, 2009).

Os modelos PAR podem ser pensados como extensões periódicas dos modelos autorregressivos (AR), ou ainda, como tipos específicos de modelos AR multivariados. O PAR associa um modelo AR para cada período m considerado em um ano. Entretanto, como apresentado na seção 2.2.1, sempre que há mudança de topologia de usinas para um período, é necessário calcular novo modelo. Portanto, além de associar um modelo ao mês correspondente, torna-se necessário associá-lo também à configuração topológica a qual pertence.

Considere \mathbf{z}_t como um processo periódico, com período \mathbb{S} . Um modelo PAR de ordem $(p_{1,c}, p_{2,c}, \dots, p_{\mathbb{S},c})$ é definido para um período m , com $m = 1, 2, \dots, 12$, na configuração c_t , com $c_t \in C = \{1, 2, \dots, n_{conf}\}$, onde C é o conjunto de índices das n_{conf} configurações diferentes, da seguinte forma:

$$\left(\frac{z_{t,c_t} - \mu_{m,c_t}}{\sigma_{m,c_t}} \right) = \sum_{i=1}^{p_{m,c_t}} \phi_i^{m,c_t} \left(\frac{z_{t-i,c_{t-i}} - \mu_{m-i,c_{t-i}}}{\sigma_{m-i,c_{t-i}}} \right) + a_{t,c_t} \quad (1)$$

Sendo:

z_{t,c_t} : Série de ENA de período \mathbb{S} , na configuração c_t , com $t = 1, \dots, T$;

μ_{m,c_t} : Média sazonal da série z_t para período m e configuração c_t ;

σ_{m,c_t} : Desvio-padrão sazonal da série z_t para período m e configuração c_t ;

ϕ_i^{m,c_t} : Coeficiente AR de defasagem i , para o período m e configuração c_t ;

A série a_{t,c_t} , relativa ao período t e configuração c_t , é um ruído branco, respeitando as condições impostas na eq. (2):

$$E[a_{t,c_t}] = 0$$

$$cov(a_{m,c_t}, a_{m-i,c_{t-i}}) = \begin{cases} \sigma_m^2, & i = 0, \\ 0, & i = 1, 2, \dots, \mathbb{S} \end{cases} \quad (2)$$

Assim, a_{t,c_t} é periodicamente independente e identicamente distribuído (i.i.d), com valor esperado zero e variância constante σ_m^2 .

A associação entre o valor do índice t da série temporal de ENA, em base mensal, a algum mês do ano, dado por m , será:

$$m = (t - 1) \% \mathbb{S} + 1 \quad (3)$$

O valor de m assume valores inteiros entre 1 e 12. A operação $a \% b$ representa o resto da divisão de a por b e \mathbb{S} representa o período, de valor igual a 12.

2.2.2.1. Identificação e Estimação de Modelos PAR

Apresentam-se a seguir os métodos de identificação e estimação clássicos do modelo PAR, indicados em (Hipel e McLeod, 1994). A identificação se dá através do emprego da função de autocorrelação parcial (FACP) da série em estudo. A estimação utiliza as equações de Yule-Walker, um estimador de momentos estatisticamente eficiente para uso no modelo PAR.

Para maiores informações acerca das modificações nos processo de identificação e estimação de um modelo quando se usa o histórico dinâmico, consultar (Ferreira, 2013).

2.2.2.1.1. Função de Autocorrelação (FAC) Periódica

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

$$\gamma_k^{(m)} = E[(z_t - \mu_m)(z_{t-k} - \mu_{m-k})] \quad (4)$$

Quando $k = 0$, a autocovariância periódica é simplesmente a variância, $\gamma_0^{(m)}$, de uma variável aleatória representante das observações do período m .

A variável padronizada, que é mais conveniente para se trabalhar quando comparada a $\gamma_k^{(m)}$, é a FAC periódica teórica, que é definida para cada período m como sendo:

$$\rho_k^{(m)} = \frac{\gamma_k^{(m)}}{\sqrt{\gamma_0^{(m)} \gamma_0^{(m-k)}}} \quad (5)$$

A equação acima permite analisar a dependência linear periódica da série em análise.

Multiplicando-se ambos os membros da eq. (1), que descreve a formulação do modelo PAR, pelo termo $\left(\frac{z^t - k - \mu_{m-k}}{\sigma_{m-k}}\right)$ e aplicando o valor esperado, o desenvolvimento da equação resultante levará à expressão apresentada abaixo.

$$\rho_k^{(m)} = \phi_1^m \rho_{k-1}^{(m-1)} + \phi_2^m \rho_{k-2}^{(m-2)} + \dots + \phi_{p_m}^m \rho_{k-p_m}^{(m-p_m)} \quad (6)$$

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

Uma vez que a FAC é simétrica em relação à defasagem zero ($k = 0$), necessita-se apenas plotar a amostra da FAC para o período m da defasagem $k \geq 0$. Para discernir valores significativamente diferentes de zero, uma vez que a FAC amostral é assintoticamente distribuída segundo uma $NID\left(0, \frac{1}{n}\right)$, isto é, normalmente e independentemente distribuída em qualquer *lag*, pode-se plotar o intervalo de confiança (IC) de 95%, dado por $\pm \frac{1,96}{\sqrt{n}}$.

A partir do exposto sobre o comportamento da FAC teórica para o modelo PAR, tem-se que, se a FAC amostral em análise apresenta atenuação para um período m , é necessário que um ou mais parâmetros AR sejam utilizados no modelo PAR a ser ajustado. Se nenhum valor da FAC é significativamente diferente de zero, então, pode-se modelar a série por ruído branco, fazendo $p_m = 0$.

2.2.2.1.2. Equações Yule-Walker Periódicas

Aplicando-se $k = 1, 2, \dots, p_m$ na equação da FAC, eq. (6), obtém-se as equações periódicas de *Yule-Walker* para o período m , como apresentado abaixo:

$$\begin{aligned}
\rho_1^{(m)} &= \phi_1^m \rho_0^{(m-1)} + \phi_2^m \rho_1^{(m-2)} + \dots + \phi_{p_m}^m \rho_{k-p_m-1}^{(m-p_m)} \\
\rho_2^{(m)} &= \phi_1^m \rho_1^{(m-1)} + \phi_2^m \rho_0^{(m-2)} + \dots + \phi_{p_m}^m \rho_{k-p_m-2}^{(m-p_m)} \\
&\vdots \\
&\vdots \\
\rho_{p_m}^{(m)} &= \phi_1^m \rho_{p_m-1}^{(m-1)} + \phi_2^m \rho_{p_m-2}^{(m-2)} + \dots + \phi_{p_m}^m \rho_0^{(m-p_m)}
\end{aligned} \tag{7}$$

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

$$\boldsymbol{\phi}^{(m)} = [\mathbf{P}_{p_m}^{(m)}]^{-1} \boldsymbol{\rho}^{(m)} \tag{8}$$

onde:

$$\boldsymbol{\phi}^{(m)} = \begin{bmatrix} \phi_1^{(m)} \\ \phi_2^{(m)} \\ \vdots \\ \phi_{p_m}^{(m)} \end{bmatrix}, \boldsymbol{\rho}^{(m)} = \begin{bmatrix} \rho_1^{(m)} \\ \rho_2^{(m)} \\ \vdots \\ \rho_{p_m}^{(m)} \end{bmatrix}, \mathbf{P}_{p_m}^{(m)} = \begin{bmatrix} \rho_0^{(m-1)} & \rho_1^{(m-2)} & \dots & \rho_{k-p_m-1}^{(m-p_m)} \\ \rho_1^{(m-1)} & \rho_0^{(m-2)} & \dots & \rho_{k-p_m-2}^{(m-p_m)} \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ \rho_{p_m-1}^{(m-1)} & \rho_{p_m-2}^{(m-2)} & \dots & \rho_0^{(m-p_m)} \end{bmatrix}$$

2.2.2.1.3. Função de Autocorrelação Parcial (FACP) Periódica

Uma vez que a FAC de um modelo PAR para um período m atenua para $p_m > 0$ e não trunca em uma defasagem k específica, é interessante definir uma função que exiba esse comportamento de corte, para fins de identificação do modelo.

Para um período m , a FACP periódica é definida como o último parâmetro AR do modelo de ordem p_m . Logo, nas equações de Yule-Walker, por definição, o parâmetro $\phi_{p_m}^{(m)}$ é a FACP periódica na defasagem p_m .

Fazendo $p_m = 1, 2, \dots$ na equação (7), por definição, a FACP periódica deve ser igual a zero para defasagens maiores que p_m , quando esta é a ordem do modelo AR naquele período.

A definição da FACP é derivada da definição do modelo PAR. Em particular, assumindo que o modelo AR para um período m é de ordem p_m , a FACP desse período será $\phi_{p_m}^{(m)}$. Fazendo p_m assumir os valores $1, 2, \dots$, a FACP é definida para os lags $1, 2, \dots$. Para estimar a FACP periódica, utiliza-se a equação (7), substituindo a variável teórica $\gamma_k^{(m)}$ pelo seu equivalente amostral.

Esse processo é realizado para cada um dos S períodos, estimando $\phi_i^{(m)}$ para $i = 1, 2, \dots, p_m$, para cada período m .

Por definição, se a FACP periódica amostral não é significativamente diferente de zero depois da defasagem p_m , há indicação de que a ordem do modelo AR ajustado à série para o período m é p_m . Se nenhum dos valores da FACP é significativamente diferente de zero, então faz-se $p_m = 0$.

Avanços na identificação de modelos PAR podem ser encontrados em (Matalas, 1967), (Stedinger, 2001) e (Oliveira, 2010).

2.2.2.2. Estimação da Distribuição de Ruídos

Uma das hipóteses do modelo apresentado é a distribuição dos ruídos. Em (Hipel e McLeod, 1994), diz-se que, por vezes é conveniente assumir que os erros, a_{t,c_t} , seguem uma distribuição normal com média zero e variância σ_{m,c_t} constante. Entretanto, as simulações que adotam esse tipo de distribuição podem gerar valores negativos de afluência, oriundos de uma sequência de erros negativos.

Assim, para a abordagem PAR no modelo vigente, que necessita gerar afluências sempre positivas, dados os pré-requisitos da PDDE, ajusta-se uma distribuição lognormal de três parâmetros ao $a_{t,c_t} \sim \ln N(\mu_{\varepsilon_t}, \sigma_{\varepsilon_t}, \Delta_t)$, sendo o terceiro parâmetro, Δ_t , o deslocamento da distribuição.

Fazendo $z_t \geq 0$ na eq. (1), no intuito de obter previsões positivas, e isolando o ruído do lado esquerdo da inequação obtém-se:

$$a_t \geq -\frac{\mu_{m,c_t}}{\sigma_{m,c_t}} - \sum_{i=1}^{p_{m,c_t}} \phi_i^{m,c_t} \left(\frac{z_{t-i,c_{t-i}} - \mu_{m-i,c_{t-i}}}{\sigma_{m-i,c_{t-i}}} \right) \quad (9)$$

Faz-se o deslocamento Δ_t igual ao lado direito da inequação (9), isto é, o deslocamento é igual ao resíduo mínimo para que a previsão da afluência seja positiva. A regra para o cálculo de Δ_t faz com que este parâmetro dependa de afluências passadas, resultando apenas em valores positivos. Entretanto, a dependência do passado criada, faz com que a hipótese de independência temporal do erro, exigida pelo algoritmo PDDE, seja violada.

2.2.2.3. Verificação de Modelos PAR

A definição teórica de um modelo AR assume que os resíduos presentes nela são identicamente e independentemente distribuídos (i.i.d), com média nula e variância constante (homocedástico). A fim de se obterem estimadores de máxima verossimilhança para os parâmetros do modelo, assume-se que os resíduos seguem uma distribuição específica. Em particular, os estimadores para o modelo PAR foram desenvolvidos considerando que os resíduos desse modelo seguissem uma distribuição lognormal.

A partir do exposto acima, para comprovar a adequação do modelo PAR ajustado, deve-se analisar algumas propriedades dos resíduos de cada período. Em particular, os resíduos devem ser descorrelacionados, homocedásticos e seguir uma distribuição lognormal.

2.3. Geração de Cenários

Uma série temporal pode ser compreendida como uma das realizações de um processo estocástico. Um modelo que represente tal processo deve ser capaz de simular séries temporais diferentes, mas igualmente prováveis à série original, permitindo melhor investigação acerca do futuro desconhecido.

Com efeito, afirma-se em (Marco, Harboe e Salas, 1993) que a principal razão matemática para se modelar um processo hidrológico, correlacionado no espaço e no tempo, é o potencial do modelo de aumentar a acurácia das soluções e decisões que envolvem o uso de recursos hídricos, através da simulação de cenários de entradas e saídas do sistema avaliado.

Um processo estocástico hidrológico normalmente possui uma ou mais componentes estocásticas independentes. O modelo PAR, por exemplo, possui o erro aditivo como componente aleatória ao qual se atribui uma distribuição de probabilidade (na seção 2.2.2.2 ajusta-se a distribuição lognormal ao ruído). Assim, os métodos de simulação geram amostras dessas distribuições para que estas sejam usadas como as variáveis aleatórias para a construção dos vários cenários.

A execução de uma simulação pode ser sumariada nas seguintes etapas:

1. Gerar variáveis independentemente distribuídas que seguem a mesma distribuição uniforme no intervalo de (0,1);
2. Transformar as variáveis que seguem a distribuição uniforme em variáveis que seguem a distribuição escolhida para os resíduos e que são temporalmente descorrelacionadas.

A seguir apresentam-se a modelagem dos ruídos para simulação de cenários PAR, a técnica Monte Carlo e, por fim, o processo de construção da árvore de cenários para a PDDE.

2.3.1.

Modelagem dos Ruídos para Simulação de Cenários a partir do Modelo PAR

Seja a variável aleatória $\varepsilon_{t,s} \sim N(\mu_{\varepsilon_{t,s}}, \sigma_{\varepsilon_{t,s}})$, definida para os períodos $t = 1, 2, \dots, T$ e para os subsistemas $s = 1, 2, \dots, n_{subs}$. Define-se $a_{t,c_t,s}$ como a exponencial de $\varepsilon_{t,s}$ deslocada por um valor $\Delta_{t,s}$:

$$a_{t,c_t,s} = e^{\varepsilon_{t,s}} + \Delta_{t,s} \quad (10)$$

Note que a variável $a_{t,c_t,s}$ desta vez está em função do subsistema s . Desta forma, $a_{t,c_t,s}$ segue uma distribuição lognormal de três parâmetros, $a_{t,c_t,s} \sim \ln N(\mu_{\varepsilon_{t,s}}, \sigma_{\varepsilon_{t,s}}, \Delta_{t,s})$, sendo o terceiro parâmetro, $\Delta_{t,s}$, o deslocamento da distribuição.

Definem-se as componentes do vetor $\mathbf{b}_t = [b_{t,1} \ b_{t,2} \ \dots \ b_{t,n_{subs}}]^T$ como sendo:

$$b_{t,s} = \frac{\varepsilon_{t,s} - \mu_{\varepsilon_{t,s}}}{\sigma_{\varepsilon_{t,s}}} \quad (11)$$

Logo, o vetor \mathbf{b}_t segue uma distribuição normal multivariada, de dimensão $n_{subs} \times 1$, com valor esperado zero e matriz de covariância igual à identidade, isto é, $\mathbf{b}_t \sim \mathcal{N}(0, \mathbf{I}_{n_{subs}})$.

Desta forma, a variável $\varepsilon_{t,s}$ pode ser reescrita em função de uma componente do vetor \mathbf{b}_t como apresenta a eq. (12):

$$\varepsilon_{t,s} = b_{t,s} \sigma_{\varepsilon_{t,s}} + \mu_{\varepsilon_{t,s}} \quad (12)$$

Substituindo-se o resultado da eq. (12) em (10), tem-se:

$$a_{t,c_t,s} = e^{b_{t,s}\sigma_{\varepsilon_t,s} + \mu_{\varepsilon_t,s}} + \Delta_{t,s}$$

Em (Charbeneau, 1978), definem-se os momentos para os parâmetros da distribuição, replicados abaixo:

$$\sigma_{\varepsilon_t}^2 = \ln \theta \quad (13)$$

$$\mu_{\varepsilon_t} = \frac{1}{2} \ln \frac{\sigma_{a_t}^2}{\theta^2 - \theta} \quad (14)$$

Fazendo $z_t \geq 0$ na eq. (1), no intuito de obter previsões positivas, e isolando o ruído do lado esquerdo da inequação obtém-se:

$$a_t \geq -\frac{\mu_{m,c_t}}{\sigma_{m,c_t}} - \sum_{i=1}^{p_{m,c_t}} \phi_i^{m,c_t} \left(\frac{z_{t-i,c_{t-i}} - \mu_{m-i,c_{t-i}}}{\sigma_{m-i,c_{t-i}}} \right) \quad (15)$$

Faz-se o deslocamento Δ_t igual ao lado direito da inequação (15), isto é, o deslocamento é igual ao resíduo mínimo para que a previsão da afluência seja positiva. Assim, o parâmetro θ pode ser calculado através da equação:

$$\theta = 1 + \left(\frac{\sigma_{a_t}}{\Delta^2} \right) \quad (16)$$

Adicionalmente, para inserir no modelo a correlação espacial mensal existente entre as vazões afluentes aos reservatórios equivalentes, utiliza-se a estratégia de transformar resíduos não correlacionados em resíduos espacialmente correlacionados, W_t . Para tal, W_t é definido como:

$$W_t = D b_t$$

Onde D é uma matriz quadrada de dimensão n_{subs} , chamada matriz de carga, cuja estimação é detalhada em (Oliveira, 2010).

2.3.2. Técnica Monte Carlo

Suponha um modelo hidrológico com uma componente aleatória, como no caso do resíduo aditivo do modelo PAR. A simulação Monte Carlo, proposta em (Metropolis e Ulam, 2008), consiste na geração de uma amostra aleatória obtida a partir da distribuição de probabilidade ajustada à componente estocástica do

modelo. Esta técnica modela o comportamento do sistema usando probabilidades, portanto a única exigência é que o sistema seja descrito em termos de funções de densidade de probabilidade.

A seguir, apresentam-se as etapas da execução de uma simulação Monte-Carlo:

1. Definir uma distribuição de probabilidades para a componente aleatória do modelo do processo sob investigação.
2. Gerar amostra aleatória a partir da distribuição escolhida no passo 1. Nesta etapa, é necessário a seleção de um gerador de número aleatório apropriado para modelar a distribuição de probabilidade observada.
3. Executar o cálculo determinístico da variável desejada, a partir dos números aleatórios gerados.
4. Agregar os resultados dos cálculos individuais ao resultado final: o processo de agregação depende do objetivo da simulação, entretanto pode ser tão direto quanto calcular a média dos resultados simulados.

2.3.3.

Construção dos Cenários para o Modelo de Otimização

Conforme apresentado anteriormente, o algoritmo PDDE, que define a política ótima de operação ao longo do horizonte de planejamento, se baseia em uma árvore finita de cenários, que deve ser representativa do processo estocástico que a gerou, ao mesmo tempo em que tenha cardinalidade reduzida.

Neste contexto, a literatura apresenta vários métodos para a construção da árvore de cenários. Em (Oliveira, 2013), utiliza-se a técnica *bootstrap* conjugada com a simulação Monte Carlo para geração de cenários. Em (Penna, 2009) propõe-se a aplicação de amostragem descritiva no modelo de geração de cenários e aplicação de métodos estatísticos para agregação de elementos. A seguir, apresenta-se a alternativa adotada na implementação dos modelos descritos nesta dissertação.

A partir de uma amostra de n_{abt} elementos, por período, da componente aleatória do modelo de afluências, é possível gerar os cenários utilizados nas etapas *forward* e *backward* do processo de otimização. Essa amostra pode ser obtida, por exemplo, via método Monte Carlo ou por sorteio aleatório simples.

Com as aberturas sorteadas, calculam-se os valores das afluições em t , substituindo-se a componente aleatória na formulação do modelo de afluições pelo resíduo sorteado.

Os cenários *forward* são gerados de forma paralela, a partir da amostra *backward*. A cada período, sorteiam-se n_t^{fwd} aberturas, uma para cada uma das n_t^{fwd} séries *forward* consideradas.

Segundo (Penna, 2009), os cenários gerados podem ou não ser condicionados aos valores iniciais da série. Neste trabalho, optou-se por não condicionar a série gerada aos valores iniciais. Foram utilizadas as médias da série histórica (MLT) como valores iniciais para recursão dos primeiros cenários gerados, definidos dentro de um período chamado de pré-estudo. Os cenários do pré-estudo são descartados ao final do processo de geração.

Para a geração de cenários *backward*, em um período t , para cada uma das séries *forward* previamente definidas, utiliza-se o passado determinado pela mesma e gera-se um cenário para cada uma das aberturas definidas para t , que formam a amostra de resíduos (matriz de aberturas). O processo descrito acima é representado na Figura 6:

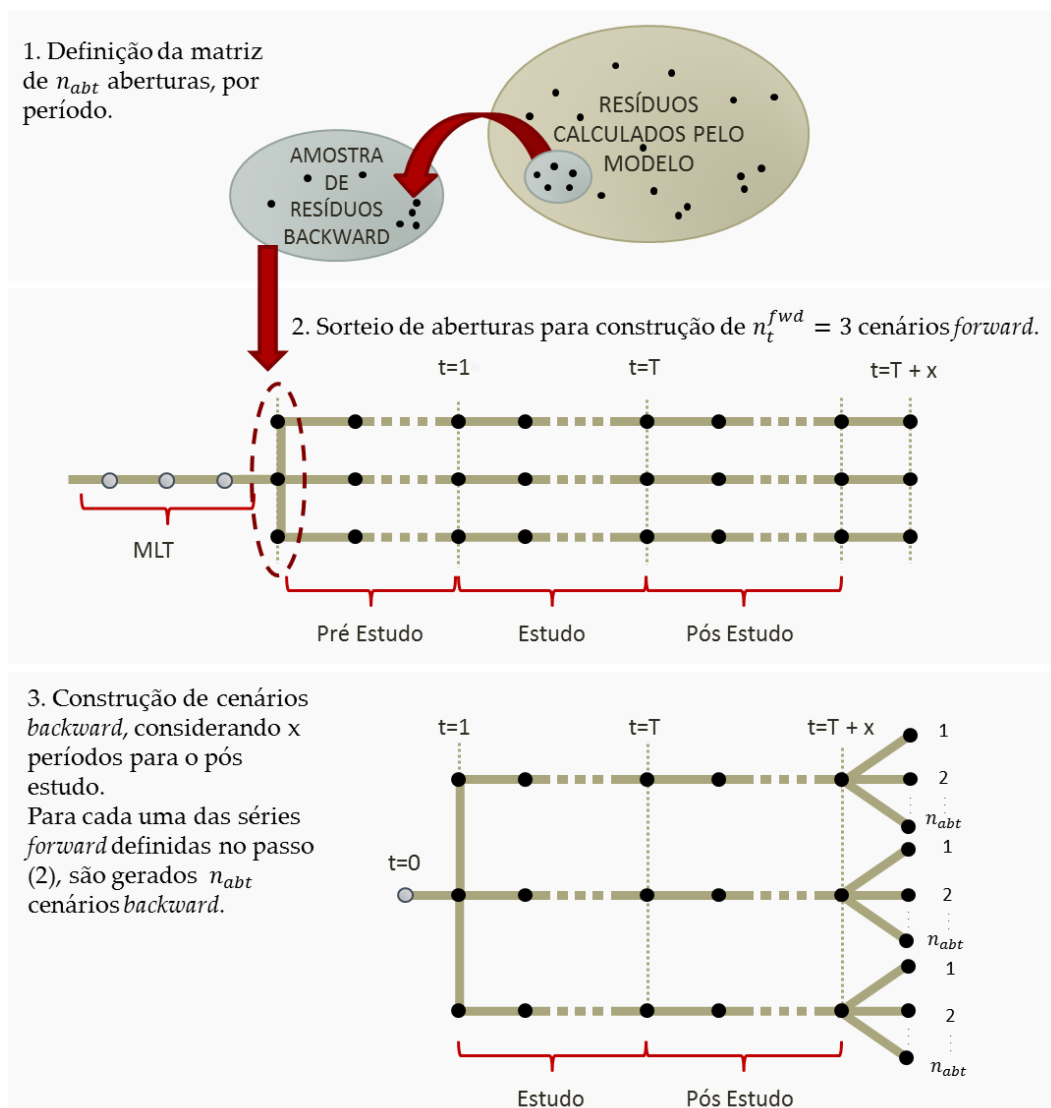


Figura 6: Construção da árvore de cenários e geração de cenários para as etapas *forward* e *backward*. Fonte: Elaborada pela autora

2.4.

Critérios de Avaliação de Desempenho de Modelos

A avaliação de desempenho de um modelo é feita com base em propriedades que são consideradas relevantes para sua aplicação. Em particular, um modelo de ENA útil deve preservar importantes características do processo estocástico que o inspirou. O desempenho desse modelo é medido pela sua capacidade em reproduzir a distribuição de probabilidade de algumas variáveis relevantes para o Planejamento Energético de Médio Prazo.

É importante notar que qualquer característica do processo estocástico que compõe a definição do modelo de ENA's não pode ser utilizada para verificar a adequação do modelo, uma vez que é sabido que tal característica será preservada. Tal comparação deve ser feita apenas para validar o programa computacional desenvolvido, não o modelo (Stedinger, 2001).

A seleção de variáveis é um assunto complexo. Alguns autores já se preocuparam em selecionar tais variáveis para a aplicação do planejamento energética, bem como em sugerir testes para tais variáveis (Kelman e Pereira, 1977), (Oliveira, 2010). Em (Kelman e Pereira, 1977), conclui-se que as variáveis relacionadas às flutuações das vazões em torno dos respectivos valores esperados são interessantes para análises energéticas, uma vez que permitem a avaliação da capacidade dos modelos em reproduzir períodos de escassez e abundância do recurso hídrico, semelhantes àqueles presentes na série histórica observada.

Em (Kelman e Pereira, 1977) definem-se os eventos “sequência positiva”, compreendendo o conjunto de valores z_t superiores aos correspondentes \bar{z}_t , precedidos e sucedidos por valores de z_t inferiores aos correspondentes \bar{z}_t (Intervalos $[t_1, t_2]$, $[t_3, t_4]$, $[t_5, t_6]$ e $[t_7, t_8]$ na Figura 7); sequência negativa, conceito análogo ao evento anterior, corresponde aos intervalos $[t_0, t_1]$, $[t_2, t_3]$, $[t_4, t_5]$ e $[t_6, t_7]$ na Figura 7.

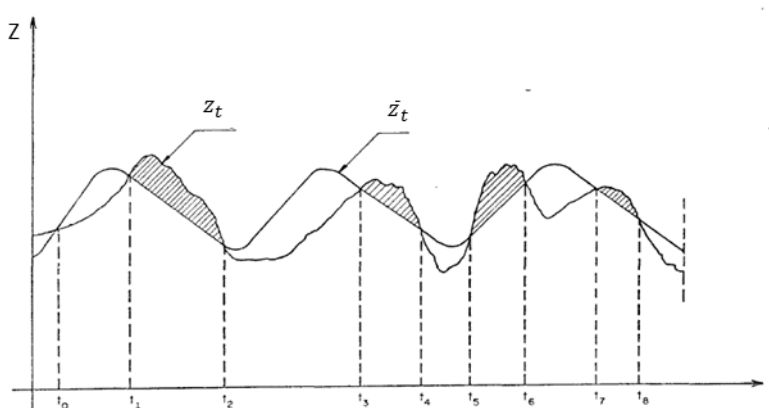


Figura 7: Ilustração do conceito de sequência. Fonte: Adaptado de (Kelman e Pereira, 1977)

Os valores \bar{z}_t compõem uma referência para a extração das sequências positivas e negativas dos cenários gerados. Uma referência intuitiva é a média de cada período.

Associadas às sequências estão as variáveis aleatórias de interesse para avaliação dos modelos de ENA:

- Comprimento de sequência
- Soma de sequência
- Intensidade de sequência

O comprimento (R_l) de uma sequência positiva ou negativa é a duração (em meses) deste evento:

$$R_l = t_f - t_i \quad (17)$$

Onde t_f é o limite superior do intervalo e t_i o limite inferior do intervalo de tempo da sequência.

A soma (R_s) de uma sequência positiva ou negativa representa o montante de energia natural afluyente, em MW médio, acima ou abaixo da média, respectivamente, durante a ocorrência do evento:

$$R_s = \sum_{t=t_i}^{t_f} (z_t - \bar{z}_t) \quad (18)$$

A intensidade (R_i) de uma sequência é definida pela razão entre a soma desta sequência e seu comprimento, isto é, representa a proporção do montante de ENA acima ou abaixo da referência no tempo:

$$R_i = \frac{R_s}{R_l} \quad (19)$$

A partir do exposto, obtêm-se amostras da série histórica e das séries sintéticas, para avaliação de seis variáveis aleatórias: comprimento, soma e intensidade de sequências positiva e negativa. Assim, cada uma das variáveis aleatórias citadas, obtidas das diferentes séries, deve ser devidamente comparada.

A comparação das variáveis será feita através dos testes de adequação de modelo. Os testes estatísticos de aderência são métodos não paramétricos e objetivam verificar se as séries sintéticas geradas seguem uma distribuição

particular. Neste trabalho, serão utilizados os testes qui-quadrado e o teste de Kolmogorov-Smirnov, explanados nas seções seguintes.

2.4.1.

Teste de Kolmogorov-Smirnov (K-S)

O teste de K-S é um teste não-paramétrico e determina se duas amostras independentes são originárias de uma mesma população contínua. A estatística K-S avalia a distância ($D_{n,m}$) entre as distribuições acumuladas empíricas de cada amostra, como segue:

$$D_{n,m} = \sup_x |F_{1,n}(x) - F_{2,m}(x)|$$

onde $F_{1,n}(x)$ e $F_{2,m}(x)$ são as distribuições acumuladas empíricas de cada uma das amostras comparadas (o histórico e a amostra o cenário gerado) e \sup_x é o supremo do conjunto de distâncias.

Rejeita-se a hipótese nula ao nível de significância de 5% se:

$$D_{n,m} > c(\alpha) \sqrt{\frac{n+m}{nm}}$$

Sendo n e m o número de elementos da primeira e da segunda amostra, respectivamente. O valor de $c(\alpha)$ é tabelado e vale, aproximadamente, 1.36 para $\alpha = 0,05$.

2.4.2.

Teste Qui-Quadrado

O teste Qui-Quadrado compara se duas amostras foram obtidas a partir de uma mesma população discreta.

A estratégia deste teste para comparar as distribuições é categorizar as amostras. Definem-se k intervalos da variável amostrada, igualmente prováveis e mutuamente exclusivos. Em seguida, calculam-se as proporções observadas em cada categoria A_i , sendo $i = 1, \dots, k$ o índice do grupo, para cada uma das duas amostras. Por fim, testa-se a seguinte hipótese:

$$H_0: p_{i,1} = p_{i,2} , \text{ para } i = 1, 2, \dots, k$$

O valor da estatística de teste é dado por:

$$Q_{k-1} = \sum_{j=1}^2 \sum_{i=1}^k \frac{(OBSERVADO_{i,j} - ESPERADO_{i,j})^2}{ESPERADO_{i,j}} \quad (20)$$

Sob a hipótese nula, a distribuição amostral de Q_{k-1} é aproximadamente uma qui-quadrada com $k - 1$ graus de liberdade. Assim, a hipótese é rejeitada se $Q_{k-1} \geq \chi^2_{\alpha}(k - 1)$, para um nível de significância α . Se a hipótese de proporções não for rejeitada, então a hipótese original, acerca da igualdade das distribuições, também não é rejeitada.

A literatura aponta algumas condições para a aplicação prática do teste qui-quadrado, a fim de garantir que sua utilização seja apropriada. Se houver mais de 20% das classes com a frequência esperada $ESPERADO_{i,j} < 5$, ou mais de uma classe com a frequência esperada $ESPERADO_{i,j} < 1$, deve-se reduzir número de classes e reiniciar os testes.

O Modelo Periódico Vetorial Autorregressivo com Erro Multiplicativo – PVARm

Modelos com erros multiplicativos para processos que geram sempre valores positivos são especificados como o produto entre um fator de escala, que pode ser considerado um valor determinístico, condicionado aos valores passados da série, e um ruído, com suporte positivo e valor esperado unitário. A literatura acerca de tais modelos é relativamente recente.

Acerca deste tema, os trabalhos (Lanne, 2006), (Hautsch, Malec e Schienle, 2014), (Cipollini, Engle e Gallo, 2006) e (Bodnar e Hautsch, 2016) trazem propostas de modelos com erros multiplicativos aplicados a séries de valores positivos relacionadas ao mercado financeiro. Já no setor elétrico, as teses (Braga, 2011) e (Ferreira, 2013) são iniciativas nacionais que propuseram modelos periódicos de distribuição gama para os erros multiplicativos de séries de ENA. Em (Ribeiro, Baldioti e Souza, 2017), um modelo PAR univariado com erro multiplicativo é utilizado para geração de cenários para o caso do Planejamento Energético de Médio Prazo.

Em (Cabral, 2016) propõe-se a utilização do modelo PAR multivariado com erro multiplicativo (PVARm) para geração de cenários de afluência, com o intuito de contornar a dificuldade encontrada pelo uso do PAR, no que concerne à simulação de afluências positivas.

A proposta, do ponto de vista da otimização estocástica, atende aos requisitos necessários para aplicação na PDDE para solução do problema do planejamento: a equação do modelo PVARm relaciona as afluências de maneira linear e garante a independência temporal dos ruídos multiplicativos. Adicionalmente, a correlação entre as afluências dos diferentes subsistemas é representada no próprio modelo.

A seguir, apresenta-se o modelo PVARm e as modificações propostas para (i) adequar a estimação à mudança de configurações do parque gerador; (ii)

identificar a ordem dos modelos; e (iii) gerar cenários em consonância com o processo de otimização ao qual podem ser aplicados como insumo.

3.1.1. Formulação do Modelo PVARm

Considere \mathbf{z}_{t,c_t} como um processo periódico, com período \mathbb{S} . Um modelo Periódico Vetorial Autorregressivo com erro multiplicativo (PVARm) de ordem $\mathbf{p} = (p_{1,c_t}, p_{2,c_t}, \dots, p_{\mathbb{S},c_t})$ é definido para um período m , com $m = 1, 2, \dots, 12$, na configuração c_t , com $c_t \in \mathcal{C} = \{1, 2, \dots, n_{conf}\}$, onde \mathcal{C} é o conjunto de índices das n_{conf} configurações diferentes, como segue:

$$\mathbf{z}_{t,c_t} = \left(\boldsymbol{\zeta}_{m,c_t} + \sum_{i=1}^{p_{m,c_t}} [\Phi_{m,c_t,i}] \mathbf{z}_{t-i,c_{t-i}} \right) \bullet \boldsymbol{\eta}_{t,c_t} \quad (21)$$

Sendo:

\mathbf{z}_{t,c_t} : série de ENA vetorial, com $t = 1, \dots, T$;

$[\Phi_{m,c_t,i}]$: matriz de parâmetros para lag i e período m na configuração c_t ;

$\boldsymbol{\zeta}_{m,c_t}$: vetor de nível para o período m na configuração c_t ;

$\boldsymbol{\eta}_{t,c_t}$: vetor de erros multiplicativos³ periodicamente i.i.d., onde:

$$\begin{aligned} E[\boldsymbol{\eta}_t] &= \mathbb{1}; \\ VAR[\boldsymbol{\eta}_t] &= \Sigma_t = VAR[\boldsymbol{\eta}_{t+\mathbb{S}}] \end{aligned} \quad (22)$$

Sendo: $\mathbb{1} = (1, 1, \dots, 1)^\top$ e Σ_t a matriz de covariância do vetor de erros $\boldsymbol{\eta}_t$.

O valor de m associa o índice t a algum mês do ano através da relação exposta na equação (3).

³ O operador \bullet representa o produto de Hadamard. Sejam $[A]$ e $[B]$ matrizes de dimensão $m \times n$, o produto de Hadamard das duas matrizes é igual a uma matriz $[C]$, também de dimensão $m \times n$, cujo elemento da linha i e coluna j , representado por $c_{i,j}$, é dado por:

$$[A] \bullet [B] = [C] \rightarrow c_{i,j} = a_{i,j} b_{i,j}$$

Assim como para o modelo PAR, o modelo PVARm também é calculado com o histórico dinâmico, recalculado de acordo com a entrada em operação de novas usinas hidrelétricas no sistema. Toda vez que esta alteração ocorre, os modelos dos próximos meses devem ser recalculados com base no novo histórico.

Note que o modelo PVARm pode relacionar as afluições atuais de um determinado subsistema (cada componente do vetor \mathbf{z}_{t,c_t} representa a ENA de um subsistema) com as afluições passadas de todos os subsistemas ($\mathbf{z}_{t-i,c_{t-i}}$), incorporando a correlação espacial entre os reservatórios equivalentes no próprio modelo.

Define-se como previsão de \mathbf{z}_t , dadas as afluições passadas, a equação:

$$PREV_{t,c_t} = \zeta_{m,c_t} + \sum_{i=1}^{p_m} [\Phi_{m,c_t,i}] \mathbf{z}_{t-i,c_t} \quad (23)$$

Adota-se como premissa a conjectura acerca da causalidade do modelo PVARm (Cabral, 2016), o que torna válida a independência entre as afluições passadas e o erro atual. Desta forma, a esperança condicional de \mathbf{z}_t , dadas as afluições até o período $t - 1$, isto é, dado $\mathbf{z}_{[t-1]}$, será:

$$E[\mathbf{z}_t | \mathbf{z}_{[t-1]}] = \zeta_m + \sum_{i=1}^{p_m} [\Phi_{m,i}] \mathbf{z}_{t-i} \quad (24)$$

O que torna a média condicional igual à previsão. A variância condicional é dada pela equação (25):

$$VAR[\mathbf{z}_t | \mathbf{z}_{[t-1]}] = PREV_t PREV_t^T \bullet \Sigma_t \quad (25)$$

Isto é, a variabilidade é proporcional ao produto cruzado das previsões.

3.1.2. Estimação dos Parâmetros

Assumindo que qualquer valor de afluição positiva possa ocorrer, mostra-se em (Cabral, 2016) que é possível que uma determinada combinação de

afluências e um coeficiente negativo da matriz $[\Phi_{m,c_t,i}]$ ou uma componente do vetor de nível ζ_m negativa possa produzir um valor de afluência negativo. Assim, restringir os parâmetros a valores não negativos é condição necessária e suficiente para geração de afluências sempre positivas.

O método dos mínimos quadrados é um método de estimação de parâmetros que busca minimizar o somatório dos quadrados das diferenças entre cada um dos valores observado e previsto da série. O método dos mínimos quadrados com restrição de não negatividade nos coeficientes garante a geração de afluências sempre positivas. Outros métodos como as equações de Yule-Walker (seção 2.2.2.1.2) são adequados para outras aplicações que não necessitem da garantia de geração de valores positivos.

A estimação para um modelo que se utiliza do histórico dinâmico deve considerar que, toda vez que houver uma mudança de configuração em um determinado mês m , deve-se calcular um novo modelo que será função do período m , e da configuração a qual ele está associada. Suponha a disponibilidade de uma série z_t , com $t = 1, 2, \dots, T - 1$. Deseja-se modelar o período $t = T$, associado ao mês m , com a configuração c_T . A previsão deste modelo é dada pela equação (23), desta vez, associada à configuração do parque gerador designada por c_T . O objetivo do método é calcular os parâmetros, restritos a valores positivos, através da minimização dos quadrados das diferenças entre os valores observados e os valores previstos das afluências,

O vetor de observações do processo \mathbf{z}_{t,c_T} , para um período t qualquer associado à configuração c_T , é dado por:

$$\mathbf{z}_{t,c_T} = \begin{bmatrix} z_{t,c_T,1} \\ z_{t,c_T,2} \\ \vdots \\ \vdots \\ z_{t,c_T,n_{subs}} \end{bmatrix} \quad (26)$$

Onde n_{subs} é o número de subsistemas (ou REEs) considerados. Inicialmente, agrupam-se na matriz Z_{m,c_T} os vetores das observações da série temporal associadas ao mês m , baseadas na configuração c_T :

$$Z_{m,c_T} = [\mathbf{z}_{m,c_T} \quad \mathbf{z}_{m+s,c_T} \quad \cdot \quad \cdot \quad \cdot \quad \mathbf{z}_{m+q_s,c_T}] \quad (27)$$

sendo q_m o maior inteiro tal que $m + q_m \mathbb{S} \leq (T - 1)$.

O previsor do modelo PVARm é definido pela equação (23), que é apresentada a seguir sob a forma matricial. Por uma questão de simplicidade, a formulação que segue considera $p = p_{m,c_T}$, como a ordem do modelo para o mÊs m na configuração c_T :

$$\begin{aligned}
 & \begin{bmatrix} \zeta_{m,c_T,1} \\ \zeta_{m,c_T,2} \\ \vdots \\ \zeta_{m,c_T,n_{sub}} \end{bmatrix} + \begin{bmatrix} \phi_{m,c_T,1}^{11} & \phi_{m,c_T,1}^{12} & \cdot & \cdot & \cdot & \phi_{m,c_T,1}^{1,n_{sub}} \\ \phi_{m,c_T,1}^{21} & \phi_{m,c_T,1}^{22} & \cdot & \cdot & \cdot & \phi_{m,c_T,1}^{2,n_{sub}} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ \phi_{m,c_T,1}^{n_{sub},1} & \phi_{m,c_T,1}^{n_{sub},2} & \cdot & \cdot & \cdot & \phi_{m,c_T,1}^{n_{sub},n_{sub}} \end{bmatrix} \begin{bmatrix} Z_{t-1,c_T-1,1} \\ Z_{t-1,c_T-1,2} \\ \cdot \\ \cdot \\ \cdot \\ Z_{t-1,c_T-1,n_{sub}} \end{bmatrix} \\
 & + \dots + \begin{bmatrix} \phi_{m,c_T,p}^{11} & \phi_{m,c_T,p}^{12} & \cdot & \cdot & \cdot & \phi_{m,c_T,p}^{1,n_{sub}} \\ \phi_{m,c_T,p}^{21} & \phi_{m,c_T,p}^{22} & \cdot & \cdot & \cdot & \phi_{m,c_T,p}^{2,n_{sub}} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & \cdot & & \cdot \\ \cdot & \cdot & & & \cdot & \cdot \\ \phi_{m,c_T,p}^{n_{sub},1} & \phi_{m,c_T,p}^{n_{sub},2} & \cdot & \cdot & \cdot & \phi_{m,c_T,p}^{n_{sub},n_{sub}} \end{bmatrix} \begin{bmatrix} Z_{t-p,c_T-p,1} \\ Z_{t-p,c_T-p,2} \\ \cdot \\ \cdot \\ \cdot \\ Z_{t-p,c_T-p,n_{sub}} \end{bmatrix} \quad (28)
 \end{aligned}$$

Afim de simplificar a representação, unem-se as matrizes de coeficientes, como se segue:

$$\begin{bmatrix} \zeta_{m,c_T,1} & \phi_{m,c_T,1}^{11} & \phi_{m,c_T,1}^{12} & \cdot & \cdot & \cdot & \phi_{m,c_T,1}^{1,n_{sub}} & \cdot & \cdot & \cdot & \phi_{m,c_T,p}^{11} & \phi_{m,c_T,p}^{12} & \cdot & \cdot & \cdot & \phi_{m,c_T,p}^{1,n_{sub}} \\ \zeta_{m,c_T,2} & \phi_{m,c_T,1}^{21} & \phi_{m,c_T,1}^{22} & \cdot & \cdot & \cdot & \phi_{m,c_T,1}^{2,n_{sub}} & \cdot & \cdot & \cdot & \phi_{m,c_T,p}^{21} & \phi_{m,c_T,p}^{22} & \cdot & \cdot & \cdot & \phi_{m,c_T,p}^{2,n_{sub}} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \zeta_{m,c_T,n_{sub}} & \phi_{m,c_T,1}^{n_{sub},1} & \phi_{m,c_T,1}^{n_{sub},2} & \cdot & \cdot & \cdot & \phi_{m,c_T,1}^{n_{sub},n_{sub}} & \cdot & \cdot & \cdot & \phi_{m,c_T,p}^{n_{sub},1} & \phi_{m,c_T,p}^{n_{sub},2} & \cdot & \cdot & \cdot & \phi_{m,c_T,p}^{n_{sub},n_{sub}} \end{bmatrix} \begin{bmatrix} 1 \\ Z_{t-1,c_T-1,1} \\ Z_{t-1,c_T-1,2} \\ \cdot \\ \cdot \\ \cdot \\ Z_{t-1,c_T-1,n_{sub}} \\ \cdot \\ \cdot \\ \cdot \\ Z_{t-p,c_T-p,1} \\ Z_{t-p,c_T-p,2} \\ \cdot \\ \cdot \\ \cdot \\ Z_{t-p,c_T-p,n_{sub}} \end{bmatrix}$$

O modelo é reescrito através da identificação dos blocos de matrizes:

$$[\zeta_{m,c_T} \quad \Phi_{m,c_T,1} \quad \cdot \quad \cdot \quad \cdot \quad \Phi_{m,c_T,p}] \begin{bmatrix} 1 \\ \mathbf{Z}_{t-1,c_{T-1}} \\ \cdot \\ \cdot \\ \mathbf{Z}_{t-p,c_{T-p}} \end{bmatrix} \quad (29)$$

Para $t = m, m + \mathbb{S}, \dots, m + q_m \mathbb{S}$.

Agrupam-se as observações associadas às regressões $t - 1, \dots, t - p$, sendo essas associadas às configurações anteriores ao período T modelado (c_{T-1}, \dots, c_{T-p}):

$$[\zeta_{m,c_T} \quad \Phi_{m,c_T,1} \quad \cdot \quad \cdot \quad \cdot \quad \Phi_{m,c_T,p}] \begin{bmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 \\ \mathbf{Z}_{m-1,c_{T-1}} & \mathbf{Z}_{(m+\mathbb{S})-1,c_{T-1}} & \cdot & \cdot & \cdot & \mathbf{Z}_{(m+q_m\mathbb{S})-1,c_{T-1}} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{Z}_{m-p,c_{T-p}} & \mathbf{Z}_{(m+\mathbb{S})-p,c_{T-p}} & \cdot & \cdot & \cdot & \mathbf{Z}_{(m+q_m\mathbb{S})-p,c_{T-p}} \end{bmatrix} \quad (30)$$

Note que cada linha da matriz de observações históricas corresponde a uma configuração diferente do sistema. Utilizando a seguinte notação:

$$\Phi_{m,c_T} = [\zeta_{m,c_T} \quad \Phi_{m,c_T,1} \quad \cdot \quad \cdot \quad \cdot \quad \Phi_{m,c_T,p}] \quad (31)$$

$$\mathbb{Z}_{m,c_T} = \begin{bmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 \\ \mathbf{Z}_{m-1,c_{T-1}} & \mathbf{Z}_{(m+\mathbb{S})-1,c_{T-1}} & \cdot & \cdot & \cdot & \mathbf{Z}_{(m+q_m\mathbb{S})-1,c_{T-1}} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \mathbf{Z}_{m-p,c_{T-p}} & \mathbf{Z}_{(m+\mathbb{S})-p,c_{T-p}} & \cdot & \cdot & \cdot & \mathbf{Z}_{(m+q_m\mathbb{S})-p,c_{T-p}} \end{bmatrix} \quad (32)$$

$$\mathbf{H}_{m,c_T} = [\epsilon_{m,c_T} \quad \epsilon_{m+\mathbb{S},c_T} \quad \cdot \quad \cdot \quad \cdot \quad \epsilon_{m+q_m\mathbb{S},c_T}] \quad (33)$$

e a notação inserida pela equação (27), a matriz com os vetores de erro, representada por \mathbf{H}_{m,c_T} , pode ser reescrita como:

$$\mathbf{H}_{m,c_T} = \mathbf{Z}_{m,c_T} - \Phi_{m,c_T} \mathbb{Z}_{m,c_T} \quad (34)$$

A parcela $\Phi_{m,c_T} \mathbb{Z}_{m,c_T}$ representa o valor previsto da matriz \mathbf{Z}_{m,c_T} e a expressão $\mathbf{Z}_{m,c_T} - \Phi_{m,c_T} \mathbb{Z}_{m,c_T}$ representa a matriz das diferenças entre os valores

observados e previstos para o período m , na configuração c_T . O problema de otimização equivalente será:

$$\min_{\Phi_1, \dots, \Phi_S \geq 0} \sum_{m=1}^S \|Z_{m,c_T} - \Phi_{m,c_T} \mathbb{Z}_{m,c_T}\|^2 \quad (35)$$

onde $\|X\| = \sqrt{\sum_{i=1}^a \sum_{j=1}^b x_{i,j}^2}$ é chamada norma de Frobenius para a matriz $X \in \mathbb{R}^{a \times b}$.

Uma vez que $m = 1, 2, \dots, S$ são desacoplados, podemos escrever:

$$\sum_{m=1}^S \min_{\Phi_m \geq 0} \|Z_{m,c_T} - \Phi_{m,c_T} \mathbb{Z}_{m,c_T}\|^2 \quad (36)$$

Sabendo que a norma de uma matriz é igual à soma das normas de suas linhas, reescreve-se o problema em função da soma de suas linhas, o que resultará em problemas menores:

Sejam:

$$\begin{aligned} \mathbf{z}_{m,c_T,1}^T &= [z_{m,c_T} \quad z_{m+S,c_T} \quad \cdot \quad \cdot \quad z_{m+q_S,c_T}] \\ \boldsymbol{\varphi}_{m,c_T,1}^T &= [\zeta_{m,c_T,1} \quad \phi_{m,c_T,1}^{11} \quad \phi_{m,c_T,1}^{12} \quad \cdot \quad \cdot \quad \phi_{m,c_T,1}^{1,n_{sub}} \quad \cdot \quad \cdot \quad \phi_{m,c_T,p}^{11} \quad \phi_{m,c_T,p}^{12} \quad \cdot \quad \cdot \quad \phi_{m,c_T,p}^{1,n_{sub}}] \end{aligned}$$

então:

$$\begin{aligned} \|Z_{m,c_T} - \Phi_{m,c_T} \mathbb{Z}_{m,c_T}\|^2 &= \left\| \begin{bmatrix} \leftarrow & \mathbf{z}_{m,c_T,1}^T & \rightarrow \\ \leftarrow & \mathbf{z}_{m,c_T,2}^T & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{z}_{m,c_T,n_{sub}}^T & \rightarrow \end{bmatrix} - \begin{pmatrix} \leftarrow & \boldsymbol{\varphi}_{m,c_T,1}^T & \rightarrow \\ \leftarrow & \boldsymbol{\varphi}_{m,c_T,2}^T & \rightarrow \\ & \vdots & \\ \leftarrow & \boldsymbol{\varphi}_{m,c_T,n_{sub}}^T & \rightarrow \end{pmatrix} \mathbb{Z}_{m,c_T} \right\|^2 \\ \|Z_m - \Phi_m \mathbb{Z}_m\|^2 &= \sum_{k=1}^{n_{sub}} \|\mathbf{z}_{m,c_T,k}^T - \boldsymbol{\varphi}_{m,c_T,k}^T \mathbb{Z}_m\|^2 \quad (37) \end{aligned}$$

Logo, o problema de otimização final será:

$$\begin{aligned}
\min_{\Phi_m \geq 0} \|Z_m - \Phi_m Z_m\|^2 &= \min_{\varphi_{m,1}, \dots, \varphi_{m,n_{sub}} \geq 0} \sum_{k=1}^{n_{sub}} \|z_{m,c_T,k}^T - \varphi_{m,c_T,k}^T Z_m\|^2 \\
&= \sum_{k=1}^{n_{sub}} \min_{\varphi_{m,k} \geq 0} \|z_{m,c_T,k}^T - \varphi_{m,c_T,k}^T Z_m\|^2
\end{aligned} \tag{38}$$

Isto é, o problema é desacoplado no período m e no subsistema k .

3.1.2.1. Estimação da Distribuição de Probabilidade dos Ruídos

O erro no modelo PVAR m é definido como a razão entre o valor observado e o valor previsto, dado pela equação (39):

$$\eta_t = \frac{z_t}{(\zeta_m + \sum_{i=1}^{p_t} [\Phi_{t,i}] z_{t-i})} \tag{39}$$

O resíduo, uma estimativa observável do ruído, é obtido ao substituir os parâmetros do valor previsto na equação (39) pelos seus estimadores, como mostra a equação (40):

$$\hat{\eta}_t = \frac{z_t}{(\hat{\zeta}_m + \sum_{i=1}^{p_t} [\hat{\Phi}_{t,i}] z_{t-i})} \tag{40}$$

Optou-se pela distribuição empírica como estimador dos ruídos, objetivando não determinar qualquer hipótese adicional ao modelo. Assim, a função de distribuição acumulada empírica, $\hat{F}_m(x)$, associada ao período m será:

$$\hat{F}_m(x) = \frac{1}{q_m + 1} \sum_{j=0}^{q_m} \mathbb{I}(\hat{\eta}_{m+jS} \leq x) \tag{41}$$

onde $\mathbb{I}(\cdot)$ é a função indicadora, dada por:

$$\mathbb{I}(\cdot) = \begin{cases} 1, & \text{se } \hat{\eta}_{m+jS} \leq x \\ 0, & \text{caso contrário} \end{cases} \tag{42}$$

3.1.3. Identificação da Ordem do Modelo

O modelo PVARm por sua definição, é um modelo com alto número de parâmetros, comparado ao modelo PAR. O número de parâmetros $n_m(\hat{\varphi})$ de um componente k do modelo vetorial para determinado mês m , é dado por:

$$n_m(\hat{\varphi}) = 1 + p_m n_{subs} \quad (43)$$

Sendo a soma de uma unidade correspondente ao parâmetro de nível e o último termo o produto entre a ordem p_m do modelo e o número de subsistemas, n_{subs} .

Um modelo de ordem unitária que incorpore a relação entre 9 subsistemas teria 10 parâmetros a estimar por componente do vetor. A ordem $p = 2$ elevaria o número de parâmetros para 19.

Em (Cabral, 2016), indica-se a utilização do critério BIC para definição da ordem do modelo. Entretanto, a utilização de critérios que penalizam o número de parâmetros pode não permitir uma análise mais ampla de um modelo naturalmente muito parametrizado. Nesse sentido, propõe-se a identificação do modelo através do critério de menor soma do quadrado dos desvios.

Suponha que a ordem de um modelo m na configuração c_T é limitada a P , inclusive. Calculam-se P candidatos a modelo, variando a ordem $p = 1, 2, \dots, P$.

Define-se a soma do quadrado dos erros de previsão ($SEQ_{m,c_T,p}$), em função da ordem do candidato a modelo pela equação (44):

$$SEQ_{p,m,c_T} = \sum_t (z_{t,c_T} - \hat{z}_{p,t,c_T})^2, \text{ para } t \in \{m, m + \mathbb{S}, \dots, m + q_m \mathbb{S}\} \quad (44)$$

sendo:

z_{t,c_T} : observações da série temporal para o período m ;

\hat{z}_{p,t,c_T} : valor estimado para a série através do candidato de ordem p .

O critério de identificação do modelo seleciona o candidato que possua a menor soma dos quadrados dos erros de previsão:

$$\min \{SEQ_{p,m,c_T} | p = 1, 2, \dots, P\} \quad (45)$$

O modelo eleito consiste na agregação dos correspondentes modelos de menor erro quadrático de cada componente do vetor de previsão.

4 Resultados

Neste capítulo apresentam-se os resultados obtidos da aplicação dos modelos PAR e PVARm Interconfigurações na geração de cenários para o SEB. O modelo PVARm foi baseado em dois critérios de identificação diferentes:

1. Fixa-se ordem unitária do modelo PVARm, identificado ao longo deste capítulo por PVARm ($p = 1$);
2. A partir do critério de menor SQE, identificado ao longo deste capítulo por PVARm ($1 \leq p \leq 6$).

Os algoritmos de ajuste do modelo PVARm às séries históricas de ENA e de geração de cenários foram implementados no *software* Matlab® e o pseudocódigo é apresentado no Apêndice A. Os cenários gerados pelos modelos PAR e PVARm foram verificados através de testes estatísticos não paramétricos e tiveram o desempenho comparado através de testes de sequência.

Ressalta-se que todo o desenvolvimento proposto foi realizado no âmbito dos *softwares* desenvolvidos para o Modelo de Despacho Hidrotérmico - MDDH, uma plataforma computacional desenvolvida no projeto de P&D Estratégico da Agência Nacional de Energia Elétrica (ANEEL), pelos programas de Pós-Graduação em Engenharia Elétrica da PUC-Rio e da Universidade Federal de Juiz de Fora (UFJF). A plataforma fornece todo o tratamento do despacho hidrotérmico, tanto no que concerne às incertezas, quanto na execução do processo de otimização. (Souza *et al.*, 2014).

A seguir avalia-se o conteúdo dos dados de entrada do processo de geração de cenários de ENA pelos modelos PAR e PVARm.

4.1. Análise dos Dados de Entrada

Os cenários de ENA foram gerados para atendimento às etapas *forward* e *backward* do algoritmo PDDE, a partir de um caso real do SIN, extraído do Programa Mensal de Operação, para o período de dezembro de 2015.

O Programa Mensal da Operação (PMO) é elaborado pelo ONS e contém, dentre suas informações, dados atualizados sobre o cronograma de expansão da geração, parâmetros de produção das usinas existentes, previsão de carga, o estado de armazenamento dos reservatórios e histórico de afluições aos aproveitamentos hidrelétricos (ONS, 2018). A Câmara de Comercialização de Energia Elétrica (CCEE) disponibiliza, mensalmente, o conjunto de arquivos de entrada do modelo Newave, o Deck de Preços, que contém as atualizações divulgadas no PMO.

Dentre os arquivos que compõem o conjunto do PMO de 12/2015, estão informações históricas de vazões⁴, desde janeiro de 1931, até dezembro de 2013, parâmetros sobre a eficiência das usinas do sistema, além do cronograma de expansão da geração. Para maiores informações acerca dos arquivos, consultar (CEPEL, 2006).

Para este PMO, os subsistemas ou REE's compõem grupos de hidrelétricas, com reservatório e a fio d'água, das bacias hidrográficas por região geográfica, a saber: i) Sudeste e Centro-Oeste; ii) Sul; iii) Nordeste; e iv) Norte, formando quatro REEs, como mostra a figura abaixo:

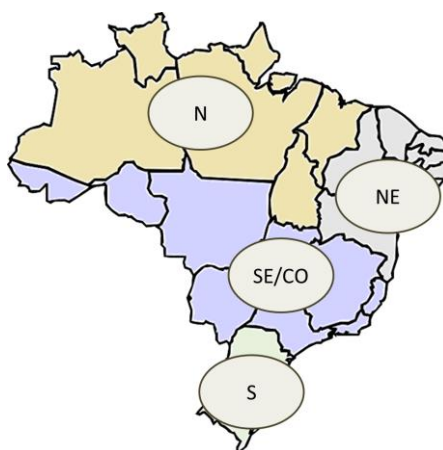


Figura 8: Regiões agrupadas para formação dos REE's. Fonte:
Adaptada de (Pixabay, 2013).

⁴ As informações acerca de dados históricos de vazões estão disponíveis no Deck de Preços em formato binário. A CCEE disponibiliza em seu site os aplicativos VazEdit e HydroEdit para conversão de tais arquivos em texto.

Cada uma das séries que compõem o histórico dinâmico é gerada em função de uma das diferentes configurações apresentadas pelo parque hidrelétrico, ao longo do horizonte de estudo. Para o PMO de 12/2015, prevê-se um horizonte que finda em dezembro de 2019. Neste horizonte, as mudanças de configurações previstas são apresentadas abaixo.

Tabela 2: Mudanças de configuração previstas no PMO de 12/2015. Fonte: Deck de Preços do Newave 12/2015. Fonte: Elaborada pela autora

ANO	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ
PRÉ ESTUDO	1	1	1	1	1	1	1	1	1	1	1	
2015	-	-	-	-	-	-	-	-	-	-	-	1
2016	2	3	4	5	6	7	8	9	10	11	12	13
2017	14	15	16	17	18	19	19	20	20	21	21	22
2018	23	24	25	26	27	28	29	30	31	32	32	33
2019	33	34	34	34	34	34	34	34	34	34	34	34
PÓS-ESTUDO	34	34	34	34	34	34	34	34	34	34	34	34

A Tabela 2 mostra que o sistema passa por 34 configurações topológicas diferentes durante o período de estudo. Além do horizonte de estudo, no âmbito do Planejamento Energético de Médio Prazo, utilizam-se ainda os períodos de pré-estudo e pós-estudo, ambos associados a uma configuração estática, do início e final do período de planejamento, respectivamente, como apresentado na Tabela 2.

O período de pré-estudo, adotado como 12 períodos anteriores ao início do estudo, é utilizado para simulação inicial do modelo ajustado no intuito de eliminar a influência dos valores iniciais nas séries geradas.

O período de pós-estudo é constituído de 5 anos à frente do período de estudo e é utilizado no processo de otimização para que a decisão do último estágio do período de estudo não seja baseada na informação de um custo futuro nulo, tornando o problema condizente com a realidade.

Acerca das mudanças de configuração, a Figura 9 apresenta a previsão da expansão hidrelétrica do SIN, por subsistema, ao longo do horizonte de 2015 – 2019. Cabe ressaltar o crescimento expressivo da potência instalada na região

Norte do país, devido à previsão da entrada de 18 unidades geradoras da Usina Hidrelétrica de Belo Monte, totalizando aproximadamente 11000 MW.

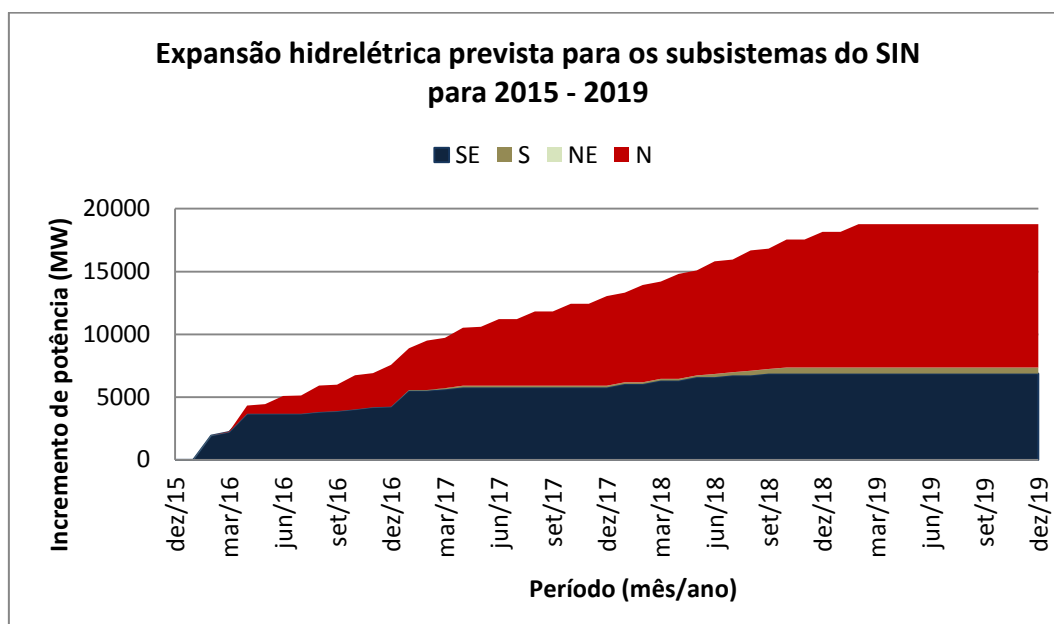


Figura 9: Cronograma do incremento da capacidade instalada de hidrelétricas do SIN para o horizonte de 2015 – 2019. Fonte: Elaborada pela autora

Dadas as informações acerca das configurações, cada um dos períodos (mês/ano) de estudo, pré e pós-estudo, apresentados na Tabela 2, é associado ao histórico de ENA's calculado com base em sua própria configuração. Esse histórico, chamado neste trabalho de histórico dinâmico, é utilizado para ajuste do modelo PVARm, que tem a implementação das etapas de identificação, ajuste e proposta para geração de cenários descrita no Apêndice A desta dissertação.

Para o estudo, a partir dos três modelos citados, foram gerados 200 cenários para as simulações *forward*, para 109 períodos, isto é, 49 meses contidos no período de estudo, somado aos 60 meses do período de pós-estudo. Foram utilizadas 20 aberturas por período.

A seguir, apresentam-se as análises dos cenários *forward* gerados, que consistem na verificação das estatísticas mensais média, desvio-padrão e assimetria e na estatística anual correlação espacial. São realizados testes não paramétricos para comparação entre os cenários gerados e os cenários históricos para igualdade de média, distribuição e variabilidade, além do teste de sequências

para aferir a capacidade de reprodução de características de interesse para a aplicação dos modelos.

4.2.

Verificação de Estatísticas Mensais e Anuais

4.2.1.

Envoltória dos Cenários

Nesta seção apresenta-se análise exploratória dos 200 cenários gerados por cada um dos modelos PAR, PVARm ($p = 1$), PVARm ($1 \leq p \leq 6$), para os subsistemas Sudeste, Sul, Nordeste e Norte. Da Figura 10 à Figura 12, apresentam-se as envoltórias dos cenários gerados (nuvem azul clara), destacando a média dos cenários (em azul contínuo) em relação à média histórica (em vermelho seccionado).

O subsistema Sul apresentou as maiores divergências em relação à média, uma vez que apresenta os menores valores de ENA, com alta variabilidade. Em particular, ressalta-se que, se a conjectura acerca da causalidade do modelo PVARm estiver correta, então a variância do modelo é proporcional ao quadrado da sua previsão, isto é, o modelo pode ser mais interessante para séries de observações baixas e variabilidade baixa, ou aquelas com observações e variabilidade altas.

Nota-se que as séries geradas pelo modelo PAR para o subsistema Nordeste têm os primeiros períodos superestimados em relação à média histórica. Para os modelos PVARm, ambos são mais pessimistas comparados ao modelo PAR.

Em particular, os cenários do subsistema Norte apresentam mudança de nível expressiva entre o segundo e o terceiro picos da série, reflexo da entrada em operação de unidades geradoras da usina de Belo Monte, prevista para o horizonte de 2015/2019. Deve-se observar que, apesar da mudança de nível, cada um desses períodos é modelado individualmente, com base em séries históricas diferentes, calculadas a partir das diferentes topologias esperadas para o sistema elétrico em cada um dos períodos.

Em suma, a partir da análise visual, os modelos reproduzem bem as médias históricas.

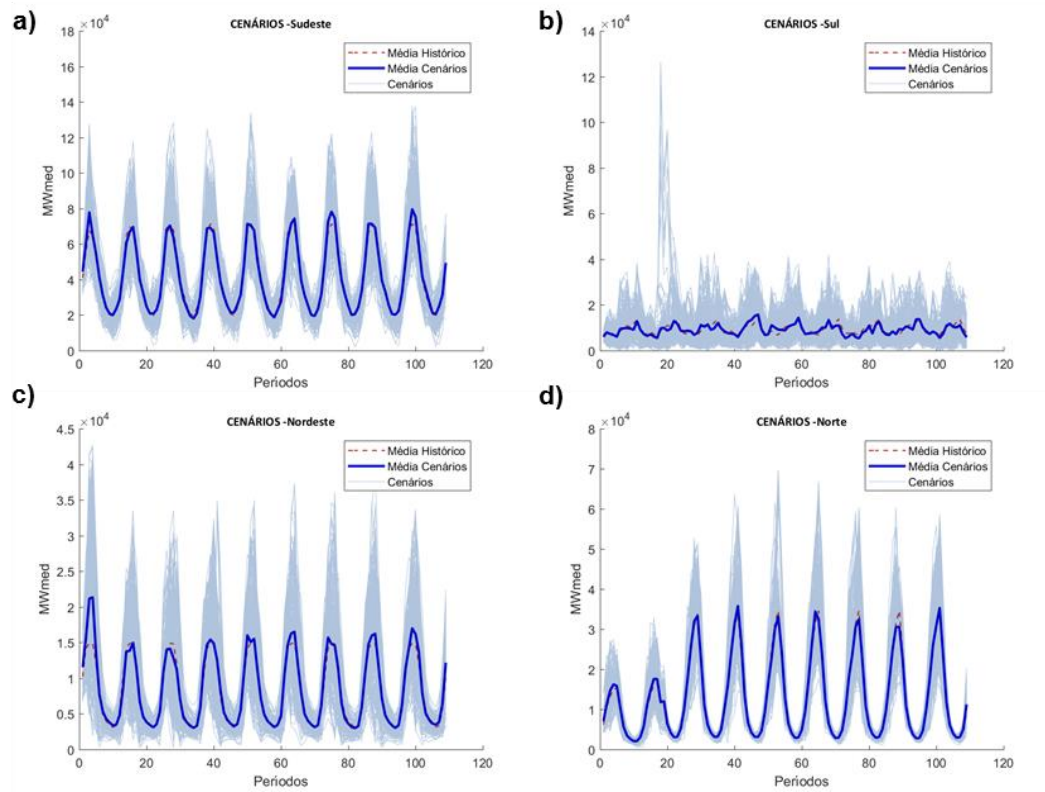


Figura 10: Envoltória dos cenários gerados pelo modelo PAR.

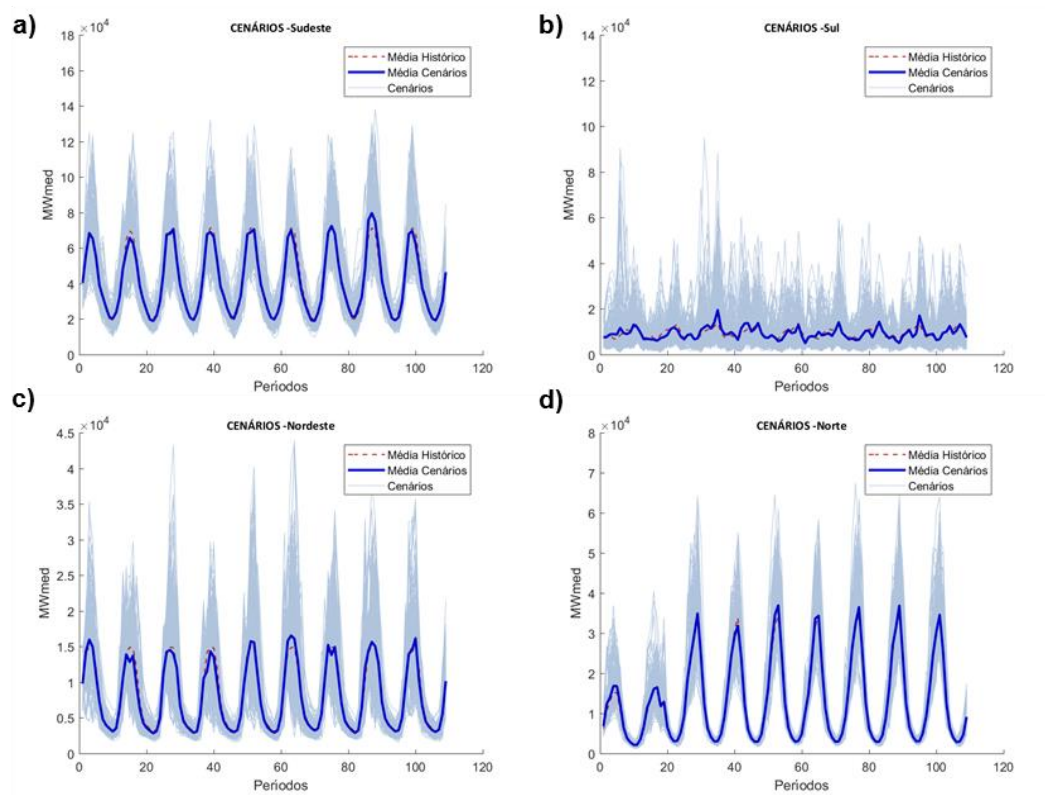


Figura 11: Envoltória dos cenários gerados pelo modelo PVARm ($p=1$)

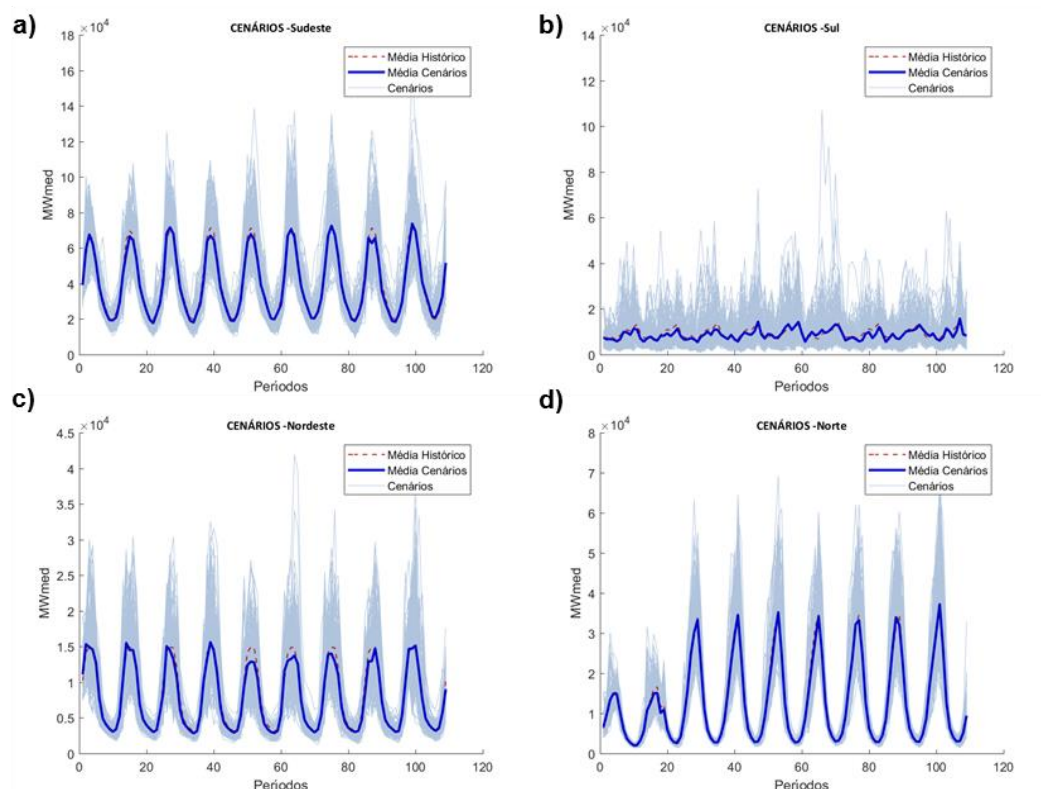


Figura 12: Envoltória dos cenários gerados pelo modelo PVARm ($1 \leq p \leq 6$)

4.2.2. Desvio Padrão

Da Figura 13 à Figura 15, apresentam-se os desvios padrão amostrais dos cenários gerados pelos modelos PAR, PVARm ($p = 1$), PVARm ($1 \leq p \leq 6$), para os subsistemas Sudeste, Sul, Nordeste e Norte, representados pela linha preta contínua, em comparação com o histórico, representado pela linha vermelha seccionada.

Observa-se, em todos os modelos, que o desvio amostral preserva a característica sazonal do histórico.

Para o subsistema Nordeste, com exceção dos primeiros períodos, o PAR teve um comportamento próximo ao histórico, enquanto os modelos PVARm mostraram maior descolamento em relação à curva histórica.

Para o subsistema Norte, nota-se que, assim como no caso das médias dos cenários, há também mudança expressiva de nível entre o segundo e terceiro picos do gráfico do desvio, devido à correspondente mudança de nível dessas séries pela

entrada em operação da usina de Belo Monte. Uma vez que o desvio amostral é função da média das séries, justifica-se a alteração de nível.

De maneira geral, os modelos PVARm subestimam levemente os valores históricos, sendo o modelo de ordem variável aquele que mais ficou abaixo da referência histórica.

Para os subsistemas Norte e Sudeste, os três modelos exibem comportamento próximo e satisfatório. Para o subsistema Sul, aquele com a maior variabilidade, o PAR foi o melhor representante.

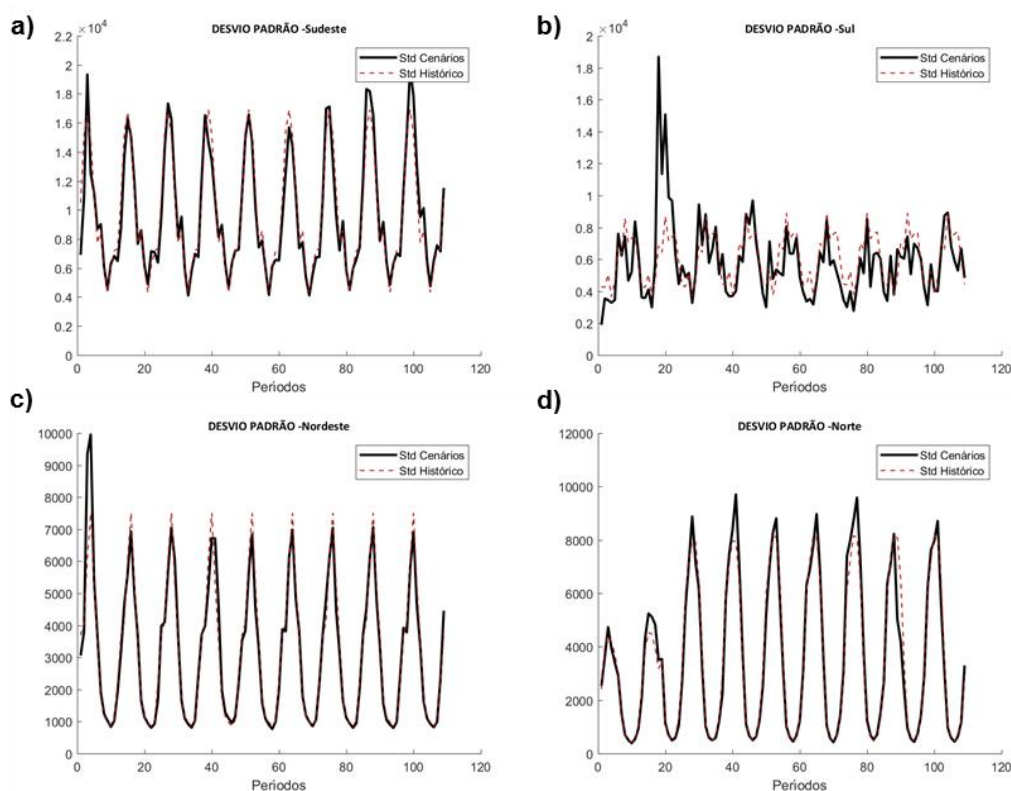


Figura 13: Desvio Padrão dos cenários gerados pelo modelo PAR

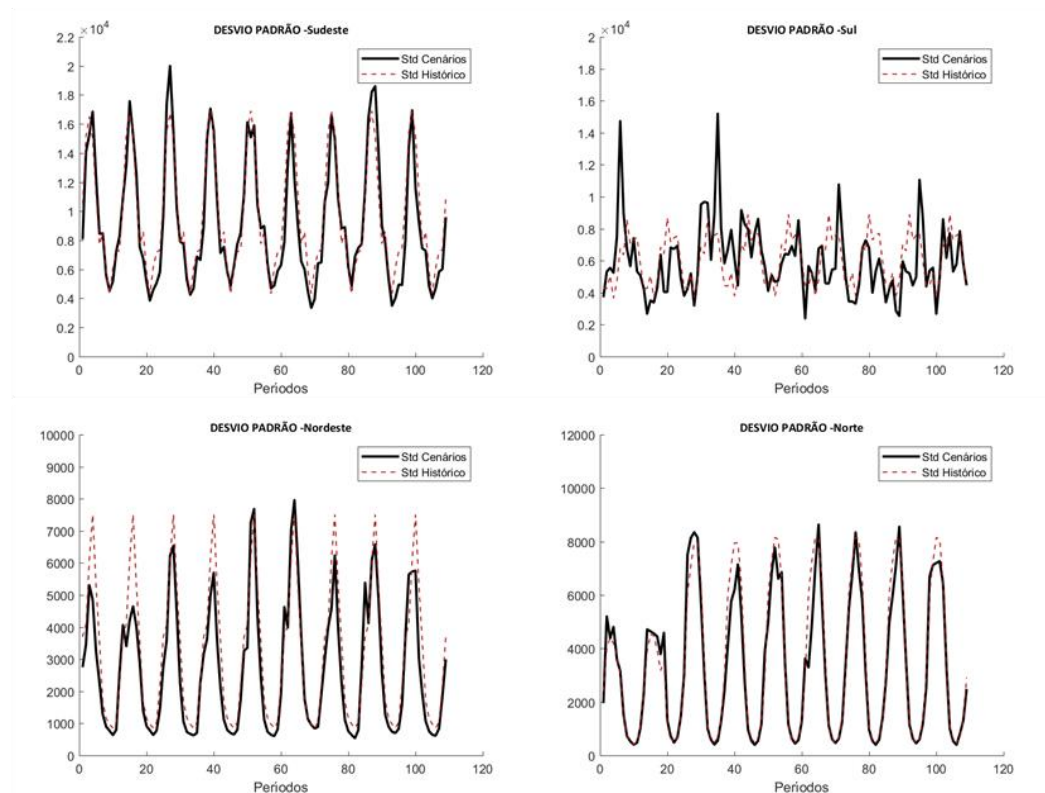


Figura 14: Desvio Padrão dos cenários gerados pelo modelo PVARm ($p=1$)

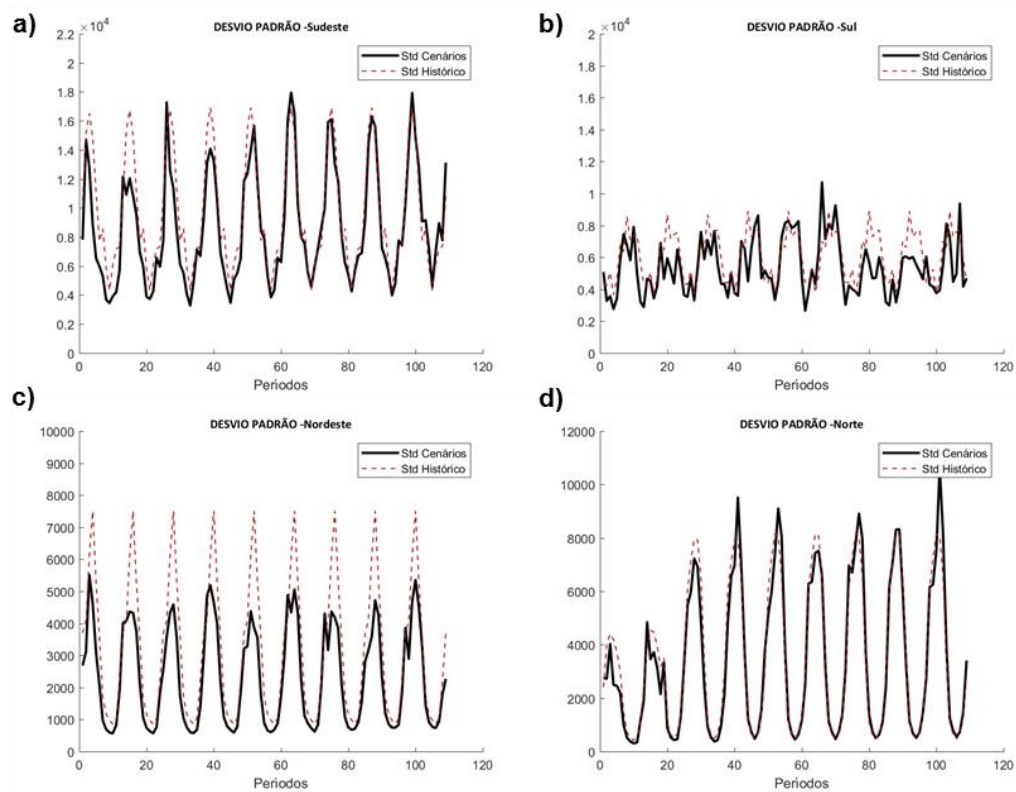


Figura 15: Desvio Padrão dos cenários gerados pelo modelo PVARm ($1 \leq p \leq 6$)

4.2.3. Assimetria

As séries de ENA são marcadas pelo forte coeficiente de assimetria. Visto que esta estatística não é parte dos modelos aplicados, sua preservação não é esperada. Entretanto, há interesse em avaliar a capacidade dos modelos em reproduzir o comportamento da assimetria histórica, já que esta característica constitui resultado positivo para o modelo.

Da Figura 16 à Figura 18, apresentam-se os coeficientes de assimetria amostrais dos cenários gerados pelos modelos PAR, PVARm ($p=1$), PVARm ($1 \leq p \leq 6$), representados pela linha preta contínua, em comparação com o histórico, representado pela linha vermelha seccionada.

Neste quesito, destacam-se as representações da assimetria dos modelos PVARm, em comparação à representatividade do modelo PAR. Para todos os subsistemas, o modelo PAR exibe o nível de assimetria sempre abaixo daquele histórico, enquanto os modelos PVARm exibem nível compatível com a referência histórica, apesar de permanecer inferior ao mesmo, sem replicar os picos do histórico, como claramente se verifica nas séries da região Sudeste.

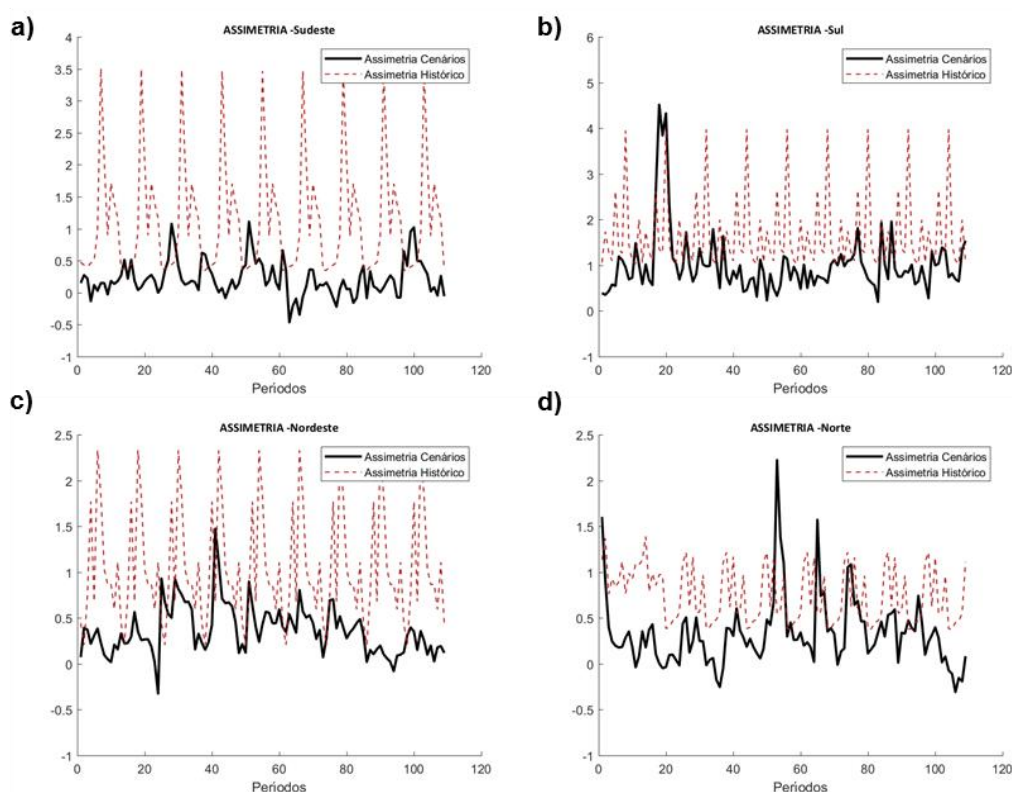


Figura 16: Assimetria dos cenários gerados pelo modelo PAR

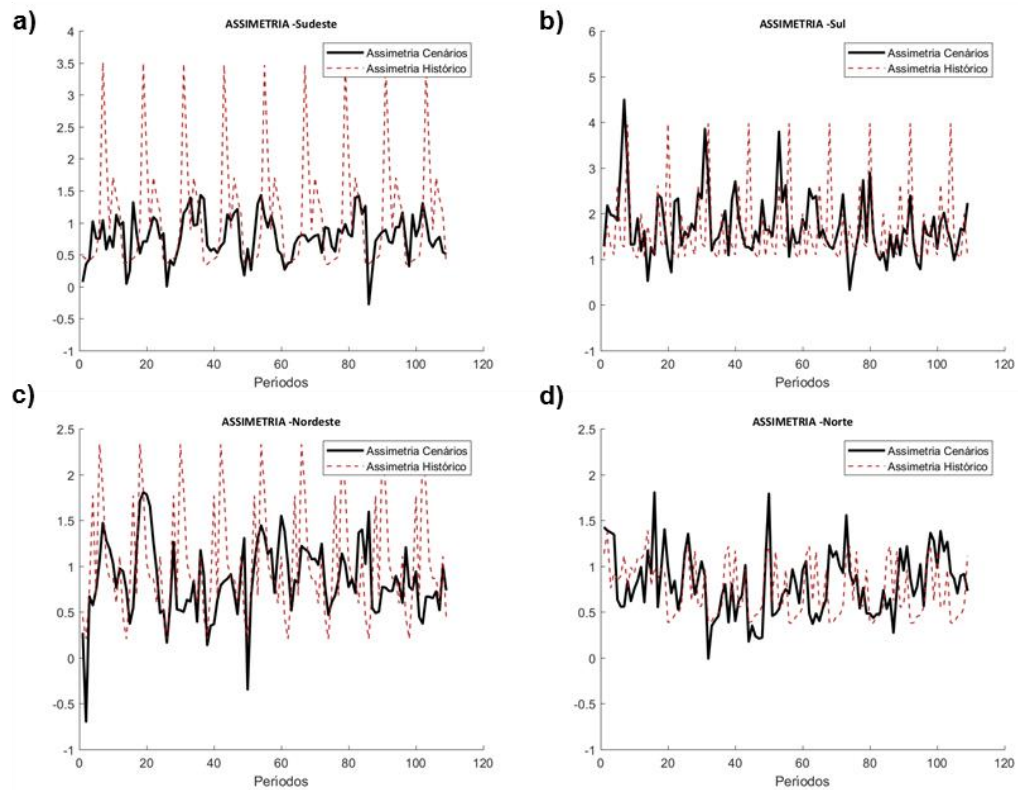


Figura 17: Assimetria dos cenários gerados pelo modelo PVARm ($p=1$)

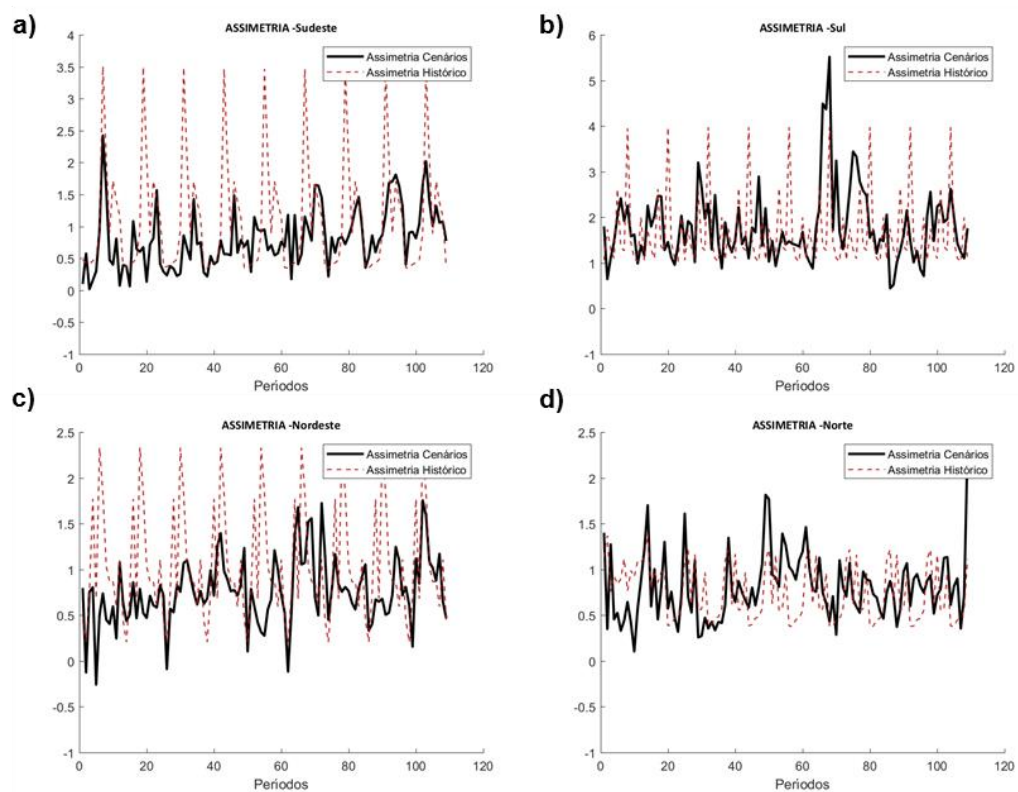


Figura 18: Assimetria dos cenários gerados pelo modelo PVARm ($1 \leq p \leq 6$)

4.2.4. Correlação Espacial entre Subsistemas

Da Figura 19 à Figura 21, apresenta-se a estatística anual coeficiente de correlação espacial amostral dos cenários gerados pelos modelos PAR, PVARm ($p = 1$), PVARm ($1 \leq p \leq 6$), para cada possível combinação dois a dois entre os subsistemas Sudeste, Sul, Nordeste e Norte, conforme sugerido em (Oliveira, 2010).

A informação é apresentada pelo gráfico *boxplot* dos cenários. O asterisco representa o coeficiente de correlação espacial entre as séries históricas dos subsistemas.

O coeficiente de correlação amostral histórico entre os subsistemas Norte e Nordeste é igual a 0.54. A mediana dos coeficientes relativos aos cenários do modelo PAR foi a que mais se aproximou do histórico. Dentre os modelos PVARm, o modelo de ordem unitária é o mais coerente com o histórico. As amplitudes alcançadas pelos modelos PVARm de ordem fixa e PAR são próximas.

O coeficiente de correlação amostral histórico entre os subsistemas Sudeste e Nordeste é igual a 0.55. A mediana do modelo PAR novamente ficou mais próxima do histórico, sendo 0.52. As medianas dos coeficientes relativos aos cenários gerados pelos modelos PVARm de ordens fixa e variável apresentaram os valores 0.33 e 0.21, respectivamente. Neste caso, o comportamento dos modelos PVARm é próximo.

Entre os subsistemas Sudeste e Norte, o coeficiente de correlação amostral do histórico é de 0.34. Neste caso, a mediana do modelo PVARm ($p = 1$) é a mais próxima do histórico, igual a 0.29. O modelo PAR teve a mediana mais afastada do valor histórico, igual a 0.23. O modelo PVARm de ordem variável foi o que apresentou maior amplitude.

Para as séries dos subsistemas Sudeste e Sul, o coeficiente de correlação amostral do histórico é igual a 0.42. Os modelos PVARm exibiram comportamento indistinguível, com mediana igual a 0.3, que mais se aproximou do histórico. O modelo PAR apresentou a mediana fora do intervalo semi-interquartilico.

Entre os subsistemas Nordeste e Sul, o coeficiente histórico anual apresenta correlação negativa de -0.1. Os cenários gerados pelo modelo PVARm variam entre cenários negativos e positivos e a mediana do modelo de ordem variável é a mais próxima do coeficiente histórico. O modelo PAR apresenta mediana de -0.31 e seus valores são majoritariamente negativos.

Entre os subsistemas Norte e Sul, a mediana do modelo PAR, igual a -0.3, é a mais próxima do histórico, que é igual a -0,25. Os modelos PVARm de ordens fixa e variável foram iguais aos coeficientes de 0.21 e 0.26.

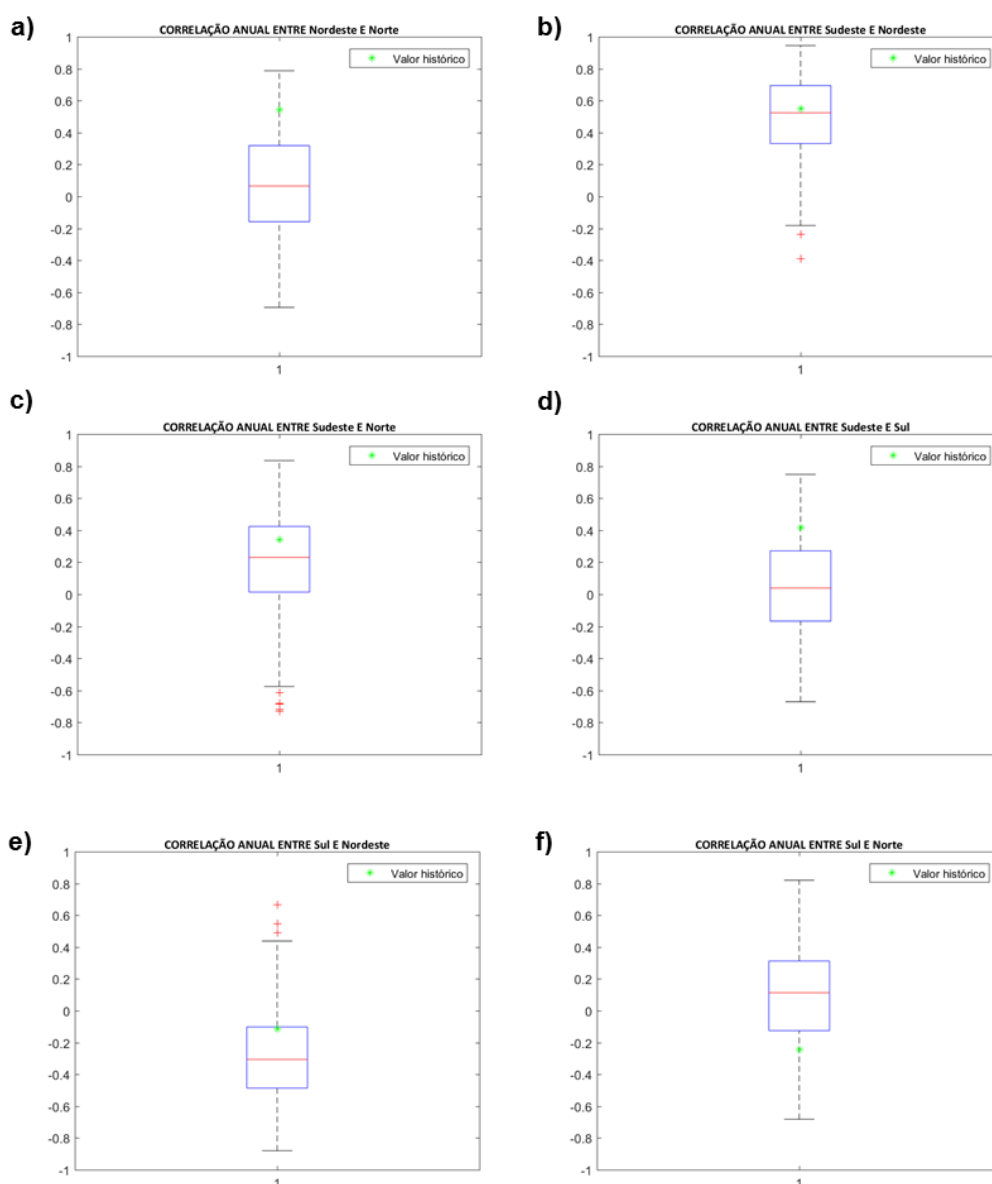


Figura 19: Correlação anual entre subsistemas dos cenários gerados pelo modelo PAR.

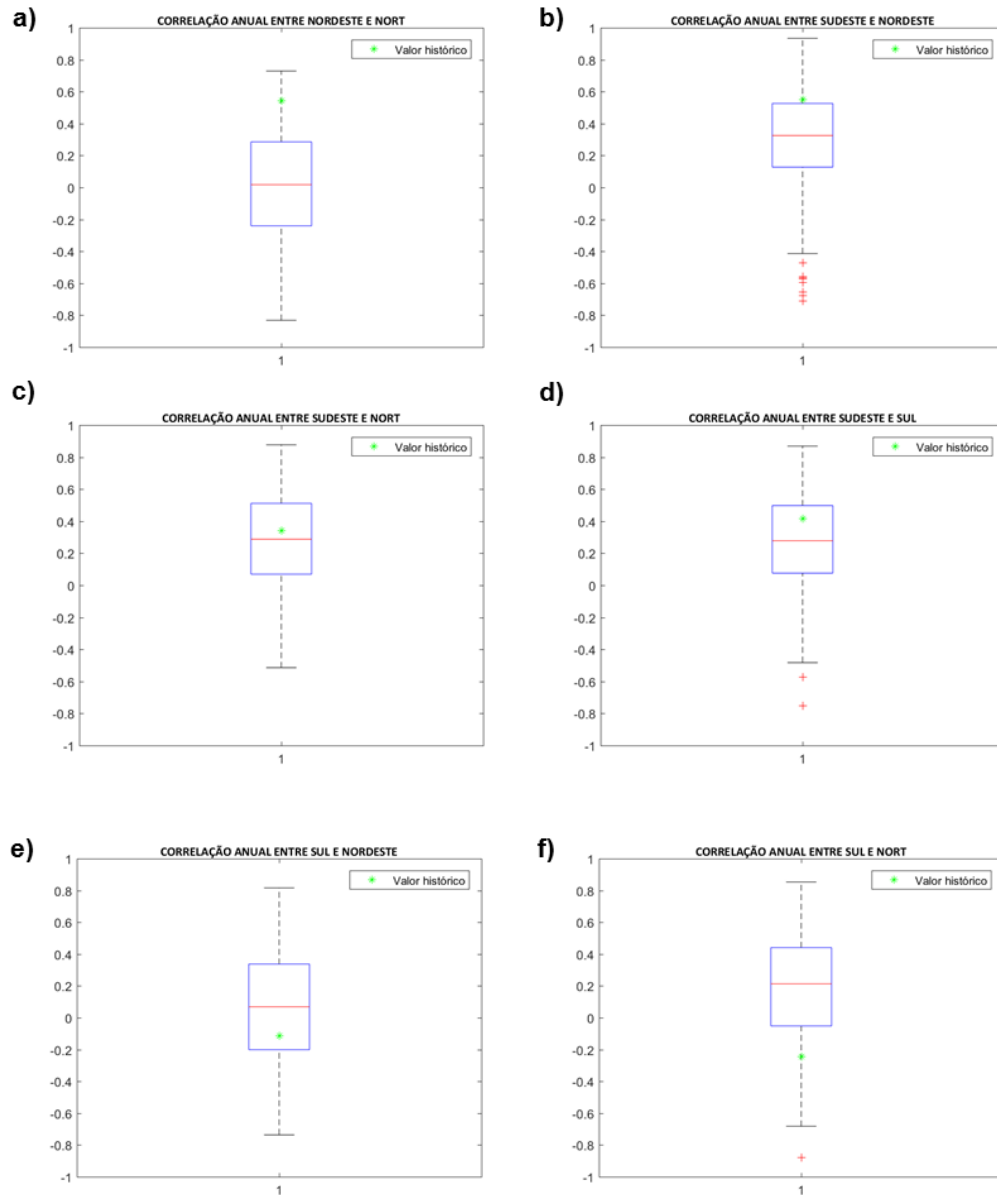


Figura 20: Correlação anual entre subsistemas dos cenários gerados pelo modelo PVARM ($p=1$)

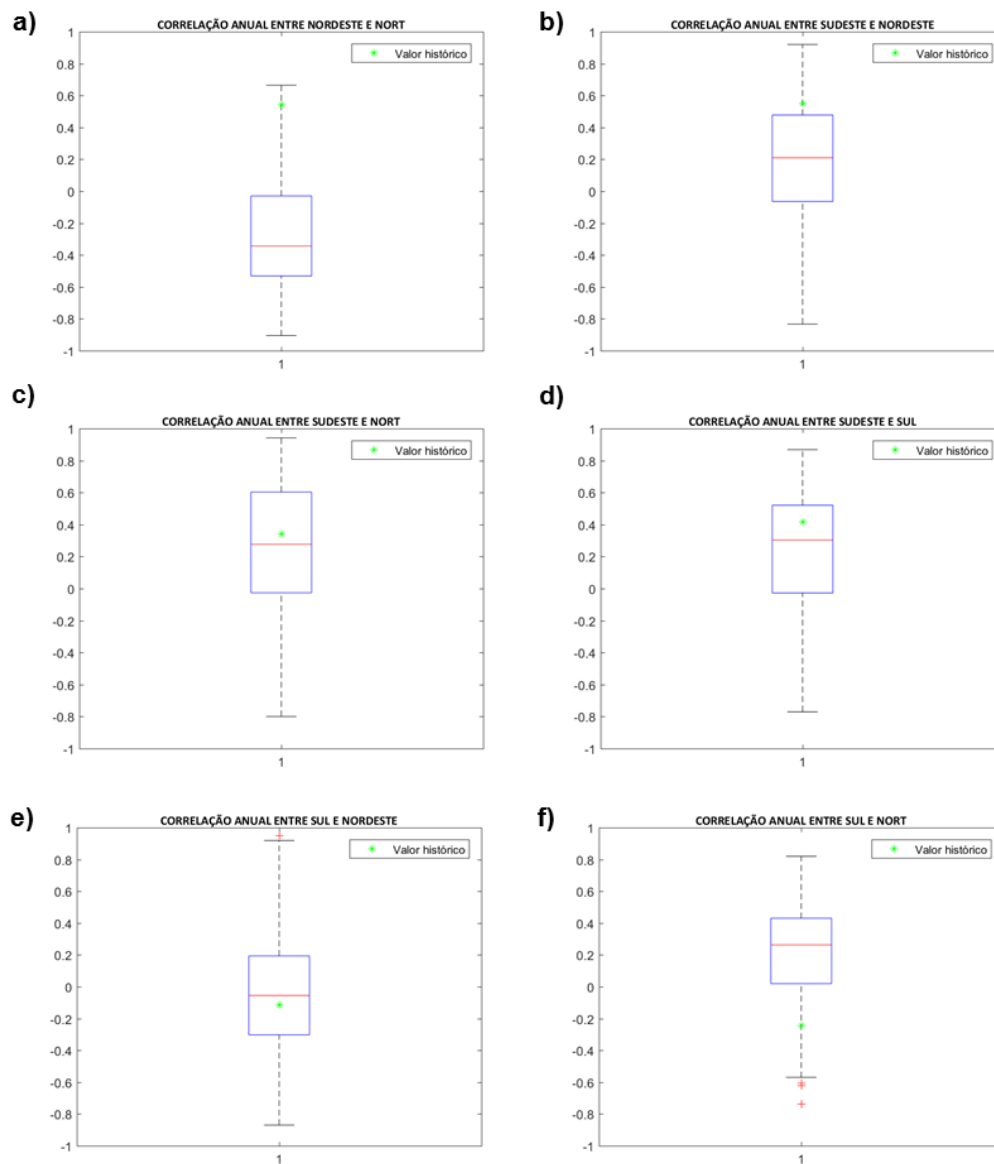


Figura 21: Correlação anual entre subsistemas dos cenários gerados pelo modelo PVARm ($1 \leq p \leq 6$)

4.3.

Testes de Avaliação e Adequação dos Modelos

4.3.1.

Teste de Kolmogorov-Smirnov (K-S)

O teste de K-S determina se duas amostras independentes são originárias de uma mesma população contínua. Avalia-se a distância entre a distribuição acumulada empírica de cada uma das amostras.

Foi calculado o p-valor para cada um dos períodos do horizonte de planejamento. Testa-se a hipótese de igualdade das distribuições a um nível de significância de 5%.

A Figura 22 apresenta o percentual de modelos de períodos aprovados pelo teste K-S, do total de 109 considerados no estudo. Avaliam-se os cenários gerados pelos modelos PAR, PVARm ($p = 1$) e PVARm ($1 \leq p \leq 6$), para os subsistemas Sudeste, Sul, Nordeste e Norte.

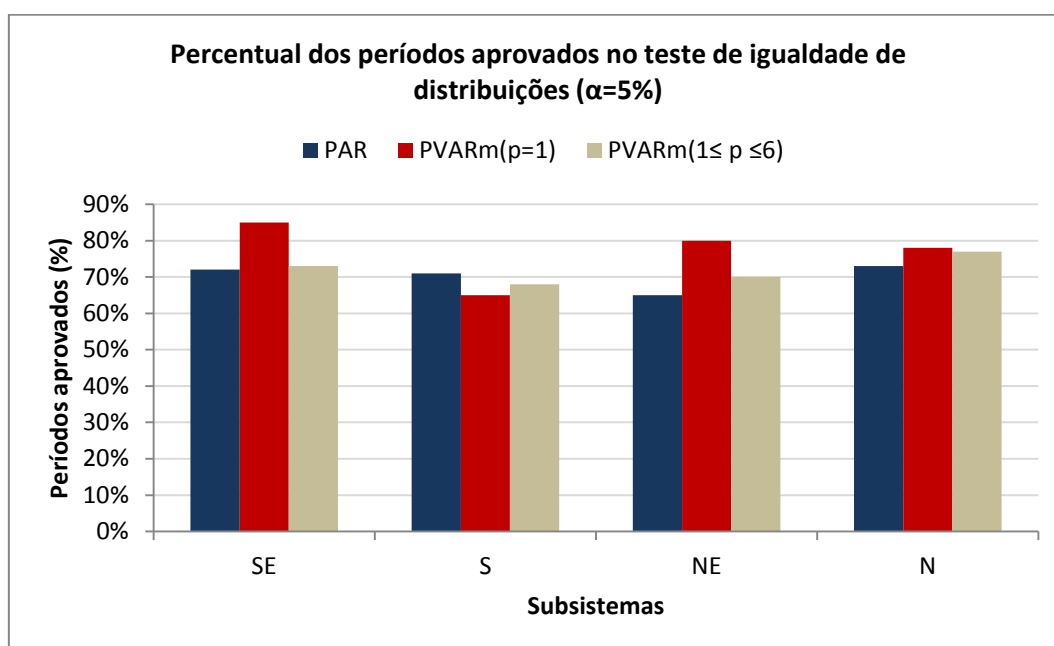


Figura 22: Percentual de períodos aprovados pelo teste estatístico de Kolmogorov-Smirnov

O modelo que recebeu aprovação na maior parte dos períodos dos subsistemas Sudeste, Norte e Nordeste, em torno de 80%, foi o PVARm de ordem fixa. Dentre os modelos PVARm, o comportamento entre eles é indistinguível para o subsistema Norte, enquanto que para os subsistemas Sudeste e Nordeste, o

modelo de ordem fixa se mostra superior. O PAR foi o modelo que teve maior aprovação de períodos para o subsistema Sul.

4.3.2. Teste de Wilcoxon

O teste de Wilcoxon é uma opção não paramétrica ao teste-t e verifica a hipótese de igualdade de mediana entre duas amostras, baseado na ordem na qual as observações das duas amostras se apresentam.

Calcula-se o p-valor de cada um dos períodos do horizonte de planejamento. Testa-se a hipótese de igualdade do parâmetro locacional a um nível de significância de 5%.

A Figura 23 apresenta a porcentagem de cenários aprovados pelo teste de Wilcoxon, do total de 109 períodos considerados no estudo. Avaliam-se os cenários gerados pelos modelos PAR, PVARm ($p = 1$) e PVARm ($1 \leq p \leq 6$), para os subsistemas Sudeste, Sul, Nordeste e Norte.

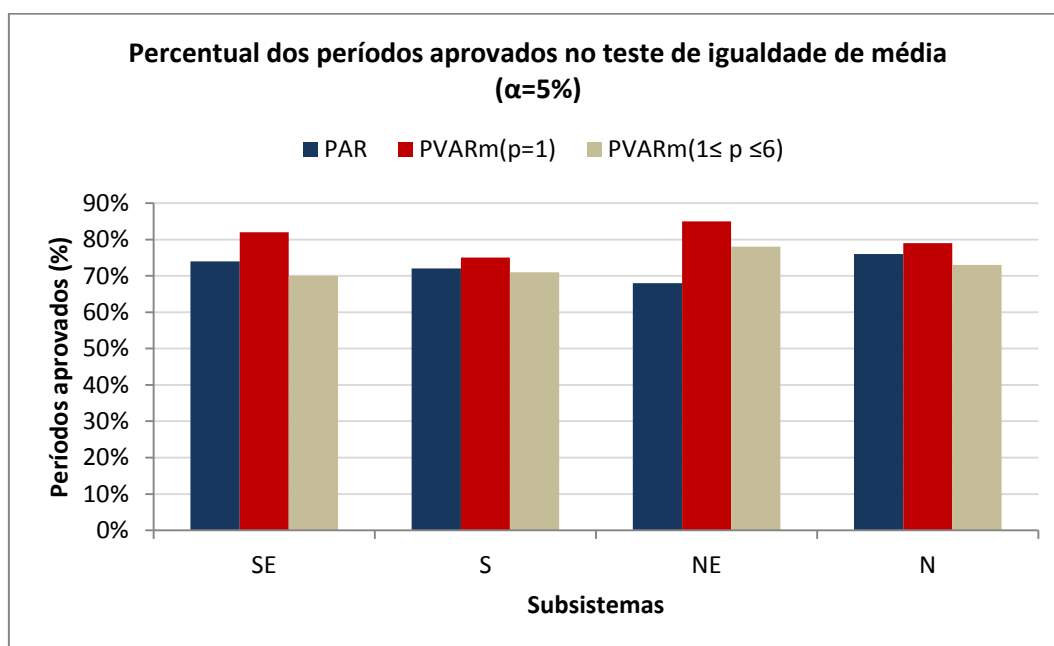


Figura 23: Percentual de períodos aprovados pelo teste estatístico de Wilcoxon.

O modelo PVARm de ordem fixa é aquele que representou de maneira satisfatória a maior parte das médias de todos os períodos, para todos os subsistemas. Nota-se

que, para os subsistemas Sul e Norte o desempenho dos três modelos é bem semelhante.

4.3.3. Teste de Ansari – Bradley

O teste de Ansari-Bradley testa a hipótese nula de que duas amostras independentes vêm da mesma distribuição, em detrimento da hipótese alternativa, de que possuem distribuições com mesma mediana, forma, mas dispersões diferentes. Assume-se que as amostras possuem medianas iguais.

Calcula-se o p-valor para cada um dos períodos do horizonte de planejamento. Testa-se a hipótese de igualdade de desvio a um nível de significância de 5%.

A Figura 24 representa a porcentagem de períodos aprovados pelo teste de dispersão de Ansari-Bradley, do total de 109 períodos considerados no estudo. Avaliam-se os cenários gerados pelos modelos PAR, PVARm ($p = 1$) e PVARm ($1 \leq p \leq 6$), para os subsistemas Sudeste, Sul, Nordeste e Norte.

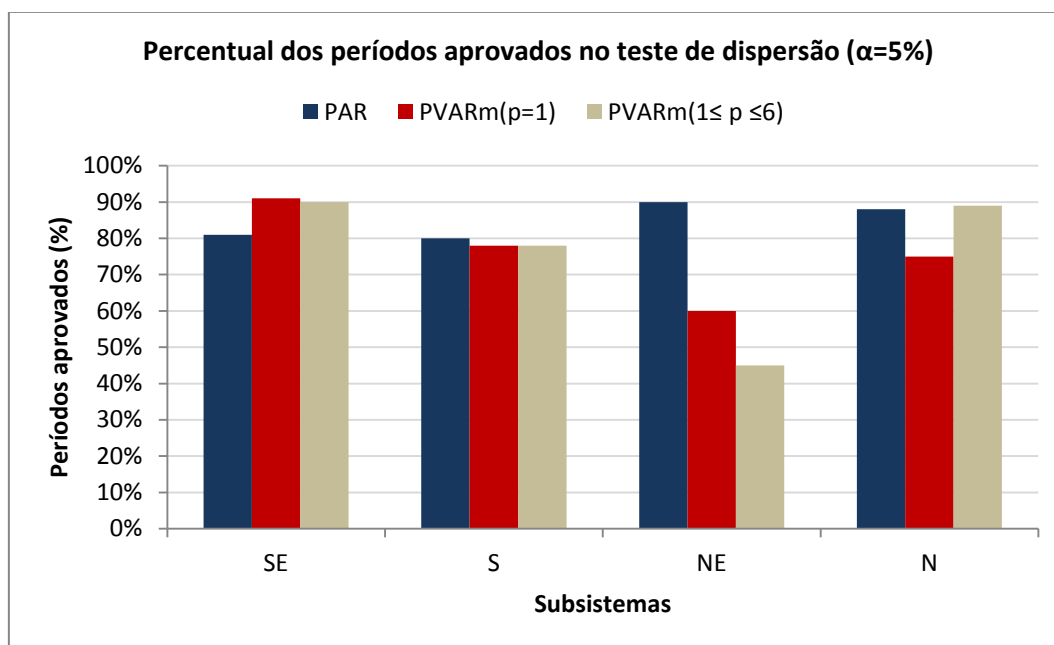


Figura 24: Percentual de períodos aprovados pelo teste estatístico de Ansari – Bradley.

Observa-se que o modelo PVARm de ordem fixa representa melhor a variabilidade da maior parte dos períodos das séries do subsistemas Sudeste e, neste caso o PVARm de ordem variável tem desempenho semelhante. Para os subsistemas Sul, Nordeste e Norte, o PAR representou melhor a maior parte dos períodos.

Atendo-se apenas aos modelos PVARm, aquele de ordem fixa apresenta melhor desempenho para os subsistemas Sudeste, e Nordeste. Em particular, para o Nordeste, o desempenho de ambos os modelos é baixo. Os dois modelos são indistinguíveis na representação do desvio para o subsistema Sul e, para o subsistema Norte, o modelo de ordem variável se destaca.

4.3.4. Análise de Sequências

Para o teste de sequências empregado em séries sintéticas baseadas em configuração estática do parque gerador, o histórico é único e compõe uma série longa, em ordem cronológica. Por exemplo, para o PMO em estudo, a série histórica se inicia em janeiro de 1931 e se estende até dezembro de 2015. Compara-se o histórico único com os valores de referência. Os eventos de sequências positivas e negativas são extraídos para comparação com os mesmos eventos observados nos diversos cenários gerados.

A análise de sequências dos cenários gerados a partir do modelo de ENA, que se baseia no histórico dinâmico, relaciona-se a diferentes históricos, calculados a cada transição de configuração do sistema de geração. Abaixo, representa-se o histórico dinâmico composto para um PMO arbitrário. Supondo que cada período contém uma configuração diferente, são 60 configurações.

	JAN/CNF1	FEV/CNF2	MAR/CNF3	ABR/CNF4	...	NOV/CNF11	DEZ/CNF12	...	JAN/CNF49	FEV/CNF50	MAR/CNF51	ABR/CNF52	...	DEZ/CNF60
1931					
1932					
1933					
.					
.					
.					
2013					
2014					
2015					

Figura 25: Histórico dinâmico para PMO de 2015, supondo 60 configurações diferentes.

Cada mês que constitui o histórico dinâmico, representado por uma coluna na imagem acima, é um trecho (um mês) extraído de um histórico estático diferente, calculado com base na configuração apresentada pelo período.

O histórico dinâmico não possui uma sequência cronológica bem definida, como o histórico estático. Tomando como exemplo o ano de 1931, nota-se que há 60 períodos diferentes associados ao ano, com diferentes meses de janeiro, diferentes meses de Fevereiro, e assim por diante. Não é possível conectar os meses do ano de 1931 do histórico interconfigurações com os meses do ano de 1932, estabelecendo uma sequência.

Assim, a estratégia adotada foi definir como referência a média do mês m para as diferentes configurações que este pode apresentar em um ano do histórico. As amostras são obtidas das séries históricas, desconectadas. Em seguida, avalia-se a amostra obtida a partir dos cenários para comparação com as amostras obtidas do histórico. As variáveis aleatórias sob análise são listadas a seguir:

1. Comprimento de Sequência Positiva;
2. Comprimento de Sequência Negativa;
3. Soma de Sequência Positiva;
4. Soma de Sequência Negativa;
5. Intensidade de Sequência Positiva;
6. Intensidade de Sequência Negativa;
7. Probabilidade de máximos de cada uma das variáveis acima.

Utiliza-se o teste de igualdade entre distribuições de K-S para avaliar as variáveis contínuas e o teste Qui-quadrado para avaliação das variáveis discretas.

A Figura 26 e a Figura 27 apresentam, através do gráfico de barras, os valores da estatística qui-quadrado para o comprimento de sequências positiva e

negativa, para os modelos PAR, PVARm de ordem fixa e variável, para os subsistemas Sudeste, Sul, Nordeste e Norte. O valor tabelado para todos os testes é igual a 3.84, para uma distribuição qui-quadrado com 2 graus de liberdade, e é destacado através da linha preta no gráfico. Aqueles modelos que apresentam valor da estatística de teste maior do que o valor tabelado são rejeitados.

No quesito reprodução do comprimento de sequências positivas históricas, os modelos PVARm tiveram a hipótese nula de igualdade das proporções rejeitada para o subsistema Nordeste. Para os demais subsistemas aceita-se a hipótese nula, ressaltando-se que o modelo PVARm ($1 \leq p \leq 6$) apresentou desempenho superior ao modelo PVARm ($p = 1$).

O modelo PAR foi rejeitado na reprodução do comprimento de sequências positivas históricas para os subsistemas Sudeste e Norte, mas apresentou o melhor desempenho, dentre todos os modelos, para o subsistema Nordeste. Para o subsistema Sul, o desempenho do PAR foi bom e ficou próximo do modelo PVARm ($1 \leq p \leq 6$).

Em relação à reprodução do comprimento de sequências históricas negativas, que se referem a períodos em que as afluências permanecem abaixo da referência histórica estabelecida, os modelos PVARm foram rejeitados para o subsistema Norte, o modelo PVARm ($1 \leq p \leq 6$) foi rejeitado para o subsistema Nordeste e ambos foram aceitos para os demais subsistemas. Em particular, o modelo PAR neste quesito também foi rejeitado para o subsistema Norte, mas foi o que apresentou o melhor desempenho para todos os subsistemas.

A Figura 28 e a Figura 29 apresentam as probabilidades de máximos comprimentos de sequências positiva e negativa. Isto é, avalia-se a probabilidade do máximo comprimento de sequência das séries geradas ser menor do que o maior comprimento de sequência identificado no histórico. Com exceção do subsistema Norte, mais de 80% dos máximos encontrados nos cenários, tanto de sequência positiva, quanto de sequência negativa, foram menores do que os máximos históricos, o que não é desejável.

Para o subsistema Norte, especificamente para sequências positivas, a probabilidade de ambos os modelos está em torno de 50%, enquanto que para as sequências negativas, a probabilidade está próxima de 100%, também para todos os modelos.

Para que as observações históricas sejam consideradas representativas da amostra, é necessário que tais observações estejam no meio da amostra, isto é, com probabilidades de máximos próximas a 50%, o que não ocorreu com as séries geradas para os subsistemas, a exceção do Norte, para sequências positivas.

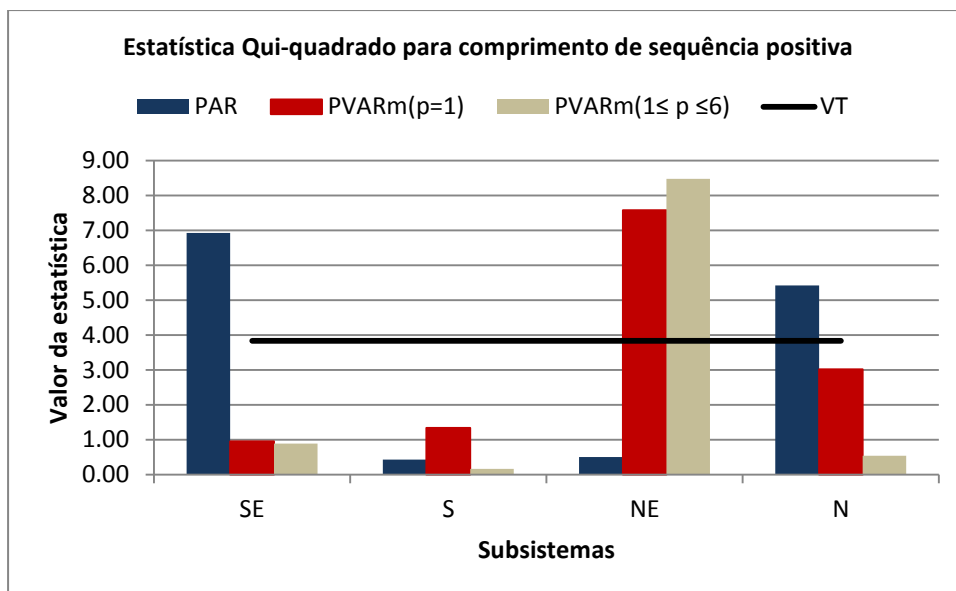


Figura 26: Estatística qui-quadrado para o teste de comprimento de sequência positiva.

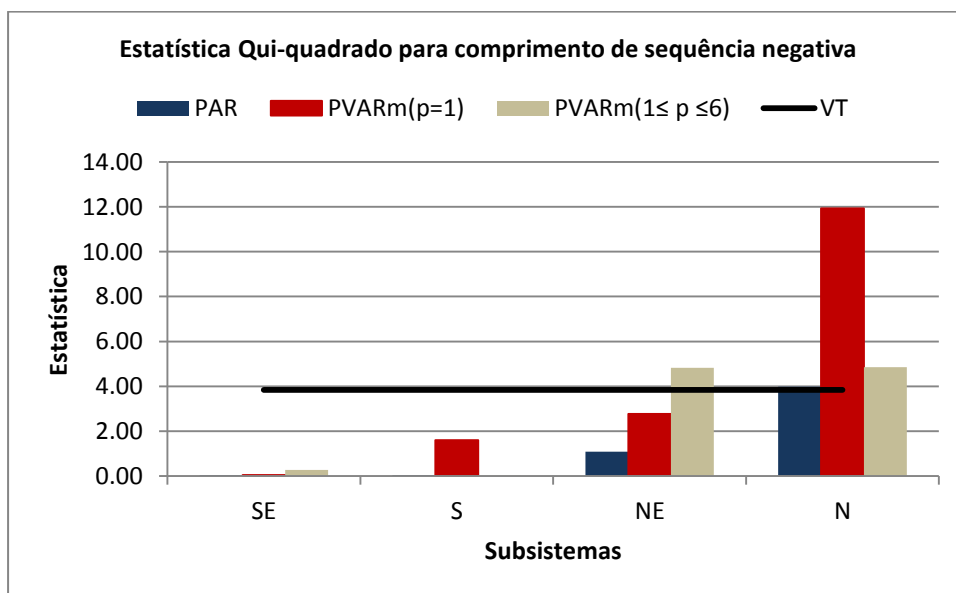


Figura 27: Estatística qui-quadrado para o teste de comprimento de sequência negativa.

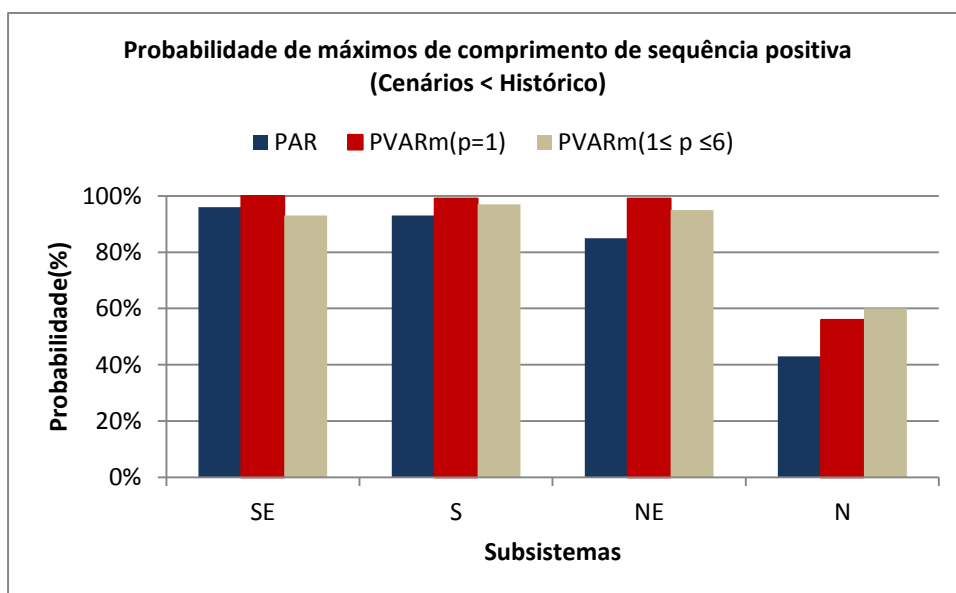


Figura 28: Probabilidade de máximos para comprimento de sequências positivas ($C < H$)

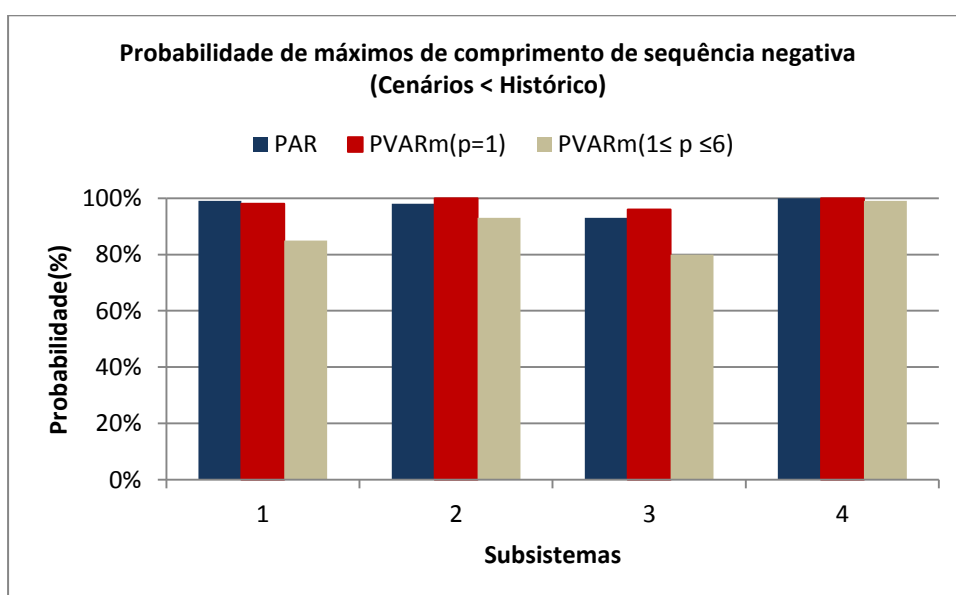


Figura 29: Probabilidade de máximos para comprimento de sequências negativas ($C < H$)

A Figura 30 e a Figura 31 apresentam os p-valores para o teste de K-S para somas de sequências positiva e negativa, para os modelos PAR, PVARm de ordem fixa e variável, para os subsistemas Sudeste, Sul, Nordeste e Norte. O nível de significância considerado é de 5%. Assim, aqueles modelos que apresentam p-valor menor do que 0.05 são rejeitados.

Na avaliação da soma de sequência positiva, atendo-se ao desempenho dos modelos PVARm, o modelo de ordem fixa é rejeitado para a reprodução de tal evento histórico para o subsistema Norte. Para os demais subsistemas, o modelo PVARm ($p = 1$) exibe melhor desempenho por apresentar os maiores p-valores em comparação com o modelo PVARm de ordem variável. Em relação ao desempenho do modelo PAR, este é rejeitado para os subsistemas Norte e Nordeste e apresenta o melhor p-valor para o subsistema Sul.

No âmbito de sequências negativas, os modelos PVARm são reprovados na reprodução de tal comportamento para os subsistemas Sudeste e Norte. O modelo PVARm ($p = 1$) apresenta melhor desempenho para o subsistema Nordeste, enquanto o PVARm ($1 \leq p \leq 6$) é o mais adequado ao subsistema Sul. Em relação ao modelo PAR, este também foi rejeitado para o Sudeste, foi o único aceito para a reprodução da soma de sequência negativa para o subsistema Norte e é o modelo de mais alto desempenho para o REE da região Nordeste.

A Figura 32 e a Figura 33 apresentam as probabilidades de máximos do histórico. Isto é, avalia-se a probabilidade da máxima soma de determinada sequência das séries geradas ser menor do que a maior soma de sequência identificada no histórico.

Para as sequências positivas, novamente o subsistema Norte é exceção e, para os demais, mais de 80% dos máximos encontrados nos cenários são menores do que o máximo histórico. Para o subsistema Norte, especificamente para sequências positivas, a probabilidade de ambos os modelos está em torno de 45%.

Para as sequências negativas, os modelos exibem comportamentos similares: todos eles possuem alta probabilidade de máximos, com valores acima de 93% em todos os casos, mostrando o baixo desempenho dos modelos na reprodução do montante histórico de energia natural afluente acumulado ao longo do período crítico identificado.

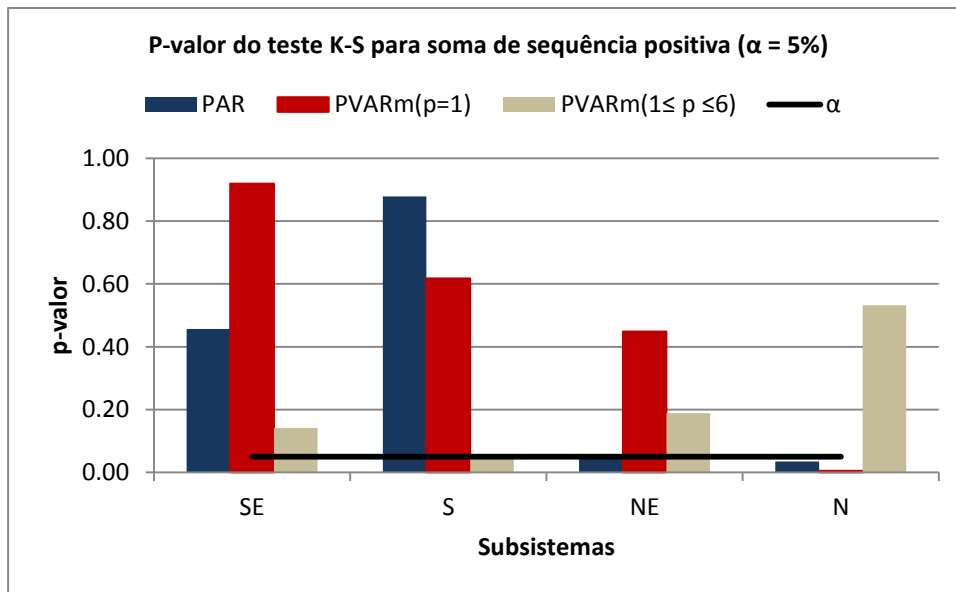


Figura 30: P-valor para o teste de hipótese de K-S para soma de sequência positiva.

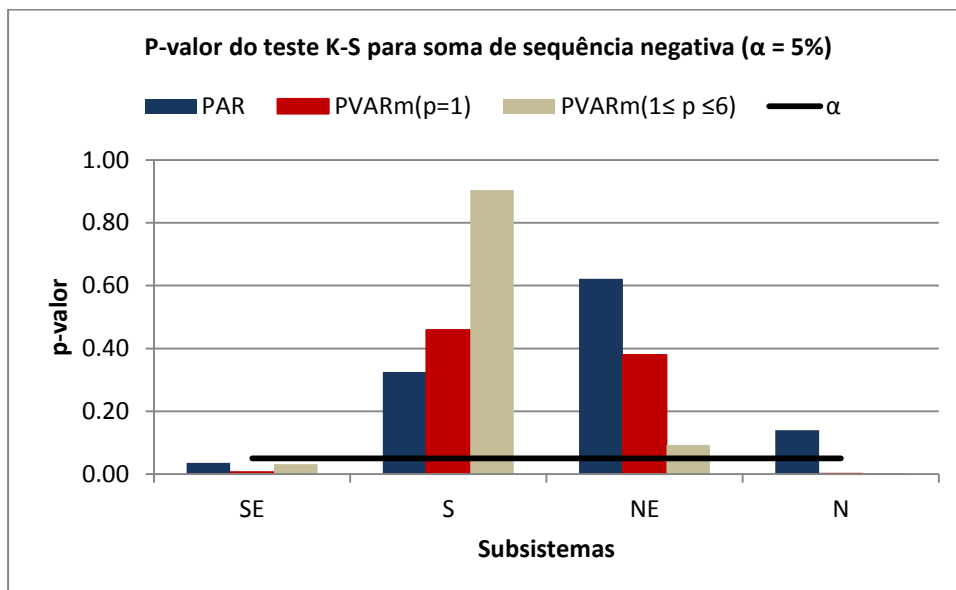
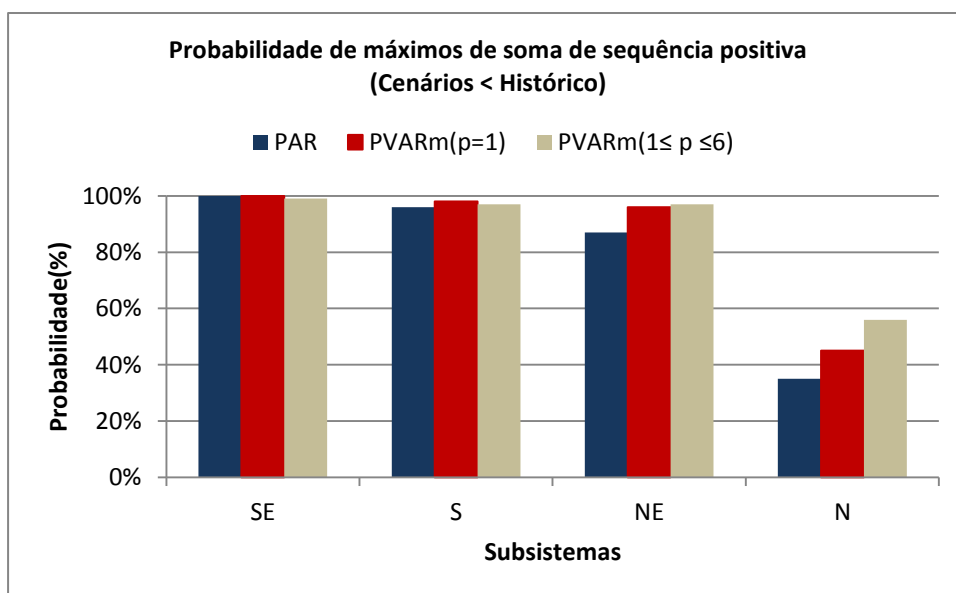
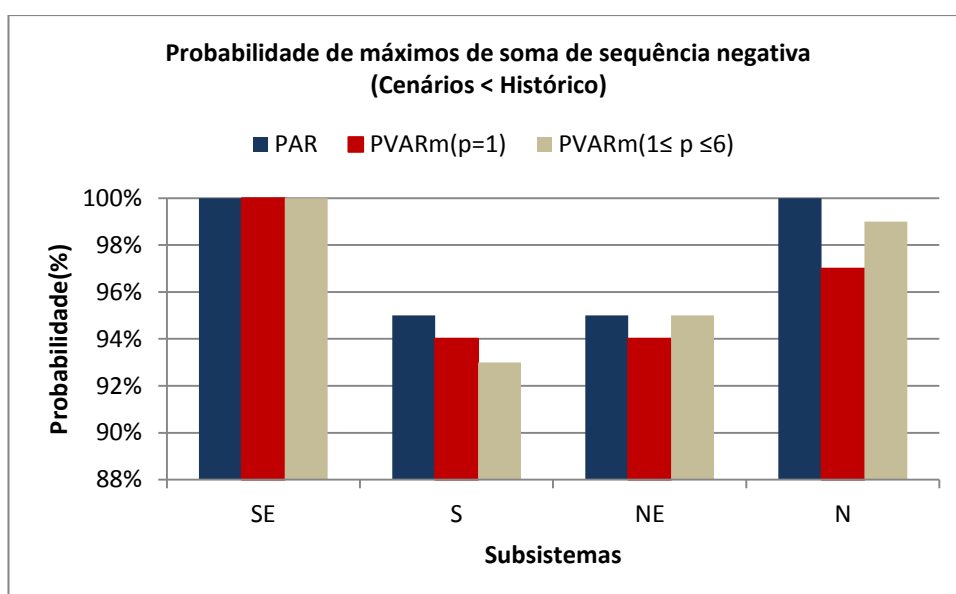


Figura 31: P-valor para o teste de hipótese de K-S para soma de sequência negativa.



**Figura 32: Probabilidade de máximos para soma de sequências positivas
(C<H)**



**Figura 33: Probabilidade de máximos para soma de sequências negativas
(C<H)**

A Figura 34 e a Figura 35 apresentam os p-valores para o teste de K-S para intensidade de sequências positiva e negativa, para os modelos PAR, PVARm de ordem fixa e variável, para todos os subsistemas. O nível de significância considerado é de 5%. Assim, aqueles modelos que apresentam p-valor menor do que 0.05 são rejeitados.

Paras as sequências positivas, apenas o modelo PVARm de ordem unitária é rejeitado na reprodução da intensidade histórica para o subsistema Norte. Em particular, o modelo PAR apresenta os maiores p-valores para os subsistemas Sudeste e Sul neste quesito, enquanto o modelo PVARm de ordem variável apresenta os melhores p-valores para os demais subsistemas.

A capacidade de reproduzir a intensidade histórica de sequências negativas é rejeitada para todos os modelos para os subsistemas Sudeste e Norte. Em particular, todos os modelos foram rejeitados na reprodução da soma de sequências negativas para o subsistema Sudeste e todos os modelos foram rejeitados na reprodução do comprimento de sequências negativas para o subsistema Norte. Por fim, o modelo PAR é rejeitado para a reprodução destas características para o subsistema Sul. Novamente o PVARm de ordem variável apresenta o melhor desempenho nessa categoria.

A Figura 36 e a Figura 37 apresentam as probabilidades das máximas intensidades de sequências positiva e negativa dos cenários gerados serem menores do que as máximas encontradas no histórico. Observa-se que, para sequências positivas, com exceção do modelo PVARm ($p = 1$) para o subsistema Sul, é altamente provável que os máximos dos cenários sejam menores do que os máximos históricos. O PVARm ($p = 1$) está associado à probabilidade de 73%.

No âmbito das sequências negativas, é altamente provável que os máximos dos cenários sejam menores do que os máximos históricos. Todos os casos apresentam probabilidade acima de 93%.

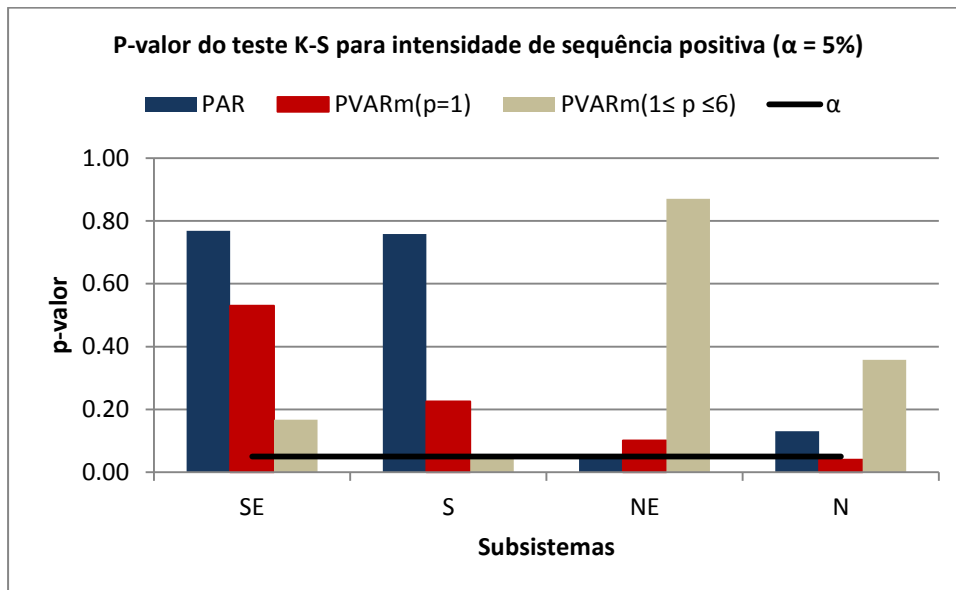


Figura 34: P-valor para o teste de hipótese de K-S para intensidade de sequência positiva.

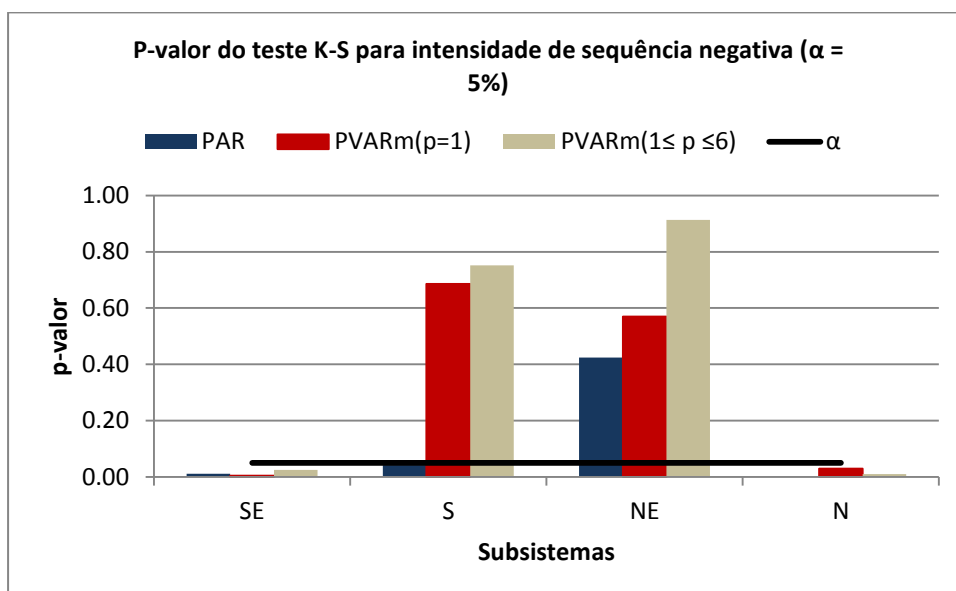


Figura 35: P-valor para o teste de hipótese de K-S para intensidade de sequência negativa.

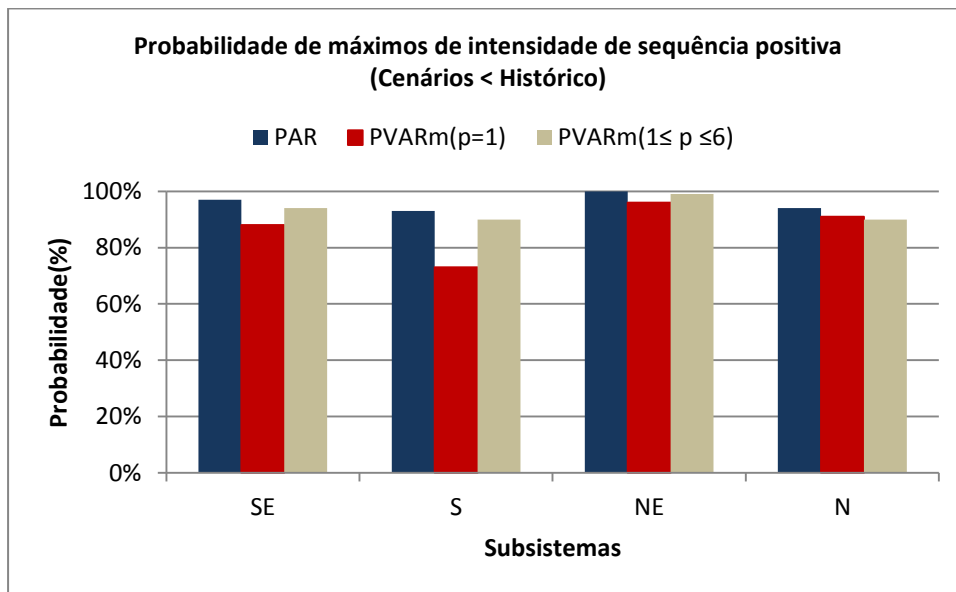


Figura 36: Probabilidade de máximos para intensidade de sequências positivas ($C < H$)

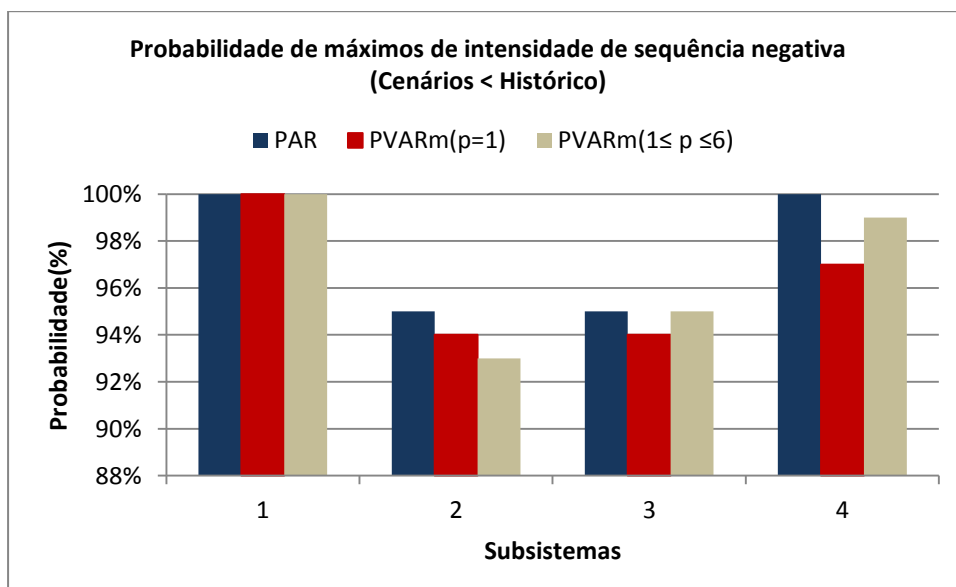


Figura 37: Probabilidade de máximos para intensidade de sequências negativas ($C < H$)

Os testes de sequência são recomendados na literatura para avaliação do desempenho de modelos aplicados ao Planejamento Energético. Nota-se que todos os modelos avaliados apresentaram seus piores desempenhos para sequências negativas. Para as sequências negativas, parte dos modelos sofreu rejeição da reprodução do comprimento dos eventos históricos de sequência e outra parte foi rejeitada nos testes de soma de sequências, por consequência, a intensidade de tais eventos também foi reprovada.

Nota-se ainda que, para os testes de probabilidade de máximos, a maior parte dos modelos não possui desempenho satisfatório, estimando probabilidades próximas de 90%, o que indica que o modelo apresenta dificuldade em reproduzir eventos históricos críticos.

5

Conclusões e Trabalhos Futuros

Este trabalho apresentou o modelo PVARm aplicado ao Planejamento Energético de Médio Prazo. Propôs-se a avaliação deste modelo em ordem unitária fixa e de ordem variada, escolhida através do critério de menor soma dos quadrados dos desvios. Ambos os modelos foram implementados e aplicados aos dados do PMO de dezembro de 2015, para comparação entre si e entre aquele modelo que representa o vigente, o modelo PAR.

Os modelos apresentados foram avaliados nos quesitos reprodução das estatísticas mensais média e desvio e na reprodução da estatística anual correlação espacial. Avaliou-se a igualdade entre as distribuições dos cenários gerados por ambos os modelos e a série histórica. Por fim, foi verificada a adequação dos modelos através da capacidade de reprodução do comprimento, da soma e da intensidade dos eventos históricos de sequências positiva e negativa.

Foi possível observar que o modelo PVARm de ordem unitária fixa se destacou na maioria dos testes selecionados, para os subsistemas Sudeste, Sul e Nordeste. Em particular, notou-se que o subsistema Norte teve representação precária por todos os modelos.

Restringindo-se à análise dos modelos PVARm explorados, de forma geral, para cada um dos testes realizados, os modelos de ordem fixa e variável ou exibiram comportamento similar, ou o modelo de ordem fixa apresentou melhor desempenho no teste. Dado o princípio da parcimônia, o acréscimo expressivo de parâmetros do modelo de ordem unitária para aquele variável não é justificado frente à comparação estabelecida entre os modelos. Assim, indica-se a melhor exploração do modelo de ordem unitária.

O modelo PAR é um modelo maduro e estabelecido no setor elétrico. Nota-se na literatura que o modelo é constantemente aprimorado para a aplicação no Planejamento Energético de Médio Prazo. A comparação entre as correlações espaciais estabelecida entre os modelos PVARm e PAR foi equivalente. Assim, nota-se que a transformação dos resíduos não correlacionados em resíduos

especialmente correlacionados se mostrou eficiente, nas condições estabelecidas para a análise.

Ressalta-se que o processo de geração de cenários, em todos os modelos se baseou em uma árvore de cenários reduzida, constituída de 20 nós em cada estágio. Os cenários gerados servem de insumo para o processo de otimização que define a política ótima de operação dos recursos energéticos do SIN.

A abordagem de modelos multiplicativos é recente, quando comparada a maturidade do modelo PAR e toda a extensa literatura acerca do mesmo. Nota-se que ainda é necessário estabelecer maiores investigações estatísticas no modelo, como aquelas propostas em (Cabral, 2016).

A maior motivação da evolução da abordagem multiplicativa no âmbito do Planejamento Energético está no acoplamento do mesmo com o modelo de otimização estocástica. O modelo PVARm relaciona as afluições entre os subsistemas, de forma linear, e é capaz de gerar afluições sempre positivas, sem criar nenhuma relação temporal entre os resíduos do modelo, garantindo a premissa de independência temporal entre as componentes aleatórias do modelo de otimização.

Nesse sentido, indicam-se, como continuidade deste trabalho, pontos propostos em (Cabral, 2016), que não foram contemplados neste trabalho, e são listados a seguir:

1. Condição de causalidade para o modelo PVARm: neste trabalho foi considerada uma conjectura. Se o modelo PVARm for causal, então a esperança condicional é igual à previsão e a variância condicional é proporcional ao produto cruzado das previsões;
2. Sensibilidade do ajuste do modelo: Avaliar variabilidade da ordem e estimativa de parâmetros do modelo, em função do tamanho da amostra;
3. Investigação de outros métodos de reamostragem dos resíduos

Além dos pontos listados acima, seguem novas propostas para continuidade deste trabalho:

4. Avaliação dos cenários gerados na etapa *backward*: O número de cenários gerados na etapa *forward* é reduzido e baseado em uma árvore de cenários de baixa cardinalidade. Investigar os cenários gerados na etapa *backward* e avaliar o tamanho da amostra utilizada.

5. Utilização de técnicas de agregação para a construção de cenários reduzida, utilizada no algoritmo PDDE.
6. Utilização de métodos de otimização alternativos na estimação dos parâmetros do modelo PVARm.
7. Investigação do espaço paramétrico dos coeficientes do modelo PVARm: este trabalho assumiu que apenas coeficientes não negativos podem garantir geração de afluências positivas.
8. Avaliação de um modelo PVAR estrutural, que permite escolha do conjunto relevante de variáveis para explicação do fenômeno modelado e permitiria redução de parâmetros do modelo.
9. Avaliação da incerteza do modelo interconfigurações: através de casos exemplo de PMO's antigos, confrontar os resultados de um modelo interconfigurações baseado em um cronograma estimado de incremento da potência instalada de hidrelétricas, com um modelo ajustado a partir das configurações que de fato foram realizadas. Avaliar o impacto de se considerar as mudanças de configurações topológicas no modelo estocástico.
10. Avaliação do modelo PVARm em conjunto com o modelo de otimização, a PDDE.

ANEEL. **Resolução Normativa Nº 425, de 1º de fevereiro de 2011**. Disponível em: <<http://www2.aneel.gov.br/cedoc/ren2011425.pdf>>. Acesso em: 2 maio. 2018.

BENDERS, J. F. Partitioning procedures for solving mixed-variables programming problems. **Numerische Mathematik**, v. 4, n. 1, p. 238–252, dez. 1962.

BODNAR, T.; HAUTSCH, N. Dynamic conditional correlation multiplicative error processes. **Journal of Empirical Finance**, v. 36, p. 41–67, 2016.

BORTOLOSSI, H. J.; PAGNONCELLI, B. K. Uma Introdução à Otimização sob Incerteza. **III Bienal da Sociedade Brasileira de Matemática**, p. 97, 2006.

BRAGA, D. B. M. **Processos Periódicos Autorregressivos Gama: Uma Contribuição para Modelagem da Energia Natural Afluente com Possíveis Usos na Determinação da Função de Custo Futuro**. Dissertação de Mestrado. UFRJ, 2011.

BRIGATTO, A.; STREET, A.; VALLADÃO, D. M. Assessing the cost of time-inconsistent operation policies in hydrothermal power systems. **IEEE Transactions on Power Systems**, v. 32, n. 6, p. 4541–4550, 2016.

CABRAL, F. G. **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. UFRJ, 2016.

CEPEL. **Manual de Referência do Modelo NEWAVE**. Rio de Janeiro: 2006.

_____. **Projeto NEWAVE: Modelo Estratégico de Geração Hidrotérmica a Subsistemas Equivalentes - Manual do Usuário**. 2013.

CHARBENEAU, R. J. Comparison of the Two and Three Parameter Log normal Distributions used in Streamflow Synthesis. **Water Resources Research**, v. 14, n. 1, p. 149–150, 1978.

CHEN, Z. L.; POWELL, W. B. Convergent cutting-Plane and Partial-Sampling

- Algorithm for Multistage Stochastic Linear Programs with Recourse. **Journal of Optimization Theory and Applications**, v. 102, n. 3, p. 497–524, 1999.
- CIPOLLINI, F.; ENGLE, R. F.; GALLO, G. M. Vector Multiplicative Error Models: Representation and Inference. **Statistics**, p. 1–53, 2006.
- DONOHUE, C.; BIRGE, J. The Abridged Nested Decomposition Method for Multistage Stochastic Linear Programs with Relatively Complete Recourse. **Algorithmic Operations Research**, v. 1, n. 1, p. 20–30, 24 jan. 2006.
- FERREIRA, P. G. C. **A Estocasticidade Associada ao Setor Elétrico Brasileiro e uma Nova Abordagem para a Geração de Afluências Via Modelos Periódicos Gama**. Tese de Doutorado. PUC-Rio, 2013.
- HAUTSCH, N.; MALEC, P.; SCHIENLE, M. Capturing the Zero: A New Class of Zero-Augmented Distributions and Multiplicative Error Processes. **Journal of Financial Econometrics**, v. 12, n. 1, p. 89–121, 1 jan. 2014.
- HIPEL, K. W.; MCLEOD, A. I. **Time Series Modelling of Water Resources and Environmental Systems**. Amsterdam: Elsevier. 1994.
- KELMAN, J. (CEPEL); PEREIRA, M. V. F. M. V. F. (CEPEL). **Crítérios de Avaliação para Modelos de Séries Hidrológicas** IV SNPTEE. **Anais...** Rio de Janeiro: 1977
- LANNE, M. A Mixture Multiplicative Error Model for Realized Volatility. **Journal of Financial Econometrics**, v. 4, n. 4, p. 594–616, 2006.
- MAASS, A. *et al.* **Design of Water-Resource Systems New Techniques for Relating Economic Objectives, Engineering Analysis, and Governmental Planning**. Harvard University Press, 1962.
- MACEIRA, M. E. P.; DAMÁZIO, J. M. The Use of the PAR(p) model in the Stochastic Dual Dynamic Programming Optimization Scheme Used in the Operation Planning of the Brazilian Hydropower System. **Probability in the Engineering and Informational Sciences**, v. 20, n. 1, p. 143–156, 2006.
- MARCATO, A. L. M. **Representação Híbrida de Sistemas Equivalentes e Individualizados para o Planejamento da Operação de Médio Prazo de Sistemas de Potência de Grande Porte**. Tese de Doutorado. Rio de Janeiro: PUC - Rio, 2002.
- MARCO, J. B.; HARBOE, R.; SALAS, J. D. (EDS.). **Stochastic Hydrology and its Use in Water Resources Systems Simulation and Optimization**. Peñíscola, Espanha: 1993

- MATALAS, N. C. Mathematical Assessment of Synthetic Hydrology. **Water Resources Research**, v. 3, n. 4, p. 937–945, dez. 1967.
- MATOS, V. L. DE; PHILPOTT, A. B.; FINARDI, E. C. Improving the performance of Stochastic Dual Dynamic Programming. **Journal of Computational and Applied Mathematics**, v. 290, p. 196–208, 2015.
- METROPOLIS, N.; ULAM, S. The Monte Carlo Method. **Journal of American Statistical Association**, v. 44, n. 247, p. 335–341, 2008.
- OLIVEIRA, F. L. C. **Nova abordagem para geração de cenários de afluições no planejamento da operação energética de médio prazo**. Dissertação de Mestrado. PUC-Rio, 2010.
- _____. **Modelo de Séries Temporais para Construção de Árvores de Cenários Aplicadas à Otimização Estocástica**. Tese de Doutorado. PUC-Rio, 2013.
- ONS. **Plano da Operação Energética 2017/2021**. Rio de Janeiro, 2017.
- _____. **O que é o SIN**. Operador Nacional do Sistema Elétrico, 2018. Disponível em: <<http://www.ons.org.br/pt/paginas/sobre-o-sin/o-que-e-o-sin>>. Acesso em: 24 jan. 2018.
- PENNA, D. D. J. **Definição da Árvore de Cenários de Afluições para o Planejamento da Operação Energética de Médio Prazo**. Tese de Doutorado. PUC-Rio, 2009.
- PEREIRA, I. F. *et al.* Using Stochastic Dual Dynamic Programming and a Periodic Autoregressive Model for Wind-hydrothermal Long-Term Planning. **IEEE Eindhoven PowerTech, PowerTech 2015**, 2015.
- PEREIRA, M. V. F. *et al.* Stochastic Streamflow Models for Hydroelectric Systems. **Water Resources Research**, v. 20, n. 3, p. 379–390, 1984.
- PEREIRA, M. V. F.; PINTO, L. M. V. G. **Operation Planning of Large-Scale Hydroelectrical Systems**. The Organizing Committee of the Eighth Power Systems Computation Conference, 1984.
- PEREIRA, M. V. F.; PINTO, L. M. V. G. Stochastic Optimization of a Multireservoir Hydroelectric System: A Decomposition Approach. **Water Resources Research**, v. 21, n. 6, p. 779–792, 1985.
- PEREIRA, M. V. F.; PINTO, L. M. V. G. Multi-stage stochastic optimization applied to energy planning. **Mathematical Programming**, v. 52, n. 1–3, p. 359–375, 1991.
- PIXABAY. **Mapa do Brasil**. Acesso a imagens gratuitas. Disponível em:

<<https://pixabay.com/pt/brasil-geografia-mapa-estados-153881/>>. Acesso em: 15 abr. 2018.

RIBEIRO, B. A.; BALDIOTI, H. R.; SOUZA, R. C. Modelo PAR(p) com Resíduo Multiplicativo Aplicado na Simulação de Cenários de Energia Natural Afluente. **XLIX Simpósio Brasileiro de Pesquisa Operacional**, 2017.

SHAPIRO, A. Analysis of Stochastic Dual Dynamic Programming Method. **European Journal of Operational Research**, v. 209, n. 1, p. 63–72, 2011.

____. Final Report for Technical Cooperation between Georgia Institute of Technology and ONS. 2012.

SHAPIRO, A.; PHILPOTT, A. A Tutorial on Stochastic Programming. **Manuscript**. Available at www2.isye.gatech.edu ..., p. 1–35, 2007.

SOUZA, R. C. *et al.* **Planejamento da Operação de Sistemas Hidrotérmicos no Brasil: Geração de Cenários e Otimização**. 1. ed. Rio de Janeiro: Ed. PUC-Rio, 2014.

STEDINGER, J. R. **Report on the Evaluation of CEPEL's PAR Models**. Cornell University, Ithaca, New York., 2001.

7

Apêndice A

7.1. Algoritmo para Geração de Cenários com Modelo PVARm

O algoritmo desenvolvido para geração de cenários pelo modelo PVARm Interconfigurações pode ser desmembrado em três etapas, apresentadas pela Figura 38, a seguir:

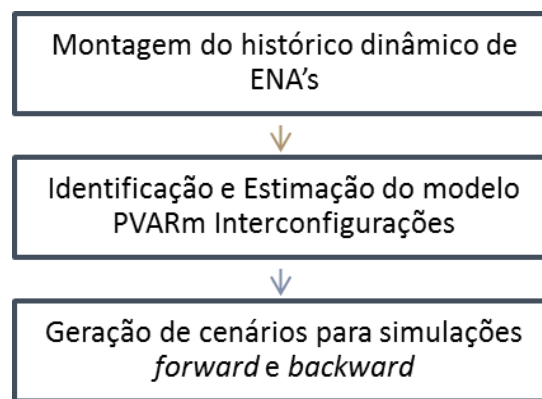


Figura 38: Etapas do algoritmo para ajuste e geração de cenários com o modelo PVARm Interconfigurações.

A primeira etapa consiste na adequação dos dados históricos disponíveis no PMO ao formato necessário para a modelagem PVARm. A segunda etapa detalha o processo de ajuste e identificação adotados para o modelo. Por fim, o terceiro passo envolve a simulação de afluências a partir do modelo ajustado na segunda etapa.

A seguir, cada uma das subseções descreve a parte do algoritmo utilizado que cabe em cada um dos passos apresentados.

7.1.1. Montagem do Histórico Dinâmico de ENAs

Dentre os arquivos que compõem o conjunto do Deck de Preços do Newave, de 12/2015, estão informações históricas de vazões, desde janeiro de 1931, até dezembro de 2013, parâmetros sobre a eficiência das usinas do sistema, além do cronograma de expansão da geração. Para maiores informações acerca desses arquivos, consultar (CEPEL, 2006).

A partir dessas informações, as hidrelétricas, com reservatório e a fio d'água, são agrupadas por região geográfica, sendo elas Sudeste, Sul, Nordeste e Norte, como mostrou a Figura 8.

Em seguida, calculam-se séries históricas de energias para os subsistemas definidos, para o período de 1931 a 2013. Cada uma das séries é gerada em função de uma das diferentes configurações apresentadas pelo parque hidrelétrico, ao longo do horizonte de estudo.

Uma vez calculadas as diferentes séries históricas de energia para cada mudança de configuração, cada um dos períodos (mês/ano) que compõem o horizonte de estudo, pré-estudo e pós-estudo é associado ao histórico de ENA's do mês e configuração correspondentes. Esse histórico, chamado neste trabalho de histórico dinâmico, é utilizado para ajuste do modelo PVARm, que tem a implementação de sua identificação, ajuste e proposta para geração de cenários descrita a seguir. O histórico dinâmico montado a partir das informações do PMO de 12/2015 tem sua estrutura apresentada a seguir, pela Figura 39.

PERÍODO	1	2	...	12	13	14	...	61	62	...	73
MÊS	DEZ	JAN	...	NOV	DEZ	JAN	...	DEZ	JAN	...	DEZ
ANO	PRÉ	PRÉ	...	PRÉ	2015	2016	...	2019	PÓS	...	PÓS
CONF	1	1	...	1	1	2	...	49	49	...	49
1931			
1932			
.
.
.
2013			
PRÉ-ESTUDO				ESTUDO				PÓS-ESTUDO			

Figura 39: Histórico dinâmico montado a partir do PMO de 12/2015

7.1.2.

Identificação do Modelo e Estimação de Parâmetros

Esta seção apresenta o algoritmo utilizado na identificação do modelo PVARm, bem como na estimação de seus parâmetros.

Especificamente para o PMO de 12/2015, que constitui o conjunto dos dados de entrada para os resultados apresentados nesta dissertação, é considerado o total de 73 períodos para modelagem: 12 meses de pré-estudo em configuração estática; 49 meses correspondentes ao período de estudo, de 12/2015 a 12/2019, em configuração dinâmica; 12 meses relacionados ao pós-estudo, em configuração estática.

Apesar do período de pós-estudo se estender por 5 anos, uma vez que sua configuração é estática, neste caso, basta ajustar o modelo para o primeiro ano e utilizar esse mesmo modelo para a simulação de séries ao longo dos 5 anos.

Uma vez que os dados estejam organizados como mostrou a seção anterior, há disponível uma série de ENAs de janeiro de 1931 ($t = 1$) a dezembro de 2013 ($t = 996$), isto é, há dados da série temporal para $t = 1, 2, \dots, 996$. Deseja-se modelar o total de 73 períodos, a começar pelo primeiro período do pré-estudo ($t = 1008$), correspondente ao mês $m = 12$, sendo o mês m relacionado a t obtido através da eq. (3). A lacuna entre $t = 996$ e $t = 1008$ foi preenchida com valores das médias mensais da série histórica, ou média de longo termo (MLT).

A proposta do algoritmo é apresentada pelo fluxograma da Figura 40 e consiste em calcular candidatos a modelo de $p_{m\acute{a}x}$ ordens diferentes por subsistema, sendo $p_{m\acute{a}x}$ a ordem máxima permitida, definida como 6 neste trabalho. Para determinado mês m , na configuração do período t , o candidato que apresentar a menor soma dos quadrados dos desvios (SEQ_{p,m,c_T}^C) será considerado o modelo daquele período, para o subsistema avaliado.

Conforme apresentado pela Figura 40, o primeiro passo do algoritmo consiste em calcular os modelos de todos os períodos do horizonte de planejamento considerando a ordem $p = 1$. Agrupam-se na matriz Z_m , eq. (32), todas as observações do histórico relativas às regressões em relação aos meses m da série, até a ordem p , isto é, as defasagens $t = m - 1, \dots, m - p, (m + S) - 1, \dots, (m + S) - p, \dots$ para todos os subsistemas. A partir de então, são

calculados os modelos para cada um dos subsistemas: agrupam-se as observações históricas relativas ao mês m , isto é, $t = m, m + \mathbb{S}, \dots, m + q_m \mathbb{S}$, calcula-se o candidato a modelo, de ordem 1, via método dos mínimos quadrados (MQ) com restrição de não negatividade nos coeficientes.

Os candidatos a modelo de cada um dos 73 períodos com ordem unitária fixa são calculados e considerados como os próprios modelos dos períodos. A seguir, o mesmo procedimento realizado para $p = 1$ é repetido para $p = 2, 3, \dots, 6$.

Para cada candidato ajustado com ordem superior à unidade, comparam-se as somas dos quadrados dos desvios, dada pela eq. (44), do candidato (SEQ_{p,m,c_T}^C) e do modelo associado ao período avaliado (SEQ_{p,m,c_T}^M). Se, em algum momento, a SEQ_{p,m,c_T}^C do candidato calculado for menor do que a SEQ_{p,m,c_T}^M , então o candidato corrente se transforma em modelo associado ao período.

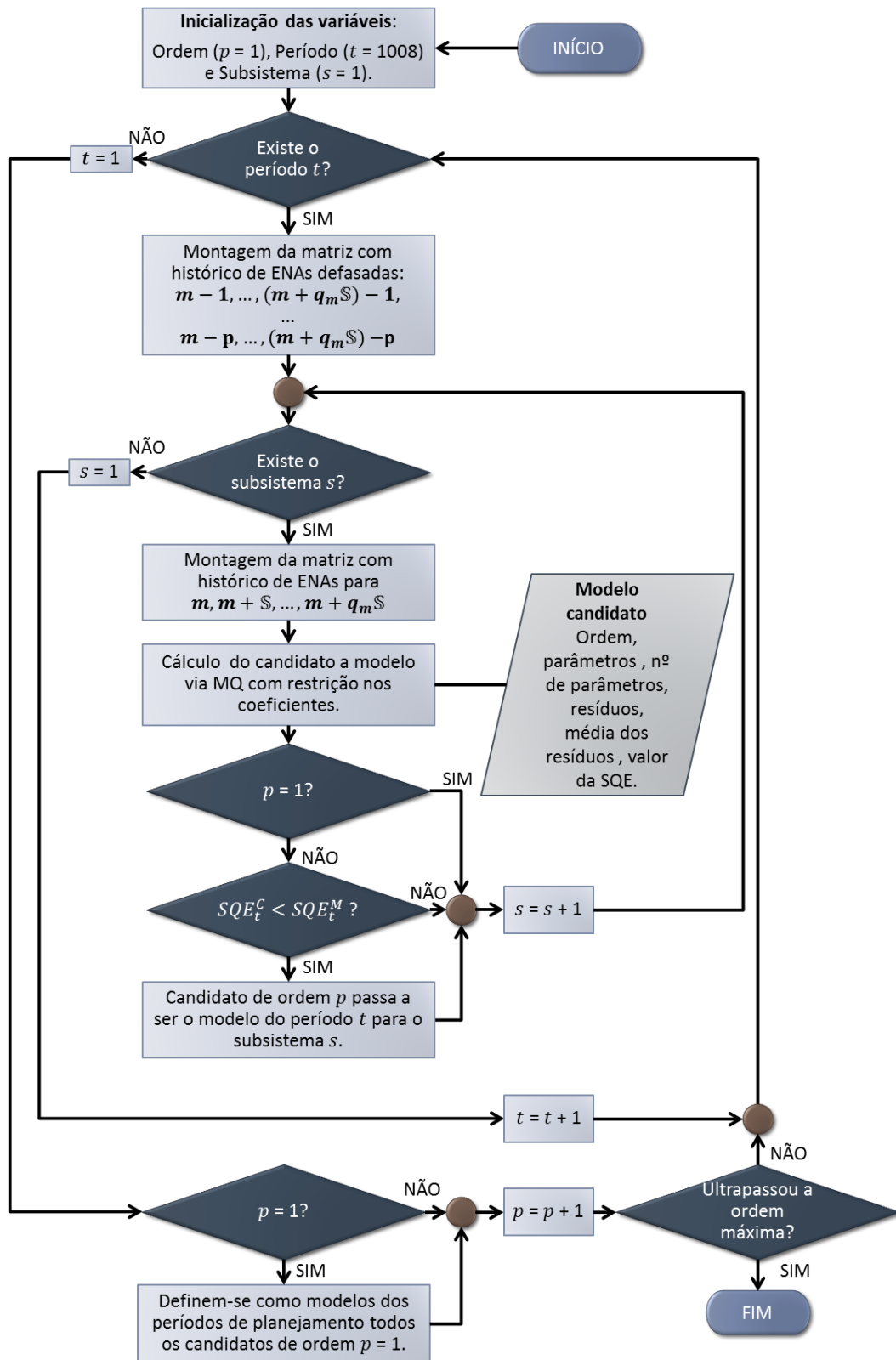


Figura 40: Fluxograma do código implementado para identificação e estimação do modelo PVARm Interconfigurações

7.1.3. Geração de Cenários

Uma vez que estejam definidos os modelos dos 73 períodos, a geração de cenários consiste em obter uma amostra de resíduos do modelo para cálculo dos cenários utilizados nas etapas *forward* e *backward* do método de otimização.

Inicialmente, a ausência de dados entre $t = 997$ e $t = 1007$ é preenchida com valores da MLT e a simulação é iniciada para o período $t = 1008$, que é o primeiro mês correspondente ao pré-estudo.

A matriz de aberturas é definida através de uma amostra de resíduos calculados na etapa de ajuste do modelo, a partir de sorteio aleatório. Foram utilizadas 20 aberturas por período, seguindo o padrão do modelo vigente para montagem dos cenários na etapa *backward*. Nota-se que a árvore de cenários é reduzida, diante do número de aberturas selecionado. Neste sentido, existem diversos métodos que propõem construir árvores de cenários reduzidas, como apresentado em (Penna, 2009).

Para os cenários utilizados nas simulações *forward*, utiliza-se a geração em paralelo de 200 séries. Para cada período t , faz-se o sorteio de um resíduo correspondente contido na matriz de aberturas daquele período. Com este resíduo calculam-se o cenário *forward* e os 20 cenários *backwards* correspondentes ao passado determinado por esta série *forward*. O procedimento é repetido para o período t até que as 200 séries *forward* e os cenários *backward* associados estejam calculados. O algoritmo construído para tal é representado na Figura 41, através de um fluxograma.

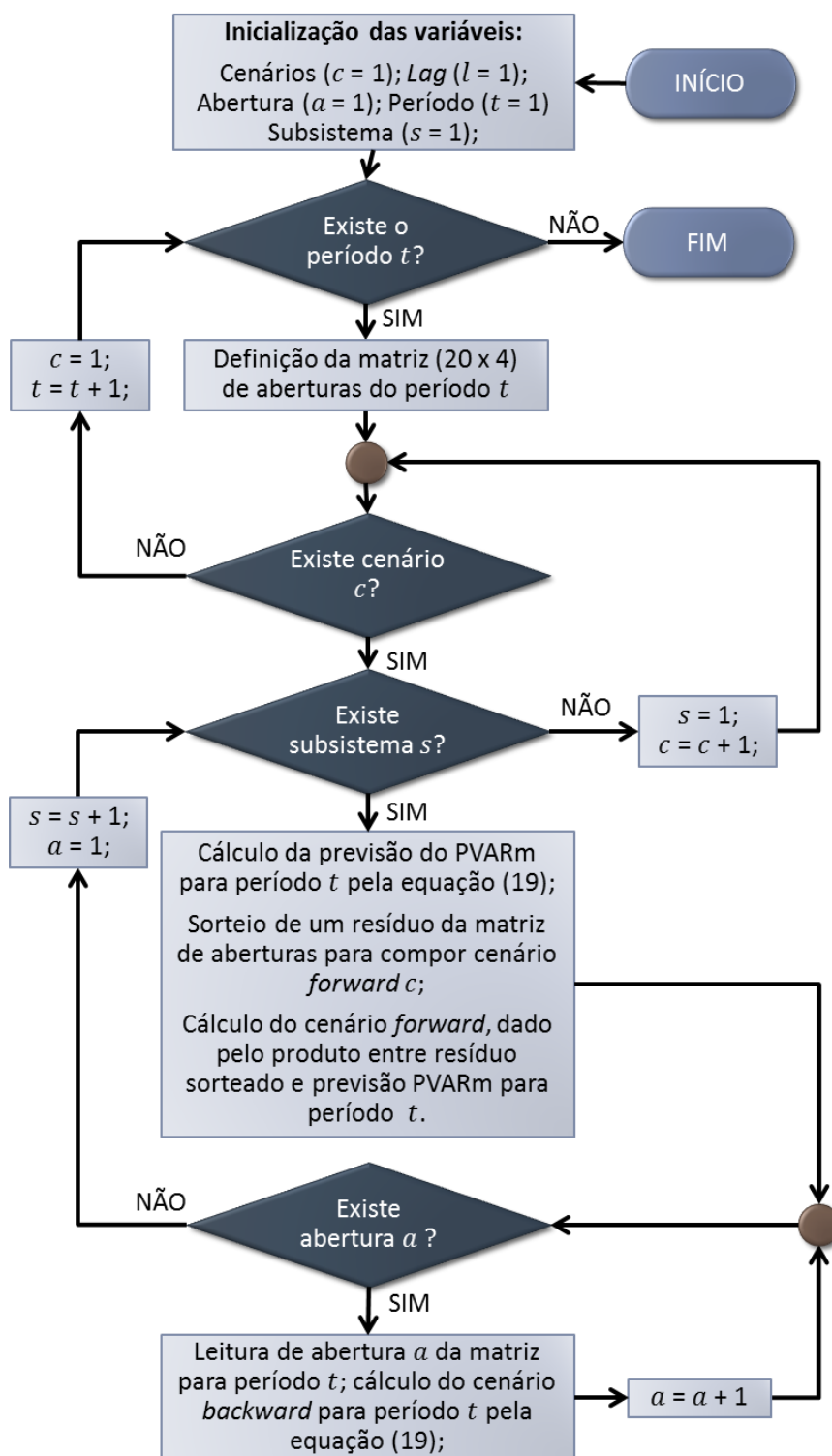


Figura 41: Fluxograma do algoritmo da geração de cenários *forward* e *backward*.

7.2.

Geração de Cenários de Afluências pelo Modelo PAR

Os cenários gerados a partir do modelo PAR foram obtidos pelo módulo de geração de cenários do MDDH. Neste módulo, o modelo PAR foi ajustado sobre o histórico dinâmico, conforme apresentado em (Ferreira, 2013). Os processos de estimação dos parâmetros e identificação do modelo são realizados conforme metodologia apresentada na seção 2.2.2.1. Além disso, utiliza-se da técnica de redução da ordem do modelo PAR caso seja identificado que este tenha potencial para gerar cortes negativos, como apresentado em (Maceira e Damázio, 2006).

A geração de cenários a partir do modelo ajustado é realizada via método Monte Carlo. Ajusta-se uma distribuição lognormal a 3 parâmetros aos resíduos, que são transformados para incorporar a correlação espacial entre as vazões aos REE's, como indicado em (Penna, 2009).

Sorteia-se uma matriz de 20 aberturas por período, a partir da distribuição ajustada, compondo a matriz de todos os resíduos consultados na fase *backward* do processo de otimização.

Os cenários *forward* são construídos a partir do sorteio aleatório dos resíduos contidos na matriz de aberturas, seguindo o procedimento de geração em paralelo, como apresentado para o modelo PVARm. Os cenários *backward* são calculados em função de cada uma das 200 séries *forward* geradas.

Para maiores informações acerca do ajuste do modelo PAR e geração de cenários para atendimento das simulações *forward* e *backward* no modelo MDDH, consultar (Oliveira, 2010).