# Centro de Pesquisas de Energia Elétrica - CEPEL

## Relatório Técnico

| | | | | | |
|---|---|---|---|---|---|
| **Nº/Ano:** | 13612/2018 | **Nº de Páginas:** 26 | | **Nº de Anexos:** | 1 de 31 páginas |

**Título:** Avaliação de alternativas para escolha do representante no processo de agregação da Amostragem Seletiva.

| | |
|---|---|
| **Departamento:** | Departamento de Otimização Energética e Meio Ambiente - DEA |

| | | | |
|---|---|---|---|
| **Área de Responsabilidade:** | B200 | **Conta de Apropriação:** | 1345 e 1798 |

**Cliente:**
Comissão Permanente para Análise de Metodologias e Programas Computacionais do Setor Elétrico - CPAMP


**Resumo:**
Neste relatório são avaliadas alternativas para escolha do representante no processo de agregação da Amostragem Seletiva com o objetivo de redução da variabilidade amostral, atividade em estudo pela Comissão Permanente para Análise de Metodologias e Programas Computacionais do Setor Elétrico (CPAMP), para o ciclo 2018/2019.



**Autores:**
Débora Dias Jardim Penna – CEPEL
Felipe Treistman – PUC-Rio
Maria Elvira Piñeiro Maceira – CEPEL



**Palavras-Chave:**
Geração de cenários. Amostragem Seletiva, Agregação de cenários

**Classificação:** CONTROLADO



**Gerentes dos Projetos**

**Nome:** Maria Elvira PiñeiroMaceira
**Tel.:** 21-2598-6454
**E-mail:** elvira@cepel.br

**Nome** Débora Dias Jardim Penna
**Tel.:** 21-2598-6450
**Email:** debora@cepel.br

**Chefe do Departamento de Otimização Energética e Meio Ambiente**

**Nome** André Luiz Diniz Souto Lima
**Tel.:** 21-2598-6046
**E-mail:** diniz@cepel.br


**Aprovação**

**Raul Balbi Soliero**
**Diretor de Pesquisa, Desenvolvimento e Inovação**
27/12/18

CEPEL

Centro de Pesquisas de Energia Elétrica

PROJETOS NEWAVE e GEVAZP

Relatório Técnico

# Avaliação de Alternativas para Escolha do Representante no Processo de Agregação da Amostagem Seletiva

Dezembro/2018

# ÍNDICE

# 1 Introdução

Durante o processo de validação da mudança de configuração do SIN, de 9 para 12 reservatórios equivalentes de energia (REEs), foram rodados os casos de PMO do ano de 2017 com a nova representação dos REEs. A partir destas análises, observou-se um comportamento de redução do custo marginal de operação obtido nas simulações finais na configuração com maior número de REEs. Foram elencadas três possíveis causas para o comportamento observado no custo marginal de operação (CMO), dentre elas, a variação amostral na representação da árvore de cenários de afluências no algoritmo da PDDE (recursão *backward*).

No Relatório Técnico intitulado "Análise da inflexão do custo marginal de operação no modelo NEWAVE entre os quarto e quinto estágios temporais ao se adotar a representação de 12 REEs para Sistema Interligado Nacional" de setembro de 2018[1], o CEPEL realizou uma avaliação detalhada dos cenários de energia naturais afluentes (ENAs), desde a etapa da geração dos ruídos independentes até a obtenção dos cenários de afluências propriamente ditos. Foi mostrado que a redução abrupta do CMO foi ocasionada pela geração de cenários de afluência com valores atipicamente superiores, gerados pelo sorteio de ruídos aleatórios com média bastante elevada, cuja probabilidade de ocorrência era bem pequena (~0.27%).

Durante os estudos da Comissão Permanente para Análise de Metodologias e Programas Computacionais do Setor Elétrico (CPAMP) sobre a reamostragem de cenários *forward*, foram realizadas sensibilidades com relação à árvore de cenários utilizada na solução do problema de planejamento da operação (árvore completa). Observou-se que a variabilidade dos resultados era grande e, em princípio, a reamostragem de cenários da simulação *forward* não era suficiente para reduzir a variabilidade amostral associada à árvore completa, definida pelos cenários da recursão *backward*.

A partir destas análises, o CEPEL investigou aprimoramentos no processo de geração de cenários de afluências de forma a minimizar a ocorrência de cenários tão atípicos e a variabilidade amostral. O presente relatório tem como objetivo apresentar uma alternativa para a escolha do representante do processo de agregação do processo de Amostragem Seletiva (AS), que resultou em uma redução na variabilidade amostral observada nos resultados do planejamento da operação de médio e longo prazos.

Esse relatório é organizado da seguinte forma: nesta seção, apresentou-se a motivação para o aprimoramento no processo de AS. A seção 2 mostra como é feita a representação da incerteza hidrológica no problema de planejamento da operação. Na seção 3 é apresentada uma revisão da

metodologia e um histórico do desenvolvimento da AS utilizada para a geração dos cenários de afluência e como o processo de agregação é utilizado. Na seção 4 são apresentados resultados que confirmam a redução da variabilidade amostral obtida com a nova alternativa proposta neste relatório e, por fim, a seção 5 traz as conclusões e possíveis trabalhos futuros.

# 2 REPRESENTAÇÃO DA INCERTEZA NO PLANEJAMENTO DA OPERAÇÃO

O objetivo principal do planejamento da operação energética de médio e longo prazo é definir a alocação de recursos hídricos e térmicos de forma ótima considerando a minimização do valor esperado do custo total de operação e a segurança energética do sistema. Atualmente, o problema de planejamento da operação energética do sistema interligado nacional de médio e longo prazos, resolvido pelo modelo NEWAVE [2], [3], é representado por um problema de programação estocástica linear multiestágio e o método aplicado para construção da política é a Programação Dinâmica Dual Estocástica (PDDE) [4], levando-se em conta a correlação temporal das afluências aos reservatórios e/ou usinas hidráulicas [5]. Assim, consideram-se como variáveis de estado do problema o armazenamento no início do período e as afluências passadas (tendência hidrológica). A incerteza hidrológica é representada explicitamente através de cenários de afluências construídos sinteticamente empregando-se um modelo autorregressivo periódico [6] e um processo de Amostragem Seletiva [7]. O modelo NEWAVE faz parte da cadeia de modelos de otimização e simulação desenvolvida pelo CEPEL para o planejamento da expansão e operação energética [8].

Os primeiros métodos de decomposição desenvolvidos para resolver problemas de programação linear estocástico percorriam a árvore de cenários em sua totalidade (árvore completa – vide Figura 2.1a), porém para problemas onde a árvore de cenários apresenta uma cardinalidade elevada, percorrê-la em sua totalidade é impossível do ponto de vista computacional ou prático. Para contornar este problema foram desenvolvidos diversos métodos que utilizam técnicas de amostragens para selecionar uma subárvore de cenários com tamanho reduzido. O primeiro método a fazer uso da amostragem em programação estocástica foi a PDDE, cujas subárvores *forward* e *backward* utilizadas durante seu processo de solução estão ilustradas na Figura 2.1b e 2.1c, respectivamente. Usualmente, o conjunto de cenários visitados na recursão *backward* a cada estado é chamado de "aberturas".

A subárvore *forward* pode ser definida por um subconjunto de cenários amostrados diretamente da árvore completa de cenários, ou por um conjunto de cenários escolhidos a partir da distribuição original da variável aleatória. O importante é que a subárvore *forward* represente a mesma distribuição de probabilidades associada à árvore original.

A subárvore *backward* é dada pelo conjunto de cenários que serão utilizados nos problemas de programação linear do passo *backward* do algoritmo da PDDE para produzir cortes de Benders para a função de custo futuro (FCF) em cada estágio do horizonte de planejamento. De forma geral, para

cada estágio de tempo são visitados os NSIM[1] estados, e para cada estado são resolvidos NLEQ[2] problemas de despacho ótimo resultando em um corte de Benders médio por estado a ser incluído na FCF.

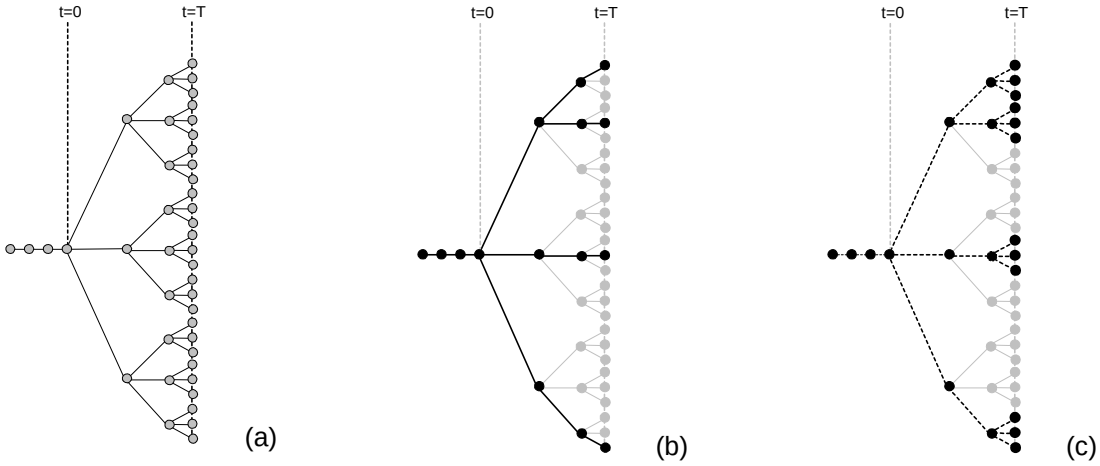



Figura 2.1: (a) árvore completa (b) subárvore *forward* (c) subárvore backward

O processo de solução iterativa da PDDE compreende dois passos principais que são executadas ao longo de todo o horizonte de estudo:

- passo *backward*: realiza uma simulação recursiva cujo objetivo é produzir os cortes de Benders, que correspondem a uma aproximação inferior da função de custo futuro para cada estágio, a partir dos estados fornecidos pelo passo *forward*.

- passo *forward*: realiza uma simulação direta e resolve um conjunto de problemas de programação linear para cada estágio de tempo. O objetivo deste passo é produzir alguns caminhos de operação possíveis (estados de armazenamento), que cumprem todas as restrições do problema de programação linear estocástica. O custo médio ao longo de todos os caminhos da operação fornece o valor esperado do custo total de operação. O passo *forward* fornece a convergência do algoritmo PDDE.

Depois de calcular a política ótima de operação, são simulados 2.000 cenários sintéticos de afluências para estimar os índices de desempenho probabilístico do sistema, como risco de déficit, energia esperada não suprida, intercâmbios entre sistemas, entre outros.

Do ponto de vista matemático, na prática é o problema de árvore completa que se está resolvendo, embora tendo sempre a preocupação de que este represente suficientemente bem o problema real dado pela distribuição de probabilidades da variável aleatória contínua multivariada. Desta forma,

[1] NSIM corresponde ao número de cenários da subárvore *forward* (estados) a cada período.
[2] NLEQ corresponde ao número de aberturas por período e estado.

espera-se que, ao escolher outro conjunto de cenários para representar a árvore completa, o valor ótimo obtido para o novo problema não seja muito diferente daquele obtido com a árvore completa anterior. A diferença entre os valores ótimos obtidos é devido à variação amostral da representação da árvore completa de cenários. Quanto melhor a representação da árvore completa, menor será a variabilidade nos resultados.

Para se obter resultados mais robustos, isto é, menos subjeitos à variação amostral, pode-se aplicar por exemplo técnicas, como a *Sample Average Approximation* (SAA) [9], que envolve a resolução do problema diversas vezes, com árvores *backward* diferentes, ou aumentar o número de cenários para representar a árvore completa (aumentar NLEQ). Porém, as duas opções podem tornar a solução do problema inviável computacionalmente. Outra opção é utilizar métodos de amostragem que possibilitem a representação adequada do processo estocástico de afluências com um número reduzido de cenários.

Atualmente, a amostra de ruídos aleatórios utilizada pelo modelo PAR(p) para gerar os cenários de afluências é obtida através do método de Amostragem Seletiva, método que propicia maior robustez aos resultados obtidos pela PDDE com relação a variações nos cenários hidrológicos da subárvore *backward*, em relação a uma Amostragem Aleatória Simples.

# 3 GERAÇÃO DE CENÁRIOS

O modelo GEVAZP [10]é responsável pela geração de cenários sintéticos multivariados de energia e vazão mensais. Esse cenários são essenciais para o cálculo da política de operação ótima no problema otimização multi-estágio e multi-reservatório do SIN.

Para desempenhar essa tarefa, o modelo GEVAZP adota em sua modelagem estocástica a família de modelos autorregressivos periódicos de ordem p, PAR(P) [6][11]. A escolha desse modelo é justificada pela sua capacidade de capturar o comportamento periódico da estrutura de autocorrelação observado comumente em séries de afluências mensais. O modelo PAR(p) pode ser descrito matematicamente pela equação 3.1:

$$\left(\frac{Z_t-\mu_m}{\sigma_m}\right) = \varphi_1^m\left(\frac{Z_{t-1}-\mu_{m-1}}{\sigma_{m-1}}\right) + \ldots + \varphi_{pm}^m\left(\frac{Z_{t-pm}-\mu_{m-pm}}{\sigma_{m-pm}}\right) + a_t \qquad (3.1)$$

Onde:

$Z_t$ é a série hidrológica sazonal de período T;

N é número de anos;

T é o índice de tempo, t = 1, 2, ..., sN; função do ano T (T = 1, 2, ..., N) e do período m (m = 1, 2, ..., s);

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

$\mu_m$ é a média sazonal do período s;

$\sigma_m$ é o desvio padrão sazonal do período s;

$\varphi_i^m$ é o i-ésimo coeficiente autorregressivo do período m;

$p_m$ é a ordem do operador de defasagem de período m;

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

Usualmente, assume-se que os ruídos $a_t$ na equação (3.1) possuem distribuição normal e são independentes e identicamente distribuídos. Se for constatada a não normalidade dos ruídos, pode-se aplicar a transformação Box-Cox [12]. No caso da modelagem de afluências feita pelo modelo GEVAZP, emprega-se a série histórica original sem transformação e, consequentemente, deve-se modelar ruídos que demonstram uma distribuição assimétrica, como a distribuição lognormal.

A Figura 3.1 ilustra os passos necessários para geração de ruídos lognormais correlacionados espacialmente como executado pelo modelo GEVAZP, internalizado no programa NEWAVE.

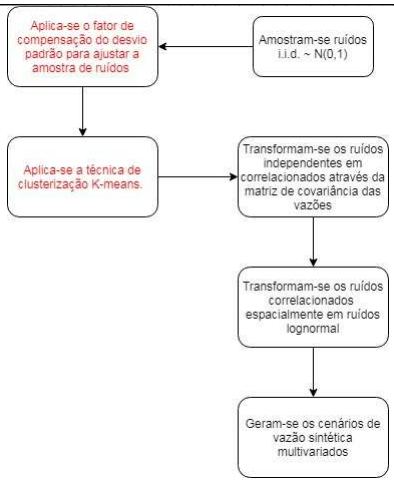



Figura 3.1: Processo de geração de ruídos como feito pelo modelo GEVAZP para a PDDE.

Atualmente, a amostra de ruídos aleatórios utilizada pelo modelo PAR(p) é obtida através do método de Amostragem Seletiva [7] a partir de uma distribuição multivariada lognormal, onde cada componente da variável aleatória representa o ruído de um reservatório equivalente de energia considerado na configuração.

A Amostragem Seletiva gera vetores de ruídos independentes que posteriormente passam a ser correlacionados espacialmente através da imposição da correlação observada nas vazões/ENAs históricas sobre os ruídos resultantes da agregação. Essa correlação imposta aos ruídos garante a multivariabilidade dos cenários sintéticos gerados de vazões e energias afluentes. Uma vez correlacionados espacialmente, os ruídos aleatórios ainda normais são transformados em ruídos log-normais com 3 parâmetros [13], com o objetivo de evitar a geração de valores negativos de afluência. A partir dessa etapa, os ruídos são usados nos processo de geração de cenários sintéticos multivariados de vazão e energia afluente de acordo com a equação (3.1). Este processo é repetido a cada período do horizonte de estudo.

Os ruídos da amostra *backward* constituem a incerteza do problema de otimização estocástica de árvore completa a ser resolvido. O conjunto de ruídos *backward* de cada período é o mesmo para todos os estados deste período, porém os cenários de vazão ou energia afluente são diferentes, dado que o passado de cada estado é diferente.

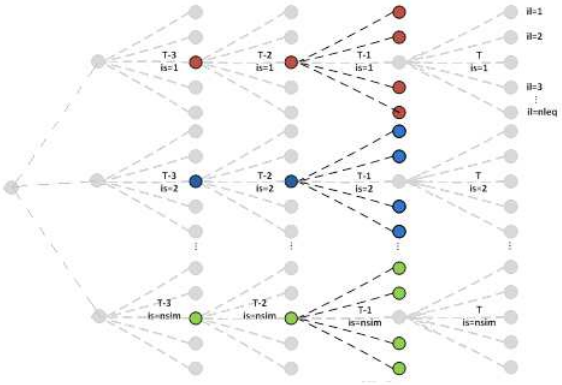



Figura 3.2: Construção da subárvore *backward*

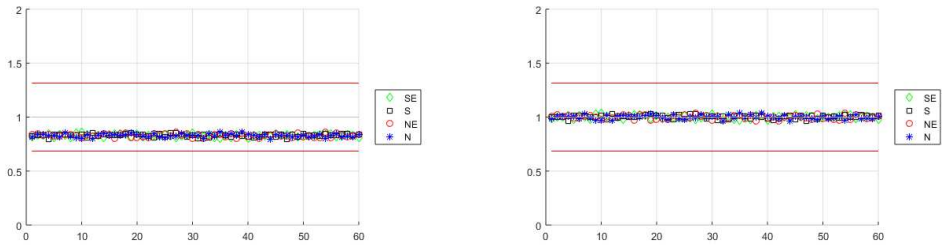
## 3.1 Amostragem Seletiva

A Amostragem Seletiva consiste em aplicar técnicas de agregação para gerar as amostras de ruídos aleatórios multivariados com o intuito de aumentar a representatividade da amostra.

Inicialmente gera-se uma amostra de 100 mil vetores de ruídos aleatórios independentes ~N(0,1), onde cada componente do vetor de ruídos independentes corresponde a uma UHE ou REE. A esta amostra inicial dá-se o nome de amostra original. Depois de gerados, os vetores de ruídos independentes são agregados através da técnica de agregação K-means [14], que servem para diminuir a dimensionalidade dos dados enquanto se preserva a exploração do espaço amostral. Após a conclusão do processo de agregação é selecionado um representante para cada grupo, e o conjunto destes representantes irá formar a amostra de ruídos *backward* e *forward*.

A aplicação das técnicas de agregação na amostra original, cujo algoritmo envolve a substituição de agrupamentos de objetos da amostra original por um único representante, resulta na obtenção de amostras agregadas com menor variabilidade do que a correspondente amostra original (a variabilidade interna nos agrupamentos é perdida). Para os estudos de PMO realizados à época da validação da Amostragem Seletiva, a forma de atenuar a redução do desvio padrão na amostra agregada foi selecionar como representante dos grupos o objeto mais próximo do centroide ao invés do próprio centroide. No entanto para os estudos de PDE, que representavam o sistema interligado nacional com um número maior de REEs, a degradação do desvio padrão na amostra agregada voltou a crescer. A solução foi ajustar um fator de correção para o desvio padrão da amostra original, com base na degradação observada e aplicá-lo à amostra original antes do processo de agregação-. Na Figura 3.3 apresenta o desvio padrão da amostra de ruídos obtidos com a Amostragem Seletiva com e sem a adoção da compensação do desvio padrão.

Mais detalhes sobre a Amostragem Seletiva e fator de compensação do desvio padrão podem ser consultados na Nota Técnica 42 do projeto NEWAVE, que se encontra no Anexo A deste relatório.




(a) (b)

Figura 3.3: Desvio Padrão dos ruídos após o processo de agregação (a) com e (b) sem a compensação do desvio padrão

## 3.2 Seleção do representante

Durante os estudos que resultaram no Relatório Técnico "Análise da inflexão do custo marginal de operação no modelo NEWAVE entre os quarto e quinto estágios temporais ao se adotar a representação de 12 REEs para Sistema Interligado Nacional" [1], foi observada a ocorrência de conjuntos de ruídos com média elevada em alguma das 12 dimensões do vetor de ruídos. A fim de reduzir a chance de ocorrência deste comportamento, foi avaliada alternativa de seleção para o representante do grupo no processo de agregação. Ao invés de utilizar o objeto mais próximo do centroide, foi utilizado o próprio centroide. A Figura 3.4 ilustra a escolha dos representantes.

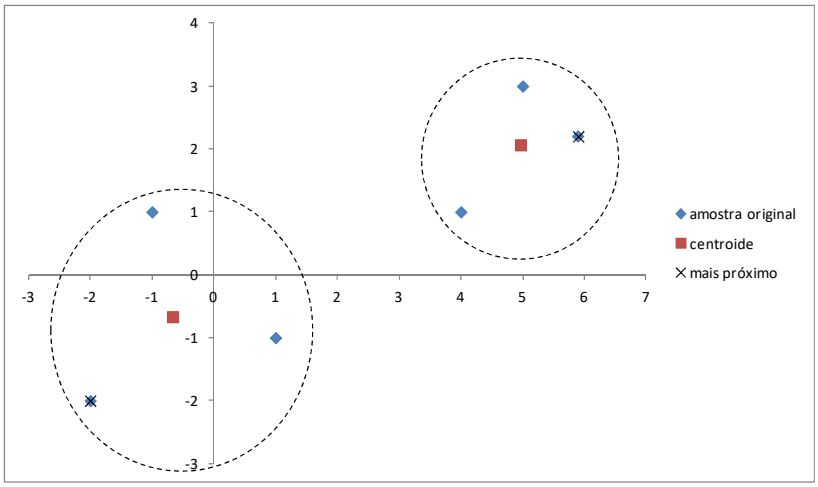



Figura 3.4: Exemplo ilustrativo da escolha do representante

Ao considerar o centroide, a média dos vetores de ruídos tende a zero, porém o desvio padrão fica mais degradado (menor do que 1). Todavia, dado que o fator de compensação do desvio padrão é calculado sob medida para o processo de agregação, este aumento da degradação resulta em um fator de compensação maior, o que ao final do processo leva a um desvio padrão tendendo a 1, conforme desejado.

Os resultados obtidos com a adoção do centroide ao invés do objeto mais próximo como representante no processo de agregação da Amostragem Seletiva, mostraram uma maior robustez nos resultados decorrentes do planejamento da operação (geração térmica, geração hidráulica, custo marginal de operação, entre outros), além de gerar conjuntos de vetores de ruídos com menor dispersão na média e desvio padrão, conforme será mostrado na seção 4.

## 3.3 Representatividade da árvore de cenários

Espera-se que quanto maior o nível de detalhamento da incerteza das afluências, isto é maior número de aberturas para a árvore completa de cenários, mais próximo do ótimo do problema contínuo estaremos. Isto pode ser verificado com o aumento do limite inferior do custo de operação (ZINF).

Esta suposição é válida quando os cenários são gerados com amostragem aleatória simples (AAS), pois quanto maior a amostra mais chances de serem sorteados cenários extremos. Porém, quando se aplica a Amostragem Seletiva não necessariamente isto ocorrerá, dado que a população, representada pela amostra original, não se altera durante a formação dos grupos. Isto quer dizer, mesmo aumentando o número de grupos, o cenário mais extremo estará limitado ao cenário mais extremo da amostra original.

Nas Figuras 3.5 e 3.6 são apresentados exemplos ilustrativos da geração de cenários, considerando AAS e AS, respectivamente. Na coluna da direita é mostrada a geração de 5 cenários para um caso com duas dimensões e na coluna da esquerda, 6 cenários.

(a) (b)

Figura 3.5: Exemplo ilustrativo AAS (a) 5 cenários e (b) 6 cenários

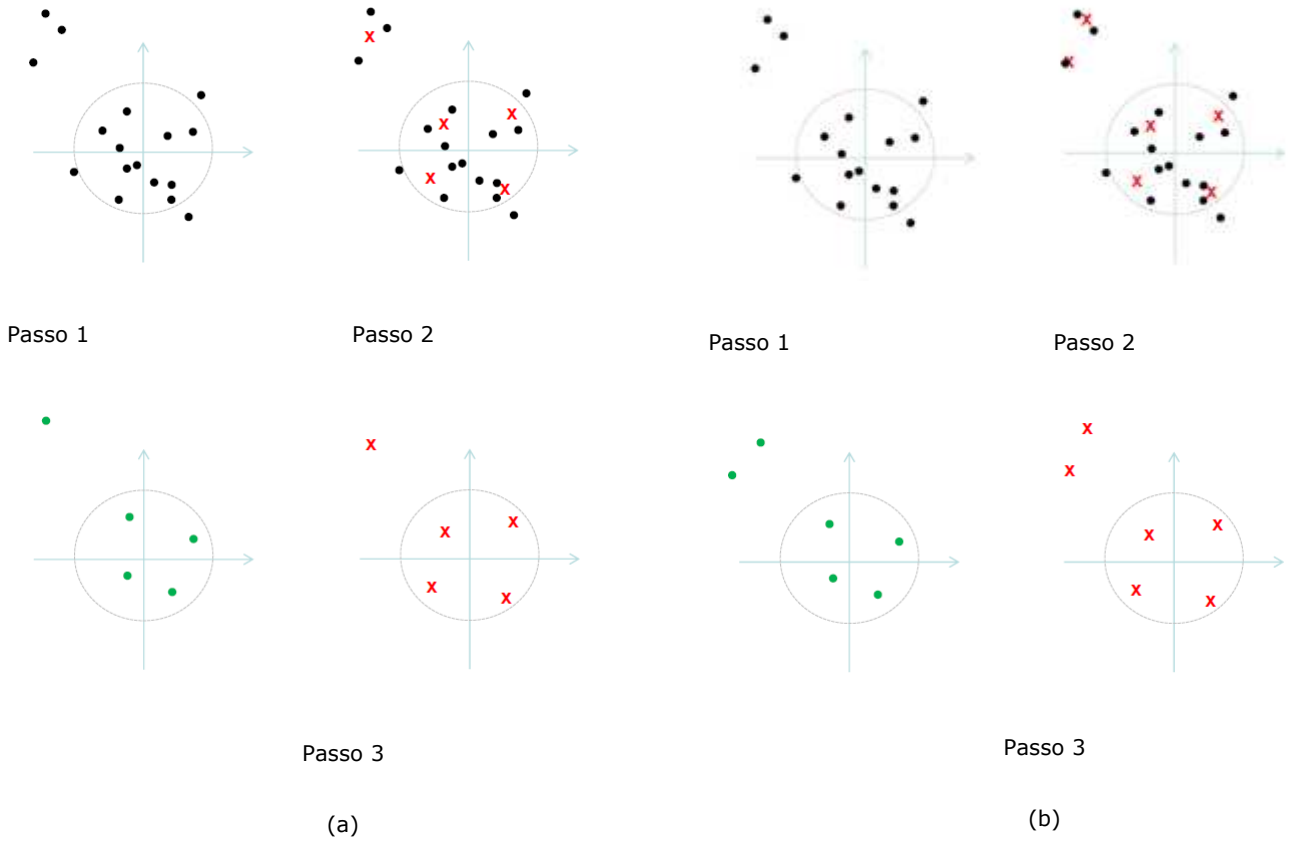



Figura 3.6: Exemplo ilustrativo AS (a) 5 cenários e (b) 6 cenários

Neste exemplo, a AS é resumida em 3 passos: no Passo 1 é gerada a amostra original, no Passo 2 são formados os grupos e, finalmente, no Passo 3 um representante é selecionado para cada grupo. Na Figura 3.6, a amostra original é representada pelos pontos pretos, o centroide de cada grupo é representado por um "x" vermelho e o objeto mais próximo do centroide, por um ponto verde. Observe que o cenário mais extremo do caso com 6 grupos é menor do que aquele representado no caso com 5 grupos, no caso cujo representante do grupo é o objeto mais próximo, o que pode levar a um custo esperado de operação mais baixo. Na solução do problema de árvore completa, este comportamento não é esperado quando se utiliza o centroide como representante do grupo. À medida que se aumenta o número de grupos, o cenário mais extremo obtido com a representação do centroide tende ao cenário mais extremo da amostra original. Já com a representação pelo objeto mais próximo, este comportamento pode ser oscilatório.

# 4 RESULTADOS COM APLICAÇÃO DO CENTROIDE

Os resultados apresentados nesta seção foram obtidos com um caso do Programa Mensal de Operação (PMO Setembro 2017), considerando 123 usinas hidroelétricas e 82 usinas termoelétricas, com aproximadamente 80GW e 19GW de capacidade instalada, respectivamente. Foi considerado um horizonte de planejamento de médio prazo de 5 anos, compreendendo os anos de 2017 a 2021, discretizado em estágios mensais. O sistema é dividido em 12 reservatórios equivalentes de energia: Sudeste, Madeira, Teles Pires, Paraná, Itaipu, Paranapanema, Sul, Iguaçu, Nordeste, Norte, Belo Monte e Manaus, conforme ilustrado na Figura 4.1.

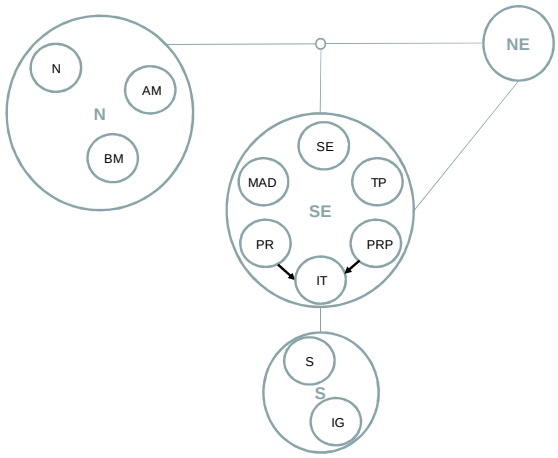
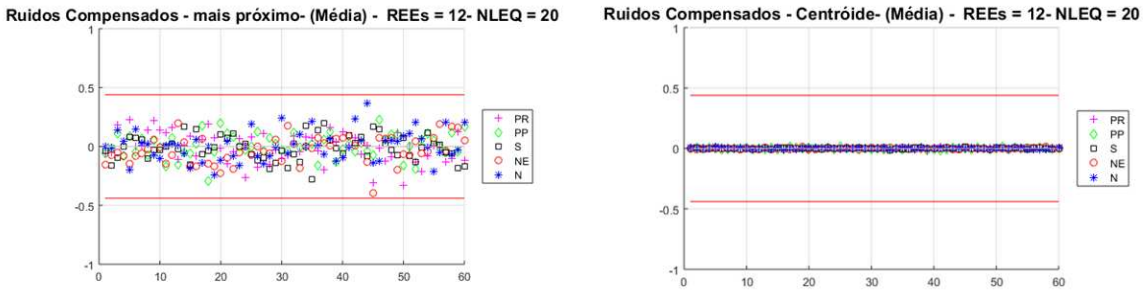



Figura 4.1 – Topologias do SIN com 12 REEs.

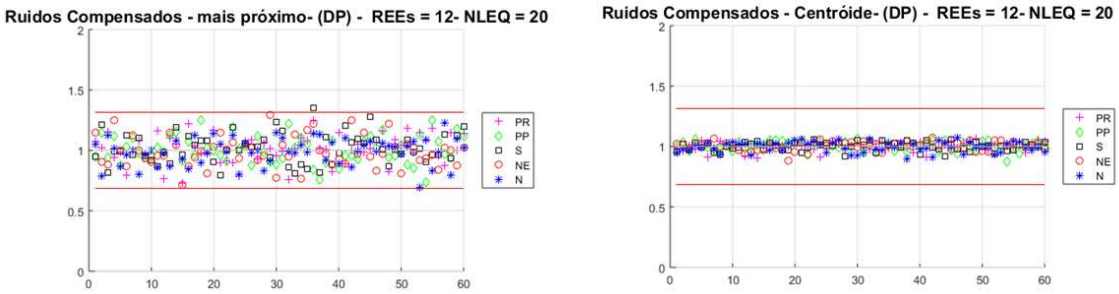
As Figuras 4.2 (a) a (d) apresentam os resultados obtidos após o processo de agregação de uma amostra original de 100.000 vetores de ruídos com 12 dimensões (número atual de REEs), que foram agregados para 20 vetores (número de aberturas *backward*) com 12 dimensões, ao longo de 60 períodos. As figuras à esquerda são referentes aos resultados obtidos quando o representante de cada grupo é o objeto mais próximo (metodologia atualmente utilizada). À direita, apresentam-se os resultados obtidos quando o representante de cada grupo é o próprio centroide do grupo. Para facilitar a visualização, escolheram-se apenas cinco dos doze REEs. As linhas em vermelho mostram o intervalo de confiança de 95%.

Dado que ambos os casos foram obtidos após o processo de agregação de uma amostra original com uma distribuição normal padrão, tem-se como resultado ideal que a média de cada vetor de ruídos seja igual a zero, e possua um desvio padrão igual a um. É possível observar que as amostras resultantes do processo de agregação com o centroide apresentam resultados bem superiores quando comparado com a amostra do mais próximo, tanto em média quanto em desvio padrão.




(a) (b)

Figura 4.2: Análise da dispersão da média dos vetores de ruídos *backward* (a) mais próximo (b) centroide




(c) (d)

Figura 4.2: Análise da dispersão do desvio padrão dos vetores de ruídos *backward* (c) mais próximo (d) centroide

Os resultados a seguir foram obtidos a partir da simulação da operação do sistema elétrico interligado com 2.000 cenários de afluências sintéticas aos REEs ao longo do período de planejamento, considerando-se a política de operação calculada por PDDE. No algoritmo de PDDE foram considerados 200 cenários para a simulação *forward* e 20 cenários para a recursão *backward*. O mecanismo de aversão a risco adotado foi o CvaR ($\alpha$=50, $\lambda$=40).

A próxima análise a ser feita, visa avaliar a reprodução das estatísticas históricas de ENA nos cenários sintéticos na amostra *backward*, que possui um total de 4.000 cenários (200 séries *forward* x 20 aberturas *backward*). Dado que a geração de cenários sintéticos é feita condicionada ao passado recente, é de se esperar que os primeiros períodos possam apresentar médias e desvios padrão diferentes do histórico. Porém, a medida que se avança no horizonte, esses valores devem passar a reproduzir o histórico. Comparando-se as Figuras 4.3 (mais próximo) com a Figura 4.4 (centroide) é possível observar uma ligeira vantagem das médias dos cenários resultantes do centroide, principalmente nos REEs 4, 6, 7 e 8.

Os resultados de desvio padrão são apresentados nas Figuras 4.5 (mais próximo) e 4.6 (centroide), onde nota-se uma ampla superioridade dos resultados obtidos com centroide. Pode-se observar que as estatísticas resultantes do centroide são mais aderentes ao histórico, quase sempre estando sobrepostas as duas curvas, enquanto as estatísticas resultantes da amostra original por algumas

vezes apresentam diferenças com relação ao histórico. A diferença observada nos primeiros períodos, deve-se ao condicionamento ao passado recente dos cenários gerados.

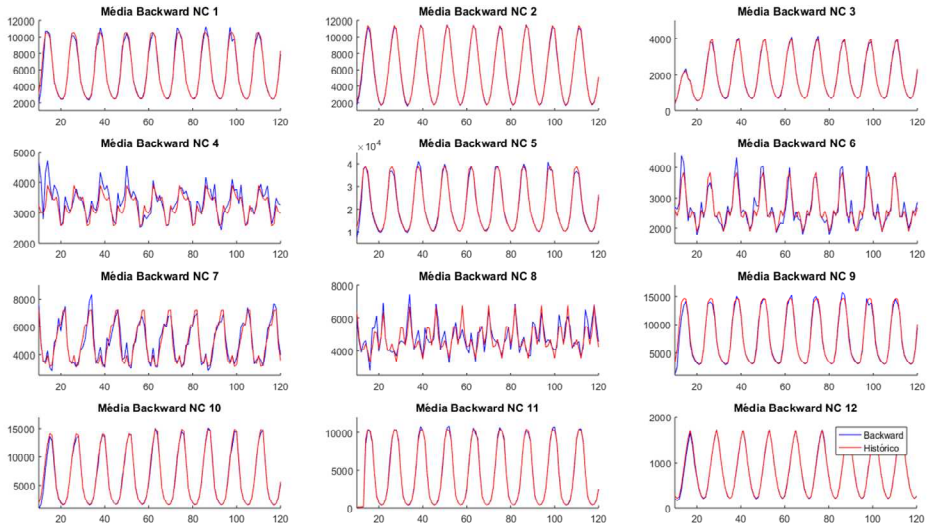



Figura 4.3: ENA média histórica x ENA média obtida na amostra *backward* (mais próximo)

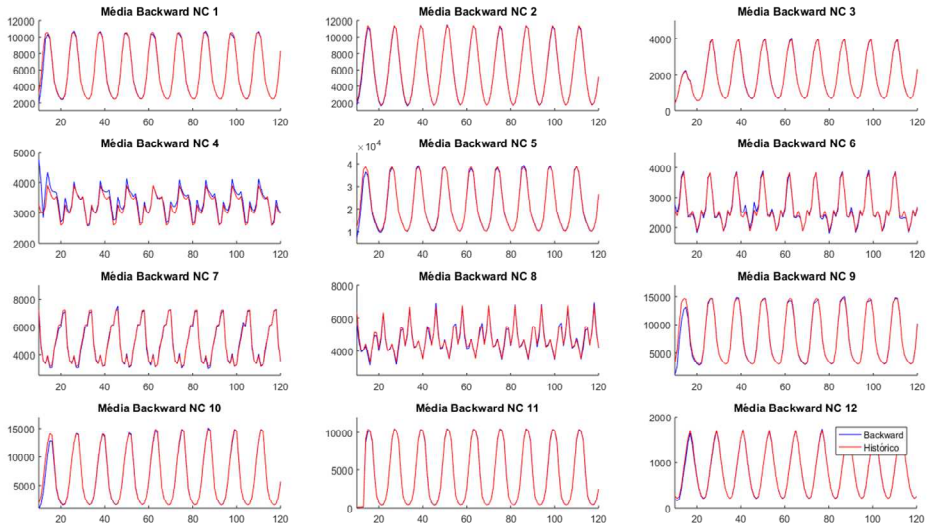



Figura 4.4: ENA média histórica x ENA média na amostra *backward* (centroide)

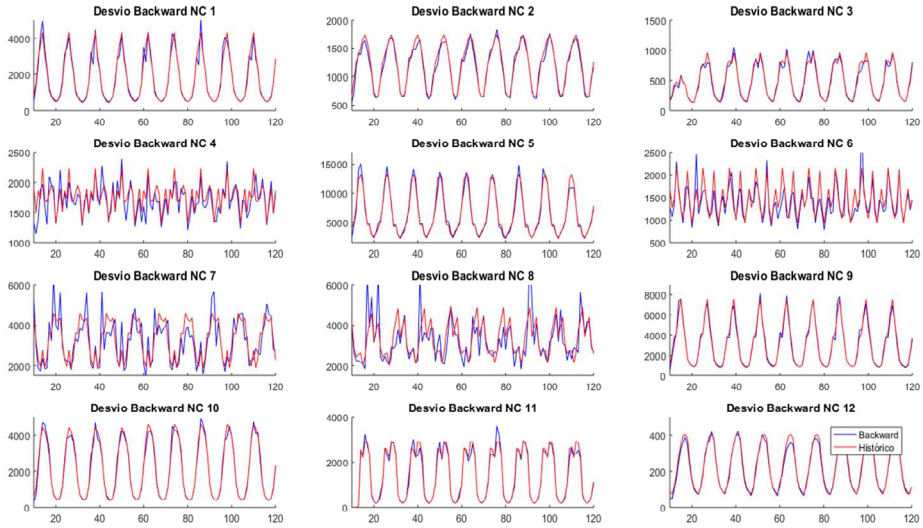



Figura 4.5: Desvio padrão histórico x desvio padrão da amostra *backward* (mais próximo)

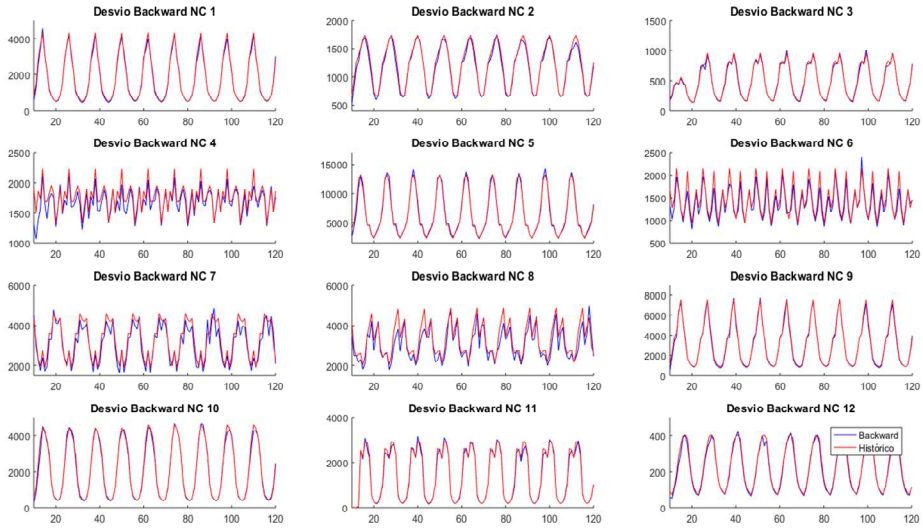



Figura 4.6: Desvio padrão histórico x desvio padrão da amostra *backward* (centroide)

A seguir apresentam-se na Figura 4.7 (a) e (b) as distribuições de frequência acumuladas dos cenários de ENA *backward* para o REE Paraná, entre janeiro e abril de 2018 para o PMO de setembro de 2017. À esquerda, apresentam-se os resultados originais, e à direita os resultados com centroide. Optou-se por mostrar esse intervalo de tempo, pois conforme o relatório técnico ”Análise do Comportamento do Custo Marginal de Operação do Modelo Newave ao se Passar de 9 para 12 Reservatórios Equivalentes de Energia” [15], em janeiro de 2018 do referido PMO havia uma inflexão da curva de evolução do CMO. As curvas em azul representam as distribuições condicionadas teóricas em cada período, obtidas através de 2.000 cenários da simulação final, enquanto em vermelho

mostram-se as distribuições de cenários *backward*. É possível observar um grande descolamento das distribuições de frequência em janeiro, devido a uma concentração de cenários mais úmidos que o esperado, o que não ocorre na amostra *backward* obtida com o centroide.

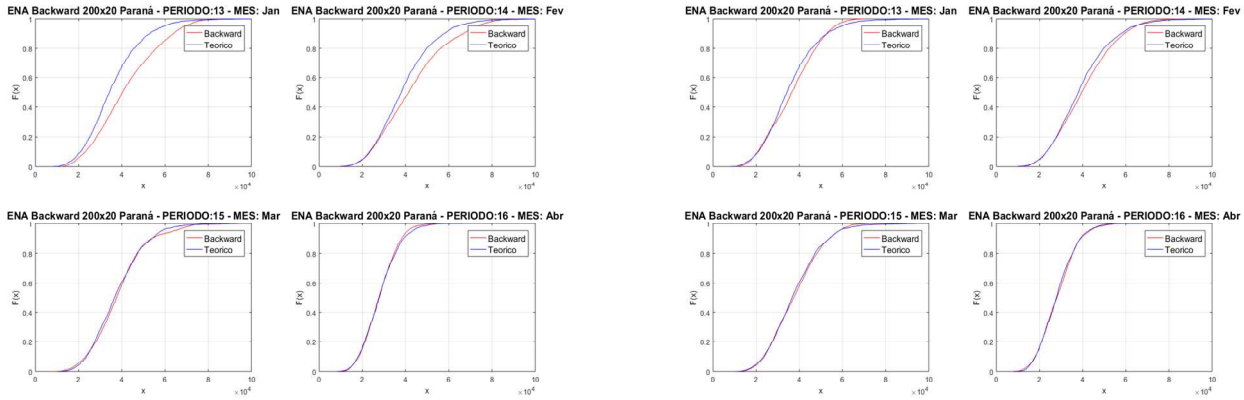



Figura 4.7: Função de distribuição acumulada ENA *backward* – Paraná (a) mais próximo (b) centroide

Passada a análise dos resultados da geração de cenários sintéticos de ENA, apresenta-se a seguir uma avaliação da variabilidade dos resultados obtidos com o modelo NEWAVE, variando-se a semente inicial da amostra *backward*, e considerando reamostragem plena dos cenários da *forward* com passo 3. Foram realizadas onze rodadas (onze sementes diferentes) do PMO de setembro de 2017 utilizando-se o objeto mais próximo como o representante do processo de agregação dos ruídos e outras onze rodadas com o centroide sendo o representante. A Figura 4.8 traz os resultados de ZINF com o mais próximo (barras em azul) e centroide (barras em vermelho) e suas respectivas médias representadas pela linha preta. Os mesmos resultados são apresentados na forma de *boxplots* na Figura 4.9.

É possível observar que há uma considerável diminuição da variabilidade dos resultados entre as diferentes sementes *backward*. Para uma dada árvore de cenários (semente *backward* fixa), não há um comportamento esperado para o valor de ZINF entre os casos centroide e mais próximo. Adicionalmente, quando se considera reamostragem de cenários *forward*, espera-se que as médias dos ZINFs entre várias sementes sejam próximas, pois uma cobertura mais exaustiva da árvore completa é obtida, seja ela adotando-se como representante no processo de agregação da Amostagem Seletiva, o centroide ou o representante mais próximo.

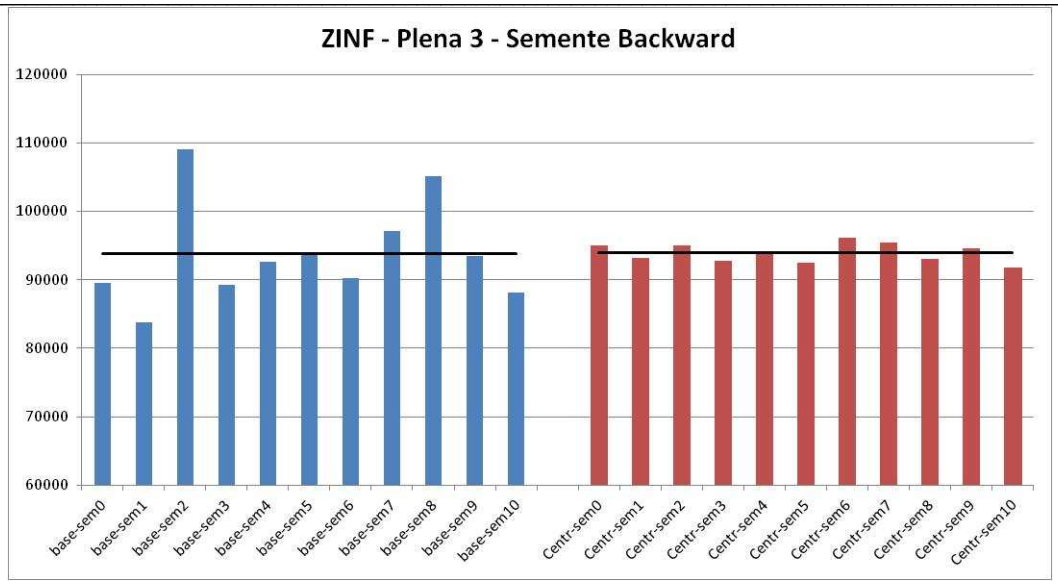



Figura 4.8: ZINF do PMO de setembro de 2017 variando-se a semente *backward* representantes mais próximo (azul) e centroide (vermelho)

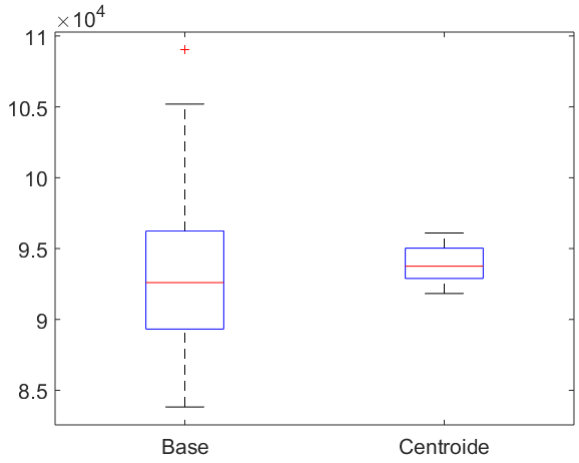



Figura 4.9: *Boxplot* do ZINF do PMO de setembro de 2017 variando-se a semente *backward*

As Figuras 4.10 e 4.11 trazem os resultados do valor esperado do custo total de operação (COPER) para os casos citados anteriormente. Novamente é possível observar uma considerável redução da variabilidade dos resultados quando utilizado o centroide como objeto representante no processo de agregação.

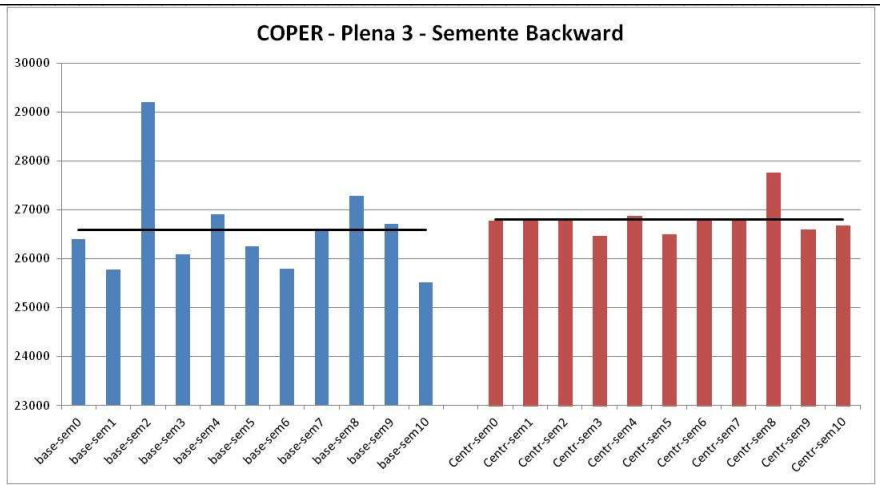



Figura 4.10: COPER do PMO de setembro de 2017 variando-se a semente *backward*, representantes mais próximo (azul) e centroide (vermelho)

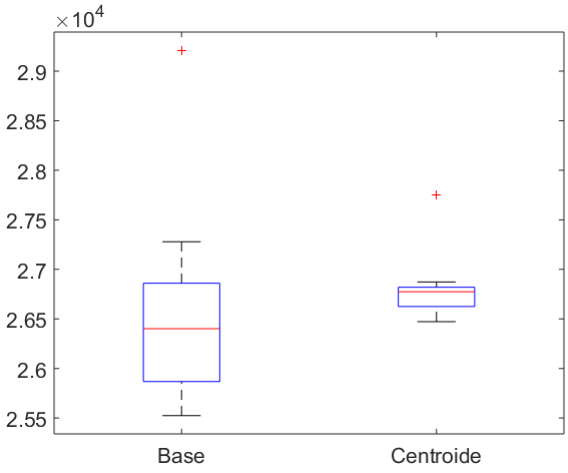
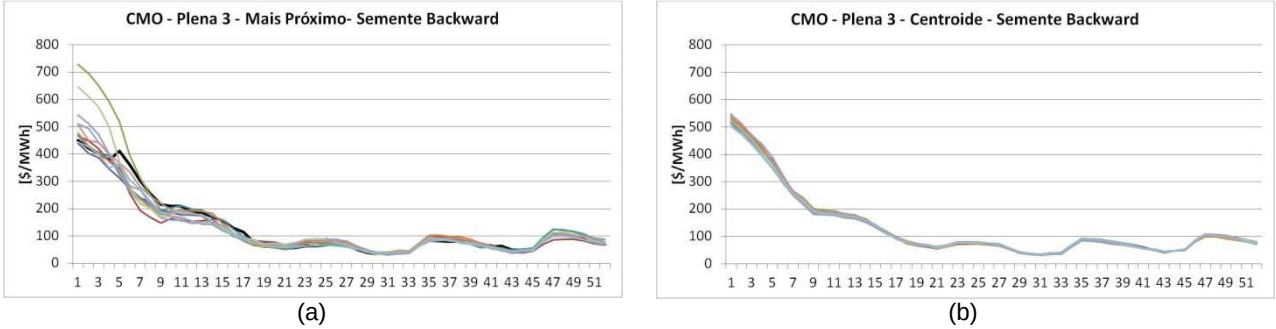



Figura 4.11: *Boxplot* do COPER do PMO de setembro de 2017 variando-se a semente *backward*

A Figura 4.12 apresenta a evolução do CMO para o subsistema Sudeste nos 11 casos, com o mais próximo (à esquerda) e com centroide (à direita). Novamente é possível observar uma redução da variabilidade dos resultados obtidos com o centroide. Além disso, verifica-se que a inflexão da evolução do CMO que ocorre na semente 0 (utilizada em casos oficiais, destacada na Figura 4.12 (a)) não é observada com o centroide.




(a) (b)

Figura 4.12: Evolução do CMO do subsistema SE variando-se a semente *backward* (a) mais próximo (b) centroide

Por último, a Figura 4.13 traz o custo térmico (eixo x) de cada caso com relação à energia não-suprida - ENS (eixo y), para cada um dos 22 casos. Nota-se uma redução na variabilidade dos resultados tanto da ENS quanto do custo de térmica.

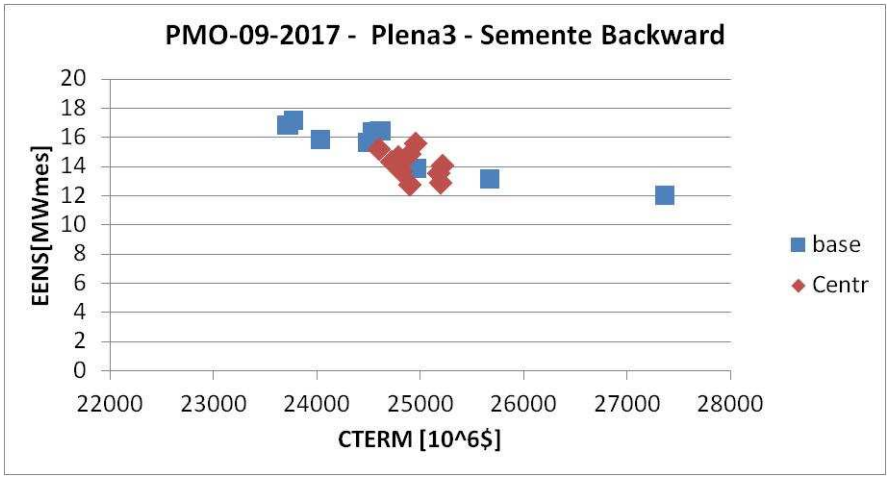



Figura 4.13: Evolução do CMO do subsistema SE variando-se a semente *backward*

A mesma estabilidade verificada nos resultados acima também pode ser obervada em variáveis associadas diretamente à operação do sistema, quando se considera o centroide como representante do grupo. Nas Figuras 4.14 a 4.16 são apresentadas as evoluções temporais da geração hidráulica total média do SIN e de cada submercado, calculada como o somatório da geração de todos os REEs que compõe o mercado, da geração térmica média e da energia armazenada final média.

Para o mercado Norte a redução da variabilidade nos resultados e geração hidráulica total, e consequentemente na energia armazenada final, não é tão evidente quanto para os demais mercados e para o SIN. Com relação à geração térmica, a redução na variabilidade pode ser facilmente verificada para todos os mercados e o SIN.

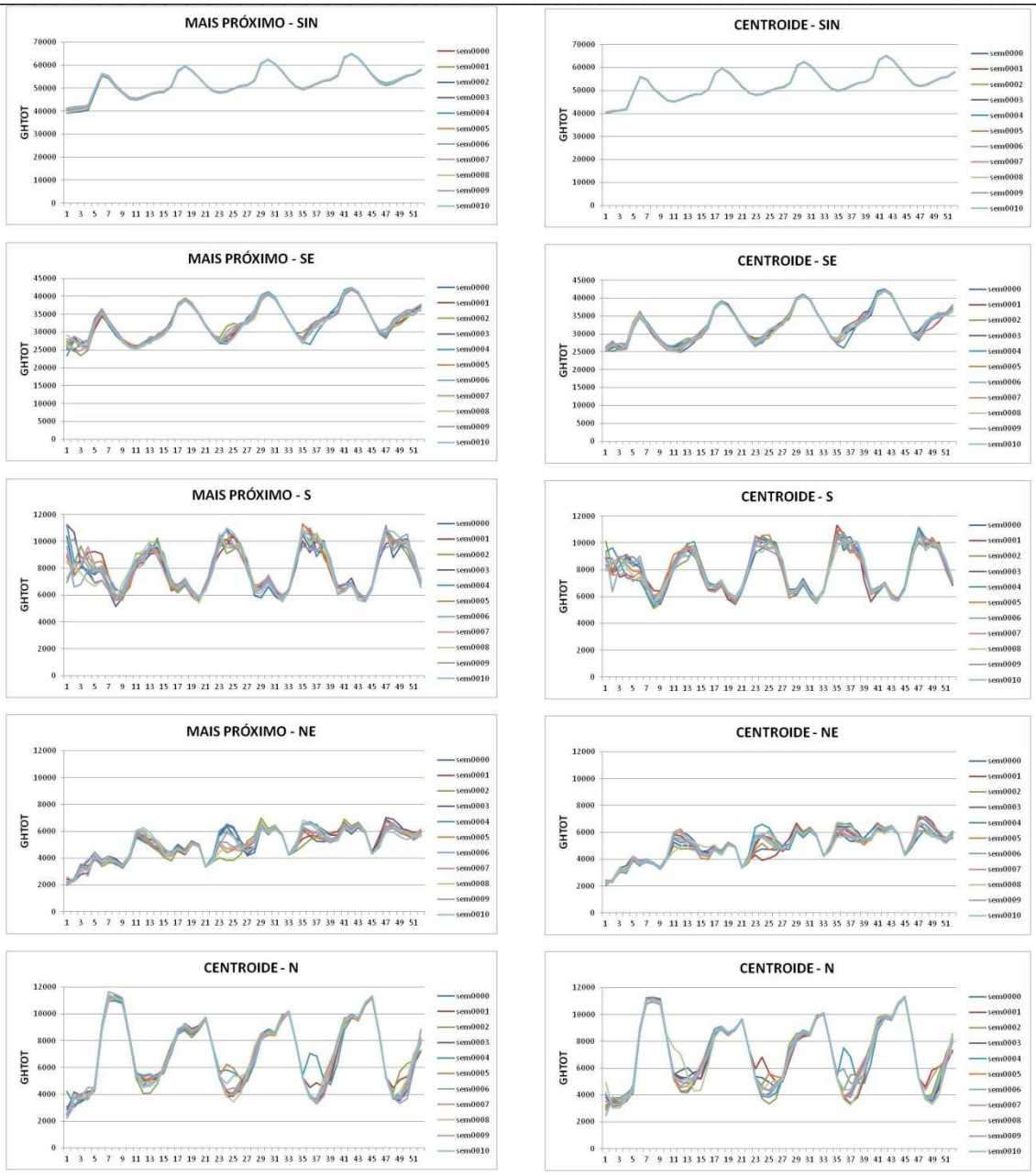



Figura 4.14: Evolução do GHTOT (MWmês) do SIN, SE, S, NE e N variando-se a semente *backward*

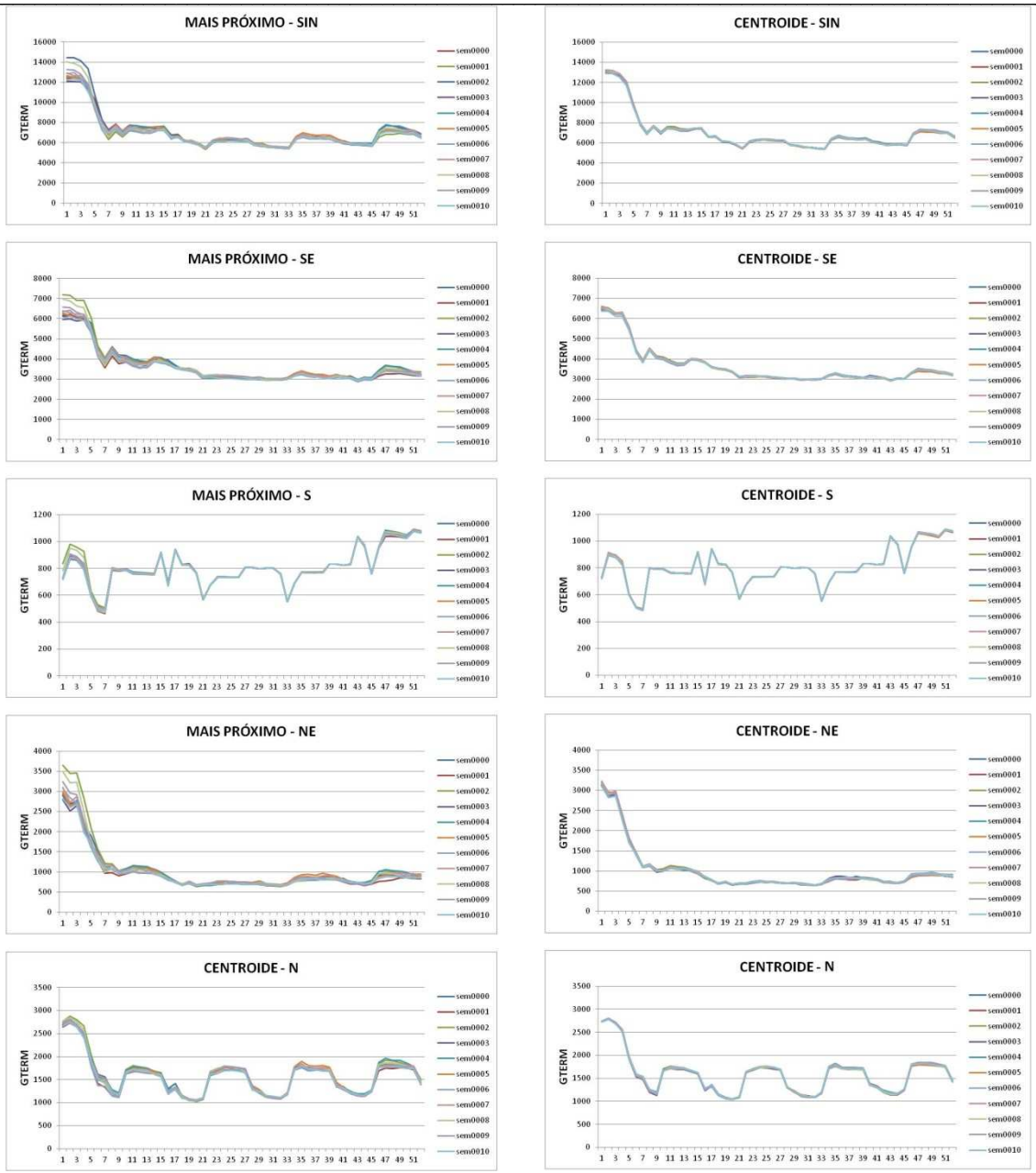



Figura 4.15: Evolução do GTERM (MWmês) do SIN, SE, S, NE e N variando-se a semente *backward*

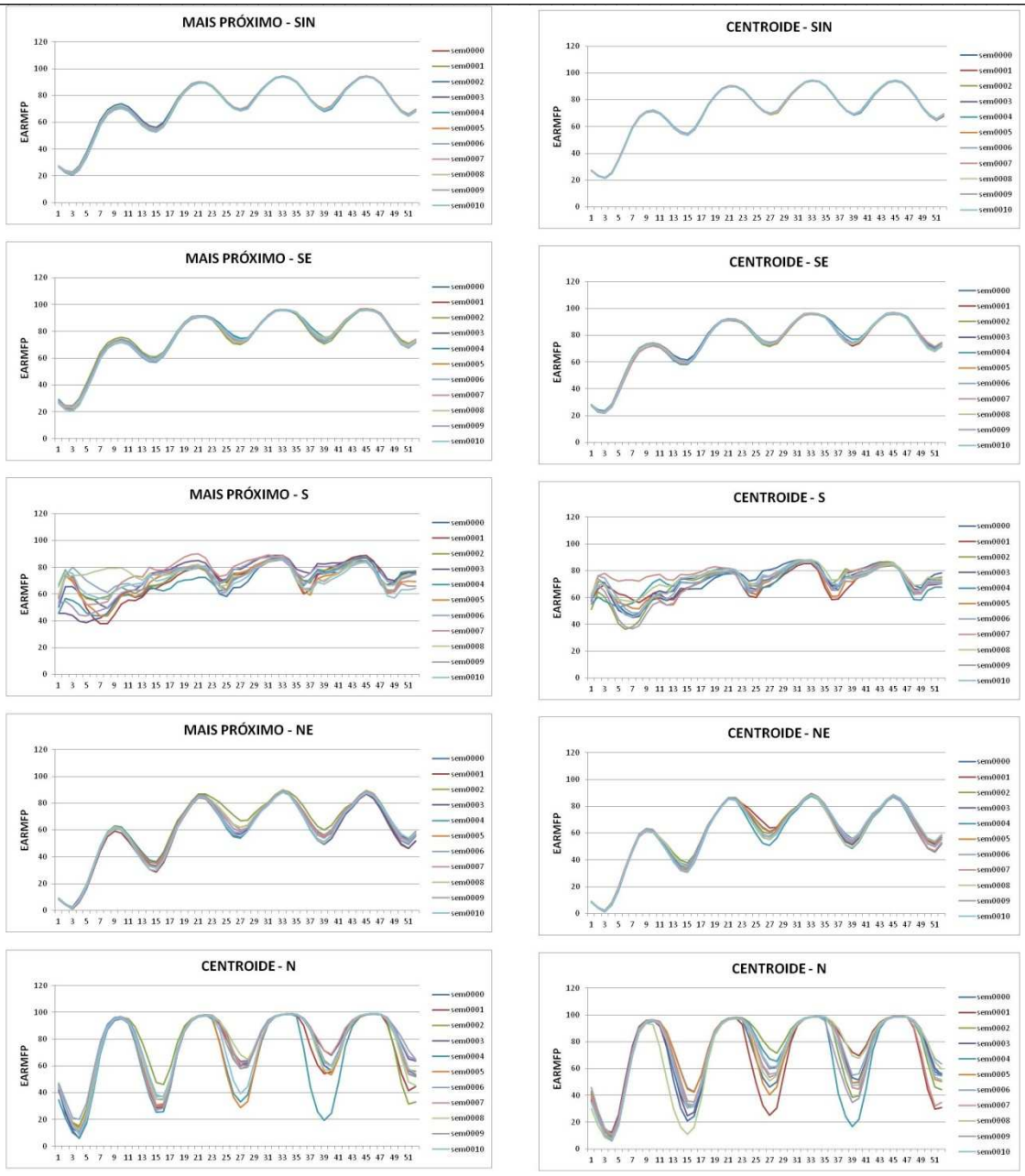



Figura 4.16: Evolução do EARMf (%) do SIN, SE, S, NE e N variando-se a semente *backward*

# 5 CONCLUSÕES

Durante o processo de validação da mudança de configuração do SIN, de 9 para 12 reservatórios equivalentes de energia (REEs), foram rodados os casos de PMO do ano de 2017 com a nova representação dos REEs. A partir destas análises, observou-se um comportamento de redução do custo marginal de operação obtido nas simulações finais na configuração com maior número de REEs. Foram elencadas três possíveis causas para o comportamento observado no custo marginal de operação (CMO), dentre elas, a variação amostral na representação da árvore de cenários de afluências no algoritmo da PDDE (recursão *backward*).

Durante o ano de 2018 foi introduzida no modelo NEWAVE a reamostragem de cenários *forward*, que reduziu a variabilidade dos resultados decorrentes do planejamento da operação em relação a diferentes sementes para a geração de cenários *forward*. No entanto, foram realizadas sensibilidades com relação à árvore de cenários utilizada na solução do problema de planejamento da operação (árvore completa) e, observou-se que ainda havia certa variabilidade dos resultados em relação a diferentes sementes para a geração de cenários *backward,*o que, em princípio, não é contornado apenas com a reamostragem de cenários da simulação *forward*.

O CEPEL investigou então aprimoramentos no processo de geração de cenários de afluências de forma a minimizar a ocorrência de cenários extremamente atípicos e a variabilidade amostral. O presente relatório apresentou uma alternativa para a escolha do representante do processo de agregação do processo de Amostragem Seletiva (AS), utilizando-se o centróide de cada grupo, o que resultou em uma redução na variabilidade amostral observada nos resultados do planejamento da operação de médio e longo prazos em relação a variações da semente para a geração de cenários *backward*. Os resultados apresentados neste relatório basearam-se no PMO de Setembro de 2017, mas foram comprovados em diferentes estudos para o planejamento da expansão e operação, apresentados durante as reuniões do GT-Metodologia/CPAMP ocorridas durante o ano de 2018 pelas Instituições participantes.

# 6 REFERÊNCIAS

[1] D.D.J. Penna, M.E.P. Maceira, J.M. Damázio, F.Treistman, H.S. Araújo, A.C.G. Melo, "Análise da inflexão do custo marginal de operação no modelo NEWAVE entre os quarto e quinto estágios temporais ao se adotar a representação de 12 REEs para Sistema Interligado Nacional", Relatório Técnico CEPEL no 10783/2018, Setembro de 2018.

[2] M.E.P. Maceira, D.D.J. Penna, A.L. Diniz, R.J. Pinto, A.C.G. Melo, C.V. Vasconcellos, C.B. Cruz, "Twenty years of application of stochastic dual dynamic Programming in official and agent studies in Brazil – Main features and improvements on theNEWAVE model", 20th PSCC – Power Systems Computation conference, Dublin, Ireland, 2018.

[3] CEPEL, Centro de Pesquisas de Energia Elétrica - Manual do Usuário do modelo NEWAVE, versão 22, Dezembro de 2015.

[4] M.V.F. Pereira, L.M.V.G. Pinto, "Multi-stage stochastic optimization applied to energy planning", Mathematical Programming, v. 52, n.1-3, pp. 359-375, Maio 1991

[5] M.E.P. Maceira, "Programação Dinâmica Dual Estocástica Aplicada ao Planejamento da Operação Energética de Sistemas Hidrotérmicos com Representação do Processo Estocástico de Afluências por Modelos Auto-Regressivos Periódicos", Relatório Técnico Cepel, Junho 1993.

[6] M. E. P. Maceira, C. V. Bezerra, "Stochastic Streamflow model for Hydroelectric Systems", In: Proceedings of 5th International Conference on Probabilistic Methods Applied to Power Systems, pp. 305-310, Vancouver, Canada, Set. 1997.

[7] D. D. J. Penna, M. E. P. Maceira, J.M. Damázio, "Selective sampling applied to long-term hydrothermal generation planning", 17th PSCC - Power Systems Computation Conferfence, Stockholm, Sweden, Ago. 2011.

[8] M.E.P. Maceira, L.A. Terry, F.S. Costa, J.M. Damazio, A.C.G. Melo, "Chain of Optimization Models for Setting the Energy Dispatch and Spot Price in the Brazilian System", 14th PSCC – Power Systems Computation conference, Sevilla, Spain, 2002.

[9] A. Kleywegt, A. Shapiro, T. Homem-de-Mello, "The sample average approximation method for stochastic discrete optimization", Siam Journal on Optimizaton, v.12, pp. 479-502, 2001.

[10] D.D.J. Penna, M.E.P. Maceira, J.M.Damázio, F. Treistman, H.S. Araújo, Manual de Referência do modelo GEVAZP, Relatório Técnico CEPEL – nº 27155/2017, 2017.

[11] J.D. Salas, J.W. Delleur, V. Yevjevich, W.L. Lane, "Applied Modeling of Hydrologic Time Series", Water Resources Publications, 1980.

[12] G.E.P.Box, D.R.Cox, "An Analysis of Transformations", Journal of The Royal Statistical Society, A127, pp. 211-252, 1964.

[13] R. J. Charbeneau, "Comparison of the two and three parameter lognormal distributions used in streamflow synthesis", Water Resources Research, Vol. 14, No. 1, pp. 149-150, 1978.

[14] J.Hartigan, M. Wong, "A K-Means Clustering Algorithm", Applied Statistics, vol.28, no. 1, pp. 100-108, 1979.

[15] D.D.J. Penna, M.E.P. Maceira, A. Diniz, A.C.G. Melo, F. Treistman, "Análise do Comportamento do Custo Marginal de Operação do Modelo Newave ao se Passar de 9 para 12 Reservatórios Equivalentes de Energia", Relatório Técnico CEPEL no 27538/2017, Dezembro de 2017.

# 7 ANEXO – NOTA TÉCNICA 42 DO PROJETO NEWAVE

**Centro de Pesquisas de Energia Elétrica - CEPEL**

**Relatório Técnico – 13612 / 2018**

Nota Técnica nº 42

# Aplicação de Técnicas de Agregação na Geração de Cenários Hidrológicos para o Planejamento de Médio Prazo

**Projeto NEWAVE**

**Abril 2010**
**(revisão 3)**

# 1- Introdução

O modelo NEWAVE, utilizado no planejamento da operação de médio prazo, define para cada mês do período de planejamento, que pode variar de 5 a 10 anos, a alocação ótima dos recursos hídricos e térmicos de forma a minimizar o valor esperado do custo de operação ao longo de todo o período de planejamento. Essa estratégia é definida por uma política de operação ótima representada por uma função de custo futuro estimada utilizando a programação dinâmica dual estocástica. Além disso, o parque hidrelétrico é representado de forma agregada e a estocasticidade das afluências é representada por um modelo estocástico periódico auto-regressivo de ordem p.

A modelagem estocástica referente às afluências é considerada de forma explícita no cálculo da função de custo futuro e de forma implícita nas simulações, através do uso de cenários hidrológicos multivariados gerados sinteticamente. O conjunto de todas as possíveis realizações do processo estocástico de afluências, ao longo de todo horizonte de planejamento, forma uma árvore de cenários. Esta árvore representa todo o universo probabilístico sobre o qual será efetuado o processo de otimização da operação energética.

Como a árvore de cenários do problema de planejamento de médio prazo possui uma cardinalidade bastante elevada, igual ao número de aberturas elevado ao número de estágios do horizonte de planejamento (normalmente igual a $20^{120}$), torna-se impossível do ponto de vista computacional percorrer completamente a árvore. Portanto, apenas uma porção da árvore (sub-árvore) é percorrida. Atualmente a sub-árvore é escolhida utilizando o método de Monte-Carlo clássico que usa a amostragem aleatória simples.

Nesta Nota Técnica é descrito um método para a definição da sub-árvore a ser visitada durante o processo de cálculo da estratégia ótima de operação com o intuito de tornar mais robusto os resultados obtidos por esta política de operação, com relação a variações no número de cenários de simulação forward e backward, e com relação à amostra de cenários hidrológicos utilizada. Para tanto se propõe que sejam aplicados ao modelo de geração de cenários hidrológicos multivariados técnicas estatísticas multivariadas capazes de elaborar critérios que possibilitam agrupar objetos similares em determinados grupos (técnicas de agregação). Estas técnicas podem ser reunidas sob o nome genérico de Análise de Conglomerados.

Usando as técnicas de agregação pretende-se escolher um conjunto representativo de cenários hidrológicos a partir de um grande número de cenários, reduzindo a variação nos resultados do modelo de médio prazo do planejamento da operação, sem deixar de representar de forma adequada o processo estocástico das afluências.

## 2- Análise de Conglomerados

### 2.1 – Considerações Gerais

A Análise de Conglomerados é usada para reduzir uma grande massa de dados, na medida em que possibilita a partição/classificação dos dados em um número menor de grupos. Também é utilizada para desenvolver hipóteses a respeito da natureza dos dados ou para examinar hipóteses previamente estabelecidas. Representa uma poderosa ferramenta com aplicações em diversos problemas de formação de grupos. Elas podem ser empregadas, por exemplo, para identificar padrões similares de demanda de energia elétrica, na construção de segmentos de mercados, para agrupar programas de TV em tipos similares de acordo com tendências registradas de audiência, etc.

A Análise de Conglomerados tem grande aplicação na pesquisa científica em diversas áreas do conhecimento. Na literatura existem vários trabalhos que utilizam técnicas de agregação. Na linha de estudos elétricos pode-se citar trabalhos que empregam as técnicas de agregação para construção da árvore de cenários hidrológicos para o planejamento da operação de curto-prazo (JARDIM et al, 2002) e na caracterização de curvas de carga (VELASQUEZ et al., 2001). Na área das Ciências da Computação, a Análise de Conglomerados está sendo amplamente utilizada para a classificação e comparação de documentos na Internet (STEINBACH et al., 2000). As Ciências Sociais também a utilizam para a realização de diversos estudos como os citados em ALDENDERFER e BLASHFIELD (1984). Existem ainda outras aplicações nas áreas de Ecologia (VALENTIN, 2000), Marketing (ZIKMUND, 1999) e Finanças (FARREL, 1997). Em HARTIGAN (1975) são mostrados diversos trabalhos em áreas distintas que empregam as técnicas de agregação.

Segundo JOHNSON e WICHERN (1998) a utilização de técnicas de agregação como procedimento exploratório é importante para o estudo da natureza complexa das inter-relações multivariáveis, indicando que a formulação de uma estrutura de agrupamentos naturais a partir de dados observados pode produzir meios informais para avaliar dimensionalidade, identificar casos marginais e sugestões de hipóteses sobre correlações. A Análise de Conglomerados difere dos métodos tradicionais de classificação porque, enquanto estes procuram associar novos itens a classes pré-determinadas, no caso de agrupamento não há nenhum conhecimento prévio acerca

da estrutura de agrupamentos, e a divisão em grupos baseia-se unicamente nas similaridades ou diferenças entre os objetos, observados através dos dados coletados.

A utilização das técnicas de agregação na pesquisa científica vem aumentando devido à disponibilidade de computadores mais capazes e poderosos. Em geral, para um conjunto de objetos que se deseja agrupar, o número de estrutura de grupos possíveis é extremamente alto, de modo que se devem procurar algoritmos que levem em consideração algumas restrições prévias como forma de se reduzir o esforço matemático e determinar uma estrutura que seja satisfatória segundo um critério escolhido.

Tendo pesquisado alguns exemplos de aplicação de técnicas de agregação ALDEANDERFER e BLASHFIELD (1984) afirmam que, apesar das diferenças quanto a objetivos, tipos de dados e métodos usados em cada experiência, os cincos procedimentos básicos necessários para caracterizar todos os estudos de Análise de Conglomerados são:

1 Seleção de uma amostra de objetos que deverão ser agrupados

2 Definição de um conjunto de variáveis que serão medidas para todos os objetos da amostra

3 Cálculo das similaridades entre os objetos

4 Utilização de um método de Análise de Conglomerados para gerar grupos de objetos similares

5 Validação da estrutura dos grupos resultantes

## 2.2 - Medidas de Similaridade

A escolha de um critério que quantifique o grau de associação entre os objetos ou variáveis tem um papel crucial nos estudos que utilizam a Análise de Conglomerados. Esta medida é chamada de coeficiente de similaridade e pode ser classificada como medida de similaridade ou de dissimilaridade. Na primeira, quanto maior for o valor da medida, mais similares são os objetos, enquanto que na segunda quanto menor o valor observado, mais parecidos são os objetos. De modo geral, medidas de dissimilaridade podem ser convertidas em medidas de similaridade através de uma relação inversa.

Apesar da aparente simplicidade, o conceito de similaridade e, especialmente, os procedimentos usados para medir a similaridade estão longe de serem simples. A similaridade entre os objetos pode ser medida de várias maneiras, dentre as quais as três que mais se destacam são as medidas de correlação, as medidas de distância, e as medidas de associação. As duas primeiras medidas requerem dados quantitativos enquanto que a última trata de dados qualitativos.

As medidas de distância representam a similaridade como a proximidade de um objeto a outro através de suas variáveis. Como é a medida mais intuitiva, as medidas de distâncias se tornaram as mais difundidas e utilizadas. Na verdade, medidas de distância representam uma medida de não similaridade, pois quanto maior a distância entre dois objetos maior a diferença entre eles. A distância é então convertida em uma medida de similaridade através do uso de uma relação inversa.

A estimação quantitativa da similaridade tem sido dominada pelo conceito de métrica. Objetos são representados como pontos no espaço e a similaridade entre eles é medida através da distância entre os pontos. A dimensionalidade do espaço é determinada pelo número de atributos (variáveis) usados para descrever os objetos, por exemplo, $R^P$ caso os objetos sejam descritos por P variáveis.

Entre uma das mais populares representações de distância está a distância Euclidiana. Sejam $X_i = (x_{i1}, \ldots, x_{ip})$ e $X_j = (x_{j1}, \ldots, x_{jp})$ dois objetos caracterizados por P atributos, então a distância Euclidiana entre os dois objetos é definida como em (1a e 1b).

$$d_{i,j} = \sqrt{\sum_{r=1}^{p} (x_{ir} - x_{jr})^2} \qquad (1a)$$

ou de forma matricial:

$$d_{i,j} = \sqrt{(X_i - X_j)'(X_i - X_j)} \qquad (1b)$$

onde $d_{ij}$ é a distância entre os objetos i e j, $X_i$ é i-ésimo objeto e $x_{ir}$ é o valor da r-ésima variável do i-ésimo objeto.

Uma outra importante métrica é a distância de Mahalanobis, também conhecida como distância generalizada. Esta métrica é definida como em (2).

$$d_{i,j} = \sqrt{(X_i - X_j)'\Sigma^{-1}(X_i - X_j)} \quad (2)$$

onde $\Sigma^{-1} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & & \sigma_{2p} \\ \vdots & & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \cdots & \sigma_{pp} \end{bmatrix}^{-1}$ é a inversa da matriz de covariância.

Quando a matriz de covariância for igual a matriz identidade, isto é a correlação entre as variáveis for nula, a distância de Mahalanobis é equivalente a distância Euclidiana. Além de ponderar pela variabilidade de cada uma das variáveis, esta medida de distância considera também o grau de correlação entre elas.

Muitas medidas de distância são sensíveis a variações na escala ou na magnitude entre as variáveis. A forma mais comum de padronização é a conversão de cada variável em valores padrão, subtraindo-se pelo valor médio (3) e dividindo-se pelo seu respectivo desvio padrão (4). A transformação (5) resulta em uma variável com média zero e desvio igual a 1. Também elimina a influência introduzida pelo uso de diferentes escalas nas variáveis usadas na análise. Não existe diferença nos valores padrão quando a escala é alterada.

$$\mu_r = \frac{1}{N}\sum_{i=1}^{N} x_{ir} \quad (3)$$

onde N é o número de objetos.

$$\sigma_r = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{ir} - \mu_r)^2} \quad (4)$$

$$y_{ir} = \frac{x_{ir} - \mu_r}{\sigma_r} \quad (5)$$

As distâncias entre dois objetos podem ser organizadas na forma de uma matriz, conhecida como matriz de distâncias ou matriz de similaridade, ilustrada na Figura 1. A matriz de similaridade é uma matriz simétrica, de ordem igual ao número de objetos N, onde o elemento $d_{ij}$ é a medida de distância entre os objetos i e j.


Figura 1 – Matriz de Similaridades

## 2.3 - Métodos de Agrupamento

O principal objetivo quando se usa a Análise de Conglomerados é encontrar grupos de objetos similares em um conjunto de dados de tal forma que as variâncias entre os grupos seja máxima, e dentro deles, mínima. Considerando-se a enorme dificuldade em examinar todas as formas de agrupamentos possíveis, foram propostos vários algoritmos que promovem a divisão de objetos em grupos sem a necessidade de testar todas as configurações.

As técnicas de agregação constituem um meio para a redução da dimensionalidade de um conjunto de dados, pois se as classes obtidas forem internamente homogêneas, pode-se associar a cada classe um objeto típico, em geral a média dos objetos da classe, e assim, ao invés de analisar todo conjunto de dados, pode-se analisar apenas um pequeno número de objetos típicos, que capturam a maior parte da diversidade, ou melhor, da variância de todo conjunto.

Os algoritmos mais comumente utilizados para problemas de agregação podem ser classificados em duas categorias: (1) métodos hierárquicos e (2) métodos não hierárquicos.

### 2.3.1- Métodos Hierárquicos

As técnicas hierárquicas podem ser aglomerativas ou divisivas. Nos métodos aglomerativos, os objetos individuais são agrupados de acordo com suas similaridades, enquanto que os métodos divisivos partem de um único grupo de objetos que é sucessivamente dividido até que cada subgrupo contenha somente um objeto.

Os resultados de ambos podem ser apresentados graficamente na forma de um diagrama bidimensional denominado dendograma, que ilustra as fusões ou divisões realizados em níveis sucessivos. A Figura 2 mostra o processo aglomerativo sendo aplicado a 5 objetos (A,B,C,D e E). A cada etapa é mostrado o centróide dos grupos que vão se formando. Na etapa inicial todos os objetos estão sós em um grupo e na etapa final todos os objetos estão reunidos no mesmo grupo. O dendograma resultante desta seqüência de fusões é mostrado na Figura 3.

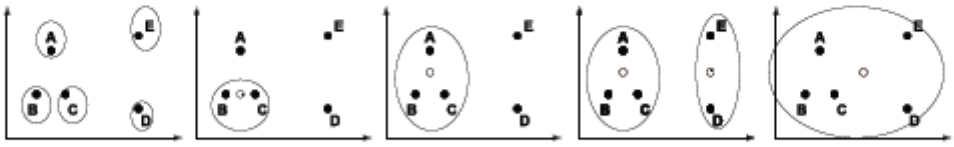



Figura 2 – Exemplo ilustrativo do processo aglomerativo


Figura 3 – Dendograma

O método é denominado hierárquico porque uma vez que dois objetos ou grupos são agrupados/separados, estes permanecem juntos/separados até o final da agregação, isto é, não há realocação dos objetos. Isto é uma desvantagem do método, pois se algum objeto for incorretamente agrupado em um estágio anterior não há possibilidade de realocá-lo em um estágio posterior. Uma outra desvantagem é a necessidade da construção e armazenamento da matriz de similaridade. A construção desta matriz pode representar uma limitação para a maioria das aplicações em microcomputadores, por este motivo os métodos hierárquicos não são indicados para conjuntos grandes de dados.

Mais detalhes sobre métodos hierárquicos, seus algoritmos e características podem ser encontrados em HARTIGAN (1975), ANDERBERGER (1973), HAIR JR. *et al.* (1998), DURAN e ODELL (1970) e JOHNSON e WICHERN (1998).

### 2.3.2- Métodos Não Hierárquicos

Nos métodos não hierárquicos os objetos são divididos em um número de grupos previamente fixado. Estes grupos são formados de modo que duas premissas básicas sejam atendidas: coesão interna e isolamento dos grupos.

Diferentemente dos métodos hierárquicos, as técnicas não hierárquicas não exigem a determinação e o armazenamento da matriz de similaridade, cuja ordem depende do número de objetos a ser analisados. Por este motivo, os métodos não hierárquicos são computacionalmente mais eficientes quando se trabalha com um grande conjunto de dados.

O caminho mais intuitivo para encontrar a melhor partição é checar todas as possíveis partições do conjunto de dados, porém o número de possibilidades é muito grande, assintoticamente de ordem de $K^{N-1}$, onde K é número de grupos e N o número de objetos que se deseja agrupar. Para resolver um problema de pequeno porte com 20 objetos e 3 grupos, é preciso investigar cerca de um bilhão de possíveis partições únicas. Dado a inviabilidade da análise de todas as partições possíveis, pesquisadores desenvolveram vários procedimentos heurísticos que investigam algumas partições com o intuito de encontrar a melhor partição, ou uma alternativa que seja quase ótima.

Dentre os procedimentos heurísticos desenvolvidos, o mais conhecido é o método K-Means. Este método, com pequenas variações, é um dos mais usados na Análise de Conglomerados quando se tem muitos objetos.

Mais informações sobre métodos não hierárquicos, suas características e sua utilização são encontradas em HARTIGAN (1975), ANDERBERGER (1973), ALDENDERFER e BLASHFIELD (1984), HAIR JR. *et al.* (1998), JOHNSON e WICHERN (1998) e BOUROCHE e SAPORTA (1980).

## 2.4 - Método K-Means

O primeiro passo deste método é formar uma partição inicial aleatória no conjunto de dados. O número de grupos deve ser estabelecido previamente. O próximo passo é o cálculo dos centróides destes grupos. Então, a distância entre cada objeto e cada centróide é calculada. Os objetos são realocados para o grupo que tiver o centróide mais próximo (menor distância). Este último passo é repetido até que não haja mais realocações de objetos. Vale a pena lembrar que toda vez que um objeto for

realocado os centróides devem ser recalculados. O algoritmo K-Means pode ser resumido nos seguintes passos:

1 Divida os N objetos em K agrupamentos através de uma partição inicial ou especificação de K centróides iniciais;

2 Realoque um objeto para o grupo cujo centróide é o mais próximo deste objeto e recalcule o centróide do grupo que recebeu e que perdeu o objeto;

3 Repita o passo 2 até que não haja mais realocações de objetos de um grupo para outro.

Com o intuito de aperfeiçoar, tornar mais rápido e mais eficiente o algoritmo apresentado, alguns procedimentos podem ser modificados, gerando assim variações deste método. A inicialização dos grupos pode ser feita de forma aleatória através do sorteio de pontos (objetos) para serem usados como semente inicial dos grupos ou pela partição aleatória do conjunto de dados. Os pontos sorteados podem ser sorteados de dentro do conjunto de dados ou não. Estes pontos também podem ser escolhidos um a um pelo especialista ou retirados de forma programada de dentro do conjunto de dados. Outra modificação que pode ser realizada é quanto à atualização dos centróides durante processo de realocação dos objetos. Esta atualização pode ser feita a cada vez que um objeto for realocado ou somente quando todos os objetos forem realocados. A primeira alternativa é a mais utilizada.

Para ilustrar como funciona o algoritmo do método não hierárquico descrito anteriormente, é utilizado um exemplo extraído de JOHNSON e WICHERN (1998). Considere um conjunto com 4 objetos (N=4) descritos por 2 variáveis (P=2), $x_1$ e $x_2$, onde se procura formar dois grupos (K=2). Os objetos são apresentados na Figura 4 e as suas coordenadas, na Tabela 1.

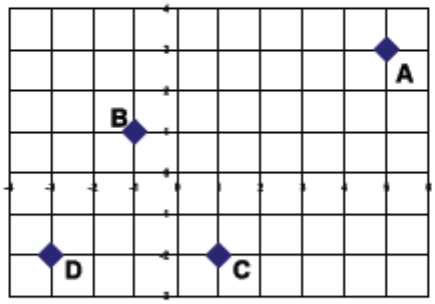



Figura 4 – Configuração dos 4 objetos

Tabela 1 – Coordenadas dos 4 objetos

| Objetos | Variáveis | |
|---|---|---|
| | $x_1$ | $X_2$ |
| A | 5 | 3 |
| B | -1 | 1 |
| C | 1 | -2 |
| D | -3 | -2 |

Para inicializar o processo de agregação, os objetos foram particionados em dois grupos AB e CD, Figura 5, e a partir deles são calculadas as coordenadas dos centróides, Tabela 2.

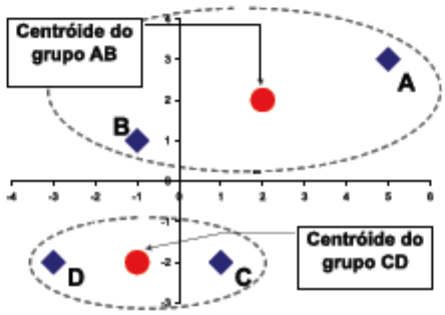



Figura 5 – Grupos AB e CD com os respectivos centróides

Tabela 2 – Coordenadas dos centróides (1ºetapa)

| Grupos | Variáveis | |
|---|---|---|
| | $x_1$ | $x_2$ |
| AB | $\frac{5+(-1)}{2}=2$ | $\frac{3+1}{2}=2$ |
| CD | $\frac{1+(-3)}{2}=-1$ | $\frac{-2+(-2)}{2}=-2$ |

Determinadas as coordenadas dos centróides, são calculadas as distâncias de cada objeto com relação aos centróides para verificar a necessidade de realocação. Neste exemplo é utilizada a distância Euclidiana.

Iniciando com o objeto A, se tem a seguinte distância:

$$D(A,(AB)) = \sqrt{(5-2)^2 + (3-2)^2} = 3.16$$

$$D(A,(CD)) = \sqrt{(5-(-1))^2 + (3-(-2))^2} = 7.81$$

Como a distância entre A e o centróide do grupo AB é menor que a distância entre A e o centróide do grupo CD, não há realocação, isto é, o objeto A permanece no grupo AB. Seguindo agora com o objeto B tem-se as seguintes distâncias ao quadrado:

$$D(B,(AB)) = \sqrt{(-1-2)^2 + (1-2)^2} = 3.16$$

$$D(B,(CD)) = \sqrt{(-1-(-1))^2 + (1-(-2))^2} = 3$$

Neste caso a distância entre B e o centróide do grupo CD é a menor distância, então o objeto B deve ser realocado para o grupo CD, Figura 6, e as novas coordenadas dos centróides são apresentadas na Tabela 3:

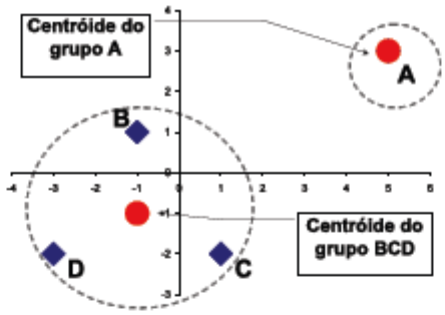



Figura 6 – Grupos A e BCD com os respectivos centróides

Tabela 3 – Coordenadas dos centróides (2ºetapa)

| Grupos | Variáveis | |
|---|---|---|
| | $x_1$ | $x_2$ |
| A | 5 | 3 |
| BCD | $\frac{-1+1+(-3)}{3} = -1$ | $\frac{1+(-2)+(-2)}{3} = -1$ |

Continuando com os objetos C e D verifica-se que não há realocação.

A Tabela 4 mostra as distâncias de cada objeto com relação ao centróides. Em todos os casos a distância do objeto ao centróide do grupo onde está alocado é sempre a menor distância, assim não há realocação de objetos e a execução do algoritmo pode ser finalizada.

Tabela 4 – Distância entre objetos e centróides

| Grupos | Distância Euclidiana | | | |
|---|---|---|---|---|
| | Objetos | | | |
| | A | B | C | D |
| A | 0 | 6.32 | 6.40 | 9.43 |
| BCD | 7.21 | 2 | 2.24 | 2.24 |

## 3- Aplicação no Modelo de Geração de Cenários

Os cenários de energia natural afluente, que são utilizados durante as simulações forward e backward do processo de definição da política ótima de operação, são obtidos através de um modelo auto-regressivo periódico de ordem *p*, PAR(p), que modela a afluência de um mês como sendo função das afluências dos *p* meses anteriores (MACEIRA e MERCIO, 1997). A amostra de ruídos aleatórios utilizada pelo modelo PAR(p), é obtida atualmente através de amostragem aleatória simples.

O método proposto nesta Nota Técnica consiste em aplicar as técnicas de agregação no procedimento de geração dos cenários de energia natural afluente das simulações forward e backward. Neste caso, as técnicas de agregação são empregadas para a geração da amostra de ruídos normais multivariados, não correlatados espacialmente, que é utilizada pelo modelo PAR(p).

Inicialmente será gerada uma amostra muito grande utilizando a amostragem aleatória simples (amostra original), onde cada objeto é um vetor de ruídos aleatórios (um ruído para cada subsistema considerado na configuração). Os vetores de ruídos que compõem essa amostra são equiprováveis. Logo após é realizada a agregação desses objetos de forma a reduzir a dimensionalidade da amostra original. A Figura 7 ilustra o procedimento proposto. Como os vetores de ruídos são sorteados a partir de uma distribuição normal padrão N(0,1), não há necessidade de padronizá-los antes do processo de agregação. A medida de distância a ser utilizada no processo de agregação é a distância Euclidiana, pois os vetores de ruídos da amostra original não possuem correlação espacial.

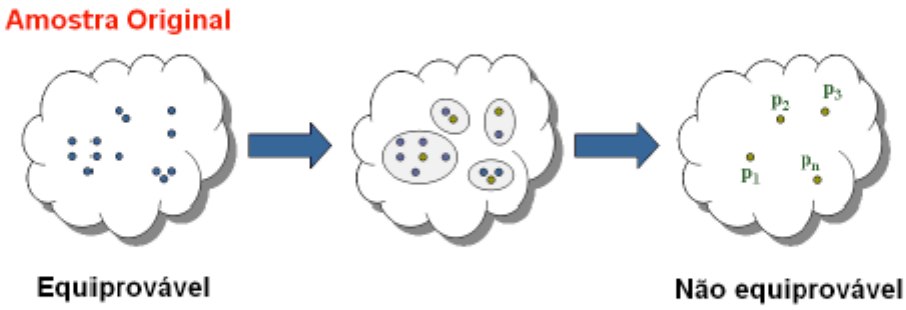



Figura 7– Aplicação do Procedimento de Agregação

O método de agregação escolhido é o método não hierárquico K-Means, pois o tamanho da amostra que é fornecida para o processo de agregação é grande. Os métodos não hierárquicos são ideais para trabalhar com grandes conjuntos de dados, pois não requerem o cálculo da matriz de similaridade.

O processo de agregação é inicializado através do sorteio aleatório de pontos iniciais para representar os centróides dos grupos. Estes pontos iniciais são objetos do conjunto de entrada, logo são vetores de ruídos pertencentes à amostra original. Desta maneira, pode-se garantir que nenhum grupo ficará vazio.

Nos passos seguintes até a convergência do processo de agregação, o centróide dos grupos será o ponto médio destes grupos. Após a convergência do processo de agregação, o centróide dos grupos será o objeto mais próximo do ponto médio deste grupo. A Figura 8 ilustra como é escolhido o representante de cada grupo formado, o ponto médio dos grupos está assinalado com um x.

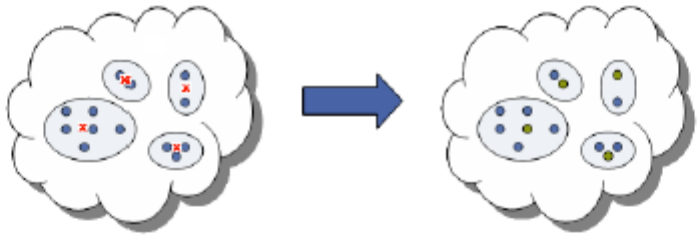


Figura 8– Escolha do Objeto Representativo

Os vetores de ruídos multivariados resultantes do processo de agregação não são mais equiprováveis. A probabilidade dos objetos representantes irá refletir a representatividade do grupo em que ele se encontra. A probabilidade $P_k$ associada ao grupo k é calculada como em (6):

$$P_k = \frac{NO_k}{N} \qquad (6)$$

onde N é tamanho da amostra antes do processo de agregação e $NO_k$ é número de objetos alocados no grupo k.

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 pode ser incorporado tanto no processo de construção da árvore do passo forward quanto do passo backward. Porém, estudos exploratórios com cenários hidrológicos não equiprováveis no passo forward mostraram resultados muito instáveis. Neste sentido, uma nova amostra de ruídos equiprováveis será construída para o passo forward a partir da amostra resultante do processo de agregação através de um sorteio condicionado.

O sorteio condicionado é baseado no teorema da transformação inversa: "Se X é uma variável aleatória de distribuição acumulada F(x), então a variável aleatória Y=F(x) tem distribuição uniforme (0,1)".

Inicialmente é calculada a distribuição acumulada empírica da amostra de ruídos não equiprováveis resultante do processo de agregação para o passo forward. Logo em seguida é sorteado um número aleatório uniforme [0,1], e a partir de uma consulta à função acumulada é identificado o ruído associado aquele valor sorteado. No exemplo da Figura 9, o número aleatório uniforme sorteado foi 0,89. O ruído cuja função acumulada corresponde a 0,89 é o ruído número 191.

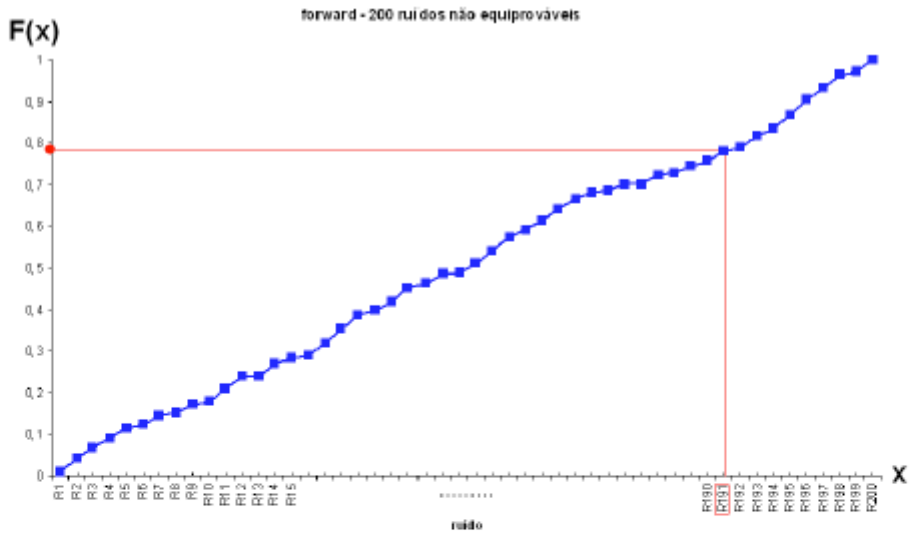



Figura 9 – Sorteio Condicionado

O sorteio condicionado é realizado tantas vezes quanto for o tamanho da amostra de ruídos do passo forward. Os ruídos que compõem essa nova amostra de ruídos são equiprováveis. Logo, os cenários hidrológicos do passo forward construídos a partir dessa amostras de ruídos também são equiprováveis.

Foram estudadas cinco alternativas de aplicação do processo de agregação na construção das árvores de cenários hidrológicos. Na primeira alternativa, chamada de **opção 0**, o processo de agregação é aplicado para definir a amostra de ruídos do passo backward. A partir dessa amostra é realizado um sorteio condicionado para definir a amostra de ruídos a ser utilizada na construção dos cenários hidrológicos a serem utilizados pelo passo foward. Esse procedimento é ilustrado na Figura 10.


Figura 10 – Opção 0

Na segunda alternativa, chamada de **opção 1**, o processo de agregação é aplicado apenas na construção da árvore de cenários do passo backward, de acordo com o descrito anteriormente. A árvore de cenários do passo forward é obtida através de amostragem aleatória simples (AAS), Figura 11.

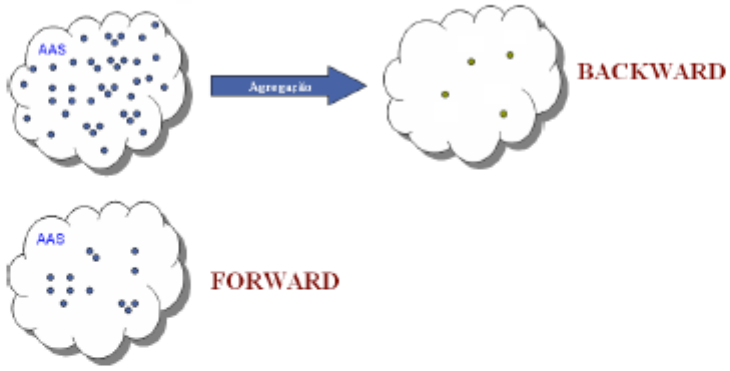



Figura 11 – Opção 1

Os ruídos aleatórios multivariados pertencentes à amostra de ruídos que será utilizada para a construção dos cenários da backward são não equiprováveis. A probabilidade do cenário de energia natural afluente da simulação backward é igual à probabilidade do ruído multivariado a partir do qual ele foi gerado. Já os cenários hidrológicos da simulação forward são equiprováveis.

Na terceira alternativa, Figura 12, o processo de agregação é aplicado para obter a amostra de ruídos do passo forward. A árvore de cenários do passo backward é obtida aplicando-se o processo de agregação na amostra de ruídos construída para o passo forward. Neste caso, o algoritmo de agregação deve levar em conta que os objetos da amostra a ser agregada são não equiprováveis. Uma amostra com objetos equiprováveis é construída para o passo forward através de sorteio condicionado. Essa alternativa é chamada **opção 2**.

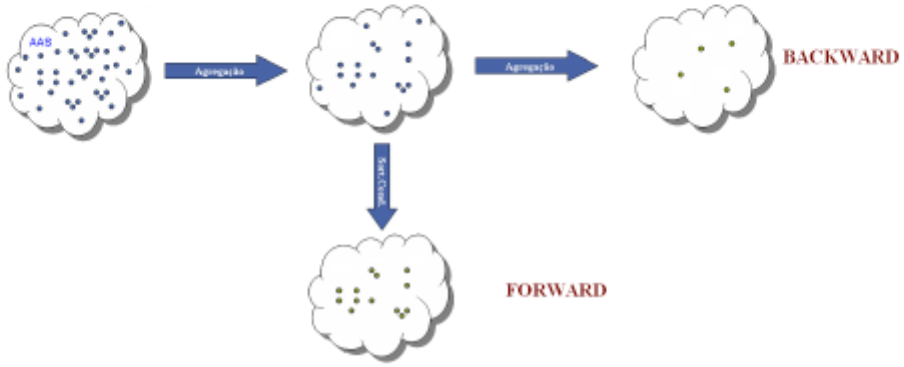



Figura 12 – Opção 2

Na quarta alternativa (**opção 3**), Figura 13, o processo de agregação é aplicado para obter a amostra de ruídos do passo forward. Uma amostra com objetos equiprováveis é construída para o passo forward através de sorteio condicionado. A árvore de cenários do passo backward é obtida aplicando-se o processo de agregação na amostra de ruídos construída para o passo forward, após o sorteio condicionado.

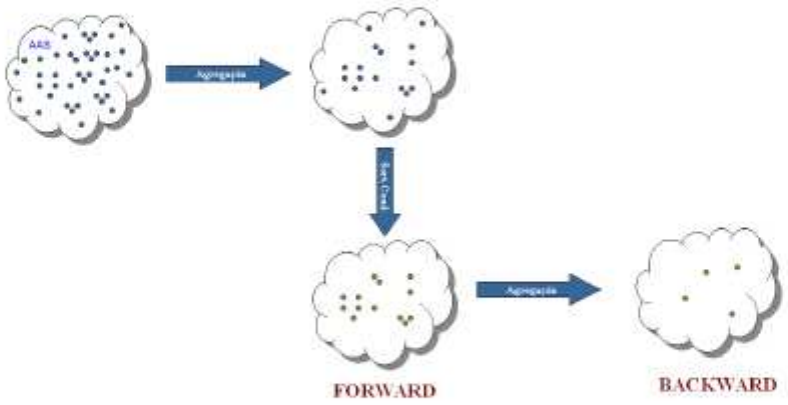



Figura 13 – Opção 3

Na quinta alternativa, o processo de agregação é aplicado na construção da árvore de cenários do passo forward. A árvore de cenários do passo backward é obtida aplicando-se novamente o processo de agregação na amostra originalmente gerada. Novamente, uma amostra com objetos equiprováveis é construída para o passo forward através de sorteio condicionado. A **opção 4** é ilustrada na Figura 14.

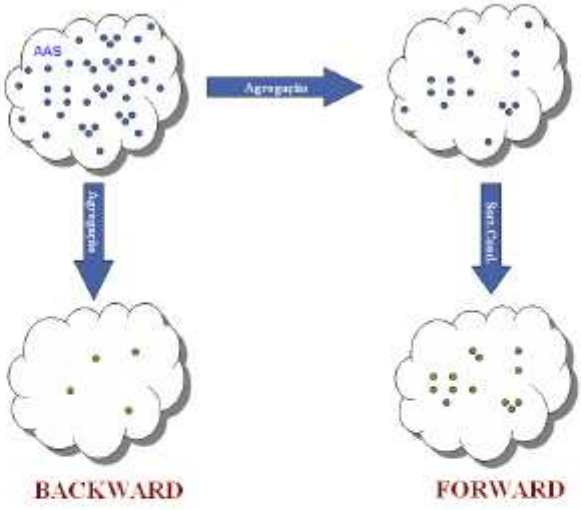



Figura 14 – Opção 4

A amostra de cenários hidrológicos do passo forward gerada para o primeiro período do horizonte de estudo utilizando a opção 0 possui no máximo *nleq* valores distintos (onde *nleq* é o número de cenários da amostra backward). Para aprimorar a

representatividade dos cenários forward utilizados no primeiro período de estudo é proposto um aperfeiçoamento na opção 0. Para o primeiro período, a amostra forward passa a ser construída conforme descrito na opção 4 e para os demais períodos a amostra é construída de acordo com o descrito na opção 0 original, Figura 15.


(a) (b)

Figura 15 – Evolução da Opção 0. (a) 1ºperíodo (b) demais períodos

# 4- Fator de Compensação para o Desvio Padrão

A aplicação das técnicas de agregação na amostra original, cujo algoritmo envolve a substituição de agrupamentos de objetos da amostra original por um único representante, resulta na obtenção de amostras agregadas com menor variabilidade do que a correspondente amostra original (a variabilidade interna nos agrupamentos é perdida). O efeito prático é que as amostras de ruídos agregadas que serão utilizadas nos passos forward e backward apresentam desvio padrão menores que o desejado, apesar das demais estatísticas estarem bem reproduzidas. Para reduzir essa degradação é proposto um fator de correção a ser aplicado previamente na geração da amostra original de ruídos para compensar a perda de variabilidade que se terá ao agregá-la.

## 4.1 – Degradação do Desvio Padrão

A degradação do desvio padrão será mais elevada quanto maior o tamanho da amostra original em relação ao tamanho da amostra após o processo de agregação, conforme ilustrado na Figura 16. Neste exemplo são consideradas duas amostras originais com tamanhos distintos, uma com 2 mil objetos e outra com 10 mil objetos. Os objetos são vetores de ruídos normal-padrão multivariados com quatro dimensões. O método de agregação foi aplicado considerando diferentes números de agrupamentos (de 10 a 2 mil grupos). Note que quanto menor a proporção entre a amostra original e a amostra agregada, mais difícil é a reprodução do desvio padrão. Além disso, considerando o objetivo de um mesmo número de agrupamentos, quanto maior o tamanho da amostra original, maior será a degradação observada no desvio padrão.

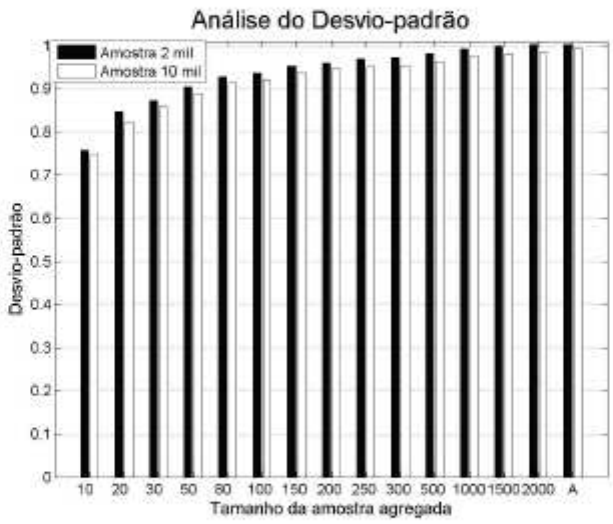



Figura 16 – Ruído Normal - DP - Objetos com 4 dimensões

Adicionalmente, quanto maior a dimensionalidade dos objetos que compõem a amostra original, mais evidente é a degradação observada no desvio padrão da amostra agregada. Na Figura 17 são apresentadas as evoluções temporal do desvio padrão histórico e das séries forward geradas com o método denominado Opção 4, para o subsistema Sudeste, considerando um caso de PMO (com 4 subsistemas) e um caso PDE (com 9 subsistemas). Note que a degradação observada no caso PDE, Figura 17b, é maior do que a verificada no PMO, Figura 17a.

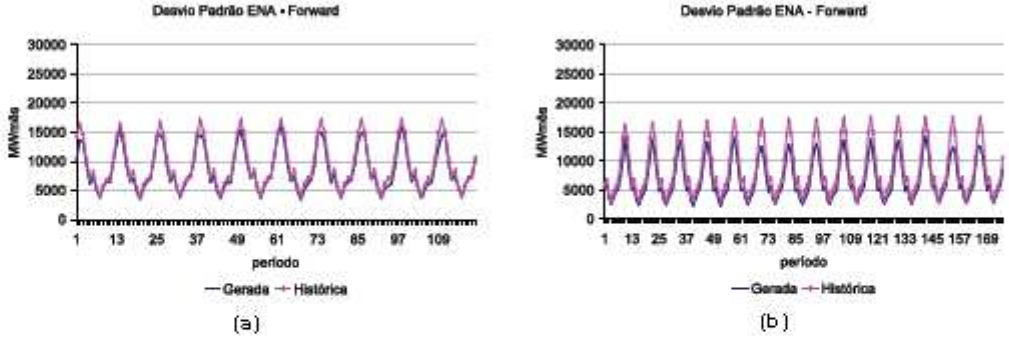



Figura 17 – ENA DP Sudeste – Amostra original = 100mil

(a) PMO (4 subsistemas) (b) PDE (9 subsistemas)

### 4.2- Redução da Degradação do Desvio Padrão

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 18.

De acordo com o exposto no item anterior, o fator de compensação é função da proporção entre as amostras original e agregada e da dimensionalidade dos objetos da amostra original. Na Figura 19 é apresentado um gráfico com fatores de compensação calculados para diversas proporções entre os tamanhos da amostra e para dois tamanhos de amostra original (2 mil e 10 mil) com objetos de 4 dimensões.

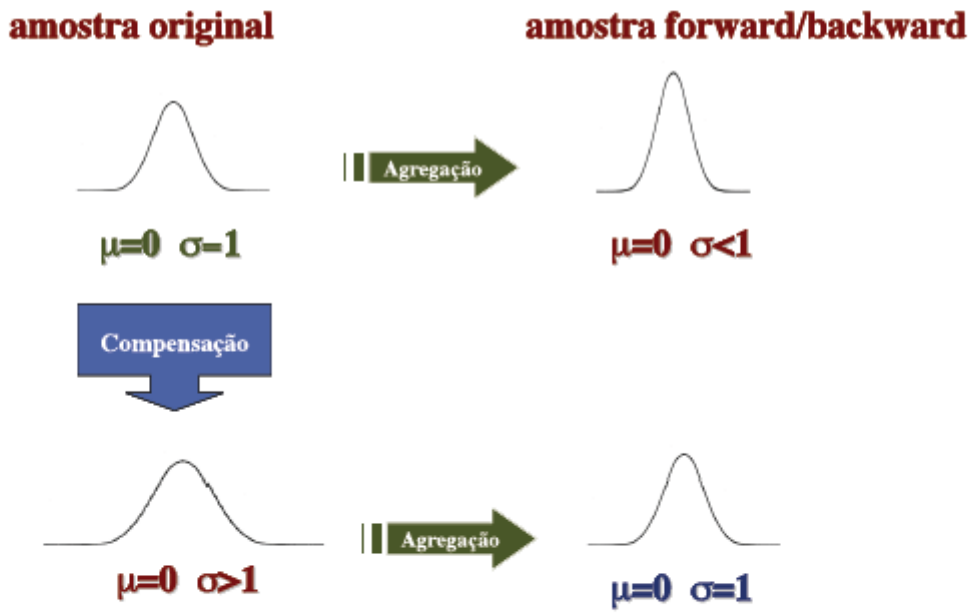



Figura 18 – Aplicação do Fator de Compensação

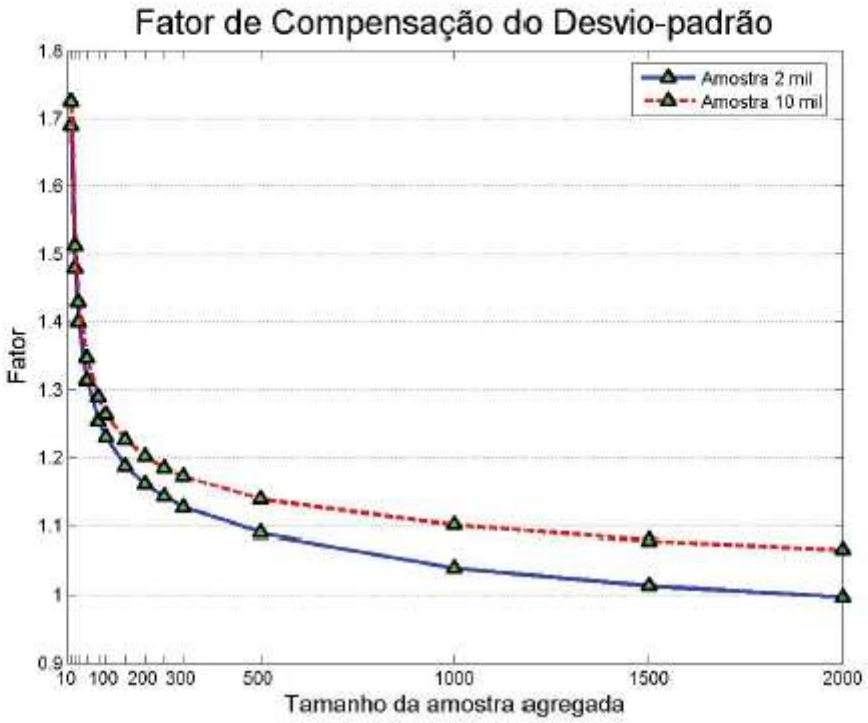



Figura 19 – Fator de Compensação – Objeto com 4 dimensões

O fator de compensação (FC) deve ser calibrado de acordo com a dimensionalidade do problema e com o fator de redução aplicado à amostra original de ruídos. Na prática, esse fator é calculado como a média dos fatores de compensação de todos os subsistemas do caso em estudo considerando doze amostras distintas.

Observando os resultados obtidos na análise dos cenários hidrológicos gerados, é verificado que a utilização do fator de compensação para o desvio padrão é eficiente para a redução da degradação do desvio padrão. Na Figura 20 são apresentados a evolução temporal e o desvio relativo da ENA média do subsistema Sudeste, com e sem a consideração do fator de compensação do desvio padrão.

Na Figura 21 são apresentados a evolução temporal e o desvio relativo da desvio padrão da ENA do subsistema Sudeste, com e sem a consideração do fator de compensação do desvio padrão. Note que o FC trouxe uma grande atenuação na degradação do desvio padrão.

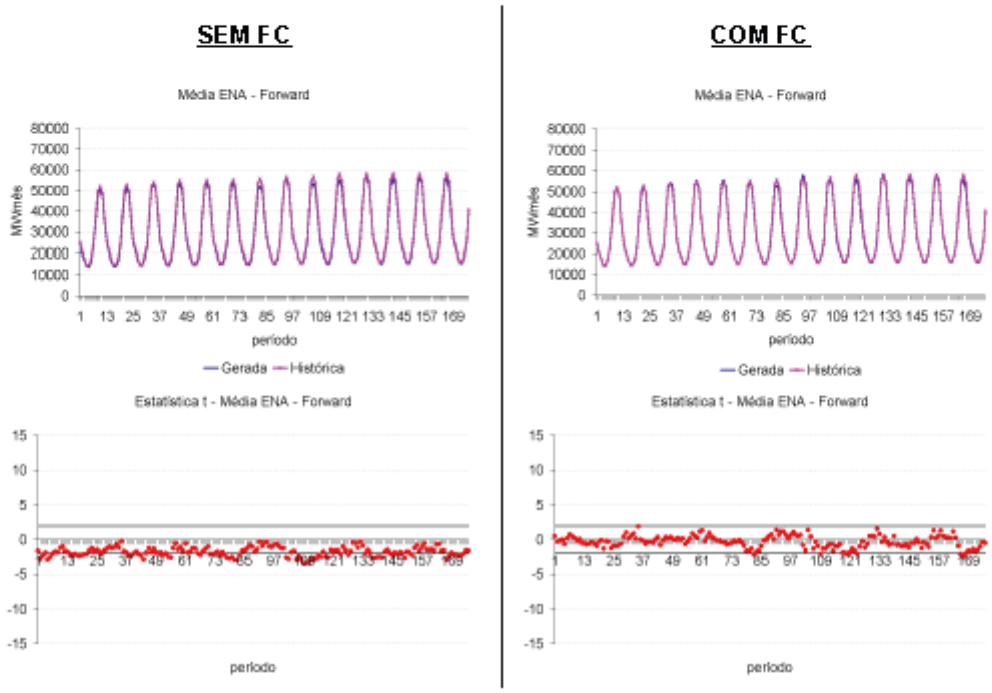
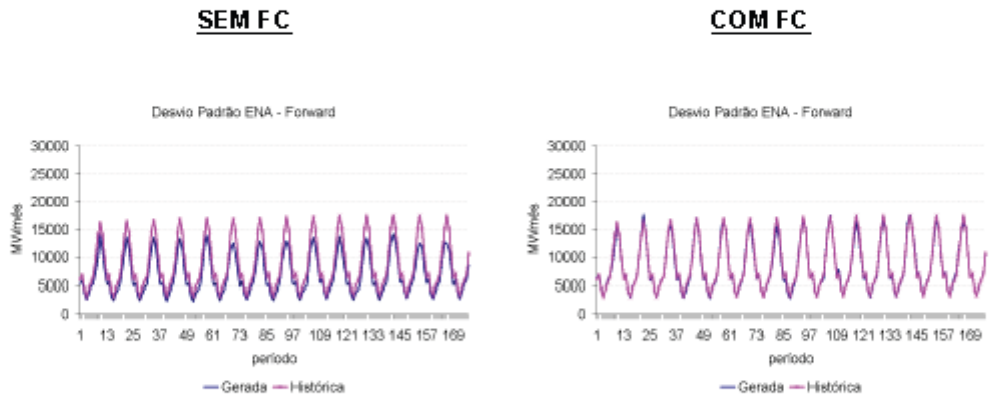



Figura 20 – ENA Média Sudeste – Amostra Original 100 mil objetos – PDE (9 subsistemas)

(a) Evolução Temporal Sem FC (b) Evolução Temporal Com FC

(c) Desvio Relativo Sem FC (d) Desvio Relativo Com FC



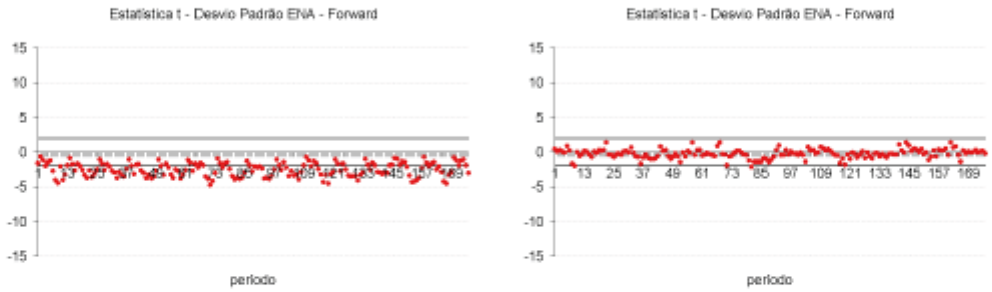



Figura 21 – ENA DP Sudeste – Amostra Original 100 mil objetos – PDE (9 subsistemas)

(a) Evolução Temporal Sem FC (b) Evolução Temporal Com FC

(c) Desvio Relativo Sem FC (d) Desvio Relativo Com FC

## 5- Cortes de Benders

No NEWAVE, a estratégia é representada pela função de custo futuro e calculada por um processo iterativo para um conjunto de estados (energia armazenada no início do estágio e tendência hidrológica). Para cada estado, o corte da função de custo futuro corresponde a uma média calculada para um conjunto de afluências utilizadas durante o cálculo da política de operação no passo backward, Figura 22.

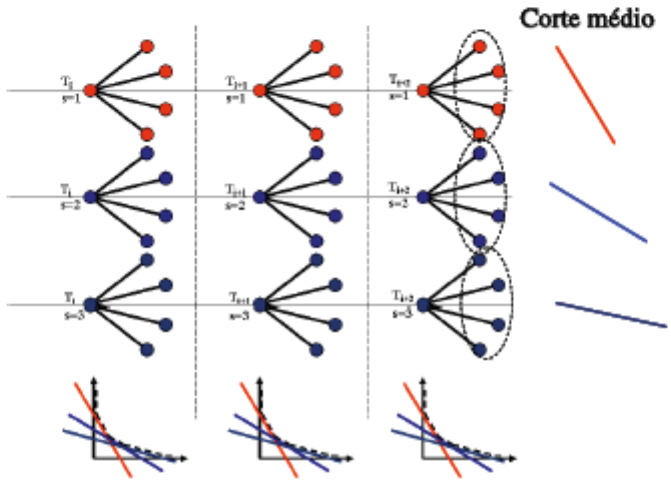



Figura 22 – Construção da FCF

A aplicação das técnicas na agregação no processo de geração dos cenários hidrológicos utilizados durante a recursão backward, resulta em um conjunto de cenários não equiprováveis. Logo, o cálculo do corte médio a ser adicionado à função de custo futuro deve ser modificado de forma a levar em conta a probabilidade de cada cenário hidrológico do conjunto backward.

$$\overline{\pi_V^{isim}} = \sum_{i=1}^{NLEQ} \pi_V^{i,isim} * P_i \qquad (7)$$

$$\overline{\pi_{Aj}^{isim}} = \sum_{i=1}^{NLEQ} \pi_{Aj}^{i,isim} * P_i \qquad j = 1, \ldots, NARP \qquad (8)$$

$$\overline{RHS^{isim}} = \sum_{i=1}^{NLEQ} RHS^{i,isim} * P_i \qquad (9)$$

onde:

NLEQ é o número de aberturas (tamanho do conjunto de afluências utilizado na recursão backward);

NARP é a ordem do modelo PAR(p);

$P_i$ é a probabilidade do i-ésimo cenário hidrológico do conjunto backward;

$\pi_V^{i,isim}$ é o coeficiente do corte de Benders associado ao estado armazenamento inicial do estágio, calculado no cenário forward isim e na i-ésima abertura;

$\overline{\pi_V^{isim}}$ é o coeficiente do corte de Benders médio associado ao estado armazenamento inicial do estágio, calculado no cenário forward isim;

$\pi_{A_j}^{i,isim}$ é o coeficiente do corte de Benders associado ao estado energia afluente passada do estágio t-j, calculado no cenário forward isim e na i-ésima abertura;

$\overline{\pi_{A_j}^{isim}}$ é o coeficiente do corte de Benders médio associado ao estado energia afluente passada do estágio t-j, calculado no cenário forward isim;

$RHS^{i,isim}$ é o termo independente do corte de Benders calculado no cenário forward isim e na i-ésima abertura;

$\overline{RHS^{isim}}$ é o termo independente do corte de Benders médio calculado no cenário forward isim.

## 6- Conclusões

Nesta Nota Técnica foram descritos procedimentos, baseados na análise de conglomerados, para a definição da sub-árvore a ser visitada durante o processo de cálculo da política ótima de operação com o intuito de tornar os resultados mais robustos, com relação a variações no número de cenários de simulação forward e backward, e com relação à amostra de cenários hidrológicos utilizada.

Foram propostas cinco alternativas para a geração dos cenários hidrológicos para os passos forward e backward. Todas as alternativas apresentadas utilizam técnicas estatísticas multivariadas capazes de elaborar critérios que possibilitam agrupar objetos similares em determinados grupos (técnicas de agregação).

Além disso, foi apresentado um aperfeiçoamento para atenuar a degradação do desvio padrão da amostra de ruídos após o processo de agregação. Para tanto é utilizado um fator de compensação para o desvio padrão na amostra original. Através dos os resultados obtidos na análise dos cenários hidrológicos gerados, foi verificado que a utilização do fator de compensação para o desvio padrão é eficiente para a redução da degradação do desvio padrão.

## Referência Bibliográfica

ALDENDERFER, M.S., BLASHFIELD, R.K., 1984, *Cluster Analysis*, Beverly Hills, Sage Publications.

ANDERBERG, M.R., 1973, *Cluster Analysis for Applications*, New York, Academic Press.

BOUROCHE, J.M., SAPORTA, G., 1980, *Análises de Dados*, Rio de Janeiro, Zahar Editores.

DURAN, B.S., ODELL, P.L., 1974, *Cluster Analysis – A Survey*, Berlin, Springer-Verlag.

FARREL, J.L., 1997, *Portfolio Management Theory & Application*, 2 ed. Orlando, Mc Graw Hill.

HAIR Jr., J.F., ANDERSON, R.E., TATHAN, R.L., BLACK, W.C., 1998, *Multivariate Data Analysis*, New Jersey, Prentice Hall.

HARTIGAN, J.A., 1975, *Clustering Algorithms*, New York, John Wiley & Sons.

JARDIM, D.L.D.D., 2002, *Modelo de Geração de Séries Sintéticas de Vazões Utilizando Técnicas de Agregação*. Tese de M. Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.

JOHNSON, R.A., WICHERN, D.W., 1998, *Applied Multivariate Statistical Analysis*, 4 ed. New Jersey, Prentice Hall.

MACEIRA, M.E.P., BEZERRA, C.V., 1997, "Stochastic Streamflow model for Hydroelectric Systems" In: *Proceedings of 5th International Conference on Probabilistic Methods Applied to Power Systems*, pp. 305-310, Vancouver, Canada, Sep.

STEINCACH, M., KARYPIS, G., KUMAR, V., 2000, *A Comparison of Document Clustering Techniques*. In: Technical Report # 00-034, Department of Computer Science and Engineering, University of Minnesota, Minnesota.

VALENTIN, J.L., 2000, *Ecologia Numérica, Uma Introdução à Análise Multivariada* de Dados Ecológicos, Rio de Janeiro, Interciência.

VELASQUEZ, R.M.G., PESSANHA, J.F.M., JARDIM, D.L.D.D., MELO, S.L., MELO, A.C.G., 2001, "Técnicas de Classificação para Caracterização da Curva de Carga de Empresas de Distribuição de Energia Elétrica – Um Estudo Comparativo" In: *Proceedings of the V Congresso Brasileiro de Redes Neurais*, pp. 133-138, Rio de Janeiro, Apr.

ZIKMUND, W.G., 1999, *Exploring Marketing Research*, 6 ed. Singapore, Dryden Press.