

**APLICAÇÃO DO MÉTODO AMOSTRAGEM SELETIVA GLOBAL NA  
CONSTRUÇÃO DA ÁRVORE DE CENÁRIOS DE AFLUÊNCIAS UTILIZADA  
NO PLANEJAMENTO DA OPERAÇÃO ENERGÉTICA DE MÉDIO PRAZO**

Túlio de Souza Silva

PROJETO SUBMETIDO AO CORPO DOCENTE DO DEPARTAMENTO DE  
ENGENHARIA ELÉTRICA DA ESCOLA POLITÉCNICA DA UNIVERSIDADE  
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS  
PARA A OBTENÇÃO DO GRAU DE ENGENHEIRO ELETRICISTA.

Aprovada por:

---

Prof. Carmen Lucia Tancredo Borges  
(Orientadora)-UFRJ

---

Débora Dias Jardim Penna  
(Co-Orientadora)-CEPEL

---

Prof. Djalma Mosqueira Falcão  
UFRJ

---

José Francisco Moreira Pessanha  
UERJ-CEPEL

RIO DE JANEIRO, RJ - BRASIL  
JUNHO DE 2010

À minha mãe Diva, pai Adir,  
irmão Gabriel e à  
minha noiva Juliana.

## **Agradecimentos**

A Deus.

Aos meus pais, Diva e Adir, que sempre estiveram ao meu lado me apoiando, incentivando e ajudando incondicionalmente.

A minha noiva, Juliana, pelo apoio e pela compreensão nos momentos difíceis.

A pesquisadora e co-orientadora Débora Dias Jardim Penna pelas oportunidades de participar de estudos na área de planejamento da operação energética e desenvolver este trabalho.

A minha orientadora Carmen Lúcia Tancredo Borges por acreditar na realização deste trabalho e pelos conselhos durante esta etapa.

Ao amigo e pesquisador Wellington Luis de Oliveira pelos ensinamentos e conversas produtivas durante a realização deste trabalho.

A Maria Elvira Piñeiro Maceira pela oportunidade de estagiar no CEPEL, fazendo parte do Departamento de Otimização Energética e Meio Ambiente.

Aos colegas integrantes do Departamento de Otimização Energética e Meio Ambiente, Michel Pompeu, Renato Cabral, Manuel Fragoso, Saulo Gomes, Francisco Thomaz e Vitor Duarte pelo incentivo, paciência e por tudo que me ensinaram.

Aos professores e funcionários do Departamento de Engenharia Elétrica da UFRJ pelas conversas e momentos felizes que compartilhamos durante essa caminhada.

Aos colegas de curso e em especial aos amigos Flávius Vinícius Caetano Bezerra, César Vilasbôa de Vasconcellos e Rosana Paiva Mota pelo apoio e ajuda durante os momentos difíceis e pelos vários momentos de alegria e descontração.

Resumo do Projeto Final em Engenharia Elétrica apresentado ao Departamento de Engenharia Elétrica da Escola Politécnica – UFRJ como parte dos requisitos necessários para a obtenção do grau de Engenheiro Eletricista:

**APLICAÇÃO DO MÉTODO AMOSTRAGEM SELETIVA GLOBAL NA  
CONSTRUÇÃO DA ÁRVORE DE CENÁRIOS DE AFLUÊNCIAS UTILIZADA  
NO PLANEJAMENTO DA OPERAÇÃO ENERGÉTICA DE MÉDIO PRAZO**

Túlio de Souza Silva

Junho de 2010

Orientadores: Carmen Lúcia Tancredo Borges

Débora Dias Jardim Penna

Os estudos de planejamento da operação energética de médio prazo são realizados utilizando o modelo NEWAVE desenvolvido pelo CEPEL. Nesse modelo a incerteza referente às afluências é considerada através da utilização de cenários hidrológicos sintéticos gerados por um modelo autorregressivo periódico. O conjunto de todas as possíveis realizações da variável aleatória afluência ao longo do horizonte de estudo formam a árvore de cenários, que tem dimensão elevada. Devido a essa característica, somente uma pequena porção dessa árvore (sub-árvore) é percorrida nos estudos de planejamento da operação energética de médio prazo.

Este trabalho propõe um método de amostragem, denominado Amostragem Seletiva Global, que utiliza técnicas de agregação para a construção da sub-árvore de forma a preservar as características da árvore original. Pretende-se, através de uma construção acurada da sub-árvore, melhorar a representação do problema proposto, tornando o modelo mais robusto com relação às variações nos cenários hidrológicos.

# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO.....</b>	<b>10</b>
1.1	Considerações iniciais .....	10
1.2	Contextualização do trabalho.....	10
1.3	Objetivos.....	12
1.4	Metodologia.....	12
1.5	Organização .....	13
<b>2</b>	<b>PLANEJAMENTO DA OPERAÇÃO ENERGÉTICA .....</b>	<b>14</b>
2.1	Considerações iniciais .....	14
2.2	Sistemas Termoelétricos .....	14
2.3	Sistemas Hidrotérmicos .....	15
2.4	Planejamento da Operação Energética no Brasil.....	19
2.4.1	Cadeia dos Modelos .....	21
2.4.2	Planejamento de Médio Prazo .....	22
2.4.3	Planejamento de Curto Prazo.....	23
2.4.4	Programação da Operação .....	23
2.5	Resumo .....	24
<b>3</b>	<b>PLANEJAMENTO DA OPERAÇÃO DE MÉDIO PRAZO E O NEWAVE .....</b>	<b>25</b>
3.1	Newave.....	25
3.2	Representação da estocasticidade hidrológica no modelo NEWAVE.....	27
3.3	Construção da árvore de cenários .....	28
3.4	Resumo .....	28
<b>4</b>	<b>MODELO DE GERAÇÃO DE SÉRIES SINTÉTICAS DE ENERGIAS.....</b>	<b>29</b>
4.1	Considerações Iniciais .....	29
4.2	Introdução a Processos Estocásticos e Séries Temporais .....	30
4.3	O Modelo Auto-regressivo periódico.....	34
4.3.1	Descrição do Modelo.....	34
4.3.2	Ajuste do Modelo.....	37
<b>5</b>	<b>MÉTODOS DE AGREGAÇÃO.....</b>	<b>41</b>
5.1	Considerações Iniciais .....	41
5.2	Técnicas de Agrupamento.....	41
5.2.1	Métodos Hierárquicos.....	44
5.2.2	Métodos Não Hierárquicos.....	46
5.3	Resumo .....	51
<b>6</b>	<b>PROPOSTA DE CONSTRUÇÃO DA SUB-ÁRVORE DE CENÁRIOS .....</b>	<b>52</b>
6.1	Introdução .....	52
6.2	Por que usar técnicas de agregação .....	52
6.3	Proposta para a construção da sub-árvore.....	53
6.4	Comparação com AS e Atual .....	55
6.5	Resumo .....	57
<b>7</b>	<b>APLICAÇÃO DA METODOLOGIA PROPOSTA .....</b>	<b>58</b>
7.1	Introdução .....	58
7.2	Padronização da matriz de energias .....	58
7.3	Utilizando distâncias máximas, mínimas e médias entre os subsistemas .....	59
7.4	Utilização da distância euclidiana convencional.....	61
7.5	Escolha dos cenários representantes .....	61
7.6	Resultados após aplicação do método K-Means .....	62
7.6.1	Inicialização com 200 cenários tomados espaçados igualmente (matriz original).....	63
7.6.2	Ordenando e tomindo cenários espaçados igualmente .....	64
7.6.3	Ordenando e extraíndo 100 menores e 100 maiores.....	66
7.6.4	Selecionando cenários próximos à média.....	67
7.7	Agregação utilizando ruídos aleatórios .....	68
7.8	Avaliação dos cenários hidrológicos gerados .....	71

7.8.1	Média dos cenários .....	71
7.8.2	Desvio-padrão dos cenários .....	73
7.8.3	Correlação Cruzada dos Cenários.....	75
7.9	Resumo do capítulo.....	78
<b>8</b>	<b>ANÁLISE DOS RESULTADOS DO PLANEJAMENTO DA OPERAÇÃO .....</b>	<b>79</b>
8.1	Considerações iniciais .....	79
8.2	Avaliação da Convergência .....	80
8.2.1	Sensibilidade com relação às variações dos cenários hidrológicos .....	81
8.2.2	Avaliação da convergência de um caso estudo de PMO .....	85
8.3	Avaliação dos índices de desempenho utilizando a política ótima calculada ....	91
8.4	Resumo .....	101
<b>9</b>	<b>CONCLUSÕES E SUGESTÕES PARA TRABALHOS FUTUROS .....</b>	<b>102</b>
9.1	Conclusões referentes à aplicação da ASG .....	102
9.2	Sugestões para trabalhos futuros.....	103
<b>APÊNDICE A - TESTE DE SANIDADE ROTINA FORTRAN .....</b>	<b>108</b>	
A.1	Teste de sanidade da rotina fortran K-Means .....	108
A.2	Agregação realizada com K-Means MATLAB.....	108
A.2.1.	Antes do uso do fator de compensação do desvio padrão .....	108
A.2.2.	Após o uso do fator de correção do desvio .....	110
A.3	Agregação realizada com K-Means rotina Fortran – somente após uso fator corretivo do desvio .....	111

## Índice de figuras

Figura 3.1 : Exemplo de árvore de cenários completa, considerando 3 estágios e 3 aberturas ....	27
Figura 3.2 : opção Atual .....	28
Figura 4.1 : Relação Demanda x Risco em um ano qualquer (CEPEL, 2002) .....	29
Figura 4.2 : Vazão Afluente Anual x Anos (CEPEL, 2002) .....	30
Figura 4.3 : Tendência hidrológica.....	39
Figura 5.1 : Etapas para a aplicação da análise de cluster.....	43
Figura 5.2 : Processos aglomerativos e divisivos.....	44
Figura 5.3 : Definição da medida de similaridade .....	45
Figura 5.4: Dendograma utilizando ligação simples.....	46
Figura 5.5: etapas do processo de agregação com K-Means .....	50
Figura 6.1: Escolha do cenário representativo.....	54
Figura 6.2: Aplicação do procedimento de agregação .....	54
Figura 7.1: inicialização 200 cenários espaçados igualmente .....	63
Figura 7.2: inicialização 200 cenários espaçados igualmente .....	64
Figura 7.3 inicialização 200 cenários espaçados igualmente de um vetor ordenado .....	65
Figura 7.4: inicialização 200 cenários espaçados igualmente de um vetor ordenado .....	65
Figura 7.5 inicialização 200 cenários tomando os extremos.....	66
Figura 7.6: inicialização 200 cenários tomando os extremos.....	66
Figura 7.7: inicialização 200 cenários próximos à média .....	67
Figura 7.8: inicialização 200 cenários próximos à média .....	68
Figura 7.9: Aplicação do Fator de Compensação.....	69
Figura 7.10: Degradacão do desvio-padrão .....	70
Figura 7.11: média dos cenários – op. Atual .....	72
Figura 7.12: média dos cenários – op. AS .....	72
Figura 7.13: média dos cenários – op. ASG .....	73
Figura 7.14: desvio-padrão dos cenários – op. Atual .....	74
Figura 7.15: desvio-padrão dos cenários – op. AS .....	74
Figura 7.16: desvio-padrão dos cenários – op. ASG .....	75
Figura 7.17: correlação cruzada dos cenários -op. Atual .....	76
Figura 7.18: correlação cruzada dos cenários - op. AS.....	76
Figura 7.19: correlação cruzada dos cenários - op. ASG.....	77
Figura 8.1: Representação subsistemas equivalentes e seus troncos de transmissão.....	79
Figura 8.2: Critério de convergência estatístico .....	81
Figura 8.3: Variabilidade de ZINF e ZSUP com variações de amostras de cenários.....	82
Figura 8.4: Variabilidade de ZINF e ZSUP com relação às variações de amostras de cenários.....	83
Figura 8.5: Média e desvio-padrão ZSUP .....	84
Figura 8.6: Média e desvio-padrão de ZINF .....	84
Figura 8.7: Convergência método Atual.....	85
Figura 8.8: Convergência método AS.....	86
Figura 8.9: Convergência método ASG.....	86
Figura 8.10: Diferença de ZINF método Atual .....	87
Figura 8.11: Diferença de ZINF método AS .....	87
Figura 8.12: Diferença de ZINF método ASG .....	88
Figura 8.13: Diferença de ZSUP - Atual.....	88
Figura 8.14: Diferença de ZSUP método AS .....	89
Figura 8.15: Diferença de ZSUP- ASG .....	89
Figura 8.16: Evolução da energia não suprida. ....	90
Figura 8.17: Evolução do percentual de risco de não atendimento à demanda. ....	<b>Erro! Indicador não definido.</b>
Figura 8.18: Variação do CMO médio relacionado à escolha do cenário hidrológico – Método Atual.....	92
Figura 8.19 : Variação do CMO médio relacionado à escolha do cenário hidrológico – Método AS. .....	92
Figura 8.20: Variação do CMO médio relacionado à escolha do cenário hidrológico - Método ASG. .....	93
Figura 8.21: Variação do custo total de operação com relação à amostra de cenário hidrológico.....	93

Figura 8.22: Variação EARMF do Sudeste.....	94
Figura 8.23: Variação CMO do Sudeste. ....	95
Figura 8.24: Variação EARMF do Sul. ....	95
Figura 8.25: Variação CMO do Sul.....	96
Figura 8.26: Variação EARMF do Nordeste.....	96
Figura 8.27: Variação CMO do Nordeste. ....	97
Figura 8.28: Variação EARMF do Norte.....	97
8.29: Variação CMO do Norte. ....	98
Figura 8.31: Coeficiente de variação do Sudeste.....	99
Figura 8.32: Coeficiente de variação do Sul.....	99
Figura 8.33: Coeficiente de variação do Nordeste. ....	100
Figura 8.34: Coeficiente de variação do Norte. ....	100
Figura A.1 : média ruídos agregados com MATLAB.....	109
Figura A.2 desvio padrão dos ruídos agregados com MATLAB .....	109
Figura A.3 : média dos ruídos agregados com MATLAB.....	110
Figura A.4 : desvio padrão dos ruídos agregados com MATLAB .....	111
Figura A.5: média dos ruídos agregados com rotina Fortran .....	112
Figura A.6 : desvio padrão dos ruídos agregados com rotina Fortran.....	112

## **Lista de siglas**

AAS	– amostragem aleatória simples
AS	– amostragem seletiva
ASG	– amostragem seletiva global
CEPEL	– Centro de Pesquisas de Energia Elétrica da Eletrobrás
CMO	– custo marginal de operação
COPER	– custo total de operação
ENA	– energia natural afluente
EARMF	– energia armazenada no final do período
FCF	– função de custo futuro
FCI	– função de custo imediato
ONS	– Operador Nacional do Sistema Elétrico
PAR(p)	– autorregressivo periódico de ordem p (sigla em inglês)
PDDE	– programação dinâmica dual estocástica
PMO	– programa mensal da operação
SIN	– sistema interligado nacional
ZINF	– limite inferior do valor esperado do custo total de operação
ZSUP	– limite superior do valor esperado do custo total de operação

# **1 Introdução**

## **1.1 Considerações iniciais**

Com o avanço da tecnologia e o crescente uso da energia elétrica no cotidiano da sociedade moderna, o problema de planejar a operação do sistema energético ganhou importância, visto que o fornecimento deve ser feito de forma ininterrupta (salvo a ocorrência de defeito) e obedecendo padrões de qualidade, disponibilidade e confiabilidade cada vez mais rígidos.

O planejamento da operação energética busca determinar em cada etapa do horizonte de planejamento, quando cada usina deve entrar em operação e qual o montante de energia a ser gerado, o intercâmbio de energia entre os subsistemas e quais troncos de linhas de transmissão devem ser utilizados etc. Essas decisões devem ser tomadas de forma que, por exemplo, o mínimo custo seja alcançado no fornecimento de energia.

## **1.2 Contextualização do trabalho**

Atualmente o ONS (Operador Nacional do Sistema Elétrico) é o responsável pelo planejamento da operação energética do sistema interligado nacional (SIN) e para isso utiliza alguns modelos matemáticos, de acordo com o horizonte de estudo, elaborados pelo Centro de Pesquisas de Energia Elétrica (CEPEL). Os principais deles são:

- NEWAVE (MACEIRA et al, 2008) – Para o Planejamento da Operação a Médio Prazo é utilizado nos PMO's (Programas Mensais da Operação) e PDE's (Planos Decenais de Expansão).
- DECOMP (CEPEL, 2010) – Para o Planejamento da Operação a Curto Prazo, é aplicado no planejamento da operação de sistemas hidrotérmicos a curto prazo e está adaptado ao ambiente de elaboração dos programas mensais de operação do sistema brasileiro
- DESSEM (CEPEL, 2010) – Encontra-se em fase de validação pelo ONS e agentes do setor, e é utilizado na programação diária da operação.

Em todos os modelos mencionados acima o objetivo principal é determinar, em cada etapa do horizonte de planejamento, metas de geração para as usinas hidroelétricas e termoelétricas, atendendo à demanda de energia elétrica, às restrições operativas das usinas e às restrições elétricas do sistema.

O Programa NEWAVE resolve o problema de planejamento da operação interligada de sistemas hidrotérmicos empregando a técnica de programação dinâmica dual estocástica (MACEIRA et al, 2002). Os vários reservatórios do parque gerador são representados de forma simplificada em reservatórios equivalentes de energia, reduzindo assim a dimensionalidade do problema. A representação das incertezas nesse modelo é feita explicitamente ao se considerar séries hidrológicas sintéticas, geradas por um modelo estocástico, que serão visitadas durante o processo de cálculo da política de operação e da simulação final.

Durante o processo de otimização, pode-se distinguir duas etapas principais: a etapa progressiva (forward) e a regressiva (backward). A cada iteração, ao percorrer os cenários construídos para essas etapas, tem-se uma estimativa para os limites superior e inferior do valor esperado do custo total de operação.

A árvore completa de cenários possui cardinalidade bastante elevada. Por isso, de forma a viabilizar computacionalmente o processo de otimização, somente uma parte dessa árvore (sub-árvore) deve ser utilizada. A sub-árvore deve ser construída de forma que se consiga uma boa representação do processo estocástico de afluências, por exemplo.

Na versão atual do NEWAVE essa árvore é “extraída” da árvore original utilizando amostragem aleatória simples (AAS). Tramita em fase de validação a Amostragem Seletiva (AS), que utiliza técnicas de agregação na construção da sub-árvore a ser visitada nas etapas forward e backward. Este trabalho propõe a Amostragem Seletiva Global, que através de técnicas de agregação com uma perspectiva diferente das anteriores, busca selecionar cenários representativos para construção da sub-árvore da etapa forward (não equiprováveis) e futuramente da simulação final.

### **1.3 Objetivos**

Esse trabalho propõe uma alternativa para a construção da sub-árvore de cenários que deve ser percorrida durante o cálculo da política de operação (durante etapa forward).

A sub-árvore a ser percorrida deverá preservar as propriedades estatísticas da árvore original e, portanto, do registro histórico de afluências.

Pretende-se, com essa árvore reduzida, obter resultados que apresentem menor variabilidade com relação à escolha da amostra de cenários, melhorando a qualidade da função de custo futuro construída pelo NEWAVE e repassada aos demais modelos da cadeia de modelos da operação energética.

### **1.4 Metodologia**

Inicialmente, será gerada através do modelo autorregressivo periódico, uma pseudo-árvore completa, que consiste em uma árvore com cardinalidade elevada, mas menor do que a árvore completa, pois essa é teórica e formada por todas as possíveis realizações do processo estocástico de afluências. A pseudo-árvore completa será utilizada no método ASG para a extração da sub-árvore que será visitada no processo de otimização do estudo de planejamento energético de médio prazo. Algumas vezes, neste trabalho, é utilizada a denominação de ‘matriz original’ ou ‘matriz de entrada’ para fazer referência à pseudo-árvore completa disposta em uma matriz.

Para a construção da sub-árvore a ser visitada, é utilizado um método de agregação não hierárquico chamado K-Means, buscando representar a matriz de energias original utilizando-se apenas alguns de seus representantes.

Com a aplicação do método ASG para a construção dos cenários da etapa forward, esses não serão mais equiprováveis. Desta forma, o modelo NEWAVE deverá considerar a probabilidade de ocorrência de cada cenário durante o processo iterativo de cálculo da política ótima de operação.

## **1.5 Organização**

Neste capítulo foi apresentada a contextualização deste trabalho dentro do setor elétrico, os objetivos preliminares da aplicação do método proposto na construção da sub-árvore a ser visitada no processo de otimização do NEWAVE, a metodologia utilizada na construção da sub-árvore e a organização do trabalho.

No capítulo 2 são abordadas as características principais do planejamento da operação energética de sistemas hidrotérmicos, as consequências das decisões presentes no futuro, o caso de planejamento da operação do SIN, os tipos de estudo e os modelos utilizados.

No capítulo 3 discute-se o planejamento da operação de médio e o modelo utilizado para este tipo de estudo – o NEWAVE. São também apresentadas as formas de construção da sub-árvore de cenários existentes no modelo atualmente.

O modelo de geração de séries sintéticas de afluências utilizadas no estudo de médio prazo é contemplado no capítulo 4. Faz-se uma breve descrição de como essas séries são geradas com o modelo PAR(p).

O capítulo 5 aborda os métodos de agrupamentos, diferenciando cada um, indicando os tipos de estudos em que são usados e dando maior atenção ao que é adotado no presente trabalho.

O capítulo 6 traz o método Amostragem Seletiva Global, que propõe mudanças na construção da sub-árvore de energias afluentes a ser utilizada no modelo de planejamento de médio prazo, bem como suas diferenças em relação a opção Atual e à Amostragem Seletiva.

O capítulo 7 traz a metodologia adotada para a aplicação da técnica de agregação, como por exemplo, considerações sobre a disposição da matriz de energias afluente original e métricas utilizadas para as comparações. Nessa etapa são analisados os cenários hidrológicos gerados através da ASG, comparando-os com os gerados pela opção Atual e AS.

No penúltimo capítulo, o capítulo 8, analisa-se resultados decorrentes da aplicação dos cenários gerados através da ASG, considerando-se aspectos relativos à operação do sistema .

No último capítulo, o capítulo 9, são relacionadas as sugestões trabalhos futuros e a conclusão.

## **2 Planejamento da Operação Energética**

### **2.1 Considerações iniciais**

A complexidade de planejar a operação de sistemas geradores de energia elétrica evoluiu com o crescimento da capacidade de geração, e com o uso cada vez mais frequente desta forma de energia, para o funcionamento e desenvolvimento das sociedades modernas. Os primeiros sistemas elétricos de potência forneciam energia basicamente para a iluminação, e eram constituídos por apenas uma unidade geradora, termoelétrica ou hidroelétrica. Os sistemas geravam a energia solicitada e, na eventualidade de não conseguirem atender a demanda, reprimiam o consumo, isto é, reduziam a carga. Pode-se afirmar que inexistia o problema do planejamento da operação. Rapidamente, a energia elétrica conquistou enorme importância para as sociedades industrializadas. As fábricas substituíram as máquinas a vapor por motores elétricos, e um grande número de equipamentos movidos à eletricidade mostrou-se indispensável às comunidades. Em função desta realidade, os sistemas de geração de energia elétrica foram obrigados a crescer muito, baixar os custos de produção, e aumentar a confiabilidade, para garantir a continuidade do suprimento. O planejamento da operação dos sistemas de geração de energia elétrica assumiu grande importância e foi se tornando progressivamente mais complexo.

O planejamento da operação de um sistema de geração de energia elétrica busca determinar quais unidades geradoras e os respectivos níveis de geração que devem ser utilizados no atendimento à demanda (mercado de energia elétrica), em cada intervalo de tempo, de forma que o custo operativo associado ao uso dessas unidades seja o mínimo possível, devendo levar em consideração as diversas restrições relacionadas ao desempenho das unidades geradoras, bem como algumas restrições do sistema tais como: requisito de reserva, comportamento do sistema de transmissão, outros usos da água etc.

### **2.2 Sistemas Termoelétricos**

Esses sistemas são constituídos de usinas que utilizam combustíveis fósseis, usinas a gás, usinas nucleares, etc. Essas usinas utilizam o calor gerado pela queima dos combustíveis para gerar vapor, que por sua vez, movimentará as turbinas que se encontram acopladas ao eixo do gerador. As usinas nucleares beneficiam-se do calor gerado durante reações químicas no núcleo também para gerar vapor e movimentar uma turbina.

Algumas características físicas e operativas dessas usinas são tomadas de forma a representá-las num estudo de planejamento, tais como: potência máxima, tipo de combustível, taxa de tomada de carga, geração mínima, dentre outras.

Nas usinas termoelétricas, o custo de operação individual depende basicamente do custo associado ao combustível a ser utilizado. Planejar a operação de sistemas com essa característica consiste em simplesmente avaliar de acordo com fatores relevantes (físicos e operativos), quais usinas, em ordem crescente de custo de operação, deverão ser postas em operação de forma a atender a carga.

Na prática, esse problema de otimização é um pouco mais complexo, já que devem ser levados em conta a taxa de tomada de carga, tempos mínimos e máximos em operação, além de tempos de resfriamento (SILVA, 2001).

A princípio, sistemas exclusivamente termoelétricos são desacoplados no tempo, isto é, a decisão de gerar energia agora não influenciará a decisão de gerar energia no futuro, pois isso dependerá apenas de fatores de mercado, como o preço e disponibilidade de combustível. Também são desacoplados no espaço, pois a geração de uma usina não interfe na geração de outra.

### 2.3 Sistemas Hidrotérmicos

Sistemas hidrotérmicos são compostos por usinas hidroelétricas e termoelétricas (convencionais ou nucleares) que se encontram conectadas às cargas por um sistema de transmissão, que pode ou não ser interligado. Vale ressaltar que um sistema interligado possui vantagens como maior confiabilidade e qualidade no fornecimento de energia, além de possibilitar o gerenciamento dos aproveitamentos entre os subsistemas de forma a minimizar os custos.

A energia disponível (volume de água) armazenada nos reservatórios das usinas hidroelétricas é limitada e depende do uso em períodos anteriores e correspondentes afluências, fazendo com que o processo de gerenciamento desses aproveitamentos seja uma tarefa complexa.

Devido à impossibilidade de se ter um conhecimento perfeito das afluências futuras aos aproveitamentos e, em certa medida, da curva de carga do sistema, sua operação torna-se um problema probabilístico. Essa incerteza proveniente da natureza probabilística afeta tanto a capacidade de produção média nos períodos (energia) quanto à máxima (ponta),

visto que esta última depende das alturas de queda nos reservatórios, que por sua vez dependem do armazenamento (afluências e regras operativas). Uma complicação adicional vem da necessidade de atendimento a restrições hidráulicas decorrentes do uso múltiplo da água (navegação, irrigação, saneamento, etc), além das regras de segurança para controles de cheias nas bacias. (FORTUNATO, 1990).

As principais características do planejamento da operação em sistemas hidrotérmicos são:

- **Acoplamento temporal:** as decisões tomadas no presente têm consequências no futuro. A solução ótima é obtida minimizando-se o benefício presente do uso da água mais o benefício futuro de seu armazenamento.
- **Estocasticidade:** as decisões tomadas são baseadas em previsões de afluências, que podem ou não ocorrer, por isso há uma incerteza associada. Essas afluências são, atualmente, geradas por modelos estocásticos calibrados pelo registro histórico das afluências.
- **Acoplamento espacial:** a decisão de deplecionamento de uma usina a montante afeta a afluência total de uma usina à jusante, pois estas podem estar dispostas em cascata.
- **Custos indiretos associados à geração de uma hidroelétrica:** o custo da geração hidroelétrica pode ser medido em termos da economia resultante nos custos das térmicas ou déficits evitados.
- **Competição entre os objetivos do problema:** utilizando-se o máximo possível de geração hidroelétrica a cada etapa, garante-se uma política mais econômica, pois os custos relativos a combustíveis são minimizados (geração termoelétrica). Mas essa política resulta, como mencionado, em maiores riscos de déficits futuros e incertezas no atendimento à demanda (risco de déficit). Portanto, a máxima confiabilidade de atendimento é conseguida conservando-se o nível dos reservatórios o mais elevado possível; o que tem como consequência a elevação do custo de operação, pois para isso deve-se elevar o uso de usinas termoelétricas.

É apresentado na Figura 2.1 o grau de acoplamento entre as decisões e consequências e futuras de gerar utilizando mais usinas termoelétricas ou utilizando mais usinas hidroelétricas.

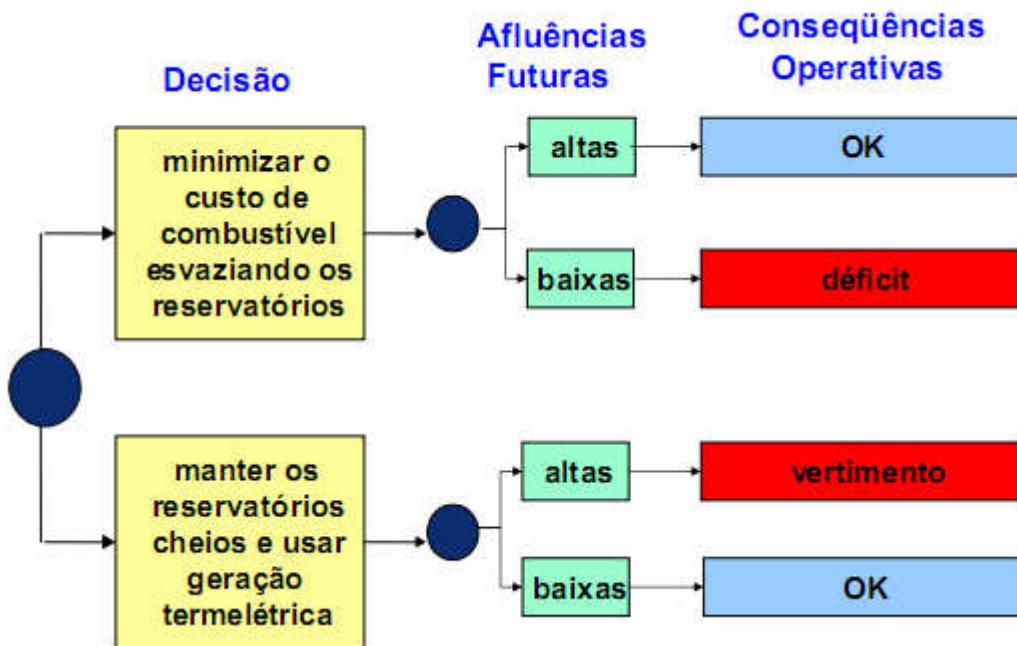


Figura 2.1 : Conseqüências das decisões presentes

Fonte : (PENNA,2009)

Pode-se observar pela Figura 2.1 a interdependência entre uma decisão presente de despacho e sua interferência no futuro. Suponha que seja decidido gerar utilizando energia hidroelétrica (maior parte) visando atender a demanda presente e que ocorram baixas afluências (períodos secos) no futuro; dessa forma usinas termoelétricas de alto custo deverão ser usadas para atender a demanda, do contrário ocorrerá déficit. Mas se for decidido gerar utilizando energia termoelétrica para atender a demanda presente e se no futuro ocorrerem períodos favoráveis (altas afluências) as usinas irão verter esse excedente de água, desperdiçando energia.

Portanto, pelo exposto acima, é crucial no momento de decidir a operação, levar em conta uma ponderação entre utilizar recursos hídricos (de baixo custo) imediatamente e armazená-los para uso futuro. O sucesso da tomada de decisão operativa está justamente na definição do montante de geração versus armazenamento. A representação do benefício do uso imediato da água é feita através de uma função chamada de Função de Custo Imediato (FCI), e a representação do benefício do seu armazenamento para uso futuro é feita a partir de uma Função de Custo Futuro (FCF). A Figura 2.2 ilustra o gráfico representativo dessas duas funções.

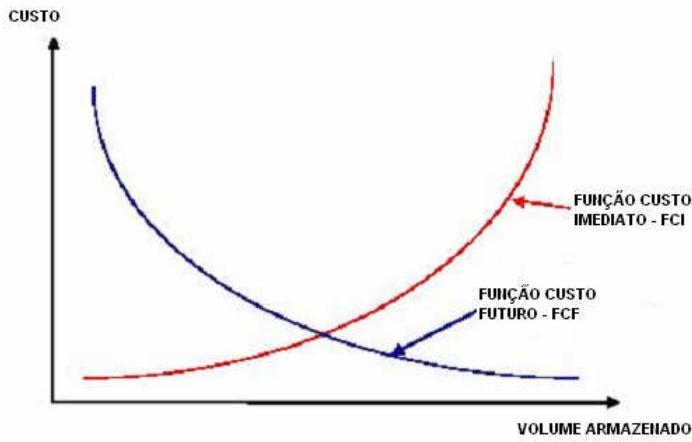


Figura 2.2 : FCI e FCF versus armazenamento

O eixo das abscissas (x) representa o volume final de água armazenado e o eixo das ordenadas (y) representa os valores das funções em unidades monetárias.

Repare que as características das funções traduzem o que foi discutido anteriormente, ou seja, a FCI cresce à medida que o volume armazenado aumenta, pois se o armazenamento é maior, implica em maior uso de usinas termoelétricas de custo elevado no estágio atual, para o atendimento à demanda de energia. Em contrapartida, a FCF diminui à medida que o armazenamento cresce porque a decisão de estocar água no presente está associada a um menor uso de geração de energia termoelétrica no futuro.

A optimização na utilização da água armazenada corresponde ao ponto que minimiza a soma dos custos associados às duas funções ( $FCI + FCF$ ). Na Figura 2.3, observa-se que o ponto de mínimo corresponde ao ponto onde as derivadas das funções de custo imediato e custo futuro em relação ao volume armazenado igualam seus módulos. As derivadas mencionadas são conhecidas como os valores da água.

$$\frac{\partial(FCI + FCF)}{\partial V} = \frac{\partial FCI}{\partial V} + \frac{\partial FCF}{\partial V} = 0 \Rightarrow \frac{\partial FCI}{\partial V} = -\frac{\partial FCF}{\partial V} \quad (2.1)$$

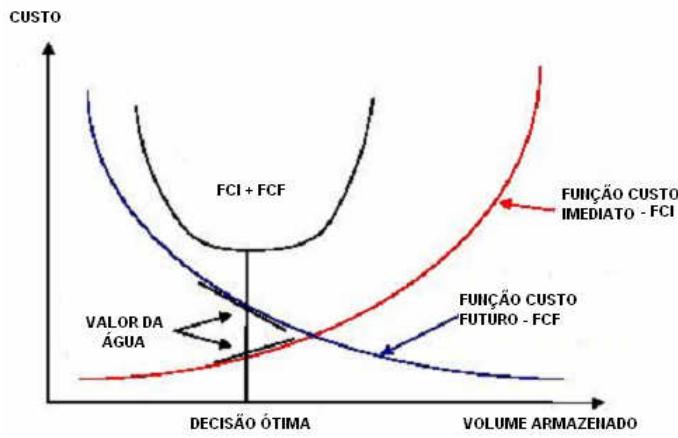


Figura 2.3 : Uso otimizado da água

O valor da energia hidroelétrica (valor da água) é o valor da geração termoelétrica mais o déficit de energia que poderia ser substituído hoje ou no futuro. Este valor é resultado do processo de determinação da política ótima de operação (MACEIRA, et al 2008). Pode-se representar, dessa forma, uma usina hidroelétrica como uma usina térmica cujo custo de operação é o valor da água. Esse valor deve ser calculado levando-se em consideração a operação interligada do sistema, pois se cada usina for tomada individualmente e de forma independente no cálculo, não resultará no valor de operação mais econômica possível.

## 2.4 Planejamento da Operação Energética no Brasil

O sistema de produção de energia elétrica brasileiro é um sistema hidrotérmico de grande porte, com forte predominância de usinas hidroelétricas, conforme pode ser observado na Tabela 2.1, e com múltiplos proprietários, formando o Sistema Interligado Nacional (SIN) com características únicas no mundo.

Fonte	Nº Usinas	kW	% Cap. Disp.
Hidrelétrica	838	78.793.231	73,63%
Gás	125	12.055.295	11,27%
Biomassa	356	6.227.660	5,82%
Petróleo	829	5.735.637	5,36%
Nuclear	2	2.007.000	1,88%
Carvão Mineral	9	1.530.304	1,43%
Eólica	37	659.284	0,62%
Solar	1	20	< 0,01
<b>Capacidade Disponível</b>	<b>2.197</b>	<b>107.008.431</b>	<b>100</b>

Tabela 2.1: Capacidade Instalada de Geração no Brasil

Fonte: ANEEL -2010

O SIN é composto por empresas das Regiões Sul, Sudeste, Centro-Oeste, Nordeste e parte da região Norte que possuem usinas hidroelétricas distribuídas em 12 diferentes bacias hidrográficas. Como as usinas hidroelétricas são geralmente construídas em locais distantes dos centros de carga tornou-se necessária a construção de um extenso sistema de transmissão. Adicionalmente, as grandes interligações oriundas do complexo sistema de transmissão possibilitam a troca de energia entre regiões, permitindo obter vantagens da diversidade do comportamento hidrológico entre as diferentes bacias hidrográficas distribuídas ao longo do extenso território brasileiro.

A capacidade de regularização dos reservatórios que compõem o SIN é plurianual e o registro histórico de energias afluentes apresenta períodos desfavoráveis de longa duração.

Portanto deve-se que operar o sistema de forma coordenada e otimizada a fim de minimizar os custos operativos globais. Dada a complexidade de operar o sistema da forma desejada, divide-se o planejamento em problemas com horizontes temporais diferentes, necessitando de criação de modelos de suporte à decisão que resolvam o problema em cada horizonte. Esses modelos são capazes de considerar ganhos energéticos com a operação integrada dos diversos subsistemas e estão em conformidade com os horizontes de tomada de decisão e níveis de incerteza associada. Portanto, torna-se necessário o uso de modelos com diferentes níveis de detalhamento para cada caso.

Como foi dito, a modelagem é dividida em diversas etapas (sub-problemas), onde em cada uma delas é adotado um horizonte de planejamento diferente, e também, uma

representação da estocasticidade das afluências e das não linearidades do problema com diferentes graus de detalhamento. A Figura 2.4 ilustra essa divisão.

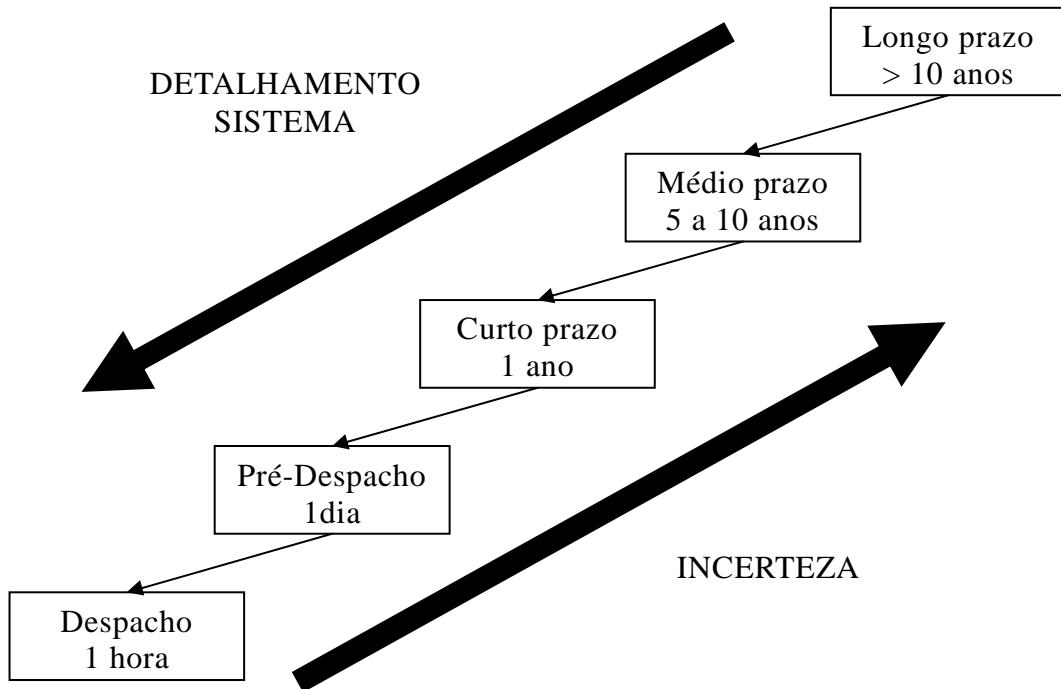


Figura 2.4 : Incertezas e detalhamento do sistema de acordo com o horizonte de estudo

#### 2.4.1 Cadeia dos Modelos

Para fazer face a esta desagregação temporal, foi necessário o desenvolvimento pelo CEPEL, de uma cadeia de modelos matemáticos e computacionais para o planejamento da expansão da geração e para o planejamento e programação da operação energética (MACEIRA et al, 2002).

Esses modelos auxiliam na coordenação da operação das usinas hidroelétricas do sistema, tentando minimizar o uso de térmicas sem prejudicar o atendimento à demanda, tornando assim a operação mais eficiente, pois reduz os custos operativos, riscos de déficits e possibilidades de vertimentos desnecessários (desperdício de energia).

Os programas utilizados para o planejamento da operação de médio prazo, curto prazo e para a programação da operação são o NEWAVE, o DECOMP e o DESSEM

respectivamente. Na figura 2.5 é mostrada a interligação desses modelos com os demais utilizados no planejamento da operação energética.

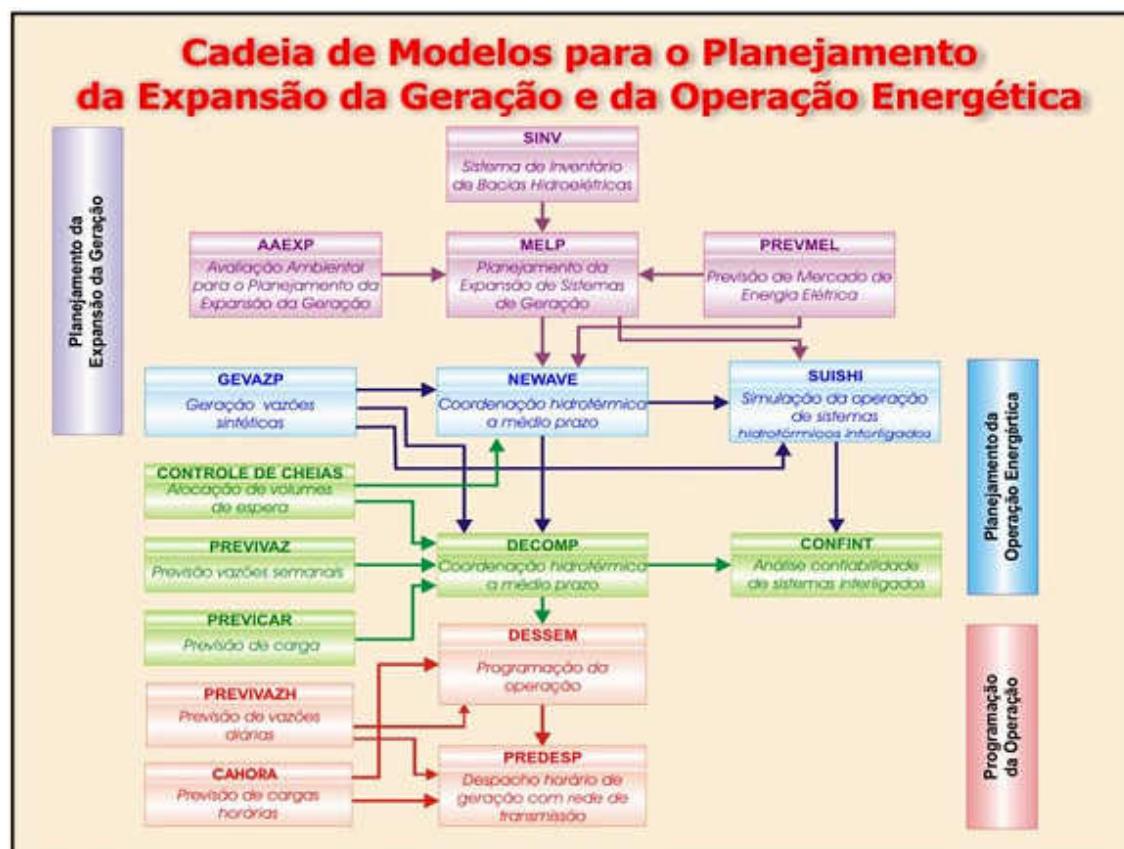


Figura 2.5 Cadeia de modelos utilizados no planejamento da operação energética

Fonte: (CEPEL, 2009)

#### 2.4.2 Planejamento de Médio Prazo

O programa utilizado no planejamento de médio prazo é o NEWAVE. Esse modelo calcula para cada etapa do período de planejamento, que pode ser de 5 anos (PMO) ou 10 anos (PDE), a distribuição ótima dos recursos hídricos e térmicos do sistema, considerando a minimização dos custos operativos e custos de déficit durante o período de planejamento. A representação do parque hidroelétrico é feita por subsistemas equivalentes e o cálculo da política de operação é baseado na PDDE.

O modelo que é utilizado em conjunto com o NEWAVE é o SUISHI-O, que simula a operação energética de sistemas interligados a usinas individualizadas. O acoplamento

desse programa com o NEWAVE se dá através da função de custo futuro para cada estágio onde podem ser consideradas outras restrições de caráter local.

O modelo GEVAZP (MACEIRA e MERCIO, 1997) gera os cenários hidrológicos que serão utilizados tanto no cálculo da política como na simulação final. Esse programa utiliza o modelo autorregressivo periódico, o PAR(p) (CEPEL, 2002), em que a afluência em um período  $t$  é função das afluências passadas ( $t-1, t-2, \dots$ ) e a estrutura da dependência temporal é sazonal. As séries geradas pelo modelo possuem características probabilísticas iguais às da série histórica, e reproduzem os períodos desfavoráveis ocorridos no passado.

#### **2.4.3 Planejamento de Curto Prazo**

O DECOMP é o modelo utilizado para o planejamento de curto prazo. O acoplamento com a etapa anterior se dá pela função de custo futuro (que representa a política ótima do horizonte de médio prazo) gerada num estágio que coincide com o final do horizonte do modelo de curto prazo. Desta forma é gerada uma função que representa o valor econômico da água armazenada nos reservatórios em função dos níveis de armazenamento destes.

O horizonte, neste caso, é de alguns meses e a incerteza relacionada às afluências dos reservatórios é representada através de uma árvore de vazões. Nesta etapa, as usinas são representadas de forma individualizada.

O objetivo desse planejamento é o mesmo que o do horizonte de médio prazo, ou seja, minimizar o valor esperado do custo de operação ao longo do período de planejamento, só que o período agora é de até um ano, com discretização semanal para o primeiro mês e mensal para os demais.

#### **2.4.4 Programação da Operação**

O programa utilizado para a programação da operação é o DESSEM. Nesta etapa, o horizonte é de apenas alguns dias, discretizados em etapas horárias ou de meia em meia hora. Não é representada a incerteza das vazões. Em contrapartida, o parque hidrotérmico é representado de forma detalhada, levando-se em conta as restrições relativas às máquinas e turbinas, tais como: tomada e alívio de carga, faixas operativas das turbinas, entre outras. A

rede de transmissão é representada por um fluxo de potência DC. A função de custo futuro gerada pelo modelo de curto prazo no estágio que coincide com último estágio do modelo de programação diária é utilizada para definir-se a meta de geração de cada unidade geradora.

## 2.5 Resumo

Neste capítulo foi abordada a complexidade do planejamento da operação energética, bem como a sua importância para a redução dos custos e otimização dos aproveitamentos hídricos ao longo dos períodos do horizonte de planejamento.

Foram descritas as peculiaridades do planejamento em sistemas termoelétricos e hidrotérmicos e como deve-se ponderar decisões presentes em função de cada tipo de sistema, de forma a evitar déficits de energias e custos elevados no futuro.

Por fim discutiu-se sobre como é feito o planejamento da operação energética no caso do SIN, abordando modelos utilizados em função de cada horizonte de planejamento e suas principais características.

### **3 Planejamento da Operação de Médio Prazo e o Newave**

#### **3.1 Newave**

O Planejamento da Operação de Médio Prazo do SIN é um problema com horizonte de alguns anos a frente, com discretização mensal, no qual a estocasticidade está representada através de séries sintéticas de energias geradas por um modelo autoregressivo, que por sua vez é calibrado pelo registro histórico de afluências.

O modelo utilizado para esse tipo de estudo é o NEWAVE, desenvolvido pelo CEPEL. No NEWAVE o problema supracitado é representado por um problema de programação estocástica linear que considera múltiplos estágios. A solução ótima é obtida por Programação Dinâmica Dual Estocástica, a PDDE, que foi proposta por Pereira e Pinto (1991) e é baseada na decomposição de Benders (BENDERS, 1962). A formulação da PDDE atualmente empregada no NEWAVE considera a correlação temporal das afluências aos reservatórios (MACEIRA, 1993). As variáveis de estado consideradas nesse problema são o armazenamento inicial do período e a tendência hidrológica, ou seja, as afluências dos últimos períodos.

O impacto de períodos de secas (afluências baixas) de longa duração bem como o efeito da capacidade de regularização plurianual dos reservatórios que compõem a bacia hidrográfica brasileira podem ser contemplados quando se trabalha com um modelo de planejamento da operação de médio prazo cuja discretização é mensal.

Os sistemas hidrotérmicos, como o caso brasileiro, são representados pelo nível de armazenamento dos seus reservatórios e pela tendência hidrológica do sistema, ou seja, o conjunto de afluências registradas dos períodos anteriores. Essas variáveis constituem o estado do sistema em cada período que compõe o horizonte de planejamento.

As muitas combinações entre estado inicial de armazenamento e tendência hidrológica têm que ser consideradas no cálculo da política de operação. O grande inconveniente é o crescimento exponencial do número de combinações, o que pode ser resolvido por uma representação adequada do sistema. Esse crescimento exponencial do número de combinações entre armazenamento inicial e tendência hidrológica é reduzido representando-se de forma agregada os reservatórios que compõem o sistema. Essa forma

de representar o sistema utilizando um único reservatório é conhecida como representação por reservatório equivalente.

Basicamente, o NEWAVE é composto pelos seguintes módulos:

**I. Módulo de cálculo do sistema equivalente** – Esta etapa consiste no cálculo dos subsistemas equivalentes de energia a partir de uma configuração definida nos dados de entrada. Dessa forma, pode-se evitar a representação individual das usinas e seus reservatórios e adotar a representação agregada mencionada anteriormente. Essa representação objetiva reduzir a dimensionalidade do problema. Cada subsistema é definido por certas características como: energias armazenáveis máximas, séries históricas de energias controláveis e energias fio d'água, parábolas de energia de vazão mínima, energia evaporada, capacidade de turbinamento, correção da energia controlável em função do armazenamento, perdas por limite de turbinamento nas usinas fio d'água, dentre outras.

**II. Módulo de energias afluentes** – Este módulo consiste em estimar os parâmetros do modelo estocástico, PAR(p), e gera séries sintéticas de energias naturais afluentes que são utilizadas no módulo de cálculo da política de operação hidrotérmica (etapas forward e backward) e para geração de séries sintéticas de energias afluentes, geralmente 2000 séries, que serão percorridas de forma a analisar o desempenho com a política de operação anteriormente definida.

**III. Módulo de cálculo da política de operação hidrotérmica** - Consiste no principal módulo do modelo, onde é calculada a política de operação , de forma a minimizar os custos em cada etapa ao longo do horizonte de planejamento. Esse cálculo é feito levando em conta as incertezas nas afluências futuras e os patamares de carga própria, de déficit e as indisponibilidades dos equipamentos. A política ótima é conseguida a partir da análise dos limites do custo de operação calculados com a função de custo futuro dentro de um determinado intervalo de confiança, que define a convergência do processo.

**IV. Módulo de simulação da operação** - Simula a operação do sistema ao longo do período de planejamento, para distintos cenários de seqüências hidrológicas. Calcula índices de desempenho, tais como o custo esperado de operação, o risco.

### 3.2 Representação da estocasticidade hidrológica no modelo NEWAVE

Os cenários de energias (ENA) que serão percorridos durante o cálculo da política de operação e simulação final são gerados pelo modelo estocástico autorregressivo periódico, o PAR(p), o que denota no NEWAVE uma representação essencialmente estocástica, visto que, esses valores de afluências não são conhecidos antecipadamente.

Um exemplo de uma árvore completa pode ser visto abaixo na Figura 3.1. Nela cada ponto corresponde a um estágio ou período e o conjunto de pontos seqüenciais interligados formam um cenário ou série de energia natural afluente.

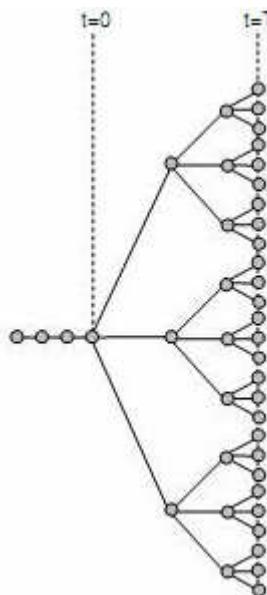


Figura 3.1 : Exemplo de árvore de cenários completa, considerando 3 estágios e 3 aberturas

Para resolver o problema completo do planejamento da operação deve-se resolver um problema para cada nó. Observa-se na árvore da Figura 3.1 que à medida que o número de aberturas e períodos aumenta (caso do PMO, por exemplo), a árvore cresce exponencialmente. Dessa forma, ficaria inviável computacionalmente percorrer todos os cenários da árvore, devendo-se buscar maneiras de reduzir a cardinalidade desta.

Na versão atual do NEWAVE a sub-árvore é construída utilizando-se amostragem aleatória simples (AAS).

### 3.3 Construção da árvore de cenários

A incerteza nas séries de energias naturais afluentes, é representada pelas séries sintéticas utilizadas no modelo de médio prazo. Essas séries formam uma sub-árvore que será visitada durante o processo de otimização.

Atualmente no NEWAVE uma única amostra é utilizada para a construção das séries das etapas forward e backward, a chamada opção Atual. Nesta alternativa, faz-se um sorteio aleatório em uma amostra maior, obtendo-se representantes que serão utilizados na construção das séries a serem percorridas na etapa backward. De posse dessa amostra, um sorteio condicionado é realizado de forma a obter os representantes que serão utilizados para sintetização das séries da etapa forward. Portanto, as séries do passo forward provém dos representantes sorteados no passo backward. O sorteio condicionado é realizado tantas vezes quanto for o tamanho da amostra de representantes necessária para a construção das séries forward. A Figura 3.2 ilustra o processo.

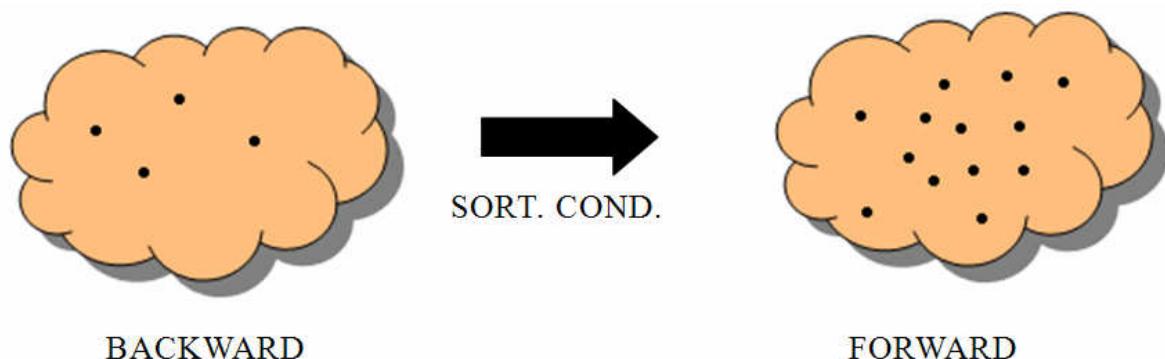


Figura 3.2 : opção Atual

### 3.4 Resumo

Nesse capítulo foi descrita a representação da incerteza nas afluências utilizadas no modelo de médio prazo e a forma de construção da sub-árvore a ser visitada chamada de opção Atual.

Também descreveu-se sucintamente o modelo NEWAVE com suas principais características e os módulos constituintes deste.

## 4 Modelo de Geração de Séries Sintéticas de Energias

### 4.1 Considerações Iniciais

No planejamento da operação energética de médio prazo necessita-se conhecer o estado do sistema e as afluências futuras a fim de calcular, utilizando modelos apropriados, a política ótima de operação para cada etapa do horizonte considerado.

Como as afluências futuras não podem ser previstas com certeza, estas são estimadas através do uso de um modelo estocástico calibrado pela série histórica.

Nos estudos energéticos, os critérios de suprimento são baseados em índices de risco, que devem ser estimados utilizando muitos cenários de energias naturais afluentes.

De posse de um grande número de cenários, pode-se utilizá-los para estimar a relação entre demanda energética e risco de não atendimento em um ano qualquer (energia garantida). É mostrado na Figura 4.1 que um sistema hidroelétrico pode suprir a demanda  $e$ , que está associada ao risco  $p$ , previamente selecionado.

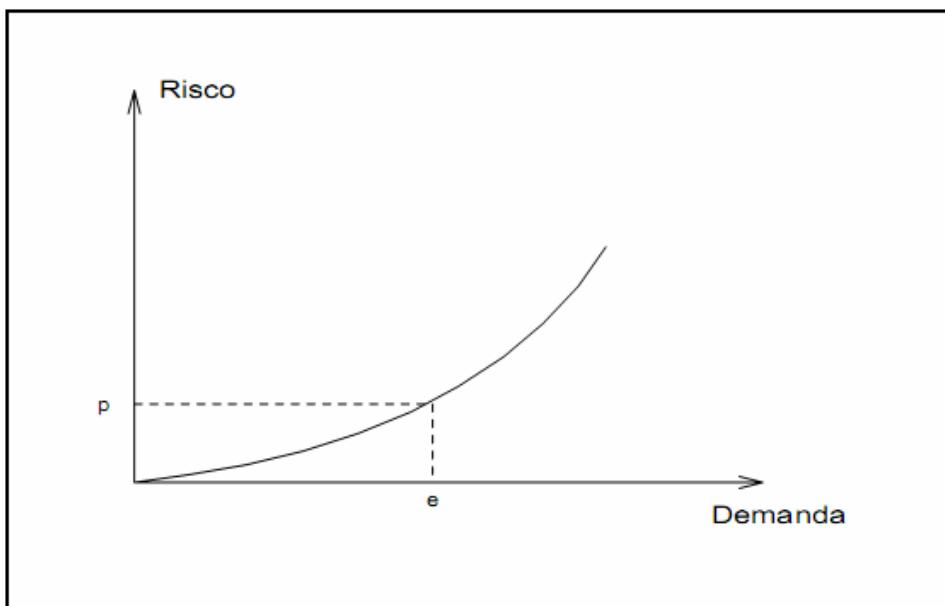


Figura 4.1 : Relação Demanda x Risco em um ano qualquer (CEPEL, 2002)

Existem registros de valores de afluências desde 1931 até 2008, chamados de série histórica de afluências. Elas podem ser divididas em seqüências e referenciadas de acordo com o primeiro ano da seqüência e de tamanho dependente do horizonte de estudo. Porém

esses cenários não são suficientes para que seja realizado o estudo de planejamento da operação energética de médio prazo de forma a obter índices de risco com incertezas aceitáveis.

Portanto, são extraídas do registro histórico de afluências suas principais características estatísticas para calibrar um modelo que se possa utilizar para gerar quantas séries quanto forem necessárias, preservando as estatísticas da população original (a série histórica). As séries assim produzidas são chamadas de séries sintéticas. Neste capítulo será descrito sucintamente o PAR( $p$ ), modelo autorregressivo periódico que calcula a energia afluente de um período  $t$  em função das energias dos  $p$  períodos anteriores.

## 4.2 Introdução a Processos Estocásticos e Séries Temporais

Para levantar-se ou registrar-se uma série histórica de afluências basta medir a vazão afluente mensal a uma dada região por um período de tempo. Por exemplo: ao medir-se a vazão afluente para um período de 5 anos pode-se obter a curva 1 da Figura 4.2 abaixo. Realizando-se as mesmas medições para outro segmento de 5 anos, pode-se obter a curva 2 da mesma Figura.

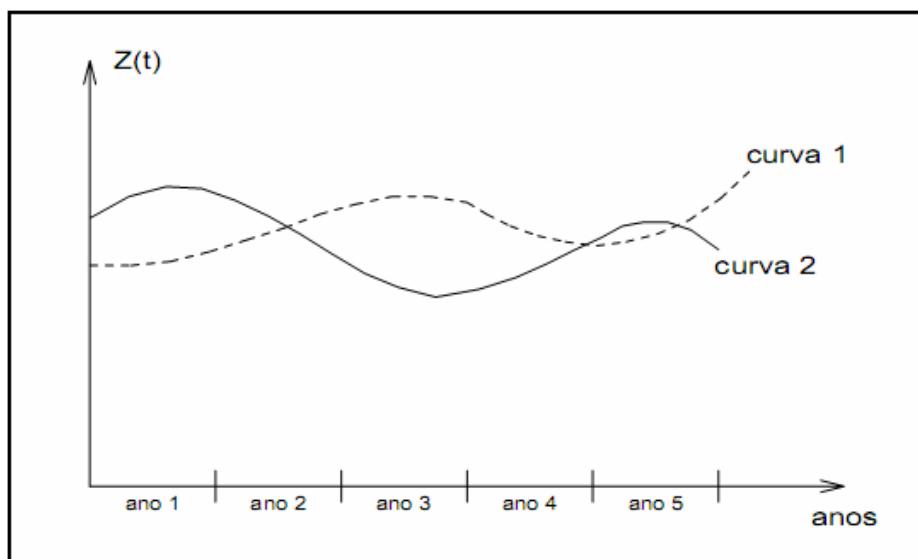


Figura 4.2 : Vazão Afluente Anual x Anos (CEPEL, 2002)

As curvas 1 e 2 são chamadas trajetórias ou realizações do processo físico que está sendo observado. Este pode ser modelado usando um processo estocástico, que consiste no

conjunto de todas as possíveis trajetórias com respectivas probabilidades que podem ser observadas. Essas trajetórias são comumente chamadas de séries temporais.

Na realidade, só está disponível uma realização do processo estocástico, que consiste na série histórica (1931-2008). Portanto, para cada instante do tempo (mês, ano, etc) o processo estocástico é uma variável aleatória. O valor da afluência observado em um instante qualquer (o valor da série histórica naquele instante), nada mais é do que uma possível realização, um valor amostrado da distribuição de probabilidade associado a variável aleatória do processo estocástico naquele instante de tempo.

A descrição de um processo estocástico se faz pelo conjunto de todas as trajetórias ou séries temporais que o constitui com suas respectivas probabilidades associadas, ou pela distribuição de probabilidade conjunta de todas as variáveis aleatórias envolvidas no processo.

Como não dispõe-se dos atributos para descrever completamente o processo estocástico, a modelagem de séries sintéticas tem por objetivo ajustar um modelo que acredita-se ter gerado a série histórica.

Existem algumas considerações que devem ser feitas a respeito das afluências anuais e mensais. Uma delas é a suposição de estacionaridade. Um processo estocástico pode ser considerado estacionário se suas estatísticas não variarem com o tempo em um determinado local, ou seja, a média, o desvio padrão, etc., não se alterarem ao longo do tempo, ou ainda, de forma genérica, significa que a distribuição de probabilidade em um instante  $t$  qualquer é a mesma para qualquer outro instante  $t$ . A intervenção direta do homem ou mudanças bruscas nos fenômenos da natureza podem causar a não estacionaridade de um processo estocástico. Em respeito ao que foi exposto, o processo de afluências mensais é considerado não estacionário, ao passo que o processo de afluências anuais é considerado estacionário.

Dada uma série temporal anual observada  $[z_1, z_2, \dots, z_N]$  pode-se estimar os seguintes índices estatísticos:

A média amostral:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N z_i \quad (4.1)$$

O desvio padrão amostral, que mede o grau de dispersão da amostra em torno da média:

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{i=1}^N (z_i - \hat{\mu})^2} \quad (4.2)$$

E a covariância de ordem um, que mede o grau de dependência linear entre duas variáveis aleatórias espaçadas de um intervalo de tempo. Neste caso, a covariância de ordem um indica a dependência linear entre a afluência de um ano com a afluência do ano imediatamente anterior:

$$\hat{\gamma}(1) = \frac{1}{N} \sum_{i=1}^N (z_i - \hat{\mu})(z_{i-1} - \hat{\mu}) \quad (4.3)$$

A covariância possui dimensão da variável Z ao quadrado. Para se obter um índice de análise mais direto divide-se a covariância de ordem um pelos desvios-padrão das variáveis envolvidas, resultando na correlação de ordem um, também conhecida por dependência serial anual.

$$\hat{\rho}(1) = \frac{\hat{\gamma}(1)}{\hat{\sigma}^2} \quad (4.4)$$

Se este valor for igual a 1 significa que a vazão de um ano pode ser perfeitamente descrita pela vazão do ano anterior. Caso este valor seja igual a -1 significa que as variáveis são perfeitamente correlacionadas, mas o crescimento de uma implica no decrescimento da outra. Se este valor é zero significa que as variáveis não têm dependência linear.

Pode-se medir também a dependência espacial entre as afluências a duas usinas hidroelétricas, se a série histórica de uma usina é linearmente dependente da série histórica de outra usina, através da correlação “cruzada” entre elas. Por exemplo, a correlação espacial anual entre as vazões afluentes às usinas hidroelétricas *a* e *b* pode ser obtida por:

$$\hat{\rho}_{ab} = \frac{\frac{1}{N} \sum_{i=1}^N (z_i^a - \hat{\mu}^a)(z_i^b - \hat{\mu}^b)}{\hat{\sigma}(a)\hat{\sigma}(b)} \quad (4.5)$$

Em geral, processos naturais em escala mensal apresentam um comportamento periódico descrito por ciclos sazonais. Cada período apresenta um conjunto de características estatísticas próprias descritas pela média, desvio-padrão e estrutura de correlações sazonais.

A média amostral de cada mês é dada por:

$$\hat{\mu}_m = \frac{1}{N} \sum_{i=1}^N z_{12*(i-1)+m} \quad m=1, \dots, 12 \quad (4.6)$$

De forma análoga, o desvio padrão amostral de cada mês é dado por:

$$\hat{\sigma}_m = \sqrt{\frac{1}{N} \sum_{i=1}^N (z_{12*(i-1)+m} - \hat{\mu}_m)^2} \quad m=1, \dots, 12 \quad (4.7)$$

Em processos mensais é comum definir um conjunto de valores que descreva a estrutura de correlação linear de um dado mês com os meses anteriores. Esses valores podem ser definidos pela correlação de ordem 1, que descreve a dependência linear da variável aleatória de um mês qualquer  $m$  com a variável aleatória do mês imediatamente anterior ( $m-1$ ); correlação de ordem 2, que descreve a dependência linear da variável aleatória do mês  $m$  com a variável aleatória do mês ( $m-2$ ); ...; correlação de ordem  $k$ , que descreve a dependência linear da variável aleatória do mês  $m$  com a variável aleatória do mês ( $m-k$ ), onde  $k$  é um valor qualquer. A esse conjunto de valores dá-se o nome de função de autocorrelação do mês  $m$ , também conhecida por dependência serial mensal.

Os valores amostrais desses índices estatísticos podem ser obtidos da seguinte forma:

$$\hat{\gamma}_m(k) = \frac{1}{N} \sum_{i=1}^N (z_{12*(i-1)+m} - \hat{\mu}_m)(z_{12*(i-1)+m-k} - \hat{\mu}_m) \quad m=1, \dots, 12 \quad (4.8)$$

$$\hat{\rho}_m(k) = \frac{\hat{\gamma}_m(k)}{\hat{\sigma}_m \hat{\sigma}_{m-k}} \quad m=1, \dots, 12 \quad (4.9)$$

## 4.3 O Modelo Auto-regressivo periódico

### 4.3.1 Descrição do Modelo

Séries hidrológicas de intervalo de tempo menor que o ano, tais como séries mensais, têm como característica o comportamento periódico das suas propriedades probabilísticas, como por exemplo a média, a variância, a assimetria e a estrutura de autocorrelação. A análise deste tipo de séries pode ser feita pelo uso de formulações autorregressivas cujos parâmetros apresentam um comportamento periódico. A esta classe de modelos costuma-se denominar modelos autorregressivos periódicos (MACEIRA, 1989). Estes modelos são referenciados por modelos PAR( $p$ ), onde  $p$  é a ordem do modelo, ou seja, o número de termos autorregressivos do modelo. Em geral,  $p$  é um vetor,  $p = (p_1, p_2, \dots, p_{12})$ , onde cada elemento fornece a ordem de cada período.

O modelo PAR ( $p_1, p_2, \dots, p_{12}$ ) pode ser descrito matematicamente por:

$$\left( \frac{Z_t - \mu_m}{\sigma_m} \right) = \phi_1^m \left( \frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{pm}^k \left( \frac{Z_{t-pm} - \mu_{m-pm}}{\sigma_{m-pm}} \right) + a_t \quad (4.10a)$$

ou

$$\phi^m(B) \left( \frac{Z_t - \mu_m}{\sigma_m} \right) = a_t \quad (4.10b)$$

onde:

$Z_t$  é uma série sazonal de período s

s é o número de períodos ( $s = 12$  para séries mensais)

N é o número de anos

t é o índice do tempo,  $t = 1, 2, \dots, s_N$ , função do ano T ( $T = 1, 2, \dots, N$ ) e do período m

( $m = 1, 2, \dots, s$ )

$\mu_m$  é a média sazonal de período s

$\sigma_m$  é desvio-padrão sazonal de período s

$\phi^m(B)$  é o operador auto-regressivo de ordem  $p_m$

$$\phi^m(B) = (1 - \phi_1^m B - \phi_2^m B^2 - \dots - \phi_{p_m}^m B^{p_m})$$

$B^i$  aplicado a  $Z_t$  resulta em  $Z_{t-i}$  ( $B^i Z_t = Z_{t-i}$ )

$p_m$  é a ordem do operador autorregressivo do período  $m$

$a_t$  é a série de ruídos independentes com média zero e variância  $\sigma_a^{2(m)}$

Seja  $\rho^m(k)$  a correlação entre  $Z_t$  e  $Z_{t-k}$  de tal forma que  $t$  corresponda ao período  $m$ :

$$\rho^m(k) = E\left[\left(\frac{Z_t - \mu_m}{\sigma_m}\right)\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] \quad (4.11)$$

O conjunto de funções de autocorrelação  $\rho^m(k)$  dos períodos  $m = 1, \dots, s$ , descrevem a estrutura de dependência temporal da série. Estas funções são obtidas multiplicando-se ambos os lados da equação (4.10a) por  $\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)$  e tomando o valor esperado obtém-se para cada período:

$$E\left[\left(\frac{Z_t - \mu_m}{\sigma_m}\right)\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] = \phi_1^m E\left[\left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right)\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] + \dots + \phi_{p_m}^m E\left[\left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right)\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] + E\left[a_t \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] \quad (4.12)$$

Por exemplo, para  $k = 1$  a expressão (4.12) resulta em:

$$\rho^m(1) = \phi_1^m + \phi_2^m \rho^{m-1}(1) + \dots + \phi_{p_m}^m \rho^{m-1}(p_m - 1) \quad (4.13)$$

Conhecidos os parâmetros de um modelo PAR( $p$ ) as funções  $\rho^m(k)$  são dadas pela solução de (4.12) e podem ser expressas por uma combinação de decaimentos exponenciais e/ou ondas senoidais, o que faz com que cada  $\rho^m(k)$  tenda a zero à medida que  $k$  cresce. Fixando-se  $m$  e variando  $k$  de 1 a  $p_m$  em (4.12) obtem-se para cada período um conjunto de

equações comumente denominado de equações de Yule-Walker. Para um período  $m$  qualquer:

$$\begin{bmatrix} 1 & \rho^{m-1}(1) & \rho^{m-1}(2) & \dots & \rho^{m-1}(p_m-1) \\ \rho^{m-1}(1) & 1 & \rho^{m-2}(1) & \dots & \rho^{m-2}(p_m-2) \\ \rho^{m-1}(2) & \rho^{m-2}(1) & 1 & \dots & \rho^{m-3}(p_m-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1}(p_m-1) & \rho^{m-2}(p_m-2) & \rho^{m-3}(p_m-3) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_1^m \\ \phi_2^m \\ \phi_3^m \\ \vdots \\ \phi_{p_m}^m \end{bmatrix} = \begin{bmatrix} \rho^m(1) \\ \rho^m(2) \\ \rho^m(3) \\ \vdots \\ \rho^m(p_m) \end{bmatrix} \quad (4.14)$$

Chamando  $\phi_{kj}$  o j-ésimo parâmetro autorregressivo de um processo de ordem  $k$ ,  $\phi_{kk}$  é o último parâmetro desse processo. As equações de Yule-Walker para cada período m podem ser reescritas da seguinte forma:

$$\begin{bmatrix} 1 & \rho^{m-1}(1) & \rho^{m-1}(2) & \dots & \rho^{m-1}(k-1) \\ \rho^{m-1}(1) & 1 & \rho^{m-2}(1) & \dots & \rho^{m-2}(k-2) \\ \rho^{m-1}(2) & \rho^{m-2}(1) & 1 & \dots & \rho^{m-3}(k-3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{m-1}(k-1) & \rho^{m-2}(k-2) & \rho^{m-3}(k-3) & \dots & 1 \end{bmatrix} \begin{bmatrix} \phi_{k1}^m \\ \phi_{k2}^m \\ \phi_{k3}^m \\ \vdots \\ \phi_{kk}^m \end{bmatrix} = \begin{bmatrix} \rho^m(1) \\ \rho^m(2) \\ \rho^m(3) \\ \vdots \\ \rho^m(k) \end{bmatrix} \quad (4.15)$$

O conjunto de valores  $\phi_{kk}^m$ ,  $k=1,2,\dots$ , são chamados de função de autocorrelação parcial do período  $m$ . E  $\phi_{kk}^m$  para  $m=1,2,\dots,s$ , é uma outra forma de representar a estrutura de dependência do processo estocástico ao longo do tempo. Em um processo auto-regressivo de ordem  $p_m$ , a função de autocorrelação parcial  $\phi_{kk}^m$  será igual a zero para k maior que  $p_m$ . Para  $k=0$  a expressão (4.12) torna-se:

$$1 = \phi_1^m \rho^m(1) + \phi_2^m \rho^m(2) + \dots + \phi_{p_m}^m \rho^m(p_m) + E\left[a_t \left(\frac{Z_t - \mu_m}{\sigma_m}\right)\right] \quad (4.16)$$

Multiplicando-se ambos os lados da equação (4.10a) por  $a_t$  e tomindo o valor esperado:

$$E\left[a_t \left(\frac{Z_t - \mu_m}{\sigma_m}\right)\right] = \sigma_a^{2(m)} \quad (4.17)$$

Substituindo a expressão (4.17) em (4.16):

$$\sigma_a^{2(m)} = 1 - \phi_1^m(1)\rho^m(1) - \phi_2^m(2)\rho^m(2) - \dots - \phi_{p_m}^m\rho^m(p_m) \quad (4.18)$$

que vale para qualquer período m.

### 4.3.2 Ajuste do Modelo

A metodologia para ajustar o modelo que pode ser aplicada à família PAR(p) divide-se em três etapas: identificação do modelo, estimação do modelo e verificação do modelo. Essas etapas serão descritas a seguir.

#### 4.3.2.1 Identificação

A identificação do modelo consiste em determinar as ordens mais apropriadas dos operadores autoregressivos de cada período  $p_m$ ,  $m = 1, \dots, s$ . Isto pode ser feito obtendo-se estimativas  $\hat{\phi}_{kk}^m$ ,  $k=1$  até  $N/4$ , e substituindo em (4.15) as autocorrelações pelos respectivos valores amostrais de (4.8) e (4.9). Segundo MACEIRA (1989), se a ordem do operador autorregressivo de um período qualquer m é  $p_m$  então  $\hat{\phi}_{kk}^m$  para  $k > p_m$  tem distribuição aproximadamente Normal com média zero e variância  $N^{-1}$ . Para cada período m procura-se a maior ordem  $i$  tal que todas as estimativas  $\hat{\phi}_{kk}^m$  para  $k > i$  não sejam mais relevantes.

#### 4.3.2.2 Estimação

Após a etapa de identificação é necessário obter estimativas para os diversos parâmetros do modelo. Para modelos auto-regressivos os estimadores de momento são em geral bastante eficientes (HARVEY, 1981)

As médias e os desvios padrões sazonais são estimados pelas equações (4.6) e (4.7) respectivamente.

Os parâmetros  $\phi_1^m, i=1, \dots, p_m$ , são estimados substituindo-se em (4.14)  $\phi_1^{m-j}(k)$ ,  $j=0, \dots, p_m$ ,  $k=1, \dots, p_m$  por suas estimativas dadas pelas equações (4.8) e (4.9).

Observa-se que os parâmetros do modelo para o m-ésimo período podem ser estimados de maneira independente dos parâmetros de qualquer outro período.

As estimativas  $\sigma_a^{2(m)}$  podem ser obtidas de (4.18).

#### **4.3.2.3 Verificação**

Na etapa de verificação são utilizados procedimentos estatísticos que permitem verificar se o modelo estimado é adequado e, portanto, capaz de gerar as séries sintéticas. A grande maioria dos testes se baseia em uma análise detalhada dos resíduos ajustados  $a_t$ . Se o modelo estimado for adequado, a média dos resíduos deve ser igual a zero, a variância deve ser pequena e constante e a função de correlação dos resíduos,  $\rho_a^m(k)$  deve ser irrelevante para  $k > 0$ .

#### **4.3.2.4 Geração de Séries Sintéticas com o modelo PAR(p)**

Como foi dito anteriormente, um conjunto de valores de energias naturais afluentes medidas em um determinado período de tempo é conhecido como série histórica. Já que o modelo autorregressivo periódico gera séries que preservam as características estocásticas da série histórica, pode-se considerar que esta também foi gerada pelo modelo PAR(p), ou seja, a série histórica é só uma das possíveis realizações do processo estocástico que se está observando, cujas leis probabilísticas foram levadas em conta na calibração do modelo. Portanto, o PAR(p) reproduz o comportamento estocástico do processo físico considerado.

Cada série gerada pelo modelo está associada a uma série temporal e é chamada de série sintética.

A geração dessas séries no modelo possui duas parcelas:

- Determinística : energias afluentes de  $p$  períodos passados (de acordo com ordem do modelo) que serão levadas em conta na determinação da energia do período atual.
- Estocástica : ruídos aleatórios que são sorteados e somados à parte determinística do modelo a fim de representar a estocasticidade deste.

Para começar o processo de geração é necessário arbitrar valores iniciais para  $p$  períodos anteriores, de acordo com a ordem obtida no estágio. São utilizados para estimar os valores iniciais, as médias mensais ou os últimos valores amostrais.

O conjunto dos últimos valores observados de uma série hidrológica é denominado tendência hidrológica e a média mensal da série histórica é chamada de MLT mensal.

No NEWAVE é utilizada a estrutura de geração em pente para gerar as séries sintéticas. Essa estrutura é ilustrada na Figura 4.3 abaixo.

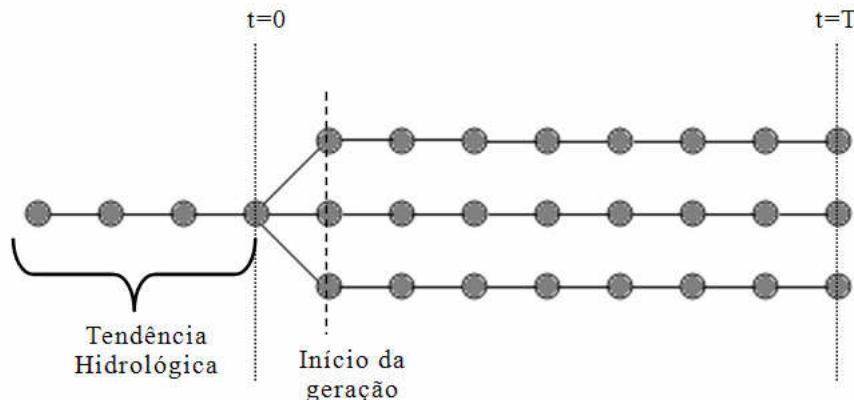


Figura 4.3 : Tendência hidrológica

#### 4.3.2.5 Ruídos Aleatórios

O modelo de geração das séries sintéticas utiliza as afluências de  $p$  períodos passados para determinar o valor da afluência do período atual, de acordo com a ordem  $p$  calculada do modelo para este período. A fim de representar a estocasticidade do processo físico, uma parcela aleatória, chamada de ruído ( $\alpha_t$ ), é somada à parte determinística do

processo, mais precisamente na equação de geração de vazão (4.10). Esse ruído é extraído atualmente através de amostragem aleatória simples.

#### **4.3.2.6 Resumo**

Neste capítulo foram abordados a importância e as principais características do modelo autorregressivo periódico usado no planejamento da operação energética de médio prazo.

Foi apresentado sucintamente a formulação matemática envolvida de forma a ilustrar as partes determinística e estocástica do processo de geração das séries sintéticas.

O próximo capítulo discorrerá sobre Técnicas de Agregação, que serão utilizadas na formulação da Amostragem Seletiva Global, que é a proposta deste trabalho.

## 5 Métodos de Agregação

### 5.1 Considerações Iniciais

Com objetivo de propor uma alternativa para a definição da sub-árvore a ser percorrida durante a etapa forward do cálculo da política, de forma a tornar os resultados obtidos mais robustos às variações de amostras, serão apresentadas nesse capítulo as técnicas de análise de agrupamentos.

Serão definidos, de acordo com as características do problema, aquela que será aplicada para redução da cardinalidade da árvore objetivando representar de forma mais precisa o processo estocástico de afluências a ser usado na etapa forward.

As técnicas de agrupamentos, quando aplicadas a um grande número de cenários hidrológicos gerados, permitem a escolha de um conjunto representativo de cenários. Esse conjunto escolhido deve preservar as características do processo estocástico que o originou, no caso o processo estocástico de afluências.

### 5.2 Técnicas de Agrupamento

O problema de construção da árvore de afluências utilizadas nos modelos de planejamento foi tratada por (JARDIM, 2002) no horizonte de curto prazo, de tal forma que a árvore selecionada preservasse as características do processo estocástico do qual as vazões são geradas e não possui cardinalidade elevada. Em (PENNA, 2009) técnicas de agregação foram utilizadas com o mesmo objetivo no problema de médio prazo.

A análise de agrupamentos é uma técnica de classificação que objetiva reduzir a cardinalidade dos dados. Agrupa um conjunto de dados em subconjuntos, utilizando um critério fixado que pode variar ligeiramente em virtude do método de agrupamento utilizado.

O conceito de análise de agrupamentos engloba uma série de diferentes métodos e algoritmos para agrupar objetos do mesmo tipo nas respectivas categorias. Em outras palavras é uma ferramenta de análise que visa a triagem de diferentes objetos em grupos, de uma forma que o grau de associação entre dois objetos é máximo se eles pertencem ao mesmo grupo e mínimo em caso contrário.

Segundo (HAIR, 1998), a análise de *cluster* ou agrupamentos, também conhecida como análise de conglomerados, é um conjunto de técnicas estatísticas cujo objetivo é agrupar objetos segundo suas características, formando grupos ou conglomerados homogêneos. Os objetos em cada grupo tendem a ser semelhantes entre si, porém diferentes dos demais objetos dos outros grupos. Os agrupamentos obtidos devem apresentar tanto uma homogeneidade interna (dentro de cada grupo), como uma grande heterogeneidade externa (entre grupos). Portanto, se a agregação for bem sucedida, quando representados em um gráfico, os objetos dentro dos grupos estarão muito próximos, e afastados se comparados com objetos de grupos distintos .

A análise de agrupamentos pode ser usada para descobrir estruturas em dados sem fornecer uma explicação/interpretação. Em outras palavras, a análise de agrupamento simplesmente descobre as estruturas de dados sem explicar por que eles existem.

Sabendo que o objetivo da análise de cluster é agrupar objetos semelhantes, é necessária uma medida da distância entre os mesmos. Os objetos com menor distância entre si são mais semelhantes, logo são agregados em um mesmo grupo. Já os mais distantes participam de grupos distintos. Existem várias formas de medir a distância entre os objetos, porém, a mais utilizada é a distância euclidiana.

De acordo com MALHOTRA (2001), as etapas para a aplicação da análise de agrupamentos são descritas na Figura 5.1. Inicialmente, é necessário definir o problema de aglomeração e as variáveis a serem tratadas estatisticamente. Escolhe-se, então, uma medida de distância dos grupos. Após, define-se o processo de agregação que dependerá das variáveis em estudo e do problema em foco. Os grupos resultantes devem ser interpretados em termos das variáveis usadas para constituí-los e de outras variáveis adicionais importantes. Finalmente a validade do processo de agregação deve ser verificada.

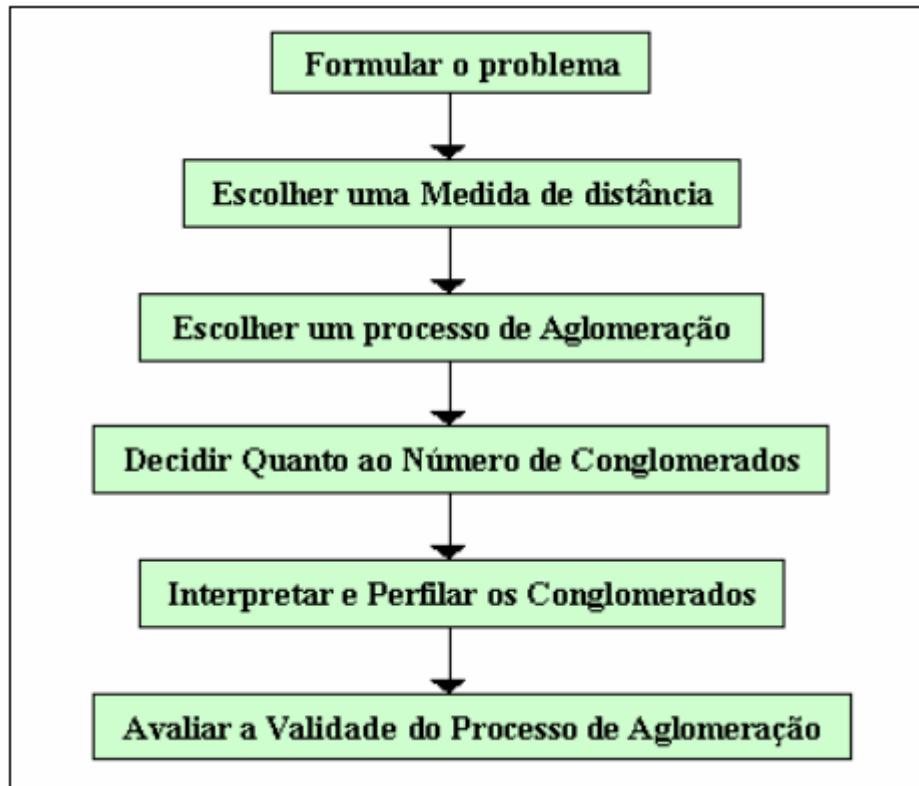


Figura 5.1 : Etapas para a aplicação da análise de cluster

Fonte: (MALHOTRA, 2001)

As técnicas de agrupamento são empregadas largamente em várias áreas de estudo. Em KRANEN (2004), algoritmos de agrupamentos são aplicados para agregar séries de potências das usinas hidroelétricas que são dependentes da aleatoriedade das afluências aos reservatórios, a fim de reduzir esforço computacional no estudo de confiabilidade de sistemas hidrotérmicos. Em estudos de planejamento da operação energética, o K-Means foi utilizado em JARDIM (2002) no problema de planejamento de curto prazo e em PENNA (2009) no problema de planejamento de médio prazo, ambos buscando construir uma sub-árvore representativa de cenários para ser visitada durante os processos de otimização. O K-Means também é aplicado na área de ciências humanas, como em GUIDINI (2008), em que o método é utilizado para analisar os estilos gerenciais das empresas.

Os métodos de agregação são divididos em duas classes: os métodos hierárquicos e os métodos não hierárquicos. A seguir é apresentada uma breve descrição destes métodos.

### 5.2.1 Métodos Hierárquicos

Os métodos hierárquicos de agrupamento são classificados em aglomerativos e divisivos. Os métodos aglomerativos partem de elementos individuais, agrupando-os de acordo com uma medida de similaridade que reflete a proximidade entre os objetos. A cada etapa novos grupos são formados e no final tem-se um único grupo. Cabe ao pesquisador definir um critério de parada do agrupamento de acordo com seu interesse. Nos métodos divisivos, parte-se de um único grupo de objetos e, também segundo uma medida de similaridade, segue-se separando os objetos em grupos até que cada grupo contenha somente um elemento. Também aqui é necessário que o pesquisador defina um critério de parada do agrupamento de acordo com seu interesse, que pode ser um número máximo de grupos, por exemplo. A Figura 5.2 ilustra esses métodos.

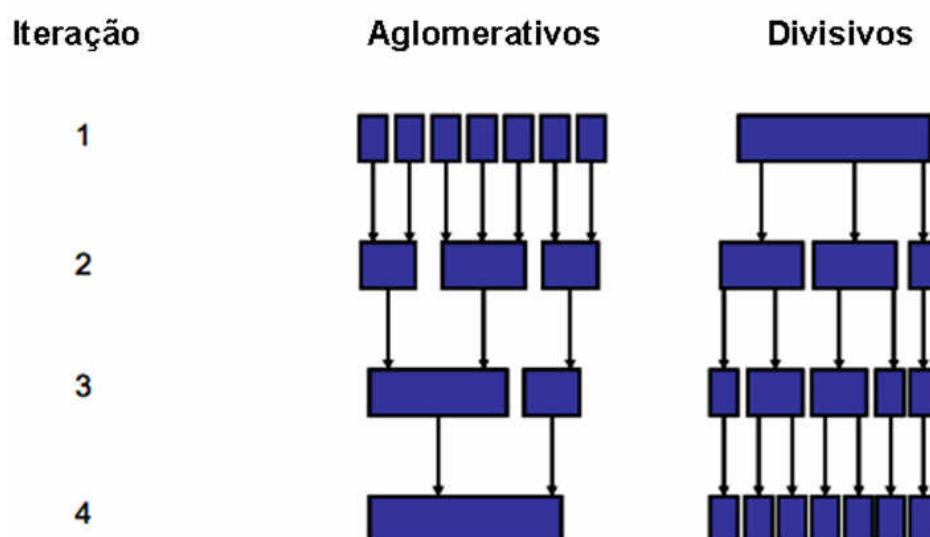


Figura 5.2 : Processos aglomerativos e divisivos

Fonte: (SCARPEL, 2010)

Os agrupamentos são gerados a partir de uma matriz de similaridades. Para cada reunião de elementos em um grupo calcula-se o centróide desse grupo, que passa a ser representado por essa nova medida para as próximas etapas de agrupamentos.

Os métodos hierárquicos mais utilizados são:

- Ligação simples ou método do vizinho mais próximo
- Ligação completa ou método do vizinho mais distante
- Ligação média
- Centróide
- Ward

De acordo com JAIN (1999), os métodos de ligação simples e ligação completa são os mais populares. Esses dois métodos diferem na forma com que caracterizam a similaridade entre os pares de grupos. No método de ligação simples, a distância entre dois grupos é a distância mínima entre todos os pares de elementos pertencentes a estes grupos. No método da ligação completa, a distância entre dois grupos é a máxima distância entre todos os pares de elementos que pertencem a estes grupos. A Figura 5.3 ilustra essas diferenças na definição da similaridade.

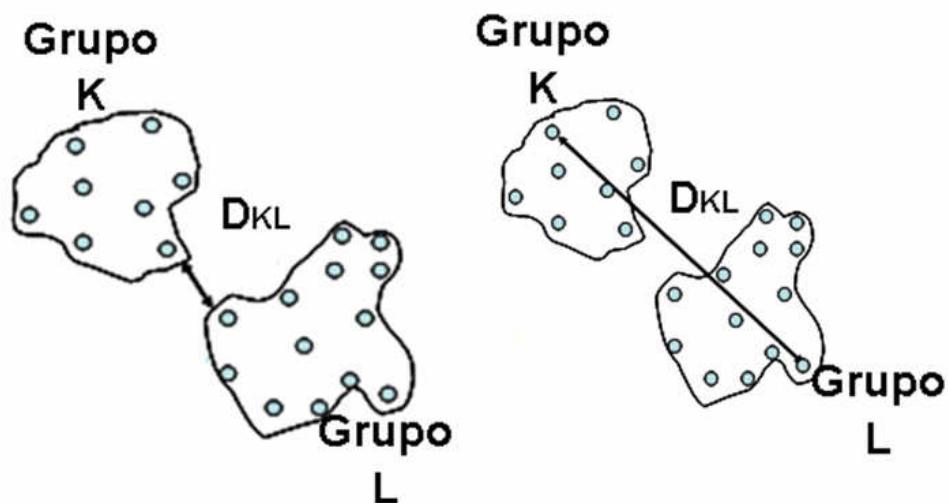


Figura 5.3 : Definição da medida de similaridade

Fonte: (SCARPEL, 2010)

Os métodos hierárquicos são assim chamados devido ao fato de que uma vez alocado o elemento dentro de um grupo, ele não poderá mais ser removido deste, mesmo que se

constate nas iterações seguintes a sua semelhança com outro grupo que não seja ao qual faz parte, o que é uma desvantagem do método. Outra desvantagem do método é a necessidade de construção e armazenamento da matriz de similaridade, o que pode ser um fator limitante (dependendo dos recursos computacionais disponíveis) para a utilização do método em aplicações com microcomputadores. Em geral, métodos hierárquicos não são, devido às desvantagens acima, recomendados para problemas envolvendo grande quantidade de dados.

Os resultados desses métodos são representados graficamente por um gráfico bidimensional chamado dendograma. No dendograma pode-se observar as etapas em que foram feitos os agrupamentos, bem como ter acesso ao valor da distância que fez com que esse agrupamento ocorresse. Na Figura 5.4, é mostrado um dendograma que representa etapas de agrupamento utilizando a ligação simples (ou vizinho mais próximo).

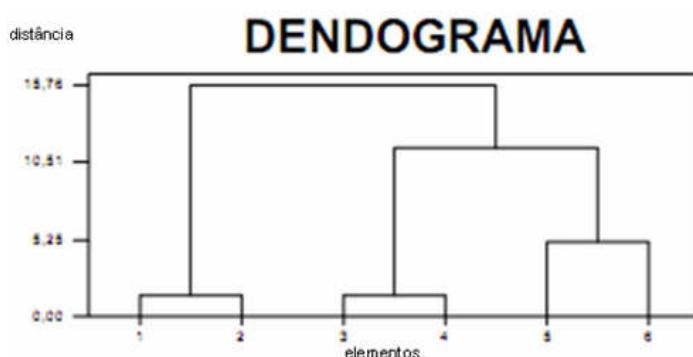


Figura 5.4: Dendrograma utilizando ligação simples

Fonte: (SCARPEL, 2010)

### 5.2.2 Métodos Não Hierárquicos

Os métodos não hierárquicos são assim chamados porque em qualquer etapa do processo de agrupamento, pode-se realocar um elemento de um grupo para outro, ao notar-se que este está mal localizado. Os grupos são formados de forma a garantir a máxima semelhança entre elementos do próprio grupo de forma que não haja outros elementos tão semelhantes que pertençam a outros grupos, o que representa uma vantagem em relação aos métodos hierárquicos. Tecnicamente falando isso se traduz em coesão interna e isolamentos dos grupos.

Em contrapartida aos métodos hierárquicos, os presentes métodos não necessitam de determinação e armazenamento da matriz de similaridade (que tem dimensões proporcionais ao número de elementos), o que representava um entrave nas aplicações computacionais. Portanto esses métodos são indicados quando é necessário reduzir a cardinalidade de um espaço amostral com grande quantidade de elementos.

Como explicitado anteriormente, o método não hierárquico precisa da definição do número de grupos a priori, ou seja, não é capaz de determinar o melhor número de agrupamentos a ser utilizado para um determinado espaço de dados. Esse problema pode ser contornado através de técnicas de validação de agrupamentos, que avaliam a qualidade dos conglomerados formados segundo um determinado índice ou fator. Uma dessas medidas é a Compacidade e Separação (CS), que avalia a compacidade dos grupos, ou seja, se as amostras constituintes de cada agrupamento encontram-se próximas e se os centróides dos agrupamentos encontram-se o mais afastados possível (KRAMEN, 2004).

O caminho mais intuitivo para encontrar a melhor partição é checar todas as possíveis partições dos conjunto de dados, porém o número de possibilidades é muito grande, assintoticamente da ordem de  $K^{n-1}$  (BUSSAB, 1990), onde ‘K’ é o número de grupos e ‘n’ o número de objetos que se deseja agrupar. Ao analisar um pequeno problema envolvendo 20 objetos separados em 3 grupos , é preciso testar/investigar aproximadamente 1000.000.000 de possíveis partições únicas. Grandes esforços vêm sendo feitos com o intuito de desenvolver procedimentos heurísticos que analisam algumas partições especiais com o intuito de encontrar a melhor ou uma quase ótima, visto que analisar todas as partições possíveis é inviável.

Um desses procedimentos heurísticos desenvolvidos é o K-Means. Este método é muito popular e aplicado quando necessita-se trabalhar com grande quantidade de elementos. Este método será o foco da aplicação deste trabalho.

### **5.2.2.1 Método K-Means**

O K-Means é um método amplamente difundido, existindo muitas variações propostas na literatura e diversos nomes (K-médias, isodata, ou migração de médias). É um método de partição baseado em realocação que necessita da definição a priori do número de agrupamentos K a que se deseja reduzir a amostra inicial.

A função ser minimizada é definida em função da distância dos elementos em relação aos centros dos agrupamentos. O elemento representativo de um cluster é o seu centróide, que possui um valor médio para os atributos considerados, relativos a todos os elementos do grupo.

A partir de uma estimativa inicial das coordenadas dos centros dos agrupamentos (centróides), o algoritmo calcula a distância de cada ponto do conjunto a estes centróides. A seguir, o algoritmo aloca cada elemento do conjunto em um grupo, de acordo com a menor distância ao centróide correspondente.

A nova estimativa das coordenadas dos centróides é calculada pela média aritmética das coordenadas dos pontos associados a cada grupo. O método K-Means é sensível ao particionamento inicial realizado, em virtude da escolha das coordenadas dos K centróides dos grupos ser feita inicialmente de forma aleatória.

O algoritmo do método K-Means é constituído dos seguintes passos básicos:

- (1) Determinar as posições iniciais dos k centróides dos grupos;
- (2) Alocar cada elemento ao grupo do centróide mais próximo;
- (3) Recalcular os centros dos grupos a partir dos elementos alocados;
- 4) Repetir os passos de 2 a 4 segundo algum critério de convergência

Como critério de convergência pode se executar o algoritmo até que os centróides não se movam mais, ou até que um determinado número máximo de iterações seja alcançado.

Uma maneira mais formal de definir o K-Means é dada por:

Sejam n pontos m-dimensionais  $x_1, x_2, \dots, x_n$ . O método K-Means consiste em separar os n pontos  $x_i$  em  $k < n$  conjuntos  $s_1, s_2, \dots, s_k$  de modo a minimizar a soma dos quadrados, como apresentado a seguir :

Seja  $n_i = |s_i|$  o número de elementos do subconjunto  $s_i$ , para  $i = 1, 2, \dots, k$ . Seja também

$y = y_1, y_2, \dots, y_k$  um ponto em  $\mathbb{R}^m$ , onde cada  $y_i$  pode ser visto como um

representante(centróide) do grupo  $s_i$ . Então o problema de encontrar a ótima repartição dos  $n$  pontos  $x_i$  em  $k$  grupos é representado matematicamente por

$$(I) \min_s \sum_{i=1}^K \min_{y_i} f(y_i; s_i),$$

onde  $s = (s_1, s_2, \dots, s_k)$  e  $f(y_i; s_i) = \sum_{x_j \in s_i} \|x_j - y_i\|^2$ .

O problema de segundo nível é dado por:

$$(II) \min_{y_i \in R^m} f(y_i; s_i)$$

O problema (II) é um problema de otimização convexa sem restrições.

Deste modo, a condição necessária e suficiente para que um ponto  $c_i$  seja a solução ótima de (II) é  $\nabla f(c_i, s_i) = 0$ , isto é,  $\sum_{x_j \in s_i} (x_j - c_i) = 0$ . Esta igualdade proporciona:

$$c_i = \frac{1}{n_i} \sum_{x_j \in s_i} x_j, \text{ ou seja, a solução ótima de (II) é o centróide do grupo } s_i \text{ para } i = 1 \dots k.$$

Deste modo, o problema (I) pode ser reescrito como

$$(I') \min_s \sum_{i=1}^K \sum_{x_j \in s_i} \|x_j - c_i\|^2$$

Na Figura 5.5 são ilustrados os estágios do processo de aplicação do método K-Means até a convergência.

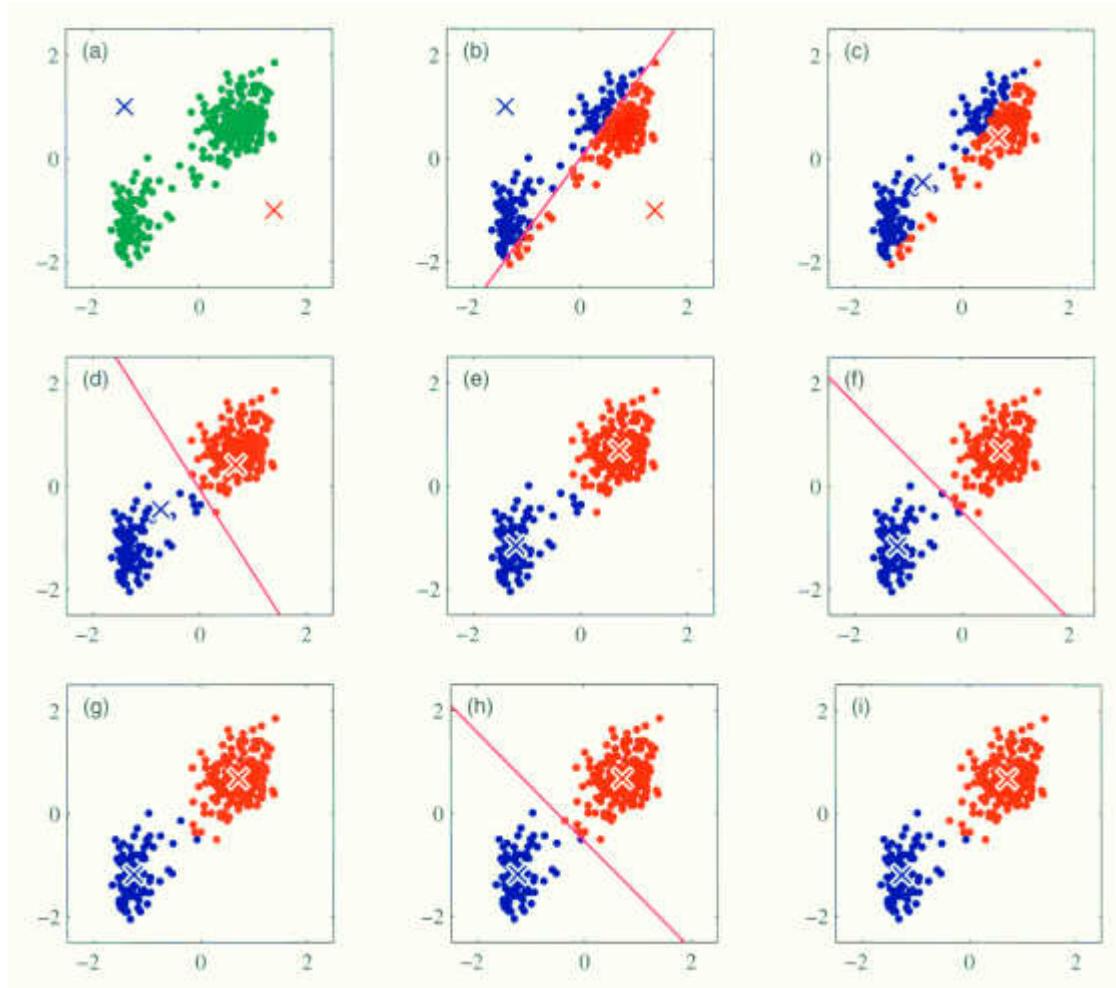


Figura 5.5: etapas do processo de agregação com K-Means

Fonte : [www.itee.uq.edu.au/~comp4702/material.html](http://www.itee.uq.edu.au/~comp4702/material.html)

As etapas da Figura 5.5 são:

- Os pontos claros (verdes) denotam a amostra original no espaço euclidiano bidimensional. As escolhas iniciais arbitrárias para os centróides estão representadas pelos pontos azul e vermelho.
- No primeiro passo cada ponto é associado ao cluster azul ou vermelho, de acordo com a medida de similaridade.
- Na etapa seguinte cada centróide é recalculado e passa a ser a média dos pontos pertencentes ao grupo.
- até (i) - Etapas consecutivas de realocações e atualizações de centróides até a convergência do método.

Vale lembrar que uma escolha mal feita para os representantes iniciais pode elevar o número de iterações necessárias até a convergência.

### **5.3 Resumo**

Neste capítulo foram abordados métodos de análise de agrupamentos, que têm como objetivo selecionar dentre um grande número de objetos, aqueles mais representativos de forma que as principais características sejam preservadas. Enfatizou-se a importância desses métodos e sua aplicabilidade, bem como suas principais características.

No próximo capítulo será apresentada a proposta de utilização do K-Means e o tratamento da matriz de dados para construção da sub-árvore de cenários a serem utilizados durante a etapa forward no modelo de planejamento de médio prazo (NEWAVE).

## **6 Proposta de construção da sub-árvore de cenários**

### **6.1 Introdução**

Como visto no capítulo 4, o modelo autorregressivo periódico PAR( $p$ ) gera as séries de afluências através de um processo estocástico em função das  $p$  afluências dos períodos anteriores e uma parcela aleatória, o ruído. Será escolhida, através do K-Means, uma porção da pseudo-árvore completa que será visitada na etapa forward do cálculo da política.

A forma com que essas afluências serão agregadas de forma a construir a sub-árvore a ser visitada, influencia diretamente na robustez do modelo, pois se as séries constituintes da sub-árvore forem incompatíveis com o registro histórico de afluências, os resultados não representarão a realidade.

### **6.2 Por que usar técnicas de agregação**

O ideal seria a existência de um registro histórico de afluências tão grande (em quantidade de séries temporais) que não se precisasse gerar as séries de afluências de forma estocástica, e esse registro (separado em seqüências) seria visitado durante os cálculos realizados no modelo de planejamento da operação de energética de médio prazo. Mas mesmo que este estivesse disponível, dependendo do número de séries, ficaria impraticável computacionalmente visitar todas pois levaria muito tempo utilizando computadores comuns.

As técnicas de análise de agrupamentos permitem que se reduza a cardinalidade de amostras para um número menor, mas mantendo a representatividade da amostra original. Um destes métodos é o K-Means. Essa técnica é muito poderosa e largamente utilizada pois permite que seja representado todo o espaço amostral utilizando apenas alguns de seus representantes.

### 6.3 Proposta para a construção da sub-árvore

O método proposto consiste em aplicar técnicas de agrupamento a fim de reduzir a dimensionalidade da árvore de cenários a ser visitada durante o cálculo da simulação forward. Inicialmente serão agrupadas afluências, que são correlacionadas temporal e espacialmente, analisando-se os resultados obtidos. Também serão agrupados os ruídos, que serão utilizados na geração das séries pelo modelo autorregressivo periódico, o PAR(p).

Primeiramente serão geradas séries de energias (pseudo-árvore completa) através do PAR(p). Em seguida essas séries serão padronizadas e utilizadas para a construção da sub-árvore através da técnica de agregação escolhida (K-Means). As séries geradas pelo PAR(p) estão dispostas em uma matriz de dimensões [Ncen NperxNsist], onde:

- Ncen representa o número de cenários hidrológicos gerados pelo PAR(p)
- Nsis representa o número de subsistemas equivalentes( sudeste, sul, nordeste e norte , por exemplo)
- Nper representa o número de períodos do horizonte de estudo.

Ou seja, são Ncen cenários, que correspondem às linhas da matriz, e Nper períodos para cada subsistema, alocados em colunas adjacentes. As primeiras Nper colunas correspondem ao subsistema Sudeste, as próximas Nper ao Sul e assim por diante.

A principal vantagem de métodos não-hierárquicos como o escolhido consiste em não ser necessário o cálculo e armazenamento da matriz de similaridades e o elemento pode mudar de grupo ao longo das iterações do método, segundo um critério de minimização escolhido.

O K-Means necessita que se conheça a priori o número K de grupos que pretende-se formar e os centróides iniciais aos quais todos os cenários serão inicialmente comparados. Portanto, para iniciar o processo de agregação serão selecionados os centróides através de amostragem aleatória simples. Estes elementos serão objetos do conjunto de entrada de energias, pertencentes à amostra original; desta maneira é garantido que cada grupo tenha pelo menos um elemento associado a ele.

A cada iteração do método novos elementos são adicionados/retirados aos grupos e nos passos seguintes até a convergência do processo de agregação, ou seja, até que nenhum elemento seja realocado. O centróide dos grupos será o ponto médio destes grupos. Terminado o processo de agrupamento e de posse dos centróides; faz-se a comparação de todos os cenários da matriz original com esses centróides a fim de capturar o cenário

representativo, que será o mais próximo ao centróide. Dessa forma, os cenários resultantes do processo de agregação são realizações geradas pelo PAR(p), modelo calibrado pelas séries históricas de afluências. A Figura 6.1 abaixo ilustra esse procedimento, onde os pontos marcados com ‘x’ representam os centróides dos grupos:

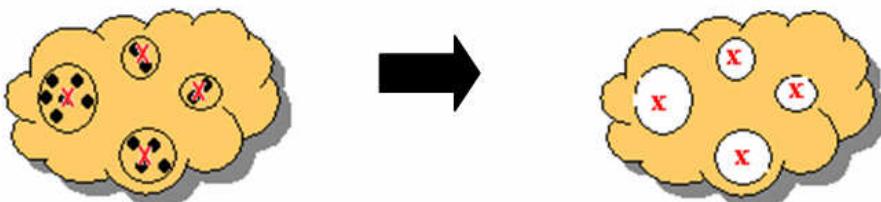


Figura 6.1: Escolha do cenário representativo

Os cenários resultantes do processo de agregação não são mais equiprováveis. A probabilidade de cada cenário leva em conta o número de elementos pertencente ao grupo. A probabilidade  $P_k$  associada ao grupo k pode se calculada da seguinte maneira:

$$P_k = N_{ek} / N \quad (6.1)$$

Onde:

$N_{ek}$  é o número de elementos/cenários pertencentes ao grupo k e  $N$  é o tamanho da amostra antes de iniciar o processo de agregação.

Na Figura 6.2 é ilustrado o procedimento completo de agregação:



Figura 6.2: Aplicação do procedimento de agregação

## **6.4 Comparação com AS e Atual**

Na versão atual do NEWAVE as séries que são geradas para o cálculo da política e simulação final são extraídas através de amostragem aleatória simples; ou seja, os ruídos (parte estocástica do modelo) são sorteados da distribuição de energias naturais afluentes e são equiprováveis. Esse procedimento é denominado opção Atual

Tramita em processo de validação uma alternativa (Amostragem Seletiva-AS) de geração da sub-árvore que também utiliza o K-Means como método de agregação (associado a técnicas de seleção de cenários). Neste método o agrupamento leva em conta que um período não interfere no outro (em termos de ruídos), ou seja, para cada estágio é gerada uma amostra grande de ruídos e em seguida procede-se ao processo de agrupamento.

Na Amostragem Seletiva, para cada mês (ou período) são geradas 100 mil amostras de ruídos aleatórios equiprováveis . Em seguida procede-se à aplicação do processo de agregação onde a amostra fica reduzida a 200 ruídos (considerando a etapa forward), agora não equiprováveis. Esses ruídos são utilizados no modelo auto-regressivo periódico para a determinação das séries de energias afluentes. A probabilidade do cenário de energia natural afluente é igual à probabilidade do vetor de ruídos resultantes a partir do qual ele foi gerado.

O processo de agregação foi incorporado tanto na geração de séries de energias afluentes para a etapa backward quanto para a etapa forward, porém estudos com cenários hidrológicos com probabilidades muito diferentes no passo forward levaram a resultados muito instáveis. A probabilidade de ocorrência de um cenário na etapa forward é calculada como o produtório das probabilidades do ruído em cada um dos períodos. Essa probabilidade é padronizada de forma que a soma da probabilidade de todos os cenários somem um. A probabilidade no último período correspondente ao cenário consiste na probabilidade deste cenário. Alguns cenários, após essa alternativa de agregação, apresentaram probabilidades muito maiores do que os demais, o que pode ser justificado devido a uma seqüência longa de ruídos com alta probabilidade. A ocorrência de cenários na etapa forward com probabilidades muito diferentes das demais trouxe problemas na convergência do processo de otimização, uma vez que o valor esperado do custo total de operação (ZSUP) é calculado como a média ponderada dos custos totais de operação em cada um dos cenários (PENNA, 2009).

Para a correção desse problema foi construída uma nova amostra de ruídos equiprováveis para a etapa forward a partir de um sorteio condicionado realizado após o processo de agregação dos ruídos, fazendo com que os cenários dessa etapa voltassem a ser equiprováveis.

Com o objetivo de aprimorar os cenários para a etapa forward, a ASG propõe agregar considerando não só o período atual, mas todos os períodos do horizonte, visando evitar seqüências longas de ruídos com alta probabilidade. Dessa forma, a ASG pode ser considerada uma agregação ‘por cenários’ e não ‘por períodos’. Na Figura 6.3 são ilustradas as etapas da ASG.

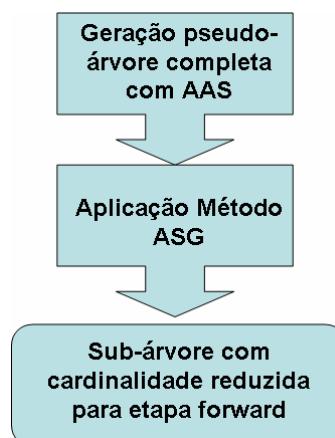


Figura 6.3: Etapas da ASG.

Desta forma, os cenários gerados a serem visitados durante a etapa de simulação forward ficarão com características semelhantes aos registros hidrológicos observados, mas com menor possibilidade de ocorrência de cenários com altíssima probabilidade. Como foi visto no caso da AS, se a probabilidade de um cenário for muito diferente das probabilidades dos demais a convergência do processo de otimização fica comprometida.

Como discutido anteriormente, a alternativa de Amostragem Seletiva, trouxe bons resultados nos cenários gerados para a etapa backward (não equiprováveis), mas a etapa forward mostrou alguns problemas quando foram considerados cenários não equiprováveis. Essa alternativa também trouxe mais robustez aos valores de ZINF (limite inferior do valor esperado do custo total de operação) frente às mudanças nos cenários hidrológicos gerados, o que não acontecia na opção Atual. Os valores de ZSUP também ficaram mais robustos, mas não tanto quanto os de ZINF.

A ASG pretende melhorar a robustez dos cenários hidrológicos gerados para a etapa forward, podendo trabalhar com cenários não equiprováveis nessa etapa sem prejuízo nos resultados do processo de otimização.

## **6.5 Resumo**

Neste capítulo foi abordada a proposta de agregação deste trabalho, a ASG, as expectativas de melhorias e a comparação com a opção atual e a alternativa de amostragem seletiva que tramita em processo de validação.

## **7 Aplicação da Metodologia proposta**

### **7.1 Introdução**

No capítulo anterior foram abordados de forma resumida os métodos de construção da sub-árvore da etapa forward a ser visitada durante o cálculo da política de operação, que são a opção Atual, a AS e a ASG .

Nesse capítulo serão descritas as formas de tratamento da matriz de energias naturais afluentes e ruídos aleatórios de forma a atender aos requisitos propostos neste projeto, o método ASG. A disposição da matriz de dados de afluências será de fundamental importância durante a aplicação do método de agrupamento.

### **7.2 Padronização da matriz de energias**

Ao utilizar uma técnica de agregação para comparar dados, além de selecionar a medida de similaridade a ser adotada no método de agrupamento, deve-se responder à seguinte pergunta: Os dados devem ser padronizados antes que as similaridades sejam calculadas?

Para responder essa pergunta é preciso analisar algumas questões. Primeiro, a maioria das medidas de distância é bastante sensível a diferentes escalas ou magnitudes entre as variáveis. Em geral, variáveis com maior dispersão (ou seja, maiores desvios padrão) têm maior impacto sobre o valor de similaridade final. Considere um exemplo para ilustrar a importância da padronização. Suponha que se queira agrupar indivíduos em três variáveis – atitudes perante um produto, idade e renda. Agora considere que mediu-se atitude em uma escala que varia de sete pontos oscilando entre os extremos gosta – não gosta, com idade medida em anos e a renda em reais. Se esses dados fossem representados num gráfico tridimensional, a distância entre pontos (e sua similaridade) seria quase totalmente baseada nas diferenças de renda. As possíveis diferenças de atitude variam de 1 a 7, ao passo que a renda pode ter um intervalo mil vezes maior. Logo, graficamente, ficaria difícil perceber qualquer mudança na dimensão associada com atitude. Portanto o pesquisador deve estar ciente do peso implícito de variáveis com base em sua dispersão relativa, o que ocorre com medidas de distância.

Antes de realizar o processo de agregação proposto será realizada a padronização da matriz de energias afluentes, visando realizar uma comparação em que os efeitos de dispersão sejam devidamente considerados. Para tanto será utilizada a forma mais comum de padronização que consiste na conversão de cada variável em escores padrão pela subtração da média e divisão pelo desvio padrão de cada variável. Feito isso os escores são convertidos em valores com média igual a zero e desvio padrão igual a 1. Esse procedimento elimina o viés introduzido pelas diferenças nas escalas das variáveis que serão utilizadas no agrupamento.

### 7.3 Utilizando distâncias máximas, mínimas e médias entre os subsistemas

A fim de construir uma sub-árvore robusta e que traga resultados no planejamento da operação energética com menor variabilidade aos cenários gerados, a matriz de energias (padronizada) será disposta tridimensionalmente, comparando os cenários originais e os centróides utilizando métricas de forma que o horizonte de estudo tenha relevância significativa, ou seja, a agregação será feita considerando todo o horizonte de estudo.

A matriz de energias será organizada desta forma, ou seja, como uma matriz chamada ENA [Ncen Nsis Nper], onde :

- Ncen representa o número de cenários hidrológicos gerados pelo PAR(p)
- Nsis representa o número de subsistemas equivalentes( sudeste, sul, nordeste e norte , por exemplo)
- Nper representa o número de períodos do horizonte de estudo, que corresponde a 60 no caso de programação mensal da operação (PMO) e 120 no caso de plano decenal de expansão (PDE).

Pretende-se evitar que se repitam cenários de alta probabilidade comparados aos demais na etapa forward, para que não haja problemas futuros quando da utilização desses cenários no processo de otimização.

Como distância a ser considerada na utilização do método de agrupamentos (K-Means) será considerada a seguinte fórmula:

$$ds_x(j,k) = \sum_{i=1}^{Nper} (s^j(i) - s^k(i))^2 \quad x = 1, \dots, nsis. \quad j = 1, \dots, ncen \quad k = 1, \dots, ngru \quad (7.1)$$

Onde:

- nsis = número de subsistemas
- ncen = número de cenários
- ngru = número de grupos a serem formados
- $ds_x$  = distância do subsistema (que pode ser sudeste, sul, nordeste ou norte)
- Nper = número de períodos do horizonte de estudo
- s = energia do cenário 'j' do subsistema considerado
- i = índice do período considerado que varia de 1 a Nper
- j = índice do cenário original a ser comparado considerado
- k = índice do cenário de referência(representante inicial) do método de agrupamento

Portanto, soma-se o valor da diferença quadrática de todos os cenários aos clusters iniciais e aos grupos recalculados a cada realocação, para cada subsistema, chegando, para cada cenário, a um vetor de quatro posições, onde cada posição corresponde a um subsistema equivalente (sudeste, sul nordeste ou norte). Mas como foi mencionado anteriormente, deve-se que definir qual desses Nsis valores representarão esse cenário para efeito comparação para posterior agrupamento.

Seguindo esse procedimento, será aplicada a técnica de agregação K-Means utilizando o valor máximo, o valor mínimo e o valor médio dentre as distâncias mencionadas acima, ou seja, cada cenário passará a ser representado pelo valor máximo entre os subsistemas, pelo valor mínimo entre os subsistemas ou pelo valor médio dos subsistemas.

Após realizada a agregação e analisados os gráficos das estatísticas das medidas de tendência central (média) e de dispersão (desvio-padrão) verificou-se por comparação das estatísticas dos resultados da agregação com as estatísticas das séries históricas, que o comportamento da média e do desvio eram semelhantes. Entretanto estas estatísticas, principalmente o desvio-padrão, não foram preservadas de forma satisfatória. Depois de investigar possíveis equívocos no tratamento do problema, chegou-se à conclusão de que não se deve utilizar métricas matriciais, pois o método de agrupamento K-Means só garante solução local ótima se utilizadas métricas vetoriais e não métricas matriciais, como foi

considerado. Ou seja, a forma com que as distâncias foram calculadas e as comparadas não eram previstas na utilização do método. Portanto, essa forma de cálculo e disposição da matriz de energias afluentes foi abandonada.

## 7.4 Utilização da distância euclidiana convencional

Visando manter as características de agrupar os cenários levando em conta todos os períodos do horizonte de estudo, a disposição da matriz de energias foi mudada, passando agora a trabalhar com uma ENA2 [Ncen Nsis\*Nper].

Ou seja, a matriz de energias, após ser padronizada, ficará com dimensão [Ncen NsisxNper], onde a primeira dimensão corresponde ao número de cenários e a segunda dimensão corresponde ao número de subsistemas multiplicados por seus correspondentes períodos.

A medida de similaridade a ser adotada será a distância euclidiana convencional.

$$dc_{icen}(j,k) = \sum_{i=1}^{NperxNsis} (s^j(i) - s^k(i))^2 \quad (7.2)$$

Onde:

icen = índice do cenário considerado

dc = distância do cenário coniderado

## 7.5 Escolha dos cenários representantes

O produto da rotina de agregação utilizada é uma matriz reduzida com dimensão K x Nsis\*Nper , onde K é a média de todos os cenários pertencentes àquele grupo. Como o centróide não foi gerado pelo modelo autorregressivo periódico PAR(p), não será utilizado como um cenário a ser visitado durante o processo de cálculo da política. Portanto, de posse da matriz dos cenários médios, recupera-se os cenários originais. Isso é conseguido comparando cada cenário médio resultante da agregação com os cenários da matriz original de energias que pertencem ao grupo desse cenário médio. Dessa forma, o cenário da matriz de energia que for mais próximo desse cenário médio será escolhido como representante desse e utilizado nos cálculos pertinentes ao planejamento da operação. Essa abordagem de

recuperação de cenário foi realizada seguindo padrões do estudo realizado por PENNA(2009).

O método K-Means é sensível à escolha dos clusters iniciais, portanto serão elaboradas algumas formas de inicialização buscando uma que obtivesse melhores resultados.

## 7.6 Resultados após aplicação do método K-Means

A seguir são mostrados os resultados da agregação de energias considerando a matriz disposta como explicado no item 7.4. Foram plotados gráficos de evolução temporal comparando estatísticas de tendência central (média) e de dispersão (desvio padrão) dos grupos formados no processo de agregação (segmento vermelho) com a série de afluências histórica (segmento azul).

No eixo das abscissas tem-se os períodos considerados no estudo, geralmente 60, no caso de PMO, e a começar no mês inicial do estudo, no caso mês de setembro. No eixo das ordenadas encontram-se as ENAs dispostas em MWmês . Repare que cada subsistema tem valores típicos de valores de ENA, que diferem demais em comparação com outros. Nas figuras que seguem, o subsistema 1 corresponde ao Sudeste, o subsistema 2 corresponde ao Sul, o subsistema 3 corresponde ao Nordeste e o subsistema 4 corresponde ao Norte.

Os gráficos devem ser analisados por subsistemas e pode-se observar que, em todos os casos, a média é bem reproduzida no conjunto agregado de cenários (após aplicado K-Means), porém o resultado foi uma pequena mas relevante degradação no desvio padrão, principalmente no subsistema 2, que corresponde ao Sul. Vale ressaltar que a manutenção do desvio padrão é extremamente importante pois ele está ligado à possibilidade de ocorrência de valores extremos de afluências, períodos secos e úmidos.

Como o K-Means é sensível à escolha dos representantes iniciais, foram testadas algumas formas de construção da matriz de centróides iniciais.

### 7.6.1 Inicialização com 200 cenários tomados espaçados igualmente (matriz original)

Nesta inicialização, foram extraídos 200 cenários, espaçados igualmente de acordo com a posição na matriz de afluências original, de forma a compor a matriz de grupos inicial. Com esta inicialização o método de agrupamento convergiu em 8 iterações .

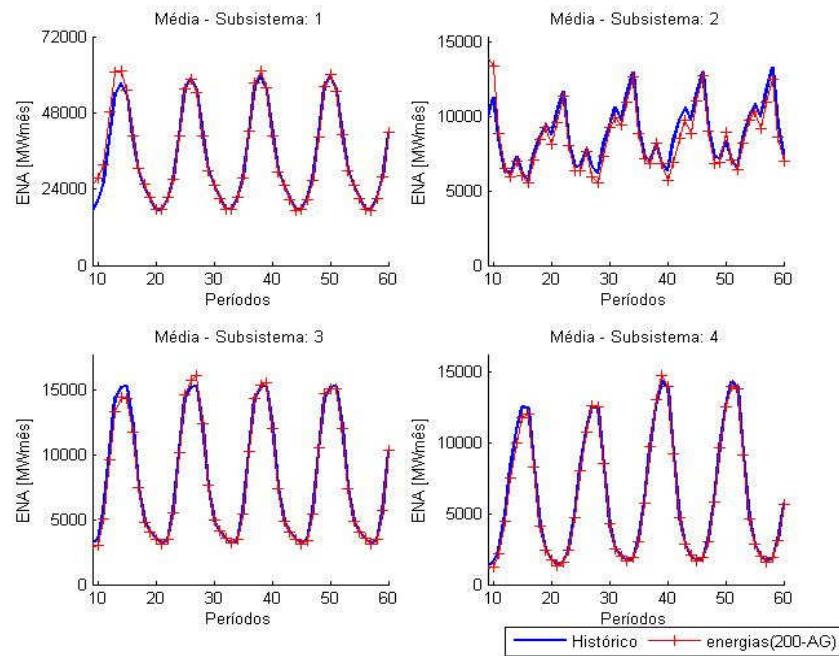


Figura 7.1: inicialização 200 cenários espaçados igualmente

Com base nos gráficos da média pode-se observar que os períodos iniciais apresentam uma pequena discrepância se comparados valores do registro histórico e do resultado da agregação. Isso pode ser explicado pelo fato dos primeiros meses serem fortemente influenciados pelo passado recente, isto é, a tendência hidrológica.

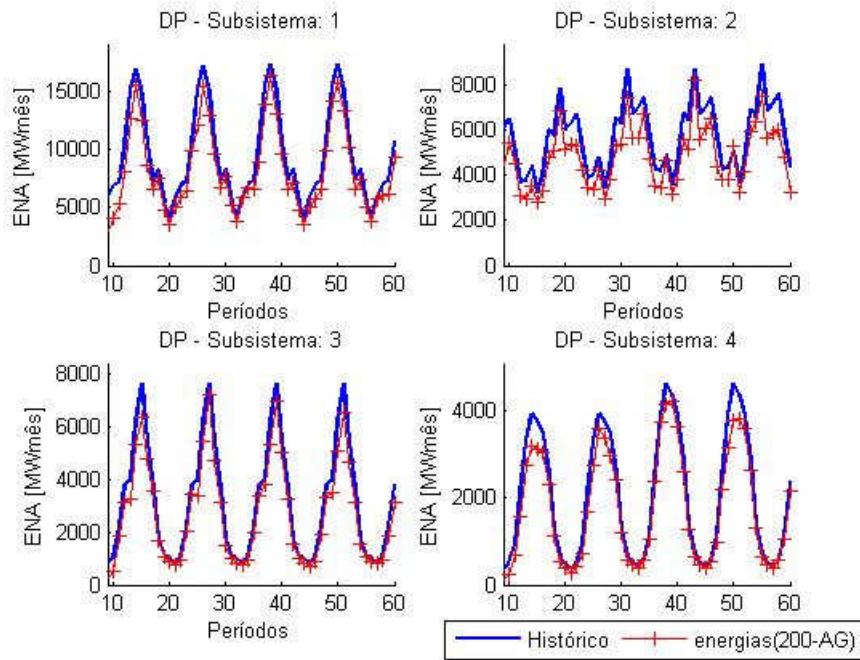


Figura 7.2: inicialização 200 cenários espaçados igualmente

Analizando o desvio padrão, nota-se que o subsistema Sul, teve uma maior discrepância nos valores de desvio padrão comparando registro histórico de afluências e os cenários agregados.

### 7.6.2 Ordenando e tomando cenários espaçados igualmente

Nesta inicialização, foram extraídos 200 cenários, espaçados igualmente de acordo com a posição na matriz de afluências original ordenada, de forma a compor a matriz de grupos inicial. Com esta inicialização o método de agrupamento convergiu em 8 iterações .

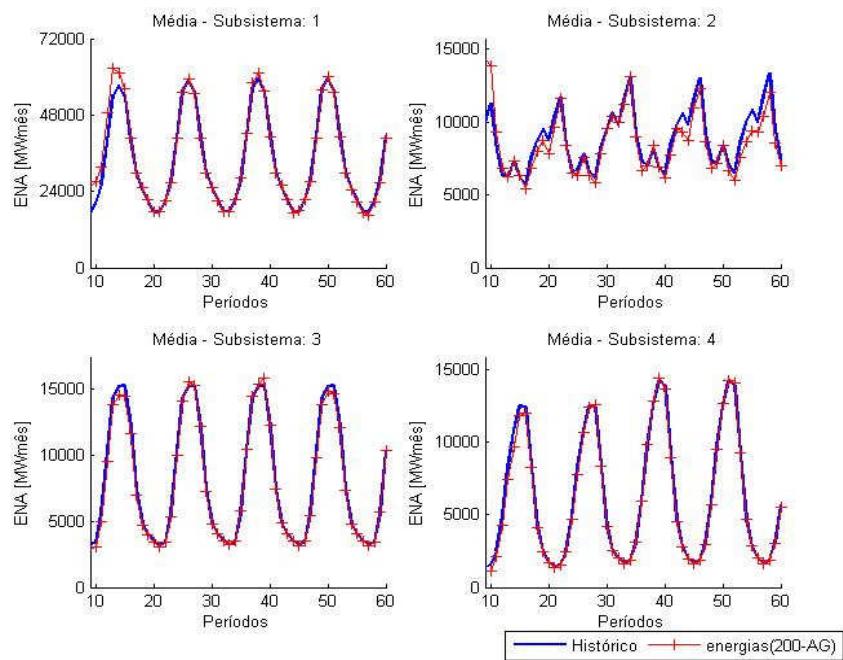


Figura 7.3 inicialização 200 cenários espaçados igualmente de um vetor ordenado

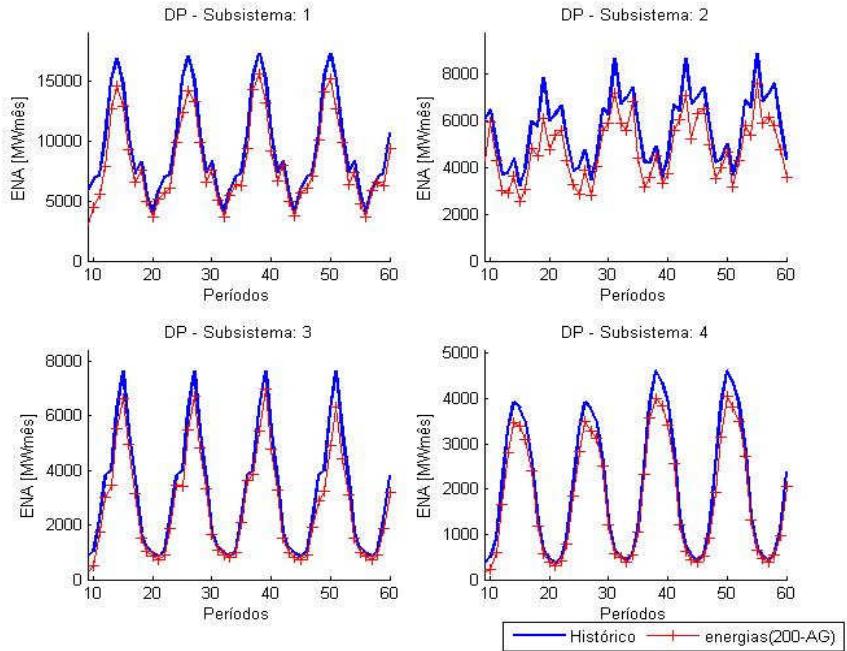


Figura 7.4: inicialização 200 cenários espaçados igualmente de um vetor ordenado

As mesmas observações da inicialização anterior podem ser feitas. Não obteve-se nenhuma melhora na preservação das estatísticas.

### 7.6.3 Ordenando e extraíndo 100 menores e 100 maiores

Nesta inicialização, realizou-se a ordenação crescente da matriz de afluências e foram extraídos 200 cenários extremos, 100 maiores e 100 menores, na expectativa de melhorar o desvio-padrão. Com esta inicialização o método de agrupamento convergiu em 9 iterações.

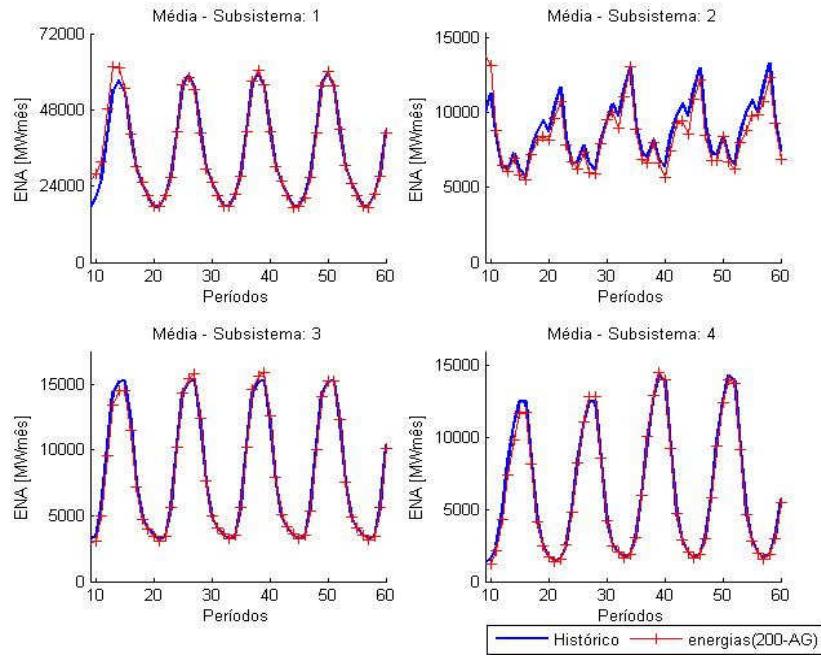


Figura 7.5 inicialização 200 cenários tomando os extremos

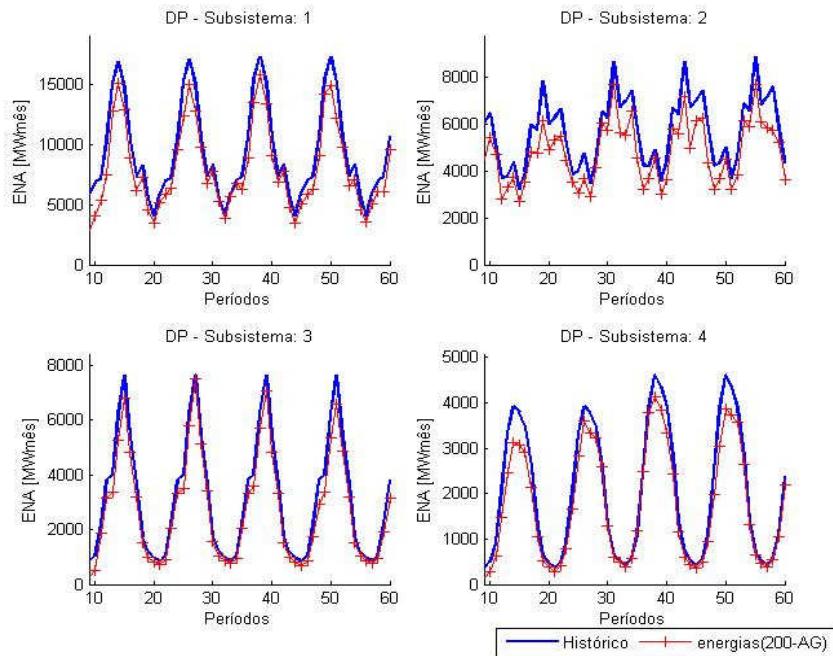


Figura 7.6: inicialização 200 cenários tomando os extremos

As mesmas observações da inicialização anterior podem ser feitas. Não obteve-se nenhuma melhora na preservação das estatísticas.

#### 7.6.4 Selecionando cenários próximos à média

Para esta inicialização foi criada uma medida de peso representativo do cenário e calculou-se a média dessa medida para tomar cenários próximos a esse valor.

Com essa inicialização o método de agrupamento convergiu em 7 iterações .

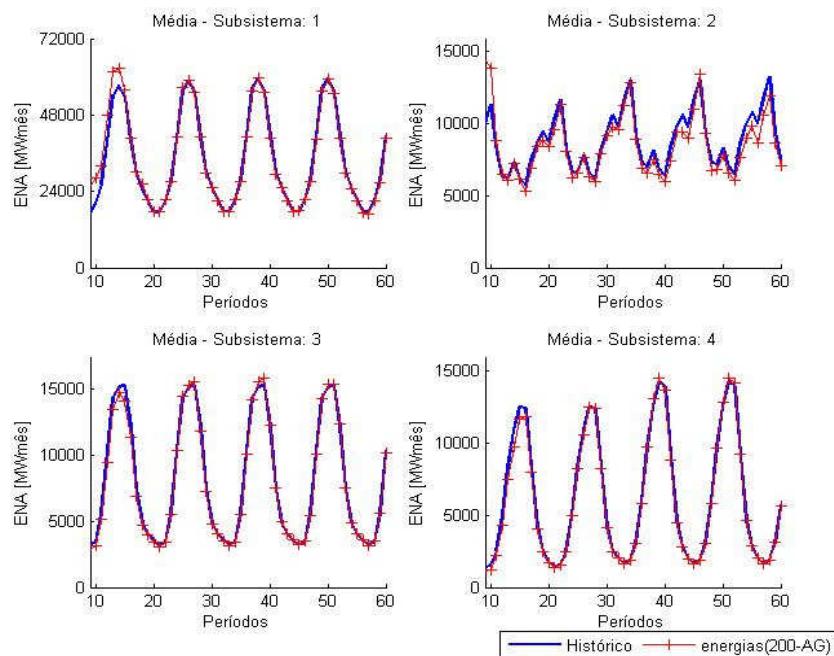


Figura 7.7: inicialização 200 cenários próximos à média

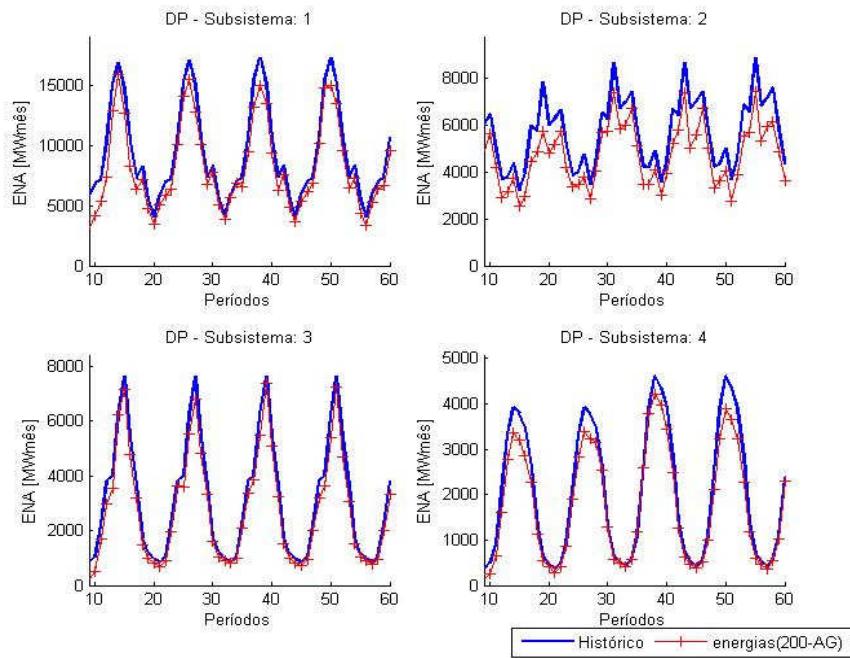


Figura 7.8: inicialização 200 cenários próximos à média

Apesar das tentativas de agregar energias, utilizando inclusive formas diferentes de inicialização dos grupos iniciais nota-se a degradação excessiva do desvio-padrão. Diante disso, verifica-se a dificuldade de agregar energias afluentes, e nessa etapa foi decidido aplicar o método ASG na parte estocástica da geração das séries pelo PAR(p), onde, através de fatores devidamente calibrados, pode-se melhorar a preservação do desvio-padrão.

## 7.7 Agregação utilizando ruídos aleatórios

Após analisadas as estatísticas acima e a percebida a degradação do desvio padrão, optou-se por estudar uma maneira de minimizar esta discrepância por meios de fatores de compensação. Decidiu-se, então, gerar uma amostra de ruídos com o desvio padrão um pouco maior, que, após agregada, pudesse preservar essa estatística. O modelo que gera as séries sintéticas de energias, o PAR(p), é calibrado de acordo com a série histórica e se fosse introduzido um fator diretamente no modelo, este deveria ser recalibrado a cada geração, o que seria inviável. Porém existe uma parte do modelo que pode ser utilizada para a introdução desse fator de compensação, que são os ruídos aleatórios. Esses ruídos representam a parte estocástica do modelo auto-regressivo e pode-se compensar a geração destes a fim de gerar energias com desvios padrão um pouco maiores. Então, com o intuito de reduzir a degradação observada no desvio padrão da amostra após o processo de agregação, é proposta a aplicação de um fator de compensação na amostra original de

forma a aumentar o desvio padrão desta. Assim, a amostra original de ruídos deve ser gerada com média zero e desvio padrão maior do que um. Esse procedimento é ilustrado na Figura 7.9.

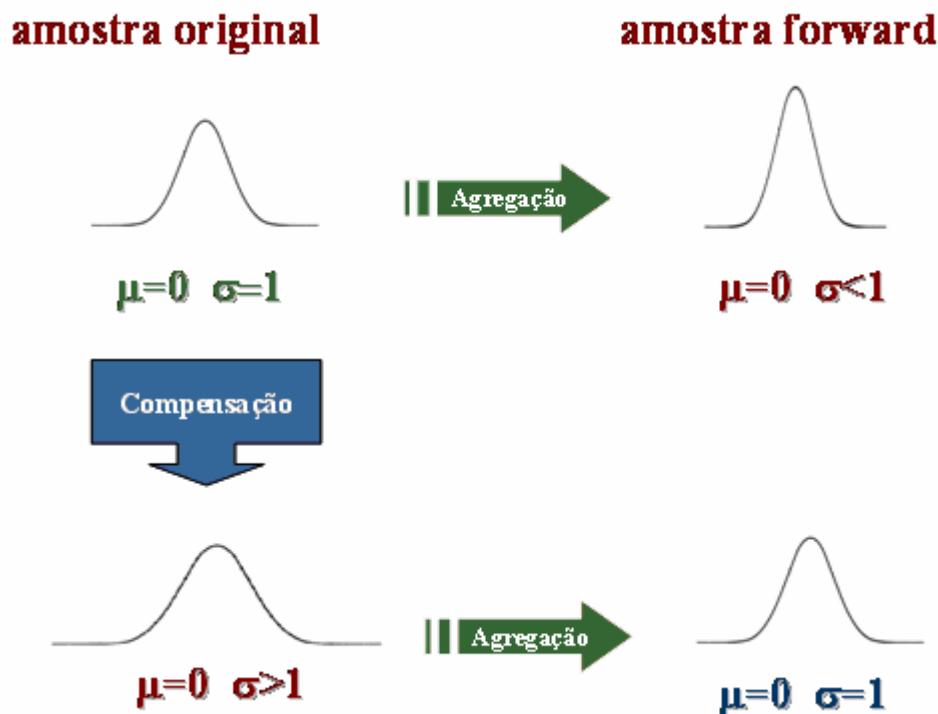


Figura 7.9: Aplicação do Fator de Compensação

Visto que a amostra de ruídos a ser agregada pode ser variada em tamanho, foi realizado um estudo para verificar o impacto do tamanho da amostra original a ser agregada na degradação do desvio-padrão. O resultado pode ser visualizado através da Figura 7.10 abaixo:

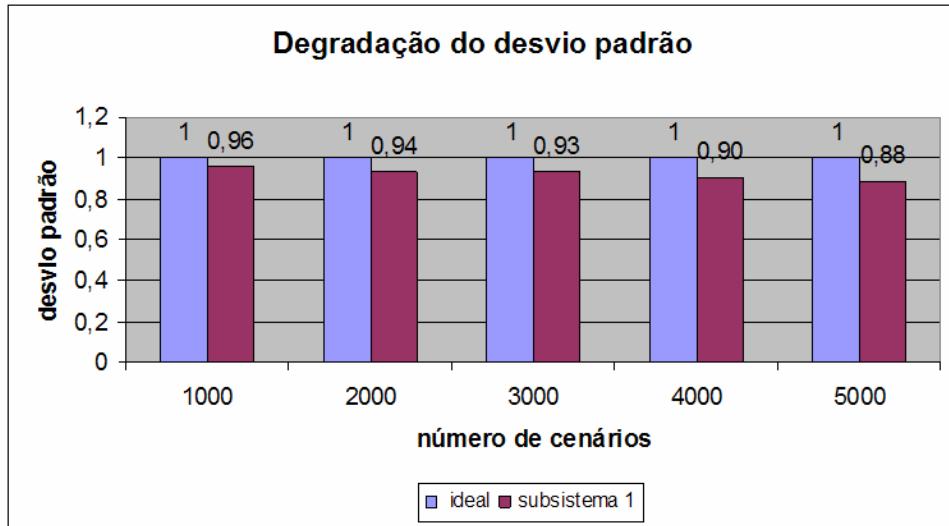


Figura 7.10: Degradação do desvio-padrão

Pode ser observado na Figura 7.10 que à medida que o número de objetos da amostra original a ser agrupado é aumentado, piora-se a representação do desvio padrão.

O fator de compensação (FC) deve ser calibrado de acordo com a cardinalidade e dimensionalidade do problema e é obtido dividindo-se o desvio-padrão ideal ( igual a um) pelo desvio-padrão obtido após aplicação da agregação nos ruídos, sendo portanto, maior que um. Na prática, esse fator de compensação é calculado como a média dos fatores de compensação de todos os subsistemas do caso em estudo considerando, por exemplo, dez amostras distintas. Na Tabela 7.1 abaixo estão os valores calculados do FC.

amostra	Subsistemas e fatores de compensação do desvio padrão			
	sudeste	sul	nordeste	norte
1	1,064169	1,071926	1,070893	1,065757
2	1,065984	1,060221	1,075963	1,058985
3	1,077818	1,071352	1,070549	1,069519
4	1,070664	1,072041	1,064623	1,080614
5	1,061458	1,062135	1,06792	1,06067
6	1,067236	1,082603	1,066212	1,06678
7	1,066553	1,066894	1,075963	1,057977
8	1,085776	1,063377	1,062925	1,082134
9	1,067692	1,061571	1,059322	1,080614
10	1,068376	1,073768	1,068604	1,068262
média	1,069573	1,068589	1,068297	1,069131

Tabela 7.1: Fatores de compensação para o desvio-padrão

## **7.8 Avaliação dos cenários hidrológicos gerados**

Nesta seção serão avaliados os cenários hidrológicos gerados pelos métodos Atual, AS e ASG, discutidos neste trabalho. São realizadas análises baseadas na estimativa da média, do desvio-padrão e da correlação cruzada.

A acurácia da árvore de cenários será avaliada através da árvore gerada para o Programa Mensal de Operação de Fevereiro de 2009.

Como mencionado em capítulos anteriores, o modelo gerador das séries sintéticas é o PAR(p), calibrado pelo registro histórico de afluências, e no qual a energia afluente de um período depende das p energias afluentes dos períodos anteriores. Nos estudos atuais, a ordem máxima considerada no modelo é igual a seis. Os primeiros períodos do estudo são fortemente influenciados pela tendência hidrológica, o que pode trazer algumas discrepâncias nas comparações das estatísticas da sub-árvore (construída com os métodos considerados) com as estatísticas do registro histórico nesses períodos iniciais.

### **7.8.1 Média dos cenários**

Neste item é analisada a média dos cenários hidrológicos gerados considerando os 3 métodos. Nas figuras 7.11 a 7.13 são apresentadas, a evolução temporal da média do histórico e dos cenários gerados para os subsistemas considerados no estudo.

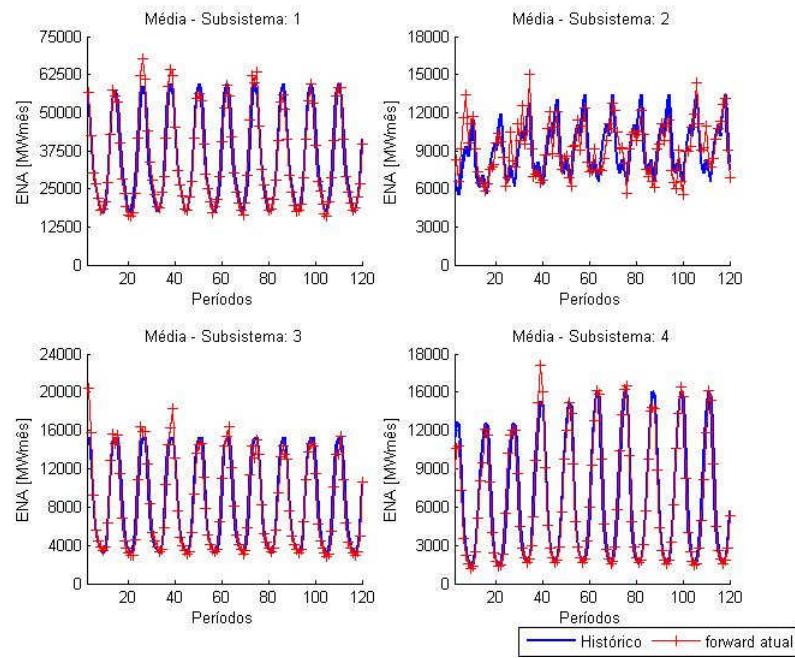


Figura 7.11: média dos cenários – op. Atual

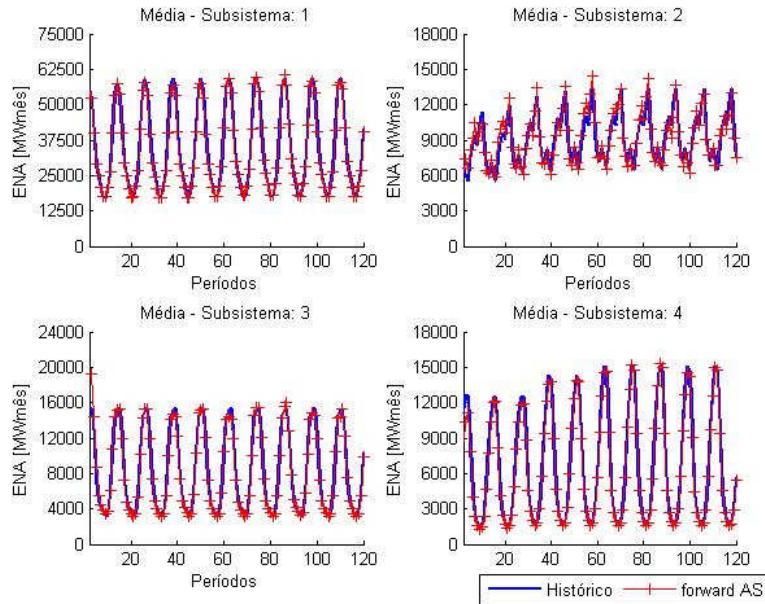


Figura 7.12: média dos cenários – op. AS

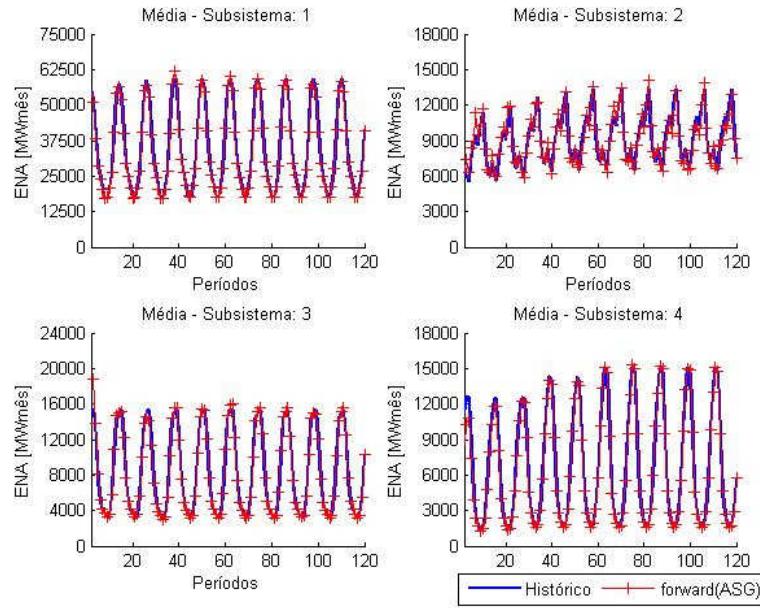


Figura 7.13: média dos cenários – op. ASG

Após analisadas as figuras 7.11 a 7.13, correspondentes aos métodos de agregação, conclui-se que a opção Atual não consegue representar a média histórica tão bem quanto as opções AS e ASG. Pode-se notar, inclusive, a ocorrência de picos de valores de afluências muito maiores do que o correspondente histórico, o que levaria a um planejamento da operação sob condições otimistas de afluências, fazendo com que se opte por maior geração hidroelétrica presente, aumentando o risco futuro de déficit. Essas são as possíveis consequências de cenários muito diferentes dos históricos.

### 7.8.2 Desvio-padrão dos cenários

Nas figuras 7.14 a 7.16 são apresentadas as estimativas do desvio-padrão para os quatro subsistemas comparando-as com o desvio-padrão histórico.

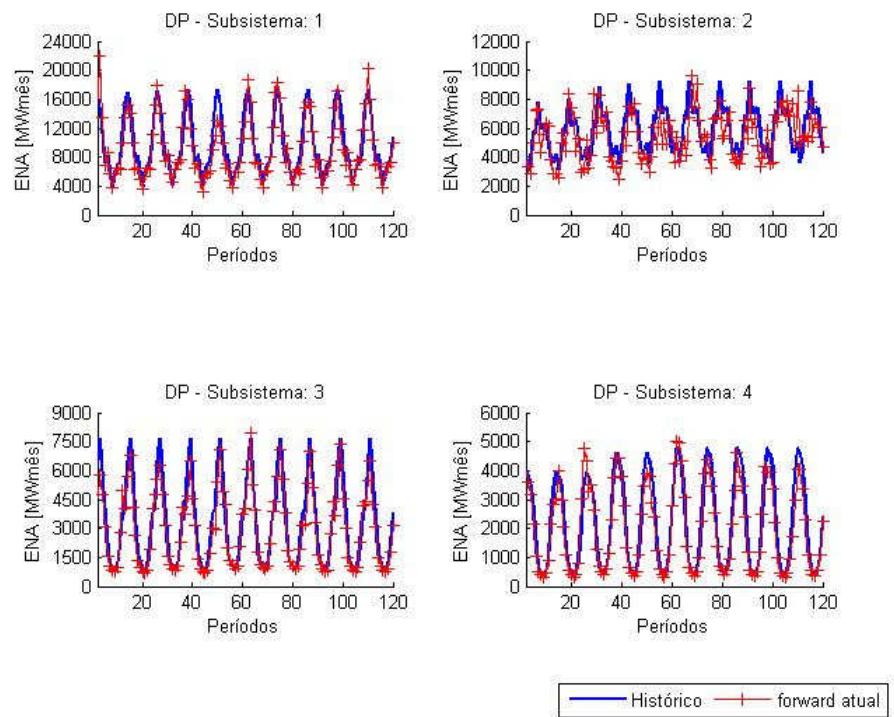


Figura 7.14: desvio-padrão dos cenários – op. Atual

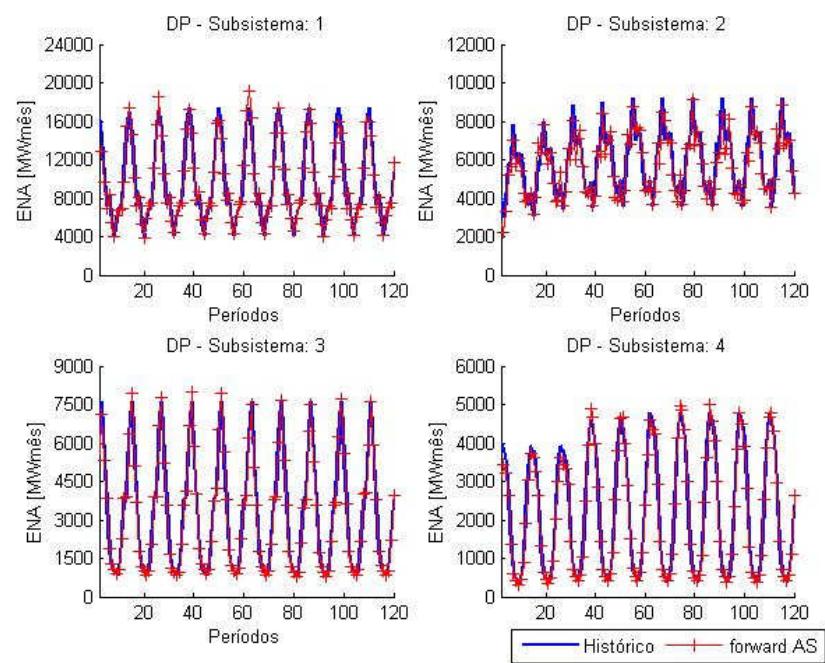


Figura 7.15: desvio-padrão dos cenários – op. AS

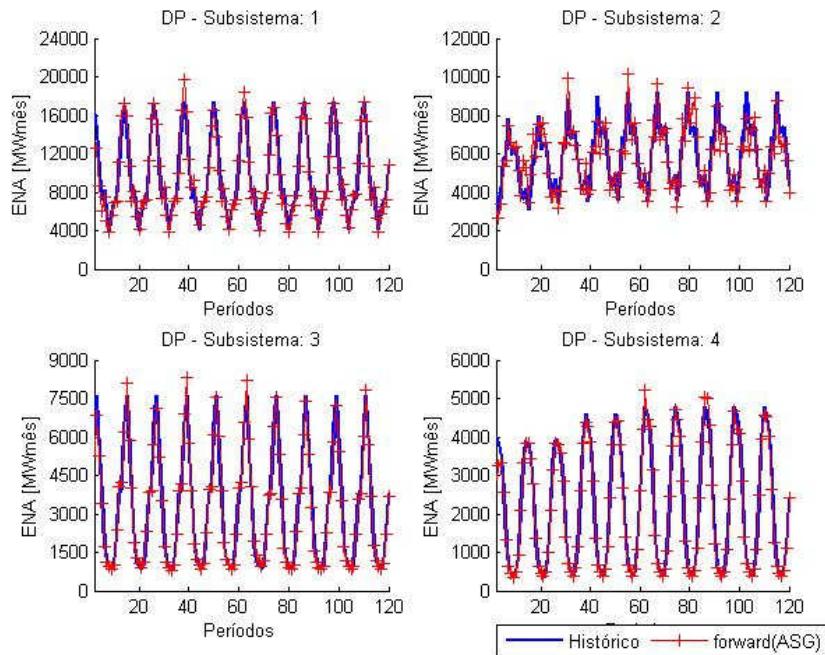


Figura 7.16: desvio-padrão dos cenários – op. ASG

Como pode ser observado, a opção atual apresenta uma degradação relevante do desvio-padrão, em comparação com os métodos AS e ASG. O desvio-padrão é uma estatística de fundamental importância para o planejamento da operação pois ele traduz a possibilidade de ocorrência de valores extremos de afluências nos cenários hidrológicos. Ou seja, preservar esta estatística significa manter a capacidade de representar corretamente períodos secos e úmidos, o que é imprescindível nos estudos deste tipo. Portanto, mais uma vez os métodos provenientes de agregação superaram a opção Atual.

### 7.8.3 Correlação Cruzada dos Cenários

Nas figuras 7.17 a 7.19 são apresentadas as correlações cruzadas do subsistema Sudeste em relação aos outros, considerando os três métodos.

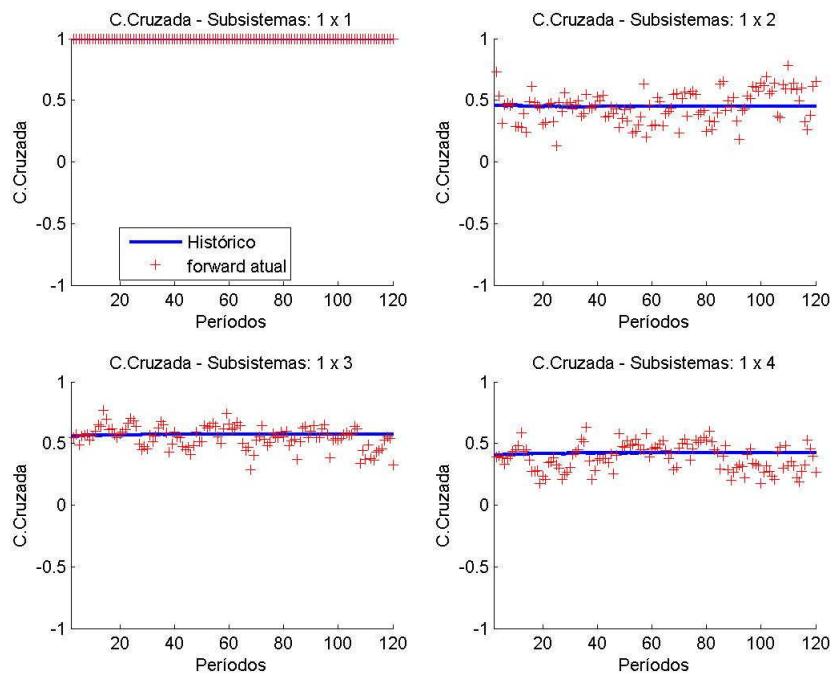


Figura 7.17: correlação cruzada dos cenários -op. Atual

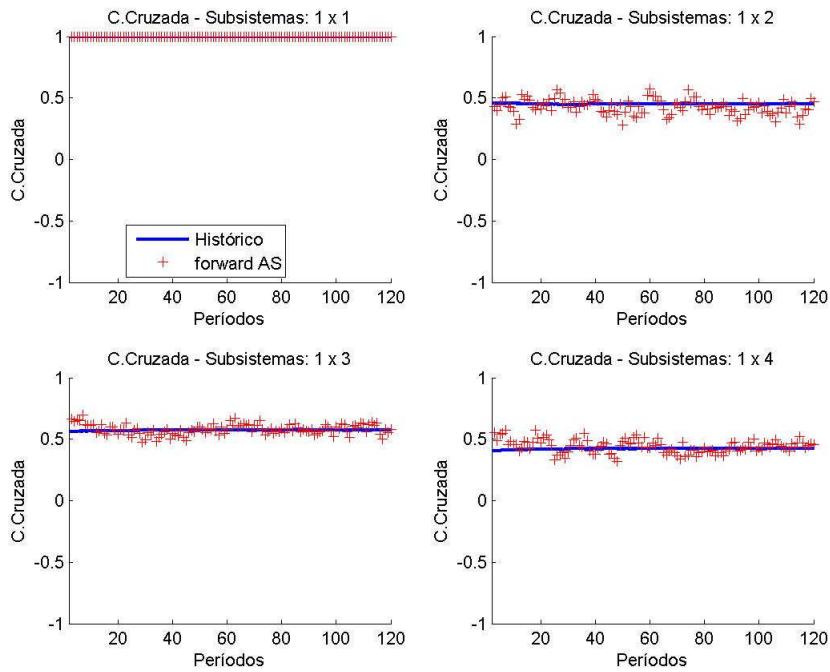


Figura 7.18: correlação cruzada dos cenários - op. AS

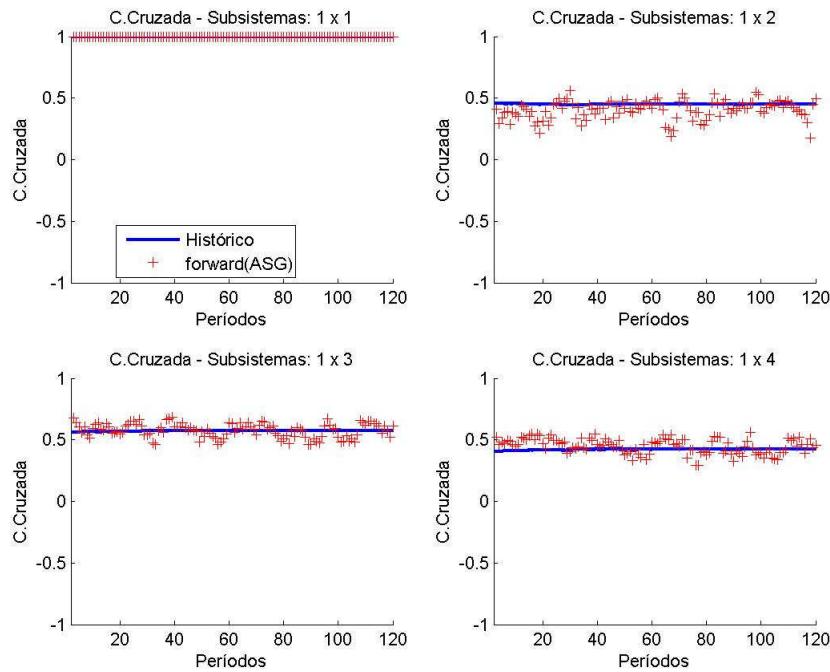


Figura 7.19: correlação cruzada dos cenários - op. ASG

Após analisadas as figuras 7.17 a 7.19 correspondentes à correlação cruzada para os três métodos, conclui-se que a opção Atual perde na representação da dependência linear espacial entre as afluências dos subsistemas considerados, em comparação com a AS e ASG. Este fato pode trazer resultados muito diferentes se comparados com resultados obtidos ao utilizar-se o registro histórico de afluências. No estudo de médio prazo, as energias afluentes aos reservatórios são agrupadas em subsistemas equivalentes e, representar mal a correlação cruzada entre estes, seria o mesmo que alterar as localizações geográficas dos reservatórios.

Por definição, a correlação temporal dos cenários hidrológicos gerados também é preservada, uma vez que os ruídos agregados (que são temporalmente independentes) são aplicados diretamente ao modelo PAR( $p$ ), e este é responsável por preservar a estrutura temporal do processo hidrológico. De outra forma, ao agregar os ruídos a estrutura de correlação temporal dos cenários não é afetada.

## **7.9 Resumo do capítulo**

Foram descritas nesse capítulo as metodologias propostas no tratamento da matriz de energias e ruídos original de forma a aplicar o método de agregação a adotado para atingir os objetivos desejados.

Comparou-se os cenários hidrológicos gerados em cada etapa com os cenários de referência, a fim de verificar a preservação das estatísticas de interesse, média, desvio padrão e correlação cruzada. Em seguida definiu-se um novo tratamento e modelagem do problema de forma a compensar as degradações ocasionadas pela redução da cardinalidade resultante do K-Means.

No próximo capítulo serão expostos e analisados os resultados segundo aspectos de operação do sistema utilizando os cenários oriundos da ASG.

## 8 Análise dos Resultados do Planejamento da Operação

### 8.1 Considerações iniciais

O presente capítulo destina-se à análise dos resultados provenientes da utilização da sub-árvore de afluências construída através do método ASG em casos reais do planejamento da operação energética de médio prazo, confrontando-os com os resultados obtidos através do uso dos métodos Atual e AS.

O horizonte de estudo é composto de 10 anos, o parque gerador possui aproximadamente 150 usinas hidroelétricas e 140 usinas termoelétricas. Como explicado em capítulos anteriores, os reservatórios constituintes do SIN são agregados em reservatórios equivalentes de energia, que constituem o Sudeste/Centro-Oeste, Sul, Nordeste e Norte. É representado na Figura 8.1 a disposição desses subsistemas e seus troncos de linhas de transmissão, que possibilitam o intercâmbio entre eles.

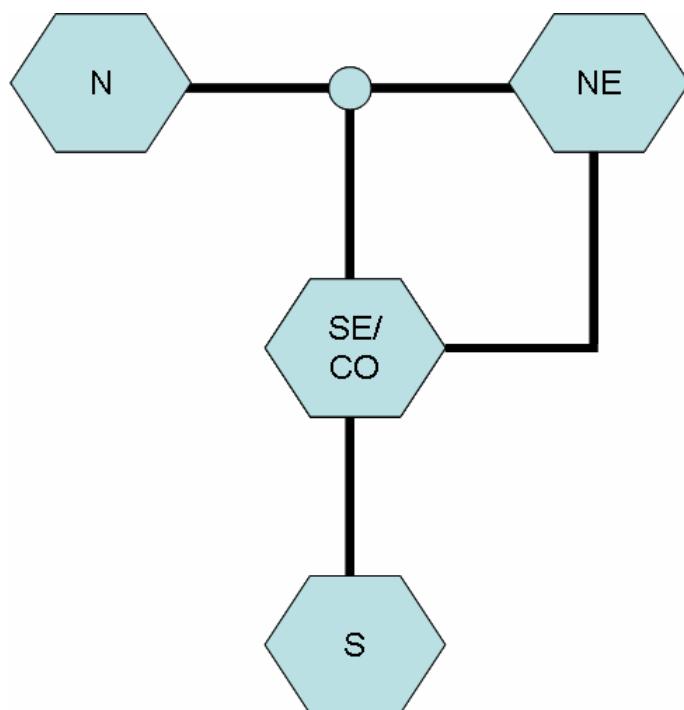


Figura 8.1: Representação subsistemas equivalentes e seus troncos de transmissão

Durante a etapa de cálculo da política de operação são usados 200 cenários para o passo forward, construídos com o método ASG. Os cenários da etapa backward são os mesmos construídos pelo método AS e são consideradas 20 aberturas. Após a definição da política ótima de operação, é realizada uma análise de desempenho dessa política no SIN, considerando 2000 cenários sintéticos gerados por AAS. Nessa etapa, chamada de simulação final, são calculados os índices de desempenho do sistema, como os riscos de déficit, valor esperado do custo total de operação, gerações hidroelétrica e termoelétrica por subsistema, armazenamentos, intercâmbios, vertimentos e etc.

Alguns testes foram feitos a fim de analisar a sensibilidade dos resultados com relação a variações das amostras de cenários forward.

Os estudos de caso foram realizados considerando o PMO referente ao mês de Fevereiro de 2009.

## 8.2 Avaliação da Convergência

O cálculo da política de operação é feito utilizando a PDDE, que através de um processo iterativo, busca aproximar o valor esperado do custo de operação ( $Z_{SUP}$ ) e o valor estimado do custo de operação ( $Z_{INF}$ ), que são calculados a cada iteração. O valor de  $Z_{INF}$  é obtido consultando a função de custo futuro, que é construída da etapa backward, e o valor de  $Z_{SUP}$  é proveniente da execução da etapa forward do processo de otimização.

Teoricamente, a convergência ocorreria quando os valores calculados para estes estimadores fossem iguais. Como o problema de planejamento da operação é um essencialmente estocástico, devido às incertezas relativas às afluências futuras, utiliza-se um critério de convergência estatístico. Este critério consiste na determinação de um intervalo de confiança para o valor esperado do custo total de operação ( $Z_{SUP}$ ). Os limites inferior e superior deste intervalo de confiança são usualmente denominados  $L_{INF}$  e  $L_{SUP}$ , respectivamente. Em outras palavras, quando o valor de  $Z_{INF}$  estiver ‘dentro’ desse intervalo de confiança, o algoritmo é dito convergido. Atualmente o intervalo de confiança é de 95%, como ilustrado na Figura 8.2.

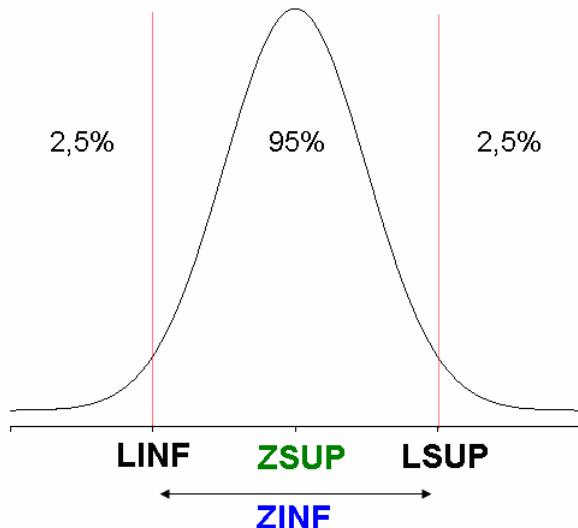


Figura 8.2: Critério de convergência estatístico

### 8.2.1 Sensibilidade com relação às variações dos cenários hidrológicos

Nesta seção são utilizadas 10 amostras diferentes de cenários no estudo de caso do PMO de Fevereiro de 2009, com o intuito de analisar a sensibilidade dos resultados frente às variações destes cenários.

Na Figura 8.3 são apresentadas as evoluções dos valores de **ZINF** e **ZSUP** ao longo do processo iterativo do cálculo da política ótima de operação. Em cada gráfico são apresentadas as trajetórias das referidas variáveis para as diferentes amostras consideradas nesta análise.

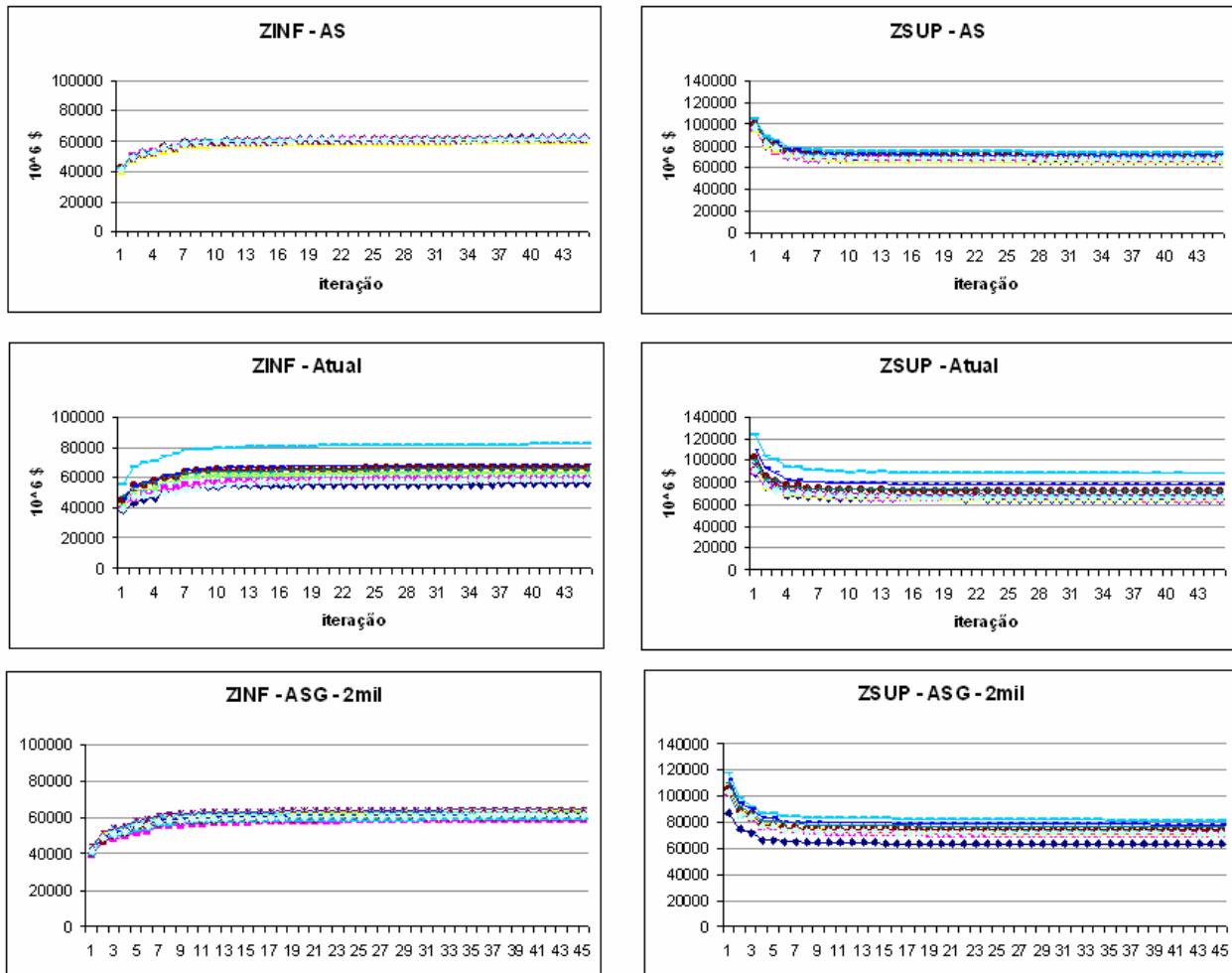


Figura 8.3: Variabilidade de ZINF e ZSUP com variações de amostras de cenários

Analizando a Figura 8.3, observa-se que o métodos AS e ASG reduzem bastante a variabilidade de ZINF com relação às alterações nos cenários hidrológicos. Vale lembrar que nas opções ASG e AS os cenários que compõem a sub-árvore da etapa backward foram construídos utilizando técnicas de agregação. Apesar dos cenários da etapa backward não terem sido alterados pelo método ASG (em comparação ao método AS), nota-se uma variabilidade no valor de ZINF ligeiramente maior se comparada com o método AS. Isto ocorre, pois o valor de ZINF também é afetado pela representação dos cenários da etapa forward (foco do trabalho).

A variabilidade de ZSUP também reduziu, em relação ao método Atual, com a aplicação do método ASG, mas não tanto quanto ZINF. Repare que na AS essa variabilidade também já tinha sido reduzida com relação ao método Atual. Com o intuito de reduzir ainda mais a variabilidade de ZSUP, foi aumentada a cardinalidade da pseudo-árvore completa,

que é dado de entrada na aplicação da ASG para construção da sub-árvore de cardinalidade inferior. O resultado é mostrado na Figura 8.4.

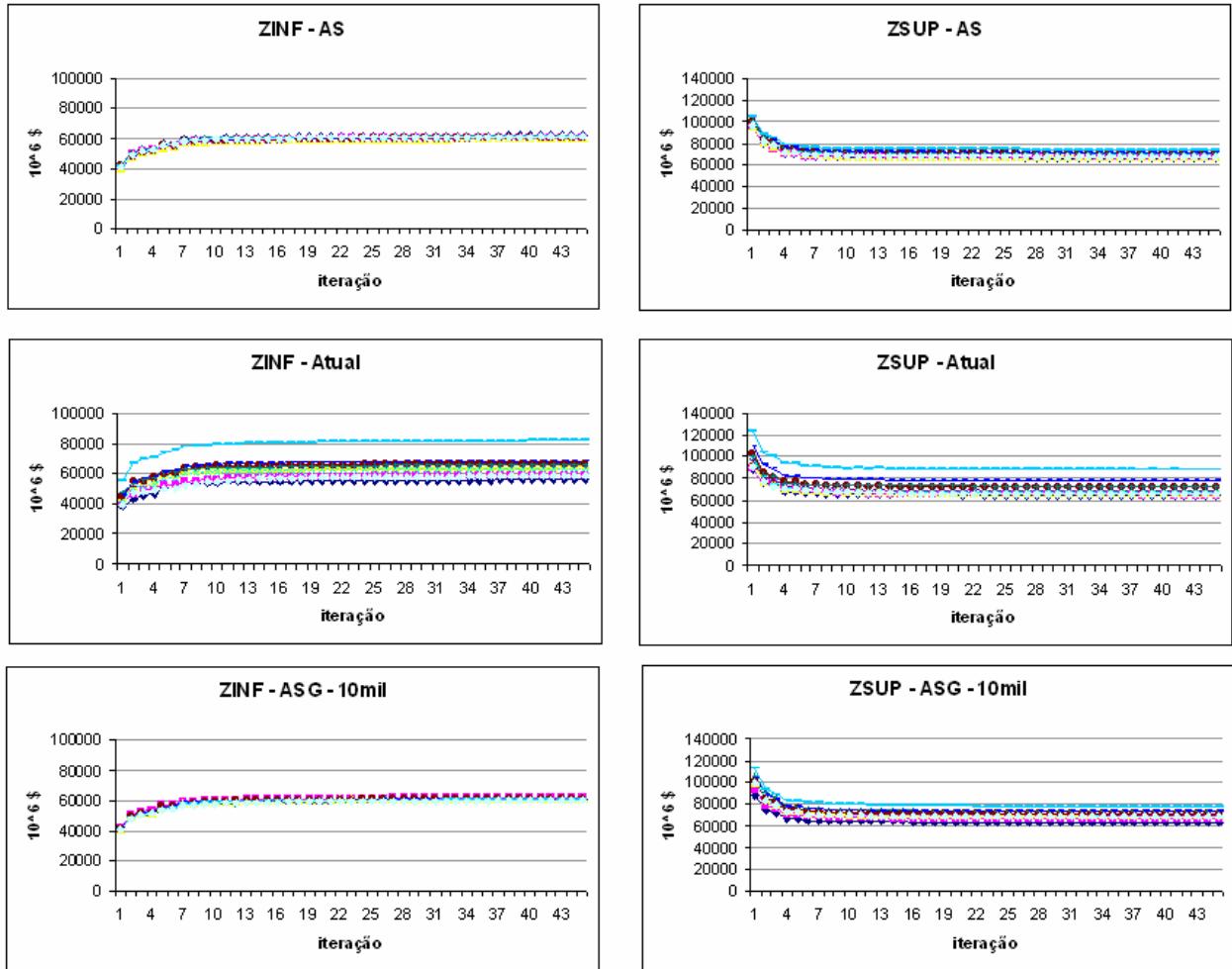


Figura 8.4: Variabilidade de ZINF e ZSUP com relação às variações de amostras de cenários

Conclui-se, baseado na Figura 8.4, que o aumento da cardinalidade da pseudo-árvore completa trouxe mais robustez aos valores de ZSUP e ZINF calculados pelo método ASG. A opção AS apresenta uma maior redução da variabilidade de ZSUP, uma possível explicação para isto é o fator desse método ser aplicado em um espaço amostral dez vezes maior do que o espaço amostral do método ASG. Ao tentar aumentar ainda mais a cardinalidade da pseudo-árvore utilizada no método proposto para construção da sub-árvore, esbarrou-se num problema de alocação de memória. Trabalhos futuros podem ser feitos buscando contornar esse problema e permitir a elevação da cardinalidade da pseudo-árvore completa, ou ainda, buscando representar seu horizonte de forma simplificada, contornando o inconveniente da alocação de memória.

Nas Figuras 8.5 e 8.6 são mostradas as estimativas da média e desvio-padrão das estimativas de ZINF e ZSUP, utilizando as diferentes amostras de cenários.

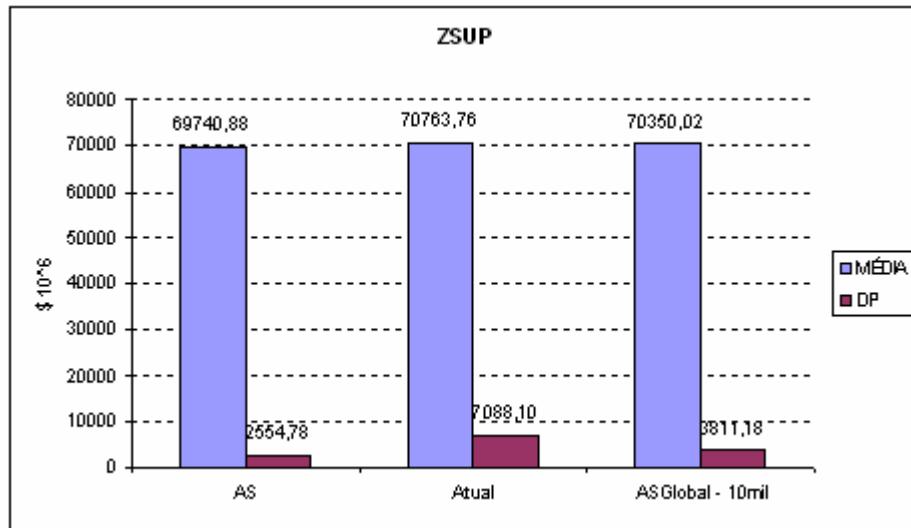


Figura 8.5: Média e desvio-padrão ZSUP

Na Figura 8.5 constata-se que a redução do valor médio de ZSUP foi pequena, porém o desvio-padrão teve uma redução muito significante, comparando os métodos AS e ASG com o Atual. Essa redução do desvio-padrão traduz o ganho de robustez frente às variações de amostras de cenários, ou seja, os valores de ZSUP calculados pela política ótima de operação estão mais próximos.

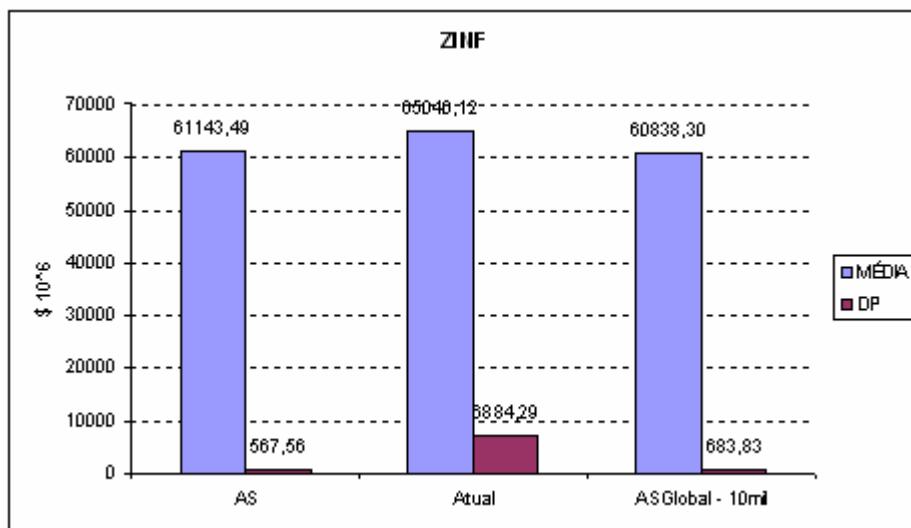


Figura 8.6: Média e desvio-padrão de ZINF

Analizando o gráfico da Figura 8.6, nota-se que, além da redução da média, tem-se uma queda notável no valor do desvio-padrão de ZINF, o que afirma ainda mais a aplicação das técnicas de agregação, representadas pelos métodos AS e ASG, frente à AAS, representada pelo método Atual.

### 8.2.2 Avaliação da convergência de um caso estudo de PMO

Neste item será avaliada a convergência do PMO de Fevereiro de 2009 utilizando os três métodos analisados. Serão considerados, além da convergência estatística, a estabilidade dos valores de ZINF e ZSUP ao longo das iterações. Nas Figuras 8.7, 8.8 e 8.9 estão ilustrados o intervalo de confiança e a evolução dos valores de ZINF, ZSUP, LINF e LSUP para os métodos Atual, AS e ASG, ao longo das 45 iterações do processo iterativo de cálculo da política ótima de operação. Nos gráficos abaixo, a iteração na qual o critério de convergência estatístico é atingido está circulada.

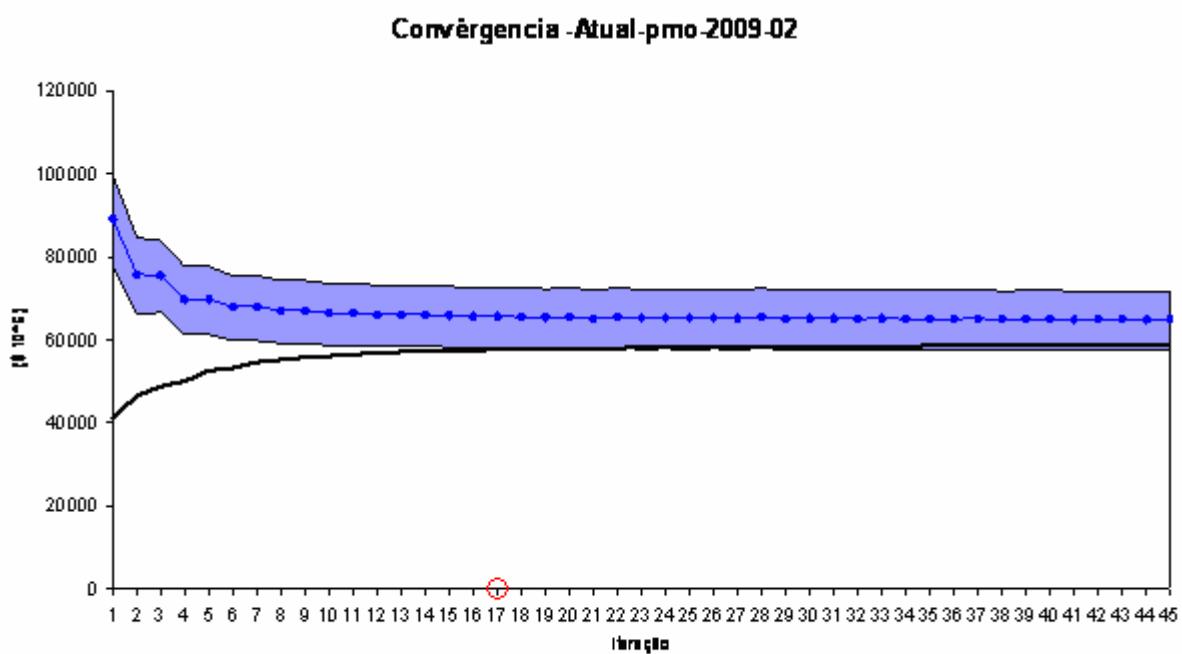


Figura 8.7: Convergência método Atual

**Convergência -AS-pmo-2009-02**

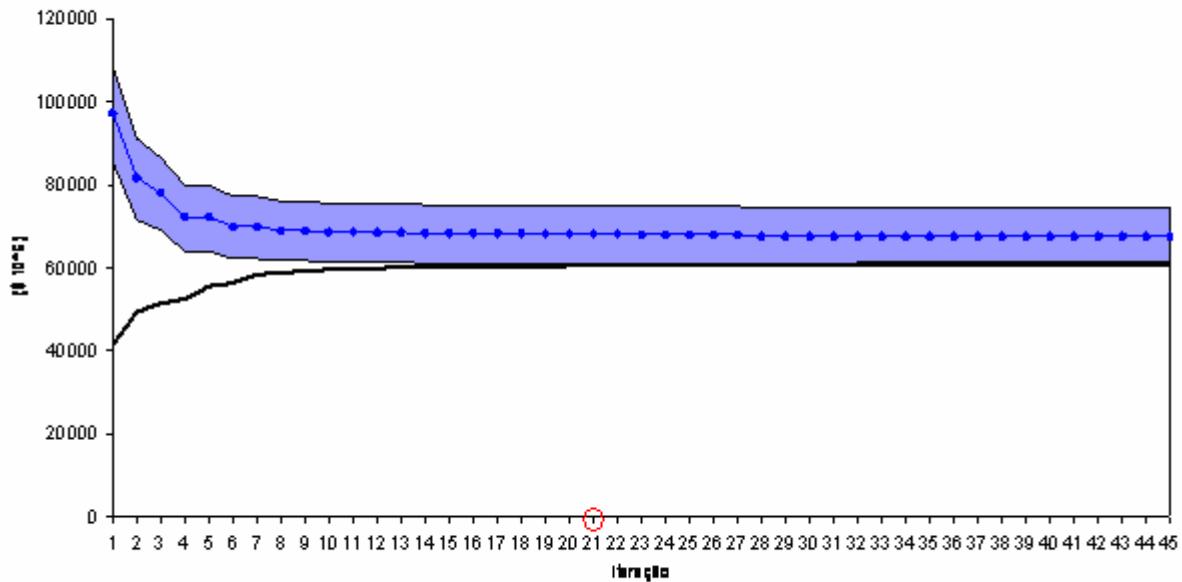


Figura 8.8: Convergência método AS

**Convergência -ASG-pmo-2009-02**

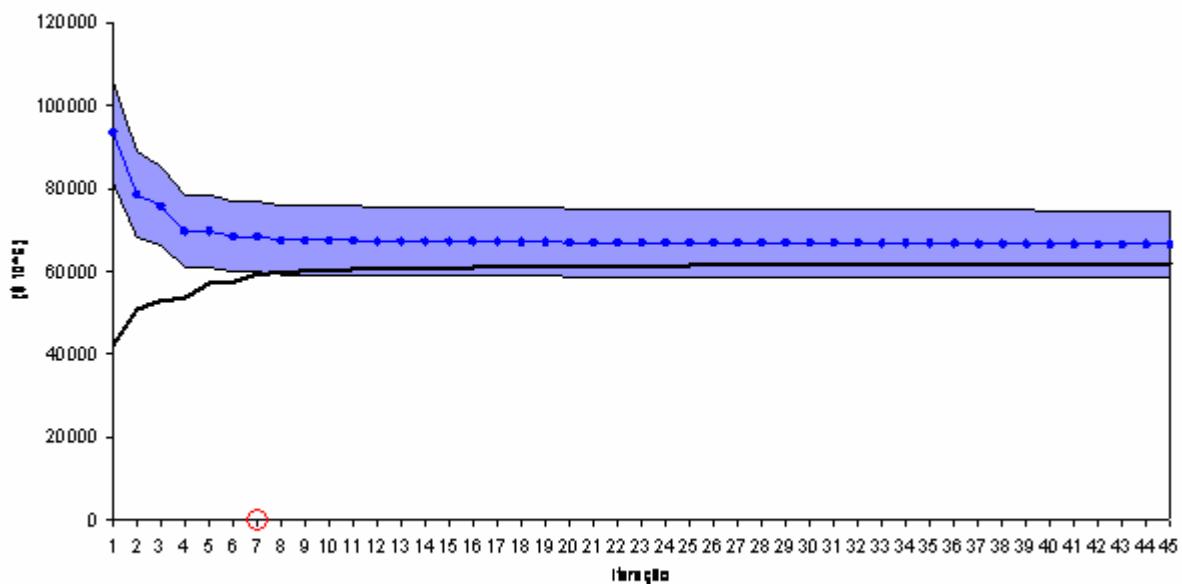


Figura 8.9: Convergência método ASG

Como pode ser constatado nas Figuras 8.7, 8.8 e 8.9, os três métodos convergem estatisticamente antes de atingir o número máximo de iterações permitido pelo

programa. Outra característica importante que pode ser notada no processo de convergência é a estabilidade de ZINF e ZSUP. Essa característica denota um ganho na qualidade da convergência. Nas Figuras 8.10 a 8.15 pode-se observar a estabilidade através da evolução da diferença percentual de ZINF e da diferença percentual de ZSUP ao longo das 45 iterações.

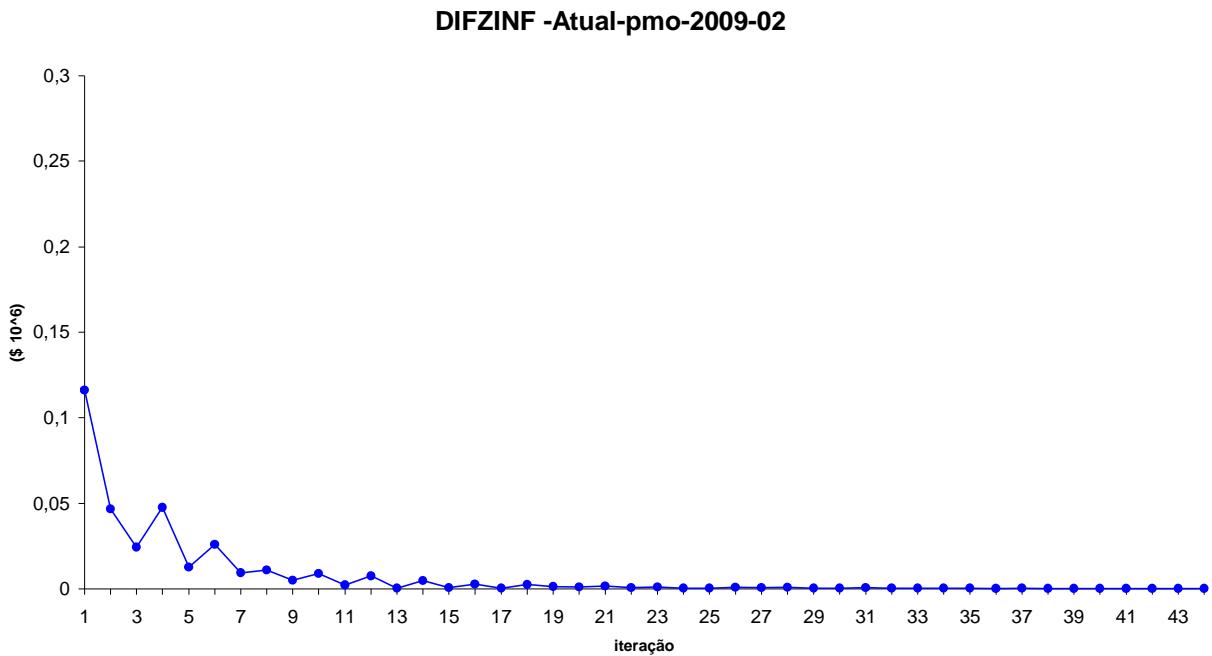


Figura 8.10: Diferença de ZINF método Atual

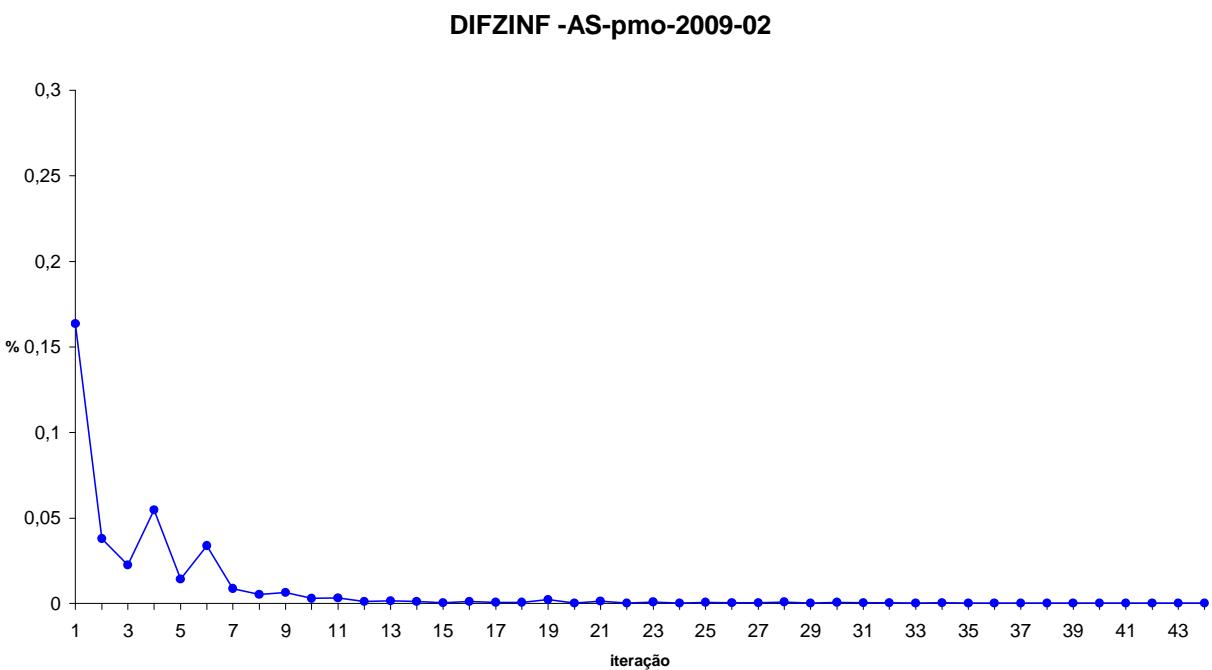


Figura 8.11: Diferença de ZINF método AS

### **DIFZINF -ASG-pmo-2009-02**

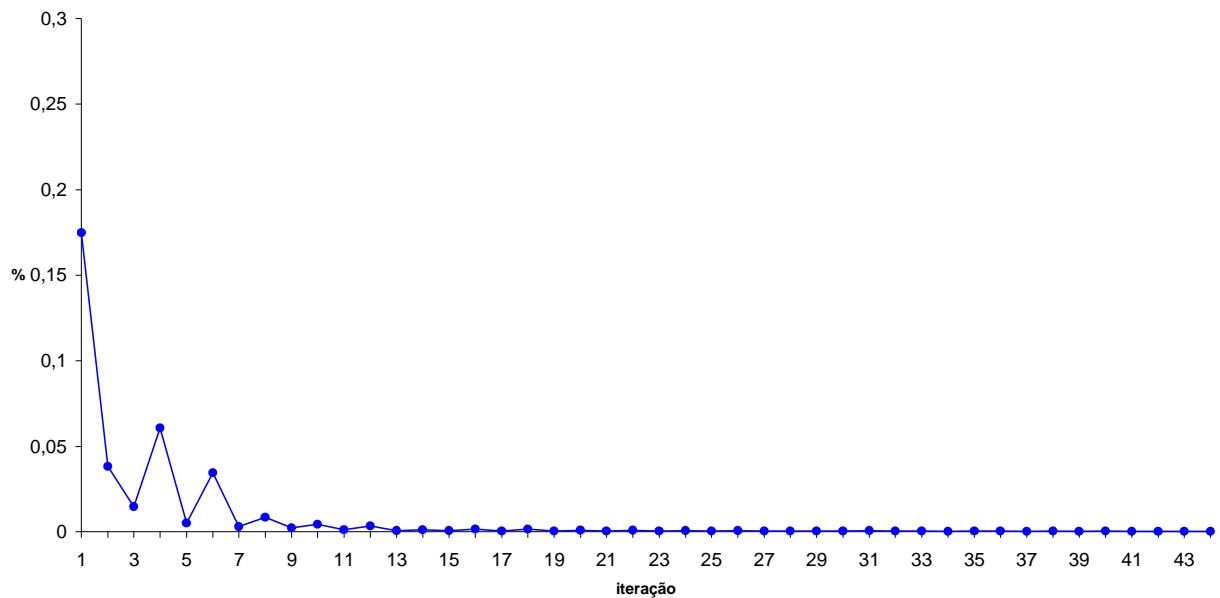


Figura 8.12: Diferença de ZINF método ASG

### **DIFZSUP-Atual-pmo-2009-02**

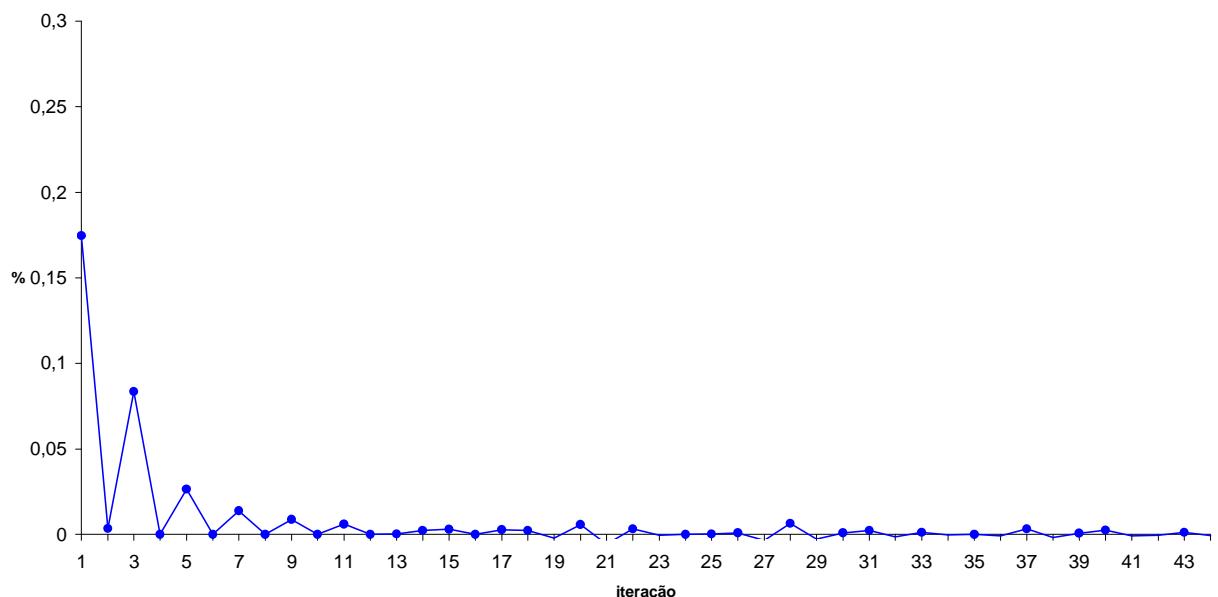


Figura 8.13: Diferença de ZSUP - Atual

### DIFZSUP-AS-pmo-2009-02

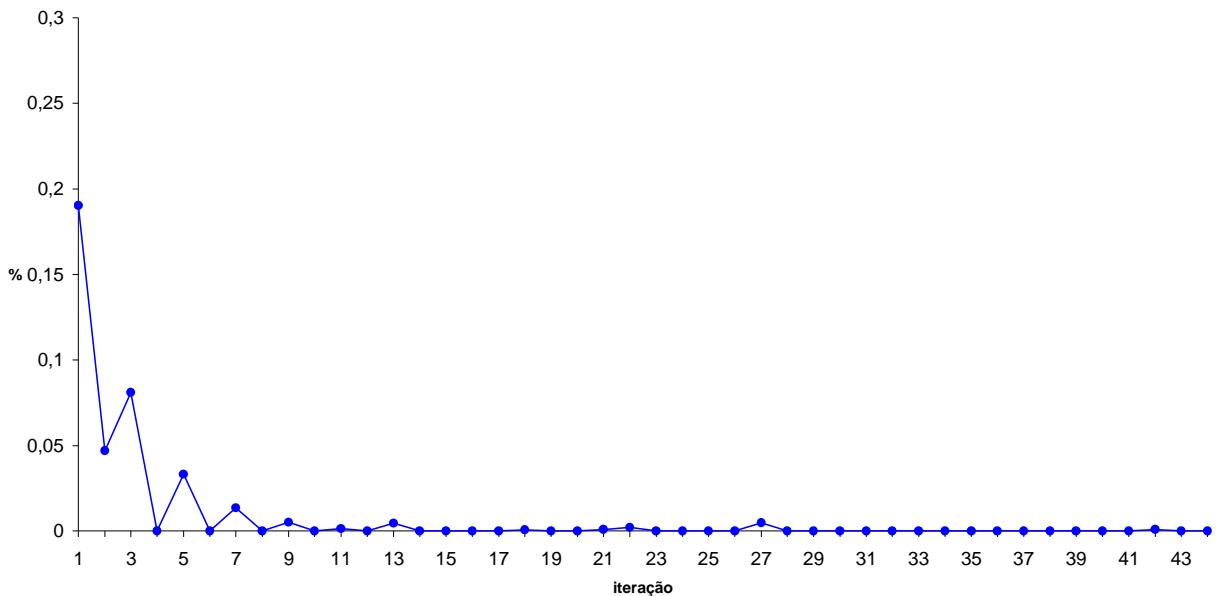


Figura 8.14: Diferença de ZSUP método AS

### DIFZSUP-ASG-pmo-2009-02

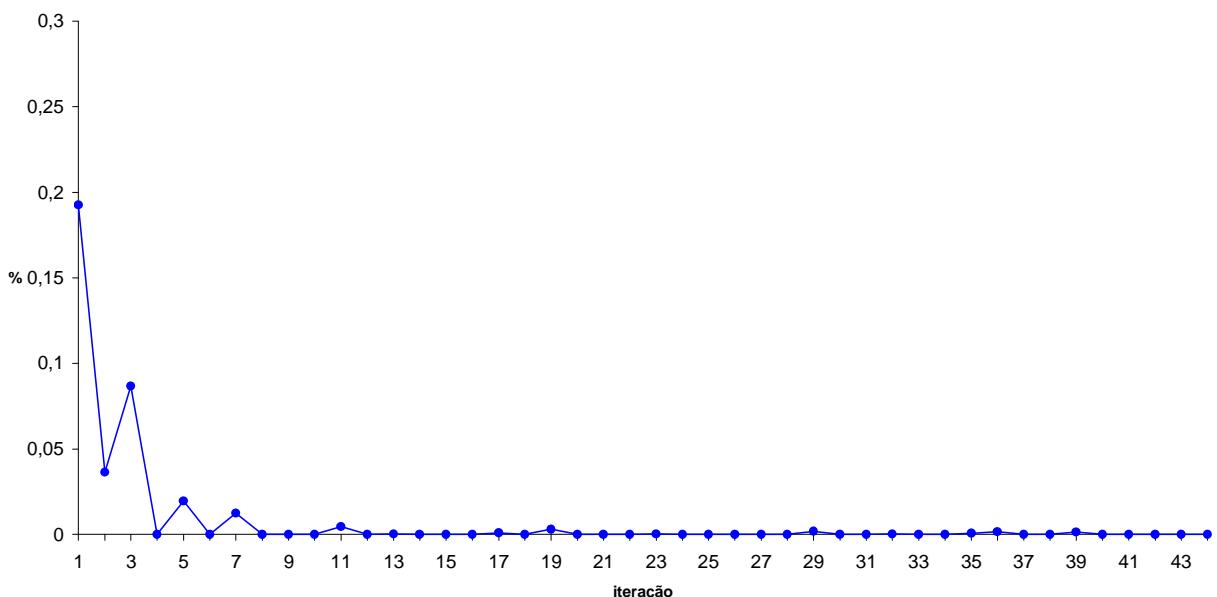


Figura 8.15: Diferença de ZSUP- ASG

Analizando as Figuras 8.10 a 8.15 pode-se observar as características de estabilidade do processo de convergência. Constata-se que nos métodos ASG e AS, que utilizam técnicas de agregação, ocorreu uma relevante redução na oscilação da diferença percentual de ZSUP

e, no método Atual, que utiliza AAS, ZSUP permanece flutuando até a última iteração. Se fosse adotado um critério de estabilidade em conjunto com o critério estatístico para a convergência, a opção Atual encontraria maior dificuldade para atendê-lo. O comportamento da diferença percentual de ZINF foi parecido, apesar do método Atual demorar um pouco mais para atenuar as oscilações.

Quando se realiza o planejamento da operação energética de um sistema hidrotérmico, deseja-se atender à demanda completamente, sem falhas, ou pelo menos, com falhas previstas, administradas e com probabilidade conhecida, ou seja, controlada. A seguir são analisados alguns índices obtidos durante as iterações do processo de convergência. Nas figura 8.16 e 8.17 é mostrada a evolução do risco de não atendimento à demanda e a energia não suprida (EENS) ao longo do processo de convergência.

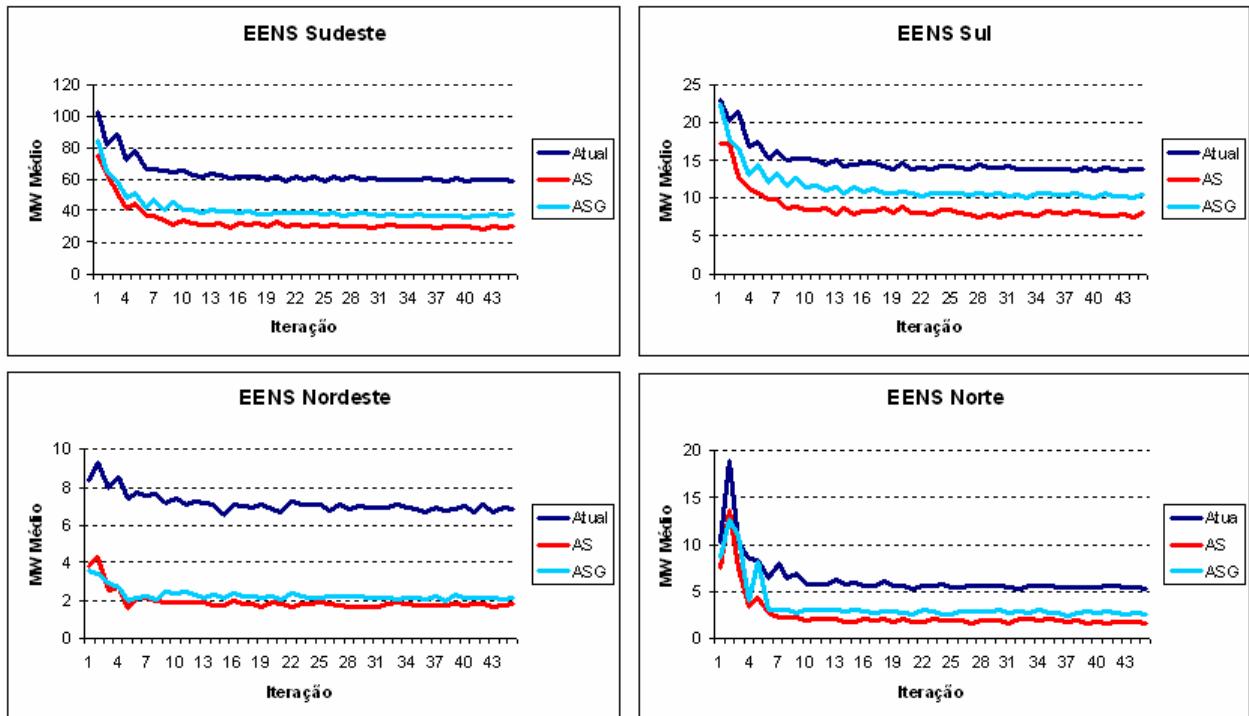
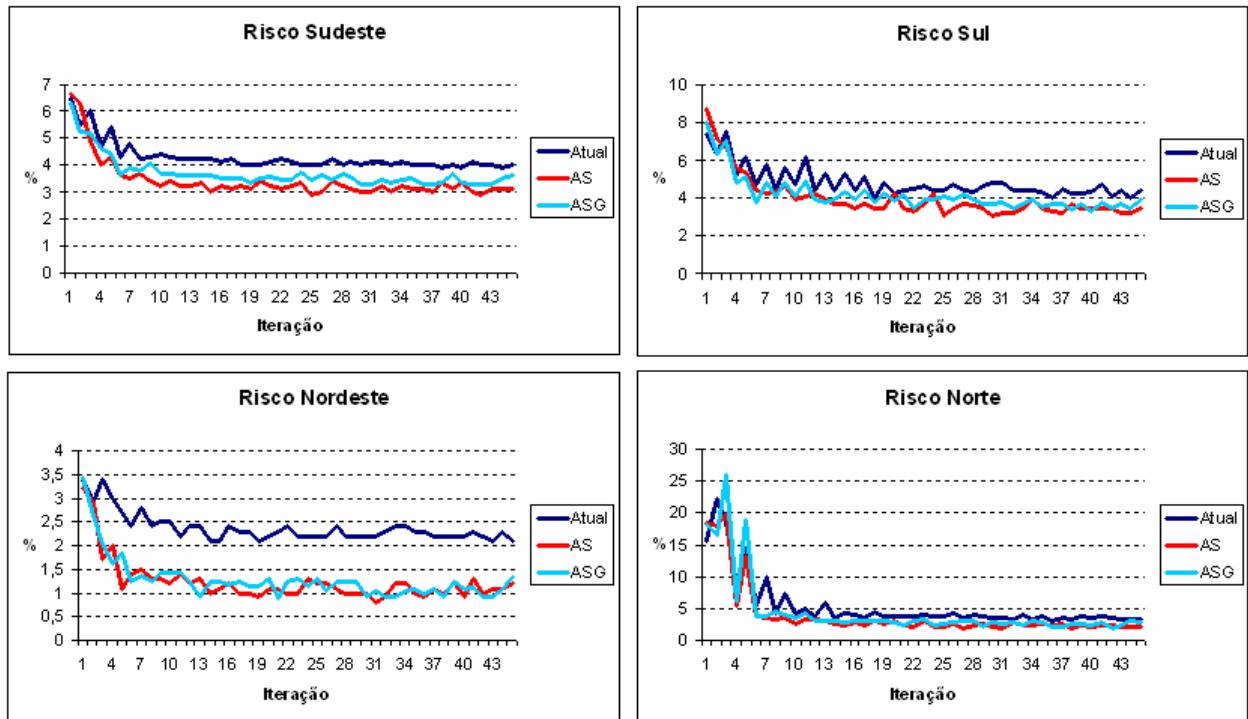


Figura 8.16: Evolução da energia não suprida.



Para o caso analisado, observa-se que os métodos que utilizam agregação possuem resultados que denotam maior capacidade de atendimento à demanda, com risco inferior ao conseguido com o método Atual.

### 8.3 Avaliação dos índices de desempenho utilizando a política ótima calculada

Nesta seção, avalia-se a robustez do custo marginal de operação (CMO) e do custo total de operação (COPER) frente às diferentes amostras escolhidas de cenários hidrológicos.

São construídos gráficos de evolução temporal para o CMO ao longo do horizonte de planejamento para duas amostras, e para cada política ótima calculada segundo cenários hidrológicos obtidos por cada método.

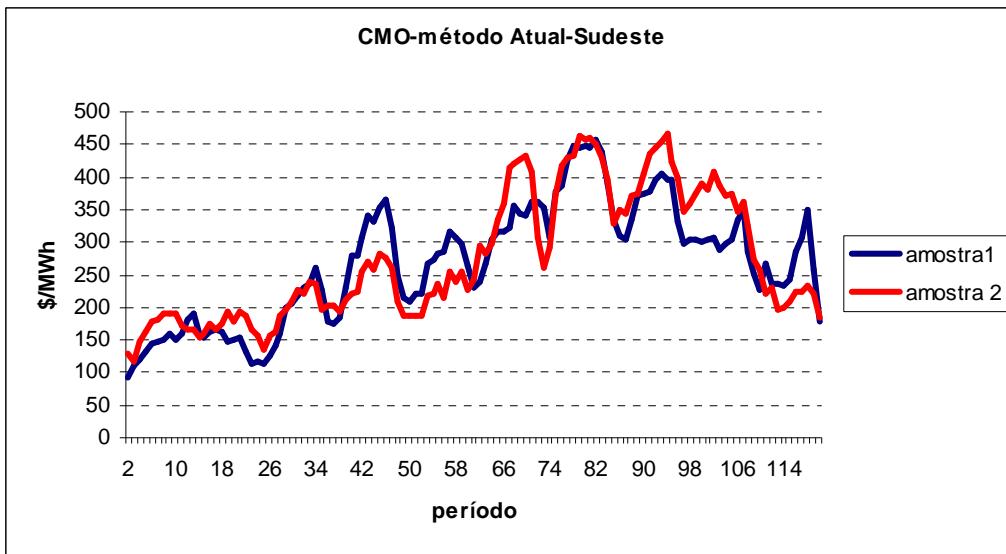


Figura 8.17: Variação do CMO médio relacionado à escolha do cenário hidrológico – Método Atual.

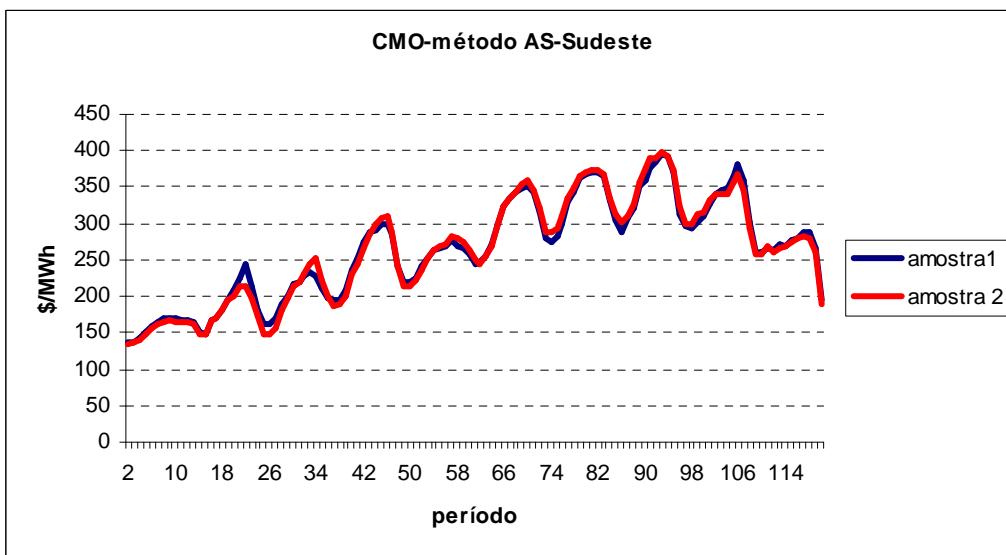


Figura 8.18 : Variação do CMO médio relacionado à escolha do cenário hidrológico – Método AS.

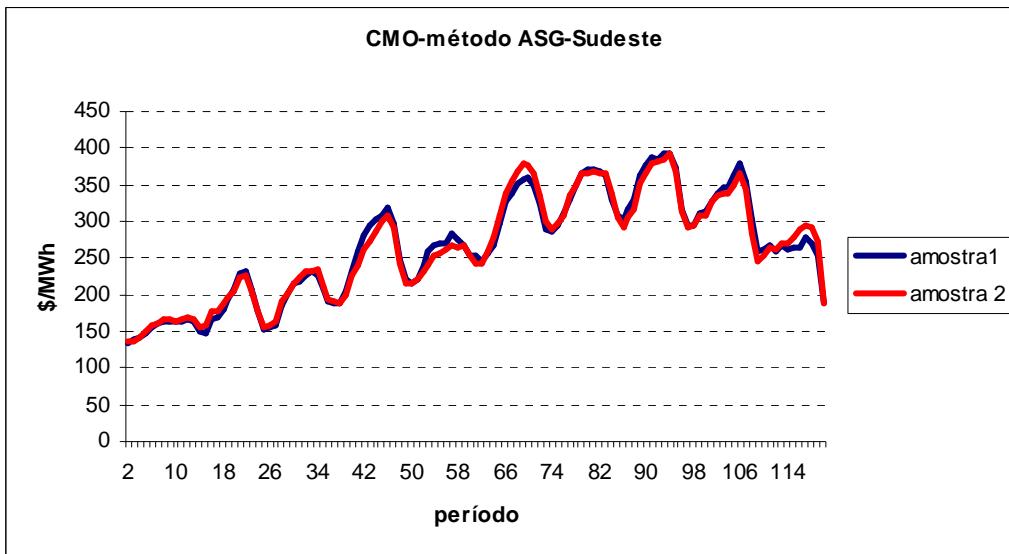


Figura 8.19: Variação do CMO médio relacionado à escolha do cenário hidrológico - Método ASG.

Os valores de CMO praticamente não variam com a amostra de cenários escolhida para o método AS e ASG, conforme notado nas figuras 8.19 e 8.20. Já o método Atual apresenta variabilidade elevada, encontrando custos muito diferentes de acordo com a amostra de cenários escolhida.

O valor esperado do custo total de operação é contemplado na Figura 8.21.

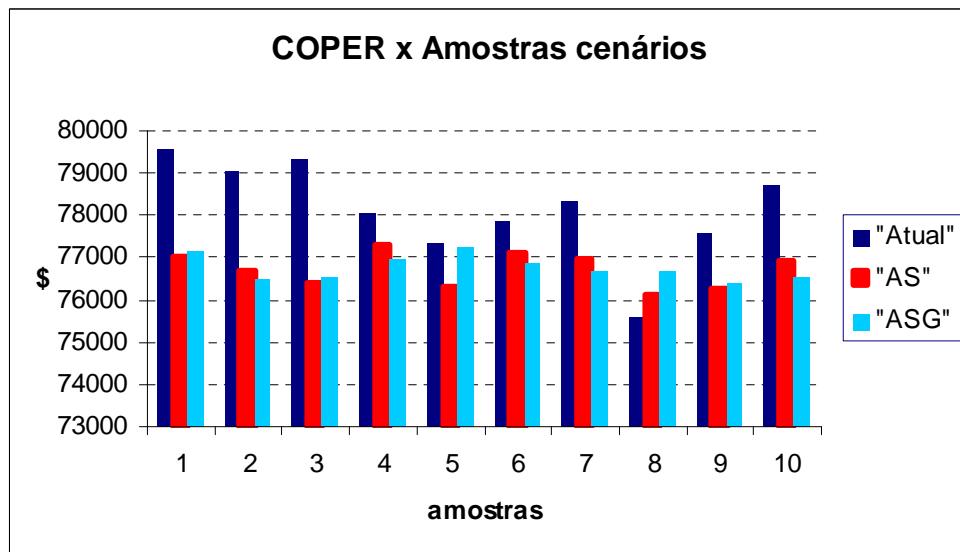


Figura 8.20: Variação do custo total de operação com relação à amostra de cenário hidrológico.

Pode-se notar que o método Atual, na grande maioria das vezes encontra um custo mais elevado e, fortemente influenciado pela amostra de cenários hidrológicos considerada. Já os métodos ASG e AS apresentam custo reduzido (exceto amostra 8), em alguns casos, a metade do custo obtido com o método Atual, mas com variação mais atenuada com relação às mudanças nas amostras de séries hidrológicas. O custo mais elevado observado nas simulações com o método Atual pode ser explicado pela maior possibilidade de geração de cenários hidrológicos críticos.

Foram analisados gráficos da EARMF e CMO para os subsistemas equivalentes que constituem o SIN. Esses índices são calculados durante a etapa de simulação final. Os resultados provenientes dos três métodos (Atual, AS e ASG) para todas as amostras de cenários podem ser observados nas figuras que seguem.

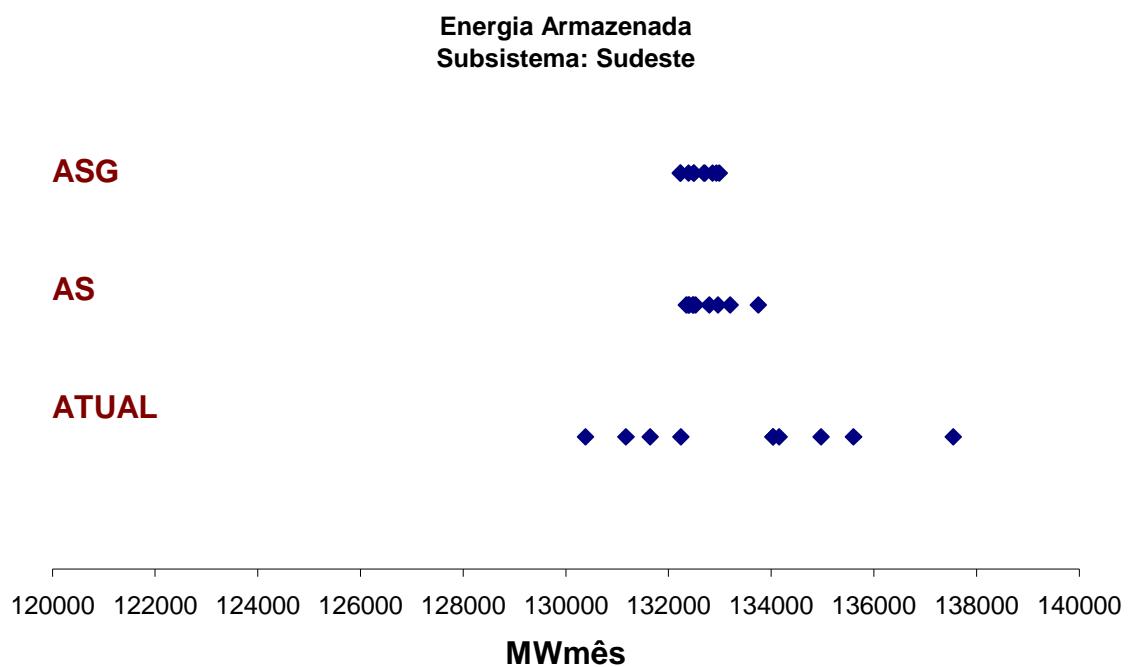


Figura 8.21: Variação EARMF do Sudeste.

**Custo Marginal de Operação**  
**Subsistema: Sudeste**

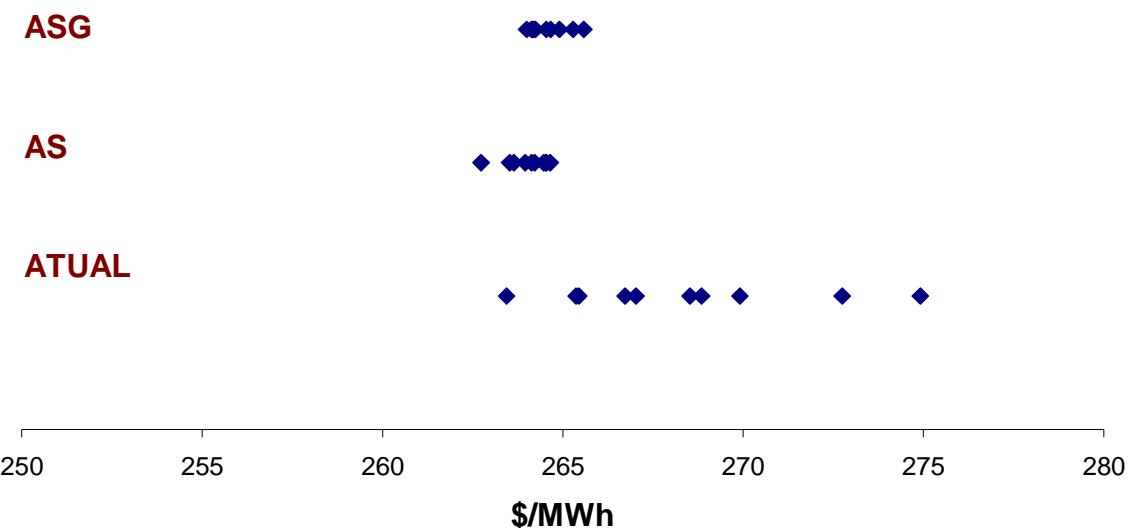


Figura 8.22: Variação CMO do Sudeste.

**Energia Armazenada**  
**Subsistema: Sul**

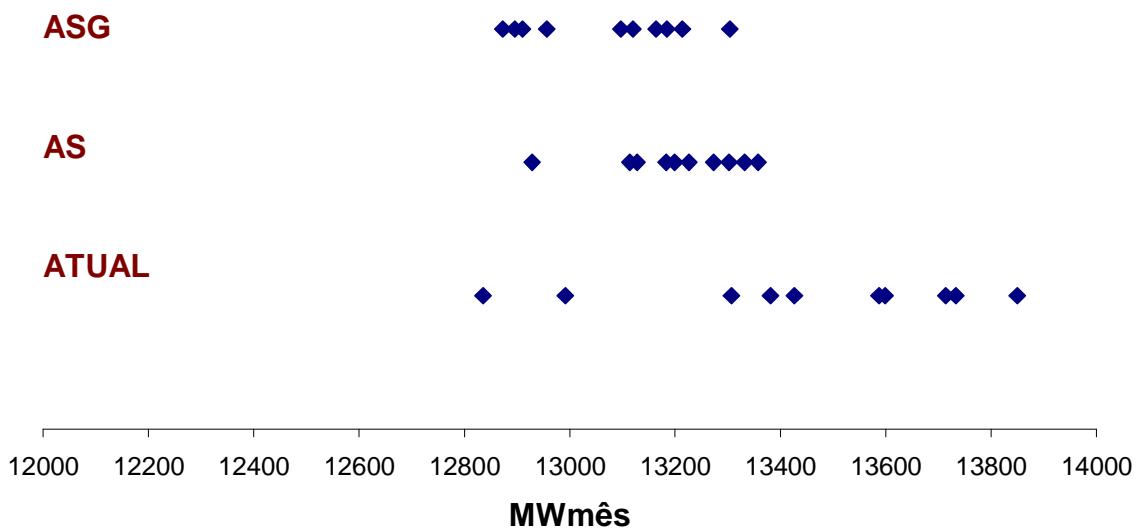


Figura 8.23: Variação EARMF do Sul.

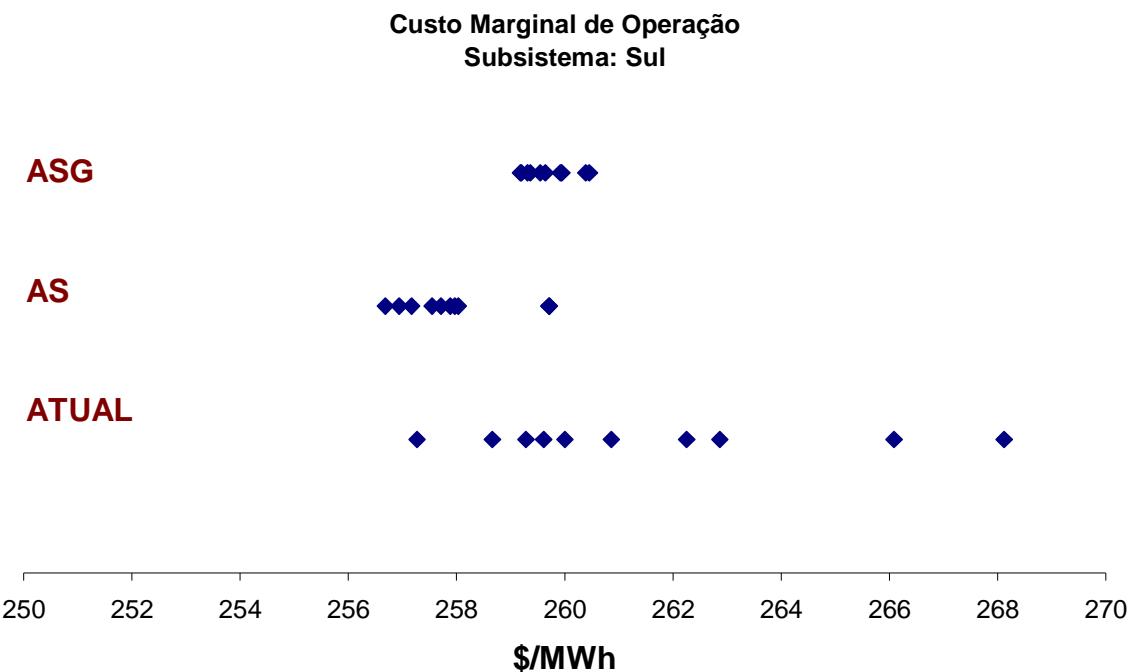


Figura 8.24: Variação CMO do Sul.

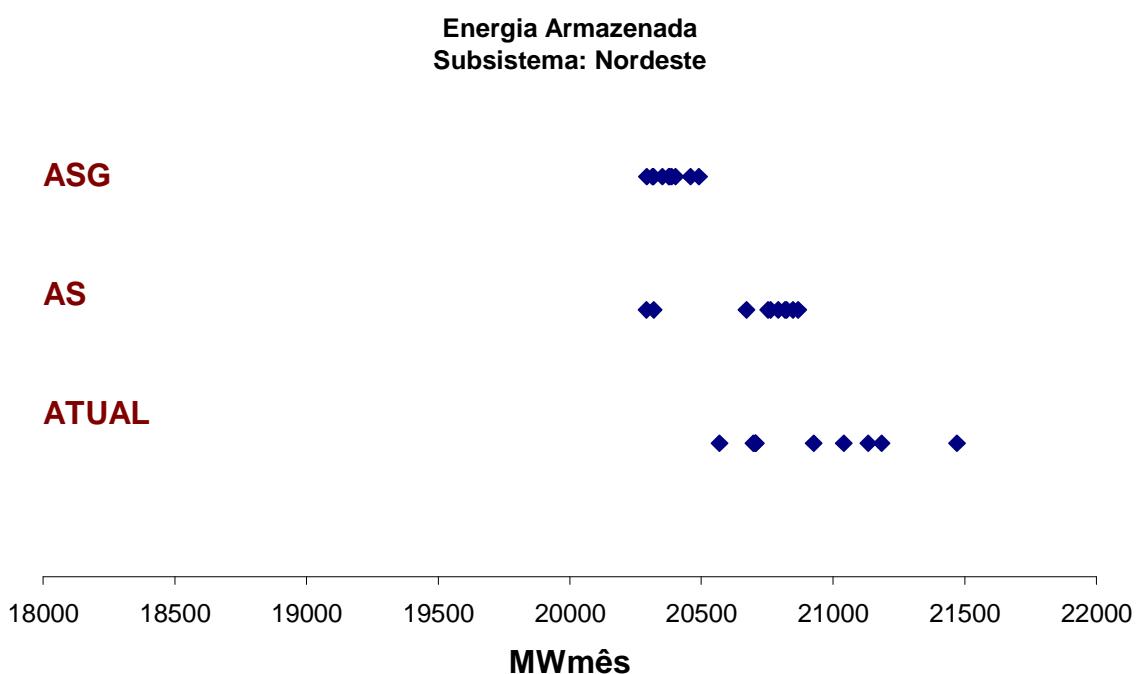


Figura 8.25: Variação EARMF do Nordeste.

**Custo Marginal de Operação**  
**Subsistema: Nordeste**

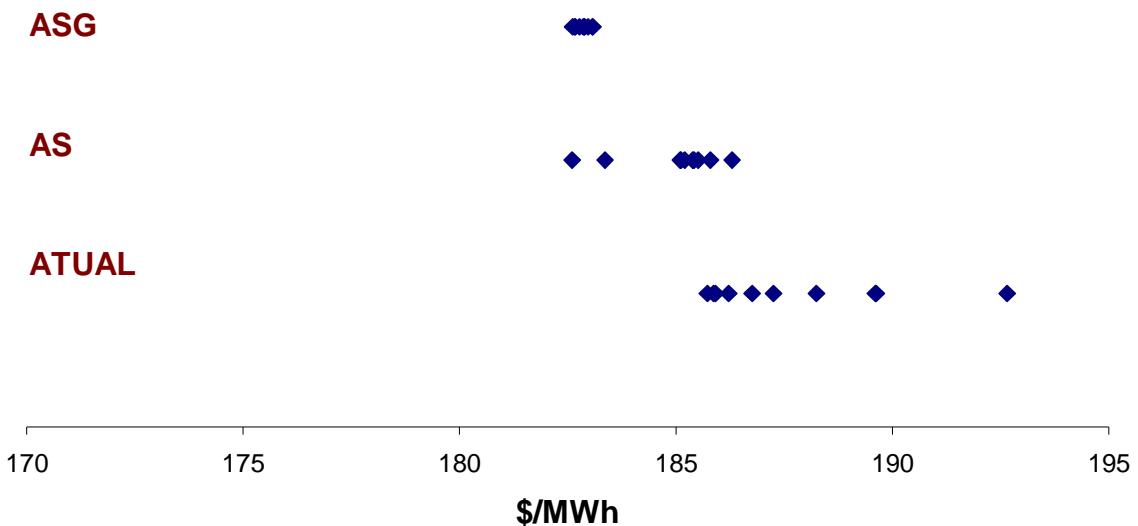


Figura 8.26: Variação CMO do Nordeste.

**Energia Armazenada**  
**Subsistema: Norte**

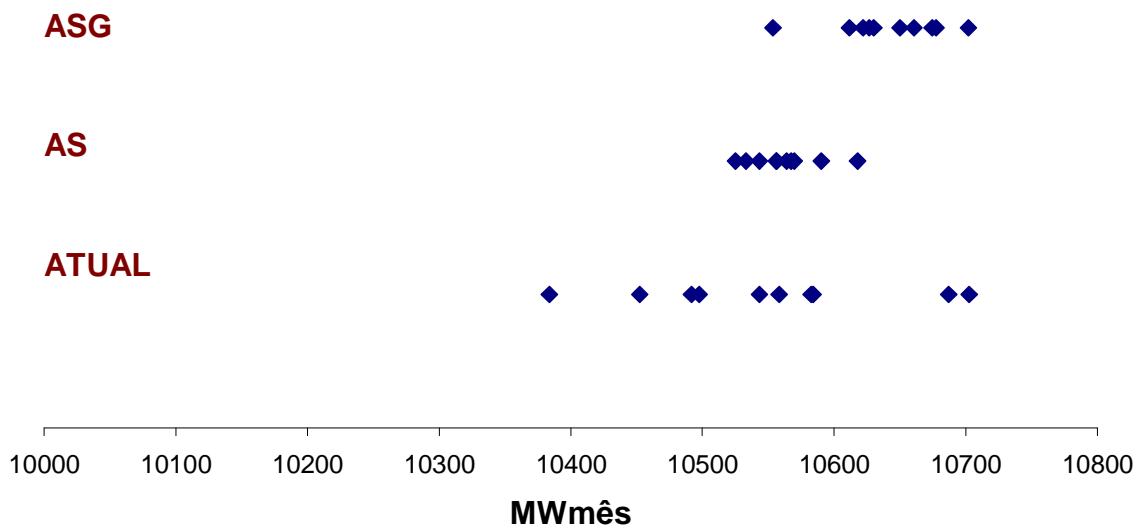
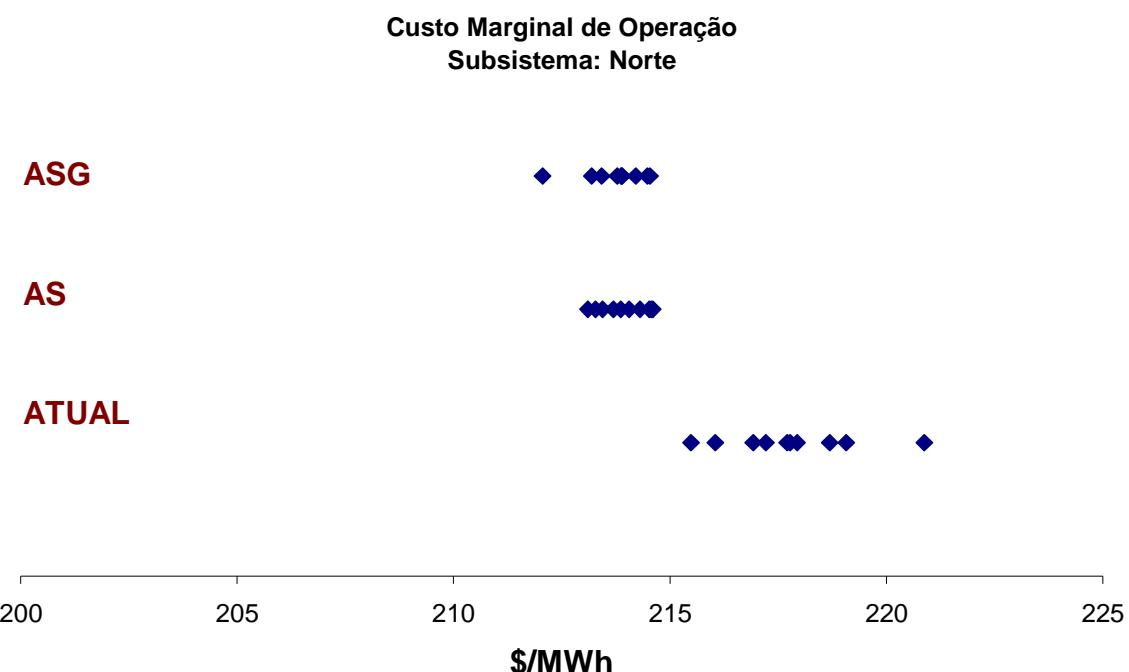


Figura 8.27: Variação EARMF do Norte.



8.28: Variação CMO do Norte.

Analizando as Figuras 8.22 a 8.30, pode ser observada a redução drástica na variabilidade dos resultados de CMO e EARMF calculados pelos métodos ASG e AS para todos os subsistemas. Ou seja, os resultados estão mais estáveis e robustos frente às variações de amostras de cenários hidrológicos. O método Atual, por outro lado, apresenta resultados muito diferentes, dependendo fortemente da amostra de cenários escolhida. A estabilidade desses resultados é uma característica desejável na operação energética.

Buscando analisar a real significância da variabilidade dos índices em comparação aos seus valores médios, foram construídos gráficos do coeficiente de variação para os seguintes índices de desempenho: risco de não atendimento à carga própria (déficit), custo marginal de operação (CMO), energia armazenada no final do período (EARMF), vertimento, geração termoelétrica (GT) e geração hidroelétrica (GH). Esses índices são calculados durante a etapa de simulação final. Os resultados provenientes dos três métodos (Atual, AS e ASG) para todos os subsistemas podem ser observados nas figuras que seguem.

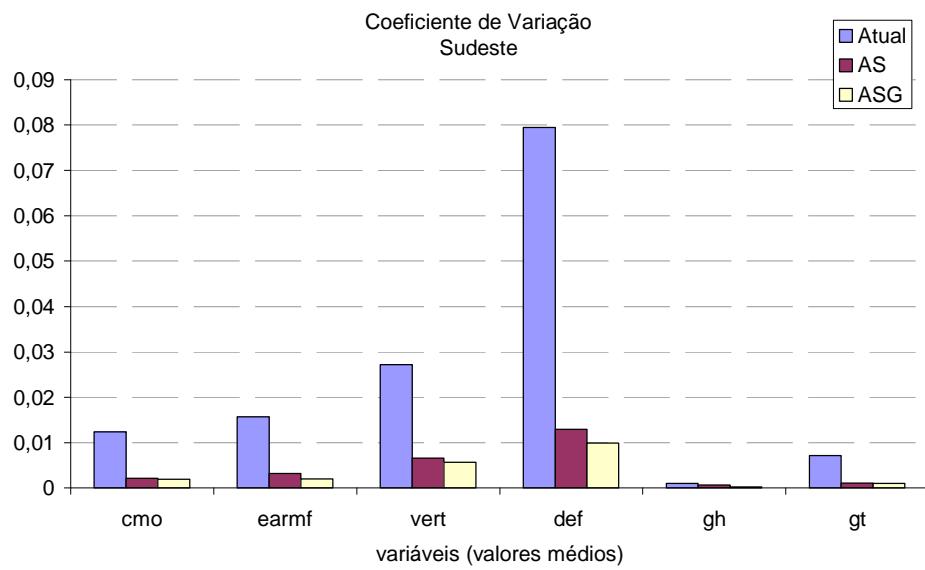


Figura 8.29: Coeficiente de variação do Sudeste.

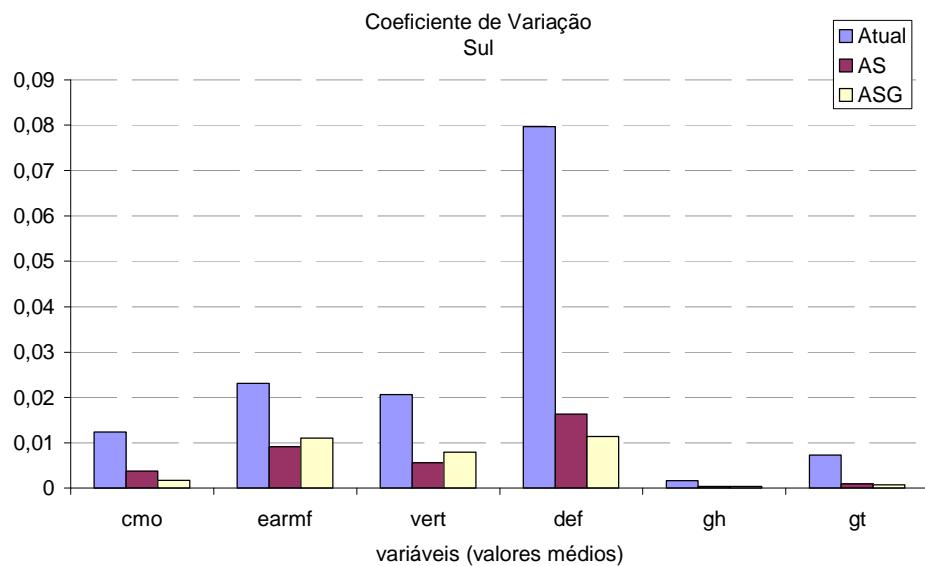


Figura 8.30: Coeficiente de variação do Sul.

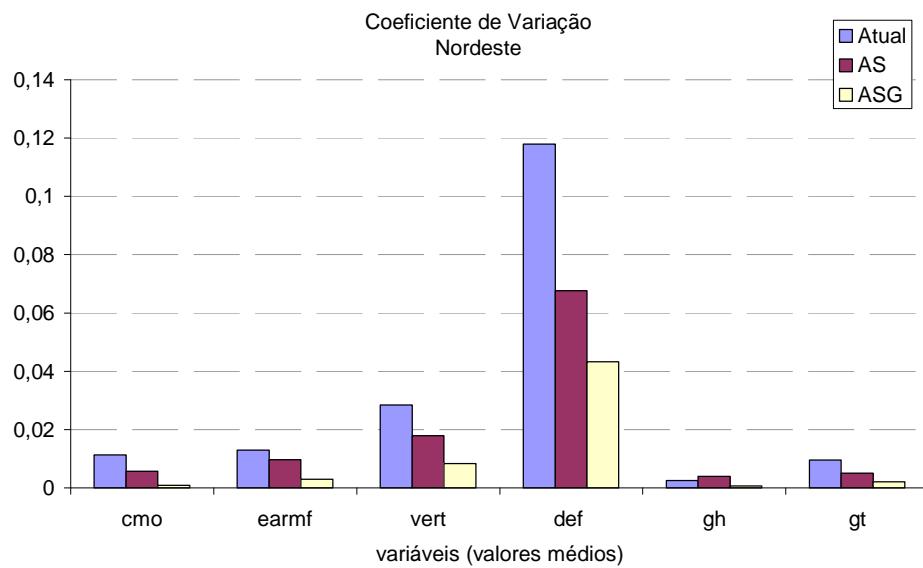


Figura 8.31: Coeficiente de variação do Nordeste.

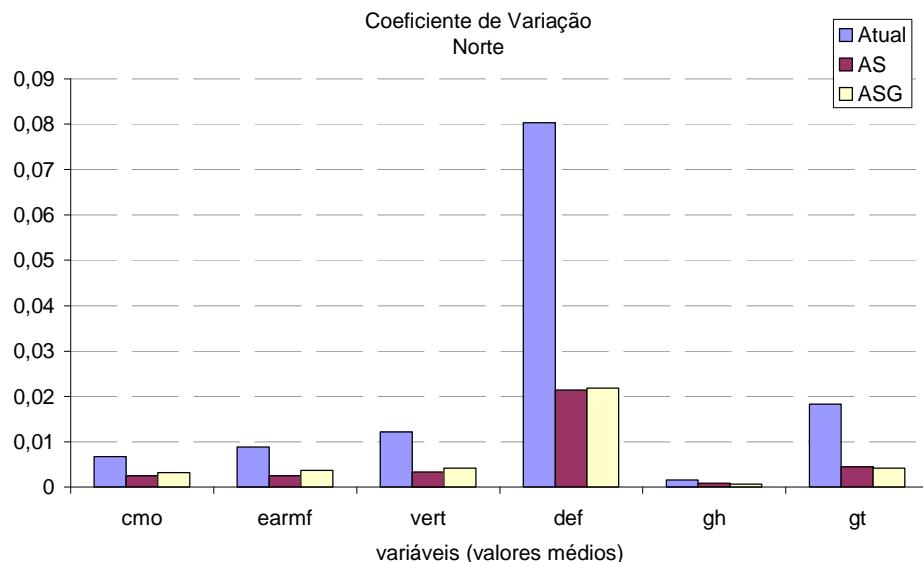


Figura 8.32: Coeficiente de variação do Norte.

Das figuras 8.30 a 8.33 pode ser observado que a variação dos índices de desempenho são mais impactantes para o método Atual, em comparação com os métodos AS e ASG, pois os valores do coeficiente de variação foram maiores utilizando este método.

De novo, isso mostra um dos pontos fracos do método Atual, um dos motivos da realização deste trabalho.

#### **8.4 Resumo**

Neste capítulo foram apresentados os principais resultados provenientes da aplicação do método ASG para a construção da sub-árvore a ser visitada durante a etapa forward do cálculo da política da operação. Realizaram-se análises do processo de convergência ressaltando a importância da estabilidade de ZINF e ZSUP.

Em seguida, os resultados dos índices de desempenho provenientes da simulação final foram analisados, confrontando os métodos ASG, AS e Atual e explicitando as falhas do método Atual com relação ao grande variabilidade dos resultados às amostras de cenários hidrológicos.

## **9 Conclusões e sugestões para trabalhos futuros**

### **9.1 Conclusões referentes à aplicação da ASG**

O Planejamento da Operação de Médio Prazo do SIN é um problema com horizonte de alguns anos a frente, com discretização mensal, no qual a estocasticidade está representada através de séries sintéticas de energias geradas por um modelo autoregressivo, que por sua vez é calibrado pelo registro histórico de afluências.

Como a cardinalidade da árvore completa é elevada, utiliza-se uma sub-árvore para os estudos de planejamento da operação energética de médio prazo, visando viabilizar computacionalmente o estudo.

Através de métodos utilizando AAS ou técnicas de agregação, pode-se construir uma sub-árvore de cardinalidade inferior para ser utilizada no processo de otimização.

Esse trabalho propôs uma alternativa para a construção dessa sub-árvore de cenários que deve ser percorrida durante o cálculo da política de operação (durante etapa forward). O objetivo principal foi obter resultados que apresentassem menor variabilidade com relação à escolha da amostra de cenários, melhorando a qualidade da função de custo futuro construída pelo NEWAVE e repassada aos demais modelos da cadeia.

Inicialmente foi aplicado o método ASG utilizando dados de energias naturais afluentes, o que rapidamente levou a uma dificuldade na preservação da estatística desvio-padrão, que é de importância fundamental em nesse estudo. Considerando que o método utiliza-se do K-Means como técnica de agregação, foram realizadas várias tentativas de melhoria dos resultados (desvio-padrão), baseado em diferentes formas de escolha dos grupos iniciais (dado de entrada para o início do processo de agregação), visto que essa técnica pode ser sensível à escolha destes. Como o desvio-padrão continuou comprometido, foi abolida a agregação de energias e adotada a agregação de ruídos aleatórios, pois dessa forma obteve-se maior controle sobre o desvio-padrão, através da introdução de um fator de correção para essa estatística. A construção da sub-árvore, através da agregação dos ruídos com posterior utilização do PAR(p), conseguiu preservar as características estatísticas do registro histórico. A partir de então, o método ASG pôde ser aplicado no modelo de planejamento da operação energética de médio prazo.

A ASG obteve uma política ótima mais robusta com relação às variações de amostras de cenários hidrológicos, se comparada com a opção Atual. Portanto, atingiu-se o objetivo preliminar que motivou este trabalho. Ganhos maiores podem ser obtidos se a pseudo-árvore completa for aumentada (Ncen).

Além da robustez, o método ASG melhorou também a qualidade do processo de convergência estatístico, fazendo com que os valores de ZINF e ZSUP ficassem mais estáveis ao longo das iterações.

A aplicação desta política (construída com a ASG) na simulação final trouxe uma operação mais estável, pois os índices calculados apresentaram variabilidade muito inferior se comparado ao método Atual.

Os resultados do método ASG foram compatíveis com o método AS, mostrando que os dois métodos conseguiram uma boa representação da sub-árvore de cenários, preservando estatísticas do registro histórico e obtendo resultados mais estáveis para convergência (ZINF e ZSUP) e para operação do sistema com relação à variações nos cenários hidrológicos.

## 9.2 Sugestões para trabalhos futuros

Neste trabalho foi utilizada uma matriz de ruídos de dimensões elevadas [Ncen NsisxNper] de entrada, de onde, após passar pelo processo de agregação, foi extraída uma matriz de cardinalidade menor [Ncen<sub>2</sub> NsisxNper]. Devido a essa característica, não se conseguiu elevar a cardinalidade da pseudo-árvore, pois foram encontradas dificuldades relacionadas à alocação de memória.

Em trabalhos futuros, pode-se explorar a redução da dimensionalidade dessa matriz através de medidas representativas de forma a utilizar valores correspondentes aos subistemas ponderados por todo o horizonte de estudo. O problema seria então de dimensões [Ncen NsisxC], onde Ncen poderia ser elevado além dos limites conseguidos nesse trabalho (dez mil) e C corresponderia a um número inteiro relativo à redução/representação do horizonte de estudo.

Dentre as formas de reduzir a dimensão mencionadas pode-se citar algumas possíveis: calcular a média de todos os períodos (C=1), utilizar médias anuais do horizonte de estudo (C=5), utilizar média dos períodos secos e úmidos de todo o horizonte (C=2) e etc.

Também podem ser utilizadas técnicas de componentes principais para reduzir a dimensionalidade do problema de agregação e viabilizar, sob o ponto de vista da alocação de memória, a elevação da cardinalidade da pseudo-árvore.

Visto que foi utilizado somente casos de PMO para estudo, trabalhos futuros podem englobar a utilização de casos de PDE; a fim de obter a política ótima e os índices de desempenho considerando diferentes configurações e horizontes de estudo.

Este estudo considerou somente a agregação objetivando obter 200 cenários para a etapa forward, mas outro foco de estudo seria elevar esta quantidade de cenários, o que deve melhorar a qualidade dos grupos formados e, portanto, a qualidade da sub-árvore forward.

Em trabalhos futuros também pode-se estender este estudo considerando a utilização de outros métodos de agregação, como por exemplo, métodos que utilizem lógica fuzzy.

## **Referências Bibliográficas**

- BENDERS, J.F., **Partitioning Procedures for Solving Mixed Variables Programming Problems**, Numerische Mathematik, v. 4, pp. 238-252, 1962.
- BUSSAB, W.O., MIAKAZI, E.S., ANDRADE, D.F., **Introdução à Análise de Agupamentos**, 9º Simpósio Nacional de Probabilidade e Estatística, São Paulo, Associação Brasileira de Estatística, 1990.
- CEPEL, Modelo DECOMP, **Determinação da Coordenação da Operação a Curto Prazo**, Manual de Referência Versão 16.1, Fevereiro/2010.
- CEPEL, MODELO DESSEM-PAT, **Otimização da Programação Diária da Operação de Sistemas Hidrotérmicos com Representação da Rede Elétrica em Patamares Cronológicos**, MANUAL DO USUÁRIO VERSÃO 6.0, Abril/2010
- FORTUNATO, L.A.M., NETO, T.A.A., ALBUQUERQUE, J.C.R., PEREIRA, M.V.F., **Introdução ao Planejamento da Expansão e Operação de Sistemas de Produção de Energia Elétrica**, Editora da Universidade Federal Fluminense 1990.
- GUIDINI, M.B., NASCIMENTO, A.M., **Aplicação do K-Means cluster para classificar estilos gerenciais**, UNISINOS, Porto Alegre, 2008.
- HAIR, J., **Multivariate Data Analysis**. New Jersey: Prentice Hall, 1998.
- HARTIGAN, J.A., **Clustering Algorithms**, New York, John Wiley & Sons, 1975.
- HARTIGAN, J., WONG, M., **A K-Means Clustering Algorithm**, Applied Statistics, vol. 28, no. 1, pp. 100-108, 1979.
- HARVEY, A.C., **Time Series Models**, London, P Allan, 1981.
- JAIN, A.K. **Data clustering: A review**. ACM Computing Surveys, V. 31, 1999.

JARDIM, D.L.D. **Modelo de geração de séries sintéticas de vazões utilizando técnicas de agregação**, Dissertação de Mestrado, COPPE-UFRJ, Rio de Janeiro, Brasil, 2002.

KELMAN, J., OLIVEIRA, G.C., MACEIRA, M.E.P., **Modelo de Séries Hidrológicas – Manual de Metodologia**. In: Relatório Técnico no. 242-92, CEPEL, Rio de Janeiro, 1992.

KRANEN, M.E., **Aplicação de Técnicas de Agrupamento sobre Séries de Potência para Avaliação de Confiabilidade em Sistemas Hidrotérmicos**, Dissertação de Mestrado, Instituto de Matemática, Universidade Federal do Rio de Janeiro, UFRJ, 2004.

MACEIRA, M.E.P., 1989, **Operação Ótima de Reservatórios com Previsão de Afluências**. Dissertação de Mestrado, COPPE/UFRJ, Rio de Janeiro, Brasil.

MACEIRA, M.E.P., **Programação Dinâmica Dual Estocástica Aplicada ao Planejamento da Operação Energética de Sistemas Hidrotérmicos com Representação do Processo Estocástico de Afluências por Modelos Autorregressivos Periódicos**. Relatório Técnico CEPEL 237-1993, 1993.

MACEIRA, M.E.P., BEZERRA, C.V., GORENSTIN, B.G., CUNHA, S.H.F., SUANNO, C., SACRAMENTO, M.C., KLINGERMAN, A.S., **Aplicação do Modelo na Avaliação Energética da Interligação dos Sistemas Norte/Nordeste e Sul/Sudeste Brasileiros** In: Proceedings of VI Simpósio de Especialistas de Planejamento da Operação e Expansão Elétrica, Salvador, May, 1998.

MACEIRA, M.E.P., JARDIM, D.L.D., **Modelo de Geração de Séries Sintéticas de Energias e Vazões – GEVAZP – Manual de Referência**, In: Relatório Técnico DPP/PEL 83/2000, CEPEL, Rio de Janeiro, 2000.

MACEIRA, M.E.P., TERRY, L.A., COSTA, F.S., DAMÁZIO, J.M., MELLO, A.C.G., **Chain of Optimization Models for Setting the Energy Dispatch and Spot Price in the Brazilian System**, In: Proceedings of 14<sup>th</sup> PSCC, Sevilla, Espanha, 2002.

MACEIRA, M.E.P., JARDIM, D.L.D., CEPEL, **Manual do Usuário – Modelo Estratégico de Geração Hidrotérmica a Subsistemas Equivalentes – PROJETO NEWAVE**, Rio de Janeiro, 2008.

MALHOTRA, N. K., **Pesquisa de Marketing: uma orientação aplicada**. Porto Alegre: Bookman, 2001.

PENNA, D.D.J., **Definição a Árvore de Cenários de Afluências para o Planejamento da Operação Energética de Médio Prazo**, Tese de Doutorado, PUC-RJ, 2009.

PEREIRA, M.V.F., PINTO L.M.V.G., **Multi-Stage Stochastic Optimization Applied to Energy Planning**. Mathematical Programming, v. 52, n. 3 (may), pp. 359-375, 1991.

SACARPEL, R.A. **Métodos Multivariados**, 2010.

Disponível em: <http://www.mec.ita.br/~rodrigo/Disciplinas/MB213/S07.pdf>

SILVA, E.L., **Formação de Preços em Mercados de Energia Elétrica**, Editora Sagra Luzzato, 2001.

TERRY, L.A., **Modelo a Sistema Equivalente – Descrição Geral**, In: Relatório Técnico CEPEL no. DPP/PEN – 1705-1265, 1980.

.

## **Apêndice A - Teste de sanidade rotina fortran**

### **A.1 Teste de sanidade da rotina fortran K-Means**

Para verificar se a agregação utilizando a rotina fortran (programada) foi feita corretamente, foi gerada uma matriz de ruídos normais no MATLAB de dimensão Ncen x Nsis\*Nper e agregada em uma de dimensão K x Nsis\*Nper. Sabe-se que a média e o desvio padrão de uma distribuição normal devem zero e um, respectivamente. Espera-se que após o processo de agregação, a média permaneça em torno de zero e o desvio próximo a 1. Como é utilizada uma técnica de agrupamento para redução da dimensionalidade da amostra original, pode-se esperar dificuldades em reproduzir valores extremos, o que não acontece quando é utilizada a amostragem aleatória simples pura, que consiste na opção Atual. Portanto, o desvio padrão deve ficar ligeiramente inferior a um.

A agregação utilizando o K-Means e a rotina em fortran deverão apresentar o mesmo comportamento. Foram plotadas a evolução temporal das estatísticas de tendência central (média) e dispersão (desvio padrão) dos ruídos aleatórios agregados com rotina fortran (K-Means) e rotina do MATLAB (K-Means) antes e após o uso do fator de compensação. Os primeiros períodos estão com valor zero nas estatísticas pois estes são anteriores ao mês inicial de estudo, o mês de setembro. A seguir seguem os resultados obtidos e as conclusões quanto à sanidade da rotina fortran utilizada.

### **A.2 Agregação realizada com K-Means MATLAB**

#### **A.2.1. Antes do uso do fator de compensação do desvio padrão**

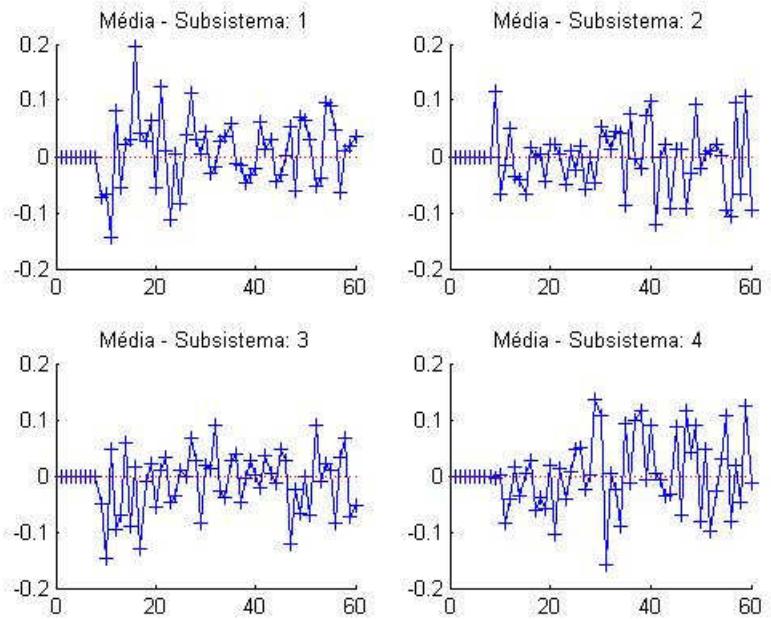


Figura A.1 : média ruídos agregados com MATLAB

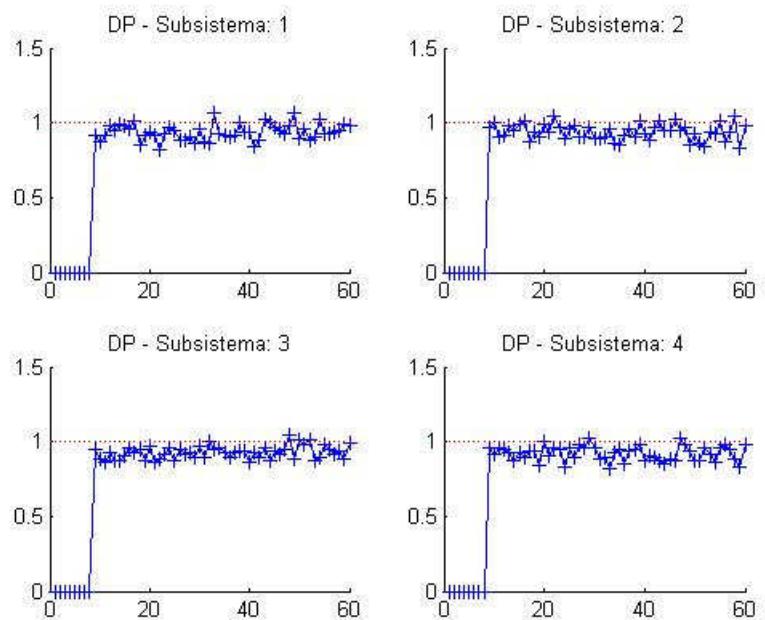


Figura A.2 desvio padrão dos ruídos agregados com MATLAB

Analisando o gráfico do desvio padrão, percebe-se que houve uma degradação em relação ao valor ideal, que seria 1. Essa degradação já era esperada, visto que ao utilizar-se

um processo de agrupamento perde-se um pouco a capacidade de reproduzir valores extremos. Porém essa característica indesejável deve ser corrigida.

### A.2.2. Após o uso do fator de correção do desvio

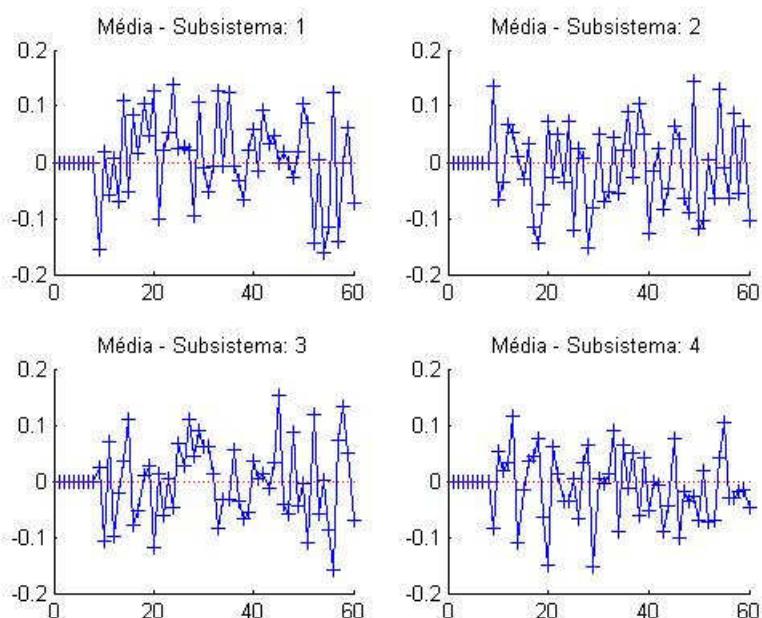


Figura A.3 : média dos ruídos agregados com MATLAB

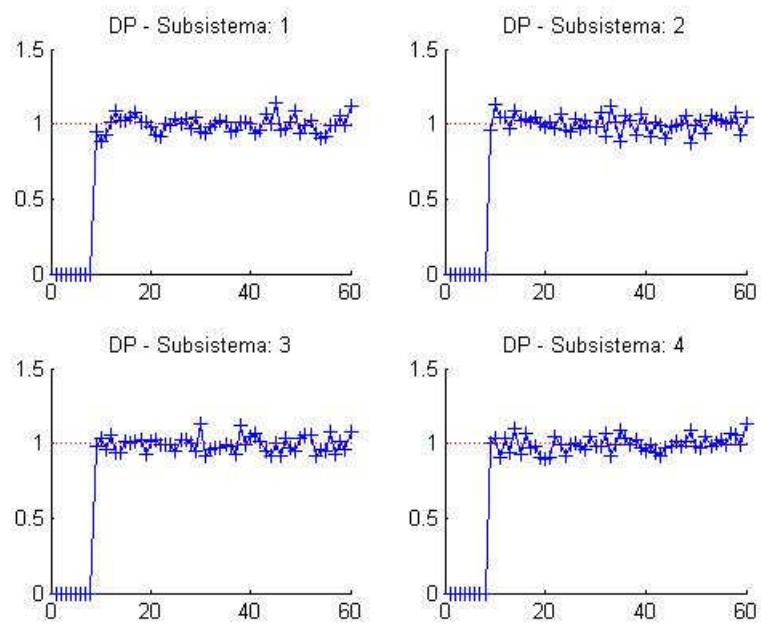


Figura A.4 : desvio padrão dos ruídos agregados com MATLAB

**A.3 Agregação realizada com K-Means rotina Fortran – somente após uso fator corretivo do desvio.**

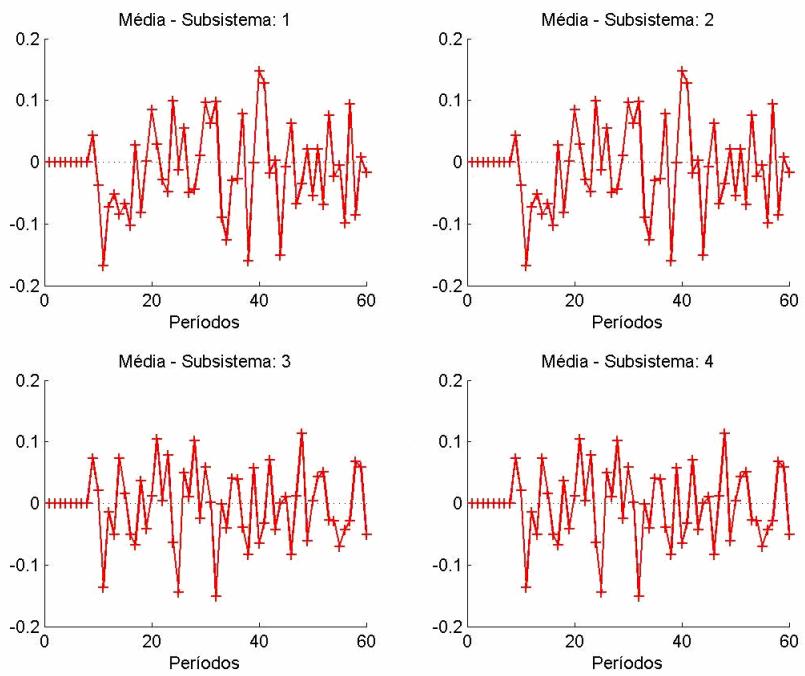


Figura A.5: média dos ruídos agregados com rotina Fortran

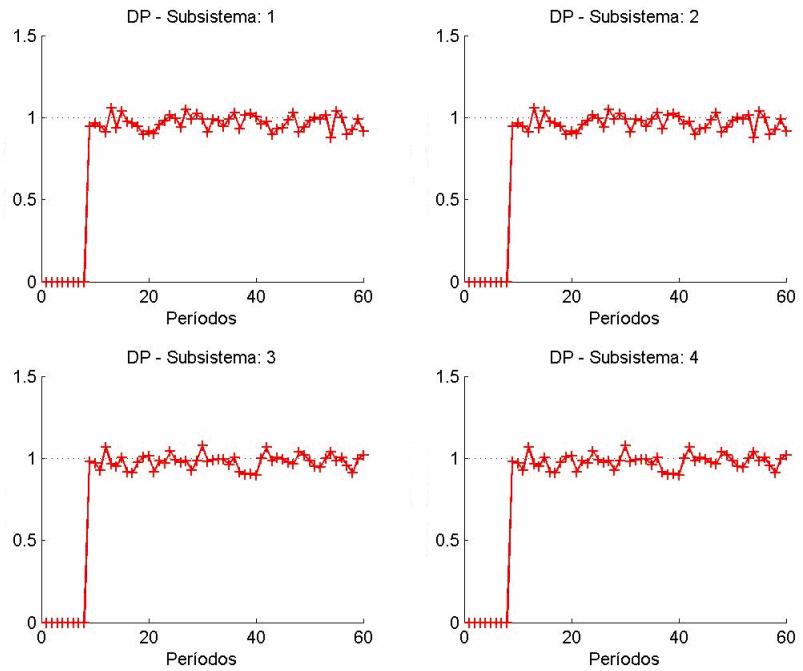


Figura A.6 : desvio padrão dos ruídos agregados com rotina Fortran

Portanto, analisando as Figuras A.5 e A.6 acima, conclui-se que a rotina fortran está funcionando corretamente, pois o comportamento das estatísticas de tendência central e dispersão analisados estão coerentes com a rotina MATLAB.