

Centro de Pesquisas de Energia Elétrica - CEPEL



Relatório Técnico

Nº/Ano: 13955/2018 **Nº de Páginas:** 40 **Nº de Anexos:** 2

Título: Modelo de Geração de Séries Sintéticas de Energias e Vazões – Manual de Referência.

Departamento: Departamento de Otimização Energética e Meio Ambiente - DEA

Área de Responsabilidade: B200 **Conta de Apropriação:** 1798

Cliente:
Centrais Elétricas Brasileiras S/A – Eletrobrás
Av. Presidente Vargas, 409, Rio de Janeiro - RJ

Atenção: Márcio Antônio Guedes Drummond

Resumo: Este relatório constitui o Manual de Referência do modelo GEVAZP. O objetivo deste modelo é produzir cenários sintéticos de vazões ou energias baseados em modelos autorregressivos periódicos.

Autores:
Débora Dias Jardim Penna – Cepel
Maria Elvira Piñeiro Maceira – Cepel
Jorge Machado Damázio - Cepel
Felipe Treistman – PUC-Rio
Hugo Santarém de Araújo – Cepel

Palavras-Chave:
Modelos estocásticos multivariados, Modelos autorregressivos periódicos, Geração séries sintéticas, Técnica de agregação

Classificação: CONTROLADO

Gerente de Projeto


Débora Dias Jardim Penna
Tel.: 21 2598-6450
E-mail: debora@cepel.br

Chefe do Departamento de Otimização Energética e Meio Ambiente


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

Aprovação


Raul Balbi Sollero
Diretor de Pesquisa, Desenvolvimento e Inovação
27/12/2019

| | | |
|-----------------|---|------------------|
| <u>1</u> | <u>APLICAÇÃO DE SÉRIES SINTÉTICAS NO PLANEJAMENTO E OPERAÇÃO DE SISTEMAS HIDROTÉRMICOS</u> | <u>4</u> |
| <u>2</u> | <u>INTRODUÇÃO A PROCESSOS ESTOCÁSTICOS E SÉRIES TEMPORAIS</u> | <u>5</u> |
| <u>3</u> | <u>ASPECTOS TÉCNICOS E PRÁTICOS DO DESENVOLVIMENTO DO MODELO</u> | <u>8</u> |
| 3.1 | GERAÇÃO DE SÉRIES SINTÉTICAS PARA OS MODELOS DE PLANEJAMENTO E OPERAÇÃO | 8 |
| 3.1.1 | REPRESENTAÇÃO DAS AFLUÊNCIAS NO MODELO DE MÉDIO PRAZO | 8 |
| 3.1.2 | REPRESENTAÇÃO DAS AFLUÊNCIAS NO MODELO DE CURTO PRAZO | 9 |
| <u>4</u> | <u>MODELO AUTORREGRESSIVO PERIÓDICO</u> | <u>10</u> |
| 4.1 | DESCRIÇÃO DO MODELO | 10 |
| 4.2 | AJUSTE DO MODELO | 12 |
| 4.2.1 | IDENTIFICAÇÃO | 12 |
| 4.2.2 | ESTIMAÇÃO | 13 |
| 4.2.3 | VERIFICAÇÃO | 13 |
| 4.3 | GERAÇÃO DE SÉRIES SINTÉTICAS COM O MODELO PAR(P) | 13 |
| 4.3.1 | VAZÕES INCREMENTAIS NEGATIVAS | 15 |
| 4.3.2 | CORRELAÇÃO ESPACIAL | 16 |
| 4.3.3 | USINAS COM VAZÕES AFETADAS POR TEMPO DE VIAGEM | 16 |
| 4.3.4 | CÁLCULO DE VAZÕES PARA POSTOS ARTIFICIAIS | 18 |
| <u>5</u> | <u>APLICAÇÃO DAS TÉCNICAS DE AGREGAÇÃO DE CENÁRIOS AO MODELO GEVAZP</u> | <u>20</u> |
| 5.1 | AGREGAÇÃO DE CENÁRIOS | 20 |
| 5.1.1 | DEFINIÇÕES | 21 |
| 5.1.2 | MÉTODO DE AGRUPAMENTO UTILIZADO | 22 |
| 5.1.3 | MEDIDA DE DISTÂNCIA | 22 |
| 5.1.4 | DETERMINAÇÃO DOS CENÁRIOS REPRESENTATIVOS | 23 |
| 5.1.5 | CORRELAÇÃO TEMPORAL E ESPACIAL | 24 |
| 5.1.6 | ESTRUTURA DO NOVO PROCESSO DE GERAÇÃO DE CENÁRIOS SINTÉTICOS EM ÁRVORE | 24 |
| <u>6</u> | <u>CRITÉRIOS PARA TESTAR A ADEQUAÇÃO DO MODELO</u> | <u>27</u> |

| | | |
|------------|---|-----------|
| 6.1 | TESTES PARA SÉRIES GERADAS EM PARALELO (PENTE) | 27 |
| 6.1.1 | VERIFICAÇÃO DAS ESTATÍSTICAS ANUAIS E MENSAIS | 27 |
| 6.1.2 | AVALIAÇÃO DO DESEMPENHO DO MODELO | 27 |
| 6.2 | TESTES PARA SÉRIES GERADAS EM ÁRVORE | 30 |
| 6.2.1 | TESTE DE MOMENTOS CONDICIONADOS DE ORDEM UM E DOIS | 31 |
| 6.2.2 | PARÂMETROS DAS DISTRIBUIÇÕES AMOSTRAIS | 31 |
| 6.2.3 | TESTES DE ADERÊNCIA | 33 |
| 6.2.4 | COMPARAÇÃO DAS ESTATÍSTICAS GERADAS E AGREGADAS | 35 |
| 6.3 | VERIFICAÇÃO DOS COEFICIENTES AUTORREGRESSIVOS PERIÓDICOS | 36 |
| 7 | <u>REFERÊNCIAS</u> | 39 |

1 APLICAÇÃO DE SÉRIES SINTÉTICAS NO PLANEJAMENTO E OPERAÇÃO DE SISTEMAS HIDROTÉRMICOS

A adoção de critérios probabilísticos em diversas atividades do planejamento e operação de sistemas hidrotérmicos criou a necessidade da modelagem probabilística de afluências a locais de aproveitamentos hidroelétricos ou a subsistemas. Em estudos energéticos, critérios de suprimento são baseados em índices de risco, estimados a partir da simulação da operação energética do sistema para diversos cenários (sequências) de afluências aos aproveitamentos hidroelétricos ou subsistemas.

Por exemplo, um grande número de cenários pode ser utilizado para estimar a relação entre demanda energética e risco de não atendimento em um ano qualquer (energia garantida). A Figura 1.1 mostra que um sistema hidroelétrico pode suprir a demanda e , que está associada ao risco p , previamente selecionado.

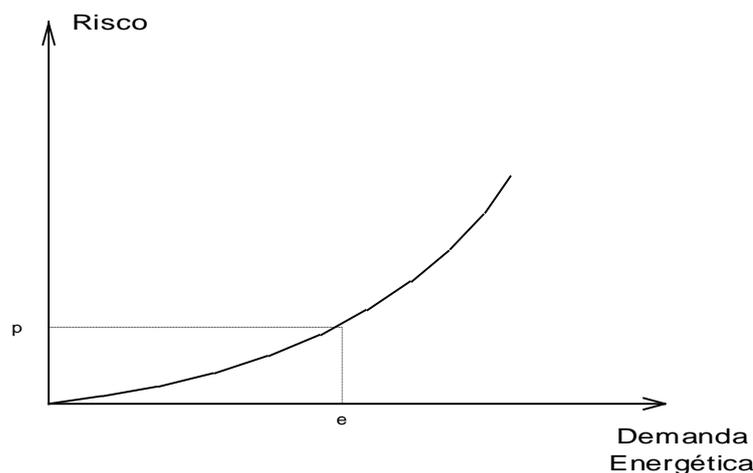


Figura 1.1 - Relação Demanda x Risco em um Ano Qualquer

O único cenário disponível na prática, o registro de afluências observado no passado (chamado de série histórica) é, no entanto, insuficiente para compor uma amostra de tamanho necessário para estimar índices de risco com incertezas aceitáveis. Entretanto, as características básicas da série histórica podem ser capturadas por modelos estocásticos capazes de produzir séries sintéticas de afluências, diferentes da série histórica, mas igualmente prováveis. Dessa forma, a informação contida na série histórica pode ser mais completamente extraída, permitindo a avaliação de riscos e incertezas pertinentes a um sistema hidroelétrico.

Neste relatório será descrito o modelo autorregressivo periódico, PAR(p), proposto para ser utilizado no Modelo Estratégico de Geração Hidrotérmica a Subsistemas Equivalentes Interligados – NEWAVE [1], no Modelo de Simulação a Usinas Individualizadas para Subsistemas Hidrotérmicos Interligados – SUISHI-O [2] e no Modelo de Determinação da Coordenação da Operação a Médio Prazo – DECOMP [3].

2 INTRODUÇÃO A PROCESSOS ESTOCÁSTICOS E SÉRIES TEMPORAIS

Medindo-se, por exemplo, a vazão afluente mensal a um dado local por um período de cinco anos, podemos obter a curva 1 da Figura 2.1. Realizando-se as mesmas medições para outro segmento de cinco anos, obtém-se outra curva, que é em geral diferente da primeira.

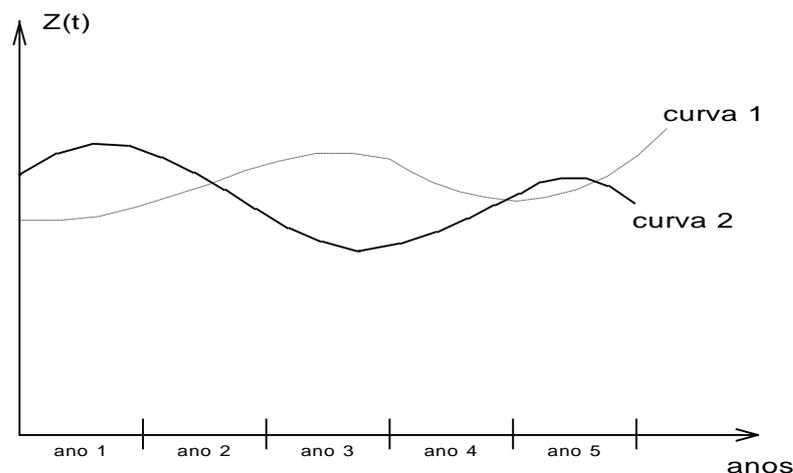


Figura 2.1 - Vazão Afluente Anual x Anos

Estas curvas são chamadas de trajetórias ou realizações do processo físico que está sendo observado. Este pode ser modelado por um **processo estocástico**, que nada mais é do que o conjunto de todas as possíveis trajetórias que podem ser observadas. Cada trajetória é também chamada de **série temporal**.

Na prática, só está disponível uma realização do processo estocástico, a **série histórica**. Assim, temos que para cada instante do tempo (dia, mês, ano, etc) o processo estocástico é uma variável aleatória. O valor observado em um instante t qualquer (valor da série histórica no instante t), nada mais é do que o valor "amostrado" da distribuição de probabilidade associado a variável aleatória do processo estocástico no instante t .

Um processo estocástico é totalmente descrito pelo conjunto de todas as séries temporais que o compõe ou pela distribuição de probabilidade conjunta de todas as variáveis aleatórias envolvidas. Como na realidade não está disponível nem uma coisa nem outra, a modelagem de **séries sintéticas** tem por objetivo ajustar um modelo pelo qual acreditamos que a série histórica tenha sido produzida e a partir dele gerar séries sintéticas que representam as séries temporais que podem ser "amostradas" pelo processo físico que se está observando, o processo estocástico.

É comum assumir algumas hipóteses simplificadoras para o problema, por exemplo, **estacionaridade**. Um processo estocástico é estacionário se ao longo do tempo as suas propriedades estocásticas não se alteram. Isto significa que a média, o desvio padrão etc, não sofrem modificações ao longo do tempo, ou de uma forma mais abrangente, significa que a distribuição de probabilidade em um instante t qualquer é válida para qualquer outro instante. A não estacionaridade de um processo estocástico pode ser causada pela intervenção direta do homem, ou da natureza, no processo físico, ou ainda pela presença de ciclos sazonais (características que se repetem dentro de um ano). Quanto ao último fator, o processo de afluições anuais é considerado estacionário e o processo de afluições mensais é considerado não estacionário.

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

A média amostral:

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N z_i \tag{2.1}$$

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

$$\hat{\sigma} = \sqrt{\frac{1}{N} \sum_{i=1}^N (z_i - \hat{\mu})^2} \tag{2.2}$$

A covariância de ordem um 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 vazão de um ano com a vazão do ano imediatamente anterior:

$$\hat{\gamma}(1) = \frac{1}{N} \sum_{i=2}^N (z_i - \hat{\mu})(z_{i-1} - \hat{\mu}) \tag{2.3}$$

A covariância possui a 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.

$$\hat{\rho}(1) = \frac{\hat{\gamma}(1)}{\hat{\sigma}^2} \tag{2.4}$$

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 vazões afluentes a duas usinas hidroelétricas, se a série histórica de uma usina é linearmente dependente da série histórica de outra usina, através da correlação "cruzada" ou correlação espacial entre elas.

Por exemplo, a correlação espacial anual entre as vazões afluentes às usinas hidroelétricas a e b pode ser obtida por:

$$\hat{\rho}_{a,b} = \frac{\frac{1}{N} \sum_{i=1}^N (z_i^{(a)} - \hat{\mu}^{(a)})(z_i^{(b)} - \hat{\mu}^{(b)})}{\hat{\sigma}^{(a)} \hat{\sigma}^{(b)}} \tag{2.5}$$

Em geral, processos físicos em escala mensal apresentam um comportamento periódico descrito pelos 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:

$$\hat{\mu}_m = \frac{1}{N} \sum_{i=1}^N z_{12*(i-1)+m} \tag{2.6}$$

$m = 1, \dots, 12$

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

$$\hat{\sigma}_m = \sqrt{\frac{1}{N} \sum_{i=1}^N (z_{12^{*(i-1)+m}} - \hat{\mu}_m)^2} \quad m = 1, \quad (2.7)$$

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:

$$\hat{\gamma}_m(k) = \frac{1}{N} \sum_{i=1}^N (z_{12^{*(i-1)+m}} - \hat{\mu}_m)(z_{12^{*(i-1)+m-k}} - \hat{\mu}_m) \quad m = 1, \dots, \quad (2.8)$$

$$\hat{\rho}_m(k) = \frac{\hat{\gamma}_m(k)}{\hat{\sigma}_m \hat{\sigma}_{m-k}} \quad m = 1, \quad (2.9)$$

3 ASPECTOS TÉCNICOS E PRÁTICOS DO DESENVOLVIMENTO DO MODELO

3.1 Geração de Séries Sintéticas Para os Modelos de Planejamento e Operação

O sistema de geração brasileiro é predominantemente hidráulico e possui acoplamento temporal e espacial. Estas características tornam o planejamento da operação energética um problema de grande porte e de difícil solução. Devido ao porte e complexidade deste problema é necessária a sua divisão em diversas etapas. Em cada etapa são utilizados modelos com diferentes graus de detalhamento para a representação do sistema, abrangendo períodos de estudos com horizontes distintos (médio prazo, curto prazo e programação diária).

A aleatoriedade das vazões também é uma característica marcante de sistemas com predominância hidráulica. Essa incerteza é tratada de diferentes formas dependendo da representação utilizada no modelo do sistema gerador. A representação dos possíveis cenários de vazões é diferenciada para cada etapa do processo de planejamento da operação.

3.1.1 Representação das Afluências no Modelo de Médio Prazo

O modelo NEWAVE [1] foi desenvolvido para o planejamento da operação de médio prazo. Ele 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. O parque hidroelétrico é representado de forma agregada e o cálculo da política de operação é baseado na programação dinâmica dual estocástica. O NEWAVE apresenta quatro módulos básicos: módulo de cálculo dos sistemas equivalentes, módulo de cálculo do modelo estocástico de energias afluentes, módulo de cálculo da política de operação hidrotérmica e módulo de simulação da operação.

O modelo SUISHI-O [2] é um modelo de simulação a usinas individualizadas da operação energética de sistemas hidrotérmicos interligados. É capaz de simular subsistemas hidrotérmicos eletricamente interligados em malha mas hidraulicamente independentes. Pode ser acoplado ao modelo de decisão estratégica NEWAVE pela função de custo futuro para cada estágio e pode considerar restrições de operação locais decorrente do uso múltiplo da água.

Os modelos descritos acima simulam um grande número de séries hidrológicas, calculando assim, índices probabilísticos de desempenho do sistema para cada estágio da simulação. Estes diversos cenários de afluências estão em uma estrutura paralela (pente), Figura 3.1, e são gerados pelo modelo GEVAZP.

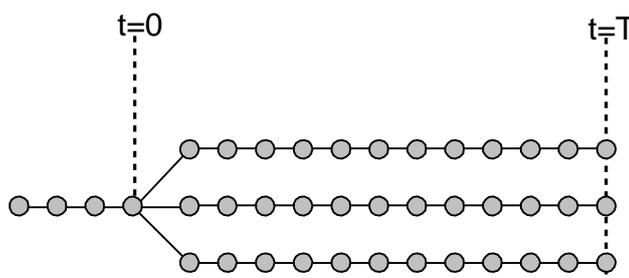


Figura 3.1 – Séries em Paralelo (ou Pente)

3.1.2 Representação das Afluências no Modelo de Curto Prazo

O modelo DECOMP [3] foi desenvolvido para aplicação no horizonte de curto prazo. Seu objetivo é determinar as metas de geração de cada usina de um sistema hidrotérmico, sujeito a afluências estocásticas, de forma a minimizar o valor esperado do custo de operação ao longo do período de planejamento. A incerteza acerca das vazões afluentes aos diversos aproveitamentos do sistema é representada através de cenários hidrológicos. Os cenários hidrológicos podem ser representados através de uma árvore de afluências, Figura 3.2, com probabilidades de ocorrência associadas a cada ramo. Estes cenários são gerados a partir do modelo GEVAZP.

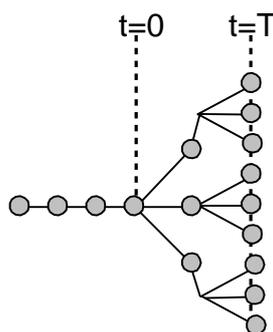


Figura 3.2 – Séries em Árvore

Devido a restrições de tempo computacional, no DECOMP é interessante trabalhar com o menor número possível de cenários hidrológicos e esta limitação não é ideal para o esquema de simulação de Monte Carlo. Um número reduzido de séries não é suficiente para caracterizar bem um processo estocástico.

Para representar bem o processo estocástico de vazões no modelo de otimização do planejamento da operação a curto prazo – DECOMP – com um reduzido número de cenários empregou-se a análise de conglomerados. Uma descrição detalhada da metodologia pode ser encontrada em [4]. Aplicadas a um grande número de cenários hidrológicos gerados, as técnicas de agrupamento proporcionam a escolha de um conjunto representativo de cenários. Este conjunto representativo de cenários hidrológicos irá conter toda a informação necessária para representar o processo estocástico de vazões. Os cenários que fazem parte deste conjunto representativo são obtidos através de agrupamento de cenários semelhantes e possuem características similares aos demais componentes do grupo em que estão localizados. Dessa forma, as correlações temporal e espacial são preservadas.

4 MODELO AUTORREGRESSIVO PERIÓDICO

4.1 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 [5]. 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:

$$\left(\frac{Z_t - \mu_m}{\sigma_m}\right) = \phi_1^m \cdot \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right) + \dots + \phi_{p_m}^m \cdot \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right) + a_t \quad (4.1a)$$

ou

$$\phi^m(B) \left(\frac{Z_t - \mu_m}{\sigma_m}\right) = a_t \quad (4.1b)$$

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, $t = 1, 2, \dots, sN$, função do ano T (T = 1, 2, ..., N) e do período m (m = 1, 2, ..., s)

μ_m é a média sazonal de período s

σ_m é desvio-padrão sazonal de período s

$\phi^m(B)$ é o operador autorregressivo de ordem p_m

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

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

p_m é a ordem do operador autorregressivo do período m

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

Seja ρ_k^m a correlação entre Z_t e Z_{t-k} , de tal forma que t corresponda ao período m:

$$\rho_k^m = E \left[\left(\frac{Z_t - \mu_m}{\sigma_m}\right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right) \right] \quad (4.2)$$

O conjunto de funções de autocorrelação ρ_k^m dos períodos $m = 1, \dots, s$, descrevem a estrutura de dependência temporal da série. Estas funções são obtidas por [6]:

Multiplicando-se ambos os lados da equação (4.1a) por $\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)$ e tomando o valor esperado obtemos para cada período:

$$\begin{aligned}
 E\left[\left(\frac{Z_t - \mu_m}{\sigma_m}\right)\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] &= \phi_1^m E\left[\left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right)\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] + \dots \\
 + \phi_{p_m}^m E\left[\left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right)\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] &+ E\left[a_t\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] + \dots
 \end{aligned} \tag{4.3}$$

Por exemplo, para $k = 1$ a expressão (4.3) resulta em:

$$\rho_1^m = \phi_1^m + \phi_2^m \rho_1^{m-1} + \dots + \phi_{p_m}^m \rho_{p_m-1}^{m-1}$$

Conhecidos os parâmetros de um modelo PAR(p) as funções ρ_k^m são dadas pela solução de (4.3) e podem ser expressas por uma combinação de decaimentos exponenciais e/ou ondas senoidais, o que faz com que cada ρ_k^m tenda a zero à medida que k cresce.

Fixando-se m e variando k de 1 a p_m em (4.3) obtemos para cada período um conjunto de equações comumente denominado de equações de Yule-Walker (4.4). Para um período m qualquer:

$$\begin{bmatrix}
 1 & \rho_1^{m-1} & \rho_2^{m-1} & \dots & \rho_{(p_m-1)}^{m-1} \\
 \rho_1^{m-1} & 1 & \rho_1^{m-2} & \dots & \rho_{(p_m-2)}^{m-2} \\
 \rho_2^{m-1} & \rho_1^{m-2} & 1 & \dots & \rho_{(p_m-3)}^{m-3} \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 \rho_{(p_m-1)}^{m-1} & \rho_{(p_m-2)}^{m-2} & \rho_{(p_m-3)}^{m-3} & \dots & 1
 \end{bmatrix} \cdot \begin{bmatrix} \phi_1^m \\ \phi_2^m \\ \phi_3^m \\ \vdots \\ \phi_{p_m}^m \end{bmatrix} = \begin{bmatrix} \rho_1^m \\ \rho_2^m \\ \rho_3^m \\ \vdots \\ \rho_{p_m}^m \end{bmatrix} \tag{4.4}$$

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

$$\begin{bmatrix}
 1 & \rho_1^{m-1} & \rho_2^{m-1} & \dots & \rho_{(k-1)}^{m-1} \\
 \rho_1^{m-1} & 1 & \rho_1^{m-2} & \dots & \rho_{(k-2)}^{m-2} \\
 \rho_2^{m-1} & \rho_1^{m-2} & 1 & \dots & \rho_{(k-3)}^{m-3} \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 \rho_{(k-1)}^{m-1} & \rho_{(k-2)}^{m-2} & \rho_{(k-3)}^{m-3} & \dots & 1
 \end{bmatrix} \cdot \begin{bmatrix} \phi_1^m \\ \phi_2^m \\ \phi_3^m \\ \vdots \\ \phi_{kk}^m \end{bmatrix} = \begin{bmatrix} \rho_1^m \\ \rho_2^m \\ \rho_3^m \\ \vdots \\ \rho_k^m \end{bmatrix} \tag{4.5}$$

Ao conjunto de valores ϕ_{kk} , $k = 1, 2, \dots$, chamamos de função autocorrelação parcial do período m . O conjunto de funções ϕ_{kk} , $m = 1, 2, \dots, s$, é uma outra forma de representar a estrutura de dependência do processo estocástico ao longo do tempo. Em um processo autorregressivo de ordem p_m , a função de autocorrelação parcial ϕ_{kk} será diferente de zero para k menor ou igual a p_m e igual a zero para k maior que p_m .

Para $k = 0$, a expressão (4.3) fica:

$$1 = \phi_1^m \rho_1^m + \phi_2^m \rho_2^m + \dots + \phi_{p_m}^m \rho_{p_m}^m + E \left[a_t \left(\frac{Z_t - \mu_m}{\sigma_m} \right) \right] \quad (4.6)$$

Multiplicando ambos os lados da equação (4.1a) por a_t e tomando o valor esperado, obtemos:

$$E \left[a_t \left(\frac{Z_t - \mu_m}{\sigma_m} \right) \right] = \sigma_a^{2(m)} \quad (4.7)$$

Substituindo (4.7) em (4.6) temos:

$$\sigma_a^{2(m)} = 1 - \phi_1^m \rho_1^m - \phi_2^m \rho_2^m - \dots - \phi_{p_m}^m \rho_{p_m}^m \quad (4.8)$$

Esta expressão é válida para qualquer período m .

4.2 Ajuste do Modelo

Box e Jenkins [7] sugeriram uma metodologia bastante elaborada para ajuste de modelos estocásticos da família ARIMA a séries temporais, que pode ser estendida para modelos da família PAR(p). Nesta metodologia a estratégia de seleção do modelo é dividida em três etapas. A primeira etapa, denominada por Box e Jenkins de **identificação do modelo**, consiste em escolher, tentativamente, a ordem do modelo baseando-se em estimativas das funções ρ_k^m e $\hat{\phi}_{kk}^m$ obtidas a partir da série amostral. Na modelagem autorregressiva periódica isso consiste em escolher o vetor p . A segunda etapa refere-se à **estimação do modelo**, ou seja, estimação dos seus parâmetros, sendo em geral recomendado o uso de estimadores de máxima verossimilhança ou suas aproximações. A terceira etapa diz respeito à **verificação do modelo**, isto é, verificar através de testes estatísticos se as hipóteses assumidas durante as etapas anteriores são atendidas. Se as hipóteses não são verificadas deve-se retornar à primeira etapa até que os resultados sejam satisfatórios.

Ressalta-se que esta estratégia pode muitas vezes resultar em mais de um modelo capaz de descrever o processo estocástico em estudo. Nestes casos a seleção do modelo mais adequado pode ser feita submetendo-os a testes de aplicação similares aos usados em Kelman *et al* [8] e em Oliveira *et al* [9].

4.2.1 Identificação

A identificação do modelo consiste em determinar as ordens mais apropriadas dos operadores autorregressivos de cada período p_m , $m = 1, \dots, s$. Isto pode ser feito obtendo-se estimativas $\hat{\phi}_{kk}^m$, $k = 1, N/4$ e substituindo em (4.5) as autocorrelações pelos respectivos valores amostrais (2.8) e (2.9). 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} [10],[11]. 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 significativas.

4.2.2 Estimação

Após a etapa de identificação é necessário obter estimativas para os diversos parâmetros do modelo. Para modelos autorregressivos os estimadores de momento são em geral bastante eficientes [12].

As médias e os desvios padrões sazonais são estimados pelas equações (2.6) e (2.7) respectivamente.

Os parâmetros ϕ_i^m , $i = 1, \dots, p_m$, são estimados substituindo-se em (4.4) ρ_k^{m-j} , $j = 0, \dots, (p_m - 1)$, $k = 1, \dots, p_m$, por suas estimativas dadas pelas expressões (2.8) e (2.9).

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

Cada um dos m sistemas resultantes pode ser resolvido por Decomposição de Cholesky.

Finalmente, as estimativas de $\sigma_a^{2(m)}$ podem ser obtidas usando-se a expressão (4.8).

4.2.3 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 numa análise detalhada dos resíduos ajustados a_t . Se o modelo estimado for adequado, a média dos resíduos deve ser igual zero, a variância deve ser pequena e constante e a função de correlação dos resíduos, $\rho_a^m(k)$, deve ser não significativa para $k > 0$.

4.3 Geração de Séries Sintéticas com o Modelo PAR(p)

Quando dispomos de uma série temporal de medições de vazões naturais, comumente a denominamos de série histórica. A série histórica é apenas uma das possíveis realizações de um processo estocástico, ou seja, pode-se imaginar que a natureza "sorteou" a série histórica segundo algum conjunto de leis probabilísticas. A um novo sorteio corresponde outra série diferente da série histórica, mas igualmente provável [11].

O modelo PAR(p) é utilizado para aproximar este comportamento estocástico, ou seja, o modelo deve permitir que, artificialmente, se façam tantos sorteios quantos forem necessários para o estudo em foco.

Cada sorteio está associado a uma série temporal, chamada de série sintética.

Seja a equação do modelo (4.1a):

$$\left(\frac{Z_t - \mu_m}{\sigma_m}\right) = \phi_1^m \cdot \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right) + \dots + \phi_{p_m}^m \cdot \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right) + a_t$$

De posse da expressão (4.8) podemos escrever (4.1a) na forma:

$$\mathbf{Z}_t = \mu_m + \phi_1^m \cdot \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{p_m}^m \cdot \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \sigma_m \sigma_a^{(m)} E_t \quad (4.9a)$$

ou

$$\mathbf{Z}_t = \mu_m + \phi_1^m \cdot \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{p_m}^m \cdot \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \sigma_m \sqrt{1 - \phi_1^m \rho_1^m - \phi_2^m \rho_2^m - \dots - \phi_{p_m}^m \rho_{p_m}^m} \cdot E_t \quad (4.9b)$$

onde a série de ruídos $\{E_t\}$ é independente com média zero e variância unitária.

A expressão (4.9b) é particularmente conveniente na geração de séries sintéticas, à medida que permite um esquema de geração sequencial no qual a cada instante t , Z_t é obtido como uma função de valores $Z_{t-\tau}$, $\tau = 1, \dots, p_m$, e de um sorteio E_t .

Para se iniciar o processo de geração é preciso arbitrar valores iniciais $Z_{t-\tau}$, onde τ pode ir de 1 a p_m . Estimativas para esses valores iniciais podem ser, por exemplo, as médias mensais ou os últimos valores amostrais. Para não condicionar a série sintética aos valores iniciais deve-se desprezar, por exemplo, os primeiros gerados.

O conjunto dos últimos valores observados ($Z_{t-\tau}$) em uma série hidrológica é denominado de tendência hidrológica, Figura 4.1, e a média mensal da série histórica é chamada de MLT mensal.

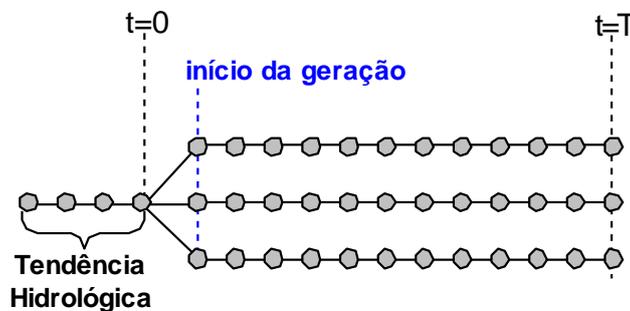


Figura 4.1 – Tendência Hidrológica

A geração das séries sintéticas pode ser realizada de forma condicionada ou não aos valores iniciais. Para não condicionar a série sintética aos valores iniciais deve-se desprezar, por exemplo, os primeiros valores gerados. Para tanto, o processo de geração da série sintética deve ser realizado em duas etapas. Na primeira etapa, a geração das séries sintéticas utiliza como valores iniciais as médias correspondentes a cada período (MLT mensal), conforme ilustrado na Figura 4.2. Os valores gerados, resultantes da primeira etapa, para os últimos períodos são utilizados como valores iniciais na segunda etapa do processo de geração, conforme Figura 4.3. Desta forma, as séries sintéticas geradas durante a segunda etapa são ditas "não condicionadas".

Para gerar séries sintéticas "condicionadas" deve-se efetuar apenas a primeira etapa do processo descrito acima, utilizando como valores iniciais uma tendência hidrológica apropriada.

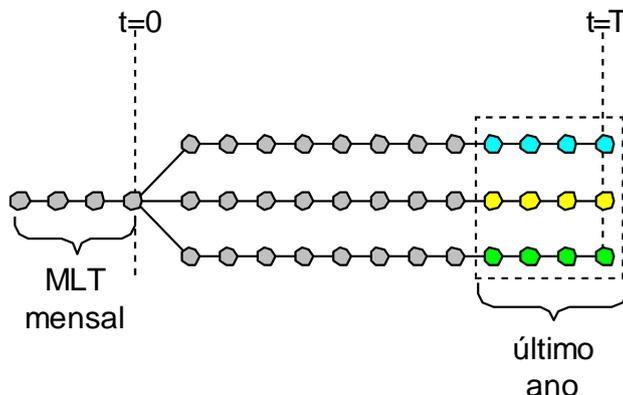


Figura 4.2 – Geração não condicionada – Primeira Etapa

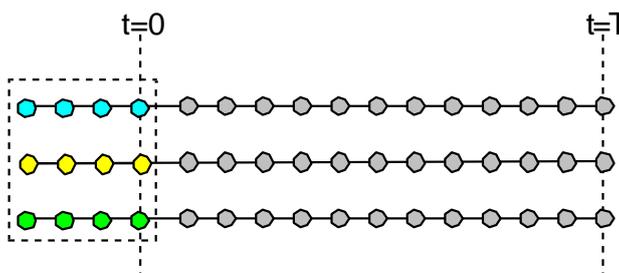


Figura 4.3 – Geração não condicionada – Segunda Etapa

4.3.1 Vazões Incrementais Negativas

Um problema frequente em sistemas com reservatórios em cascata, onde as vazões incrementais podem ser muito pequenas, é a geração de vazões mensais negativas. Para obter-se um Z_t positivo é necessário que:

$$Z_t = \mu_m + \phi_1^m \cdot \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{p_m}^m \cdot \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \sigma_m \cdot a_t > 0 \quad (4.10a)$$

ou

$$a_t > - \left(\frac{\mu_m}{\sigma_m} \right) - \phi_1^m \cdot \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) - \dots - \phi_{p_m}^m \cdot \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \quad (4.10b)$$

Muitos pesquisadores assumem que os resíduos a_t apresentam distribuição Normal e uma possível não normalidade pode ser corrigida pela transformação Box-Cox [13]. Como as séries sintéticas produzidas serão utilizadas em modelos que calculam as estratégias ótimas de operação de um sistema multireservatórios, baseados em programação dinâmica dual estocástica, o modelo de geração de séries sintéticas deve ser aplicado diretamente à série temporal original e deve ser capaz de lidar com resíduos que apresentam um forte coeficiente de assimetria.

A solução adotada foi ajustar uma distribuição Lognormal com três parâmetros aos resíduos mensais a_t , [14], cujo terceiro parâmetro Δ_t é conhecido como deslocamento. Conseqüentemente, a variável ξ_t tem distribuição Normal com média μ_{ξ_t} e desvio-padrão igual a $\sigma_{\xi_t}^2$:

$$\xi_t = \ln(a_t - \Delta_t) \tag{4.11}$$

Os parâmetros Δ_t , $\sigma_{\xi_t}^2$ e μ_{ξ_t} são estimados de tal forma a preservar os momentos dos resíduos $\sigma_a^{2(m)}$ e $\mu_a^{(m)}$, através das seguintes relações [15]:

$$\mu_{\xi_t} = \frac{1}{2} \ln\left(\frac{\sigma_a^{2(m)}}{\theta^2 - \theta}\right) \tag{4.12}$$

$$\sigma_{\xi_t}^2 = \ln(\theta) \tag{4.13}$$

$$\Delta_t = \frac{-\sigma_a^{(m)}}{\sqrt{(\theta - 1)}} \tag{4.14}$$

Fazendo Δ igual a expressão (4.10b) podemos escrever:

$$\theta = 1 + \frac{\sigma_a^{2(m)}}{(\mu_a^{(m)} - \Delta_t)^2} \tag{4.15}$$

$\sigma_{\xi_t}^2$ pode ser diretamente obtido a partir das expressões (4.5) e (4.15).

4.3.2 Correlação Espacial

Com o intuito de gerar vazões mensais multivariadas, assume-se que ruídos normais padrão não correlacionados espacialmente, b_t , podem ser transformados em ruídos espacialmente correlacionados, W_t , da seguinte forma:

$$W_t = Db_t \tag{4.16}$$

onde D é uma matriz quadrada de dimensão igual ao número de usinas hidroelétricas.

A matriz D pode ser estimada por:

$$\hat{U} = AVA' \tag{4.17}$$

$$D = A\sqrt{V}A' \tag{4.18}$$

onde \hat{U} pode ser, em princípio, a estimativa da matriz de covariâncias dos resíduos. Por exemplo, o elemento u_{ab} é a covariância espacial lag zero entre os resíduos das usinas a e b . A equação 4.17 corresponde a técnica conhecida como decomposição espectral, onde a matriz A é a matriz diagonal dos autovalores e V a matriz simétrica de autovetores. A decomposição espectral é obtida utilizando o Método de Jacobi [16]. Na prática, o comportamento dos resíduos não segue o comportamento das vazões: os resíduos não são espacialmente correlacionados. No entanto, a fim de se preservar as dependências espaciais entre as usinas utilizou-se a correlação espacial entre as vazões em substituição à correlação espacial entre os resíduos.

4.3.3 Usinas com vazões afetadas por tempo de viagem

Nesta seção introduz-se o conceito de vazão lateral que difere do conceito de vazão incremental. O termo vazão incremental é utilizado para se referir à diferença entre a vazão total em um posto de

medição num dado mês e a soma das vazões totais de todos os postos imediatamente a montante deste no mesmo mês; enquanto o termo vazão lateral é utilizado para se referir à diferença entre a vazão total no posto de medição num dado mês e a soma das vazões propagadas de todos os postos imediatamente a montante que afluem ao posto de medição no mesmo mês. A vazão lateral está assim associada exclusivamente à contribuição lateral num dado mês à vazão total afluente ao posto de medição. Para a maioria dos postos representados no SIN, a vazão lateral é idêntica à vazão incremental em escala mensal, quando se admite que o tempo de viagem da água de uma dada usina hidrelétrica até a próxima usina a jusante é nulo. Porém, quando existe um tempo de viagem elevado entre dois ou mais aproveitamentos adjacentes, essa igualdade passa a não ser verificada e é importante modelar tanto o posto de vazão total/incremental quanto o posto de vazão lateral.

No programa GEVAZP é possível gerar cenários de vazões laterais e vazões incrementais para uma usina hidrelétrica afetada pela propagação da vazão das usinas de montante. Desta forma, se faz necessário fornecer registros históricos de vazões naturais totais e laterais para todas as usinas afetadas pela propagação das vazões dos postos de montante.

4.3.3.1 Geração de cenários de vazões laterais

A geração de cenários de vazão lateral, independente da classificação da usina como principal ou satélite. Será ajustado um modelo autorregressivo periódico baseado no histórico de vazões laterais, assim como é realizado para os demais postos principais da configuração, de acordo com a metodologia descrita no item 4.

A geração dos ruídos w_t dos postos de vazão lateral é feita após a obtenção dos ruídos w_t dos cenários de vazões incrementais dos postos principais e a geração dos ruídos de cada posto de vazão lateral é feita individualmente. Em cada processo de geração de ruído w_t adota-se uma geração condicionada aos ruídos previamente gerados para os cenários de vazões incrementais de postos principais selecionados, conforme:

$$W_1 = (\Sigma_{1,2} \Sigma_{2,2}^{-1}) W_2 + E b_1 \quad (4.19)$$

$$E E^T = 1 - \Sigma_{1,2} \Sigma_{2,2}^{-1} \Sigma_{1,2}^T \quad (4.20)$$

Onde:

- O índice 1 indica o posto de vazão lateral e o índice 2 indica o conjunto selecionado de postos principais;
- $\Sigma_{1,2}$ é o vetor-linha de correlações históricas anuais entre a série de vazão lateral e as séries de vazão incremental dos postos principais selecionados;
- $\Sigma_{2,2}$ é a matriz de correlações históricas anuais entre as séries de vazão incremental dos postos principais selecionados.

A seleção dos postos principais que irão participar da geração de cenários de vazão lateral para um dado posto será feita conforme a classificação do posto de vazão total associado.

- a) Posto principal: o conjunto de postos principais selecionados se resume ao próprio posto de vazão total;
- b) Posto satélite: o conjunto de postos principais selecionados é formado pelos postos principais imediatamente a montante e o posto principal imediatamente a jusante.

4.3.3.2 Geração de cenários de vazões totais/incrementais

No programa GEVAZP são disponibilizadas duas alternativas para obtenção dos cenários de vazões totais/incrementais para as usinas afetadas por propagação de vazão: (i) gerar cenários de vazão incremental e somá-los aos cenários de vazões totais das usinas à montante; (ii) utilizar os cenários gerados de vazão lateral e somá-los aos cenários de vazões totais das usinas de montante defasadas

considerando uma regra linear de defasagem. A alternativa a ser considerada é uma escolha do usuário e pode ser definida por usina.

Na alternativa (i), os cenários gerados de vazão incremental são somados aos cenários de vazões totais das usinas de montante. A propagação das vazões dos postos de montante está representada de forma implícita, já que este fenômeno natural está refletido no histórico de vazões utilizado para ajustar o modelo.

Na alternativa (ii), os cenários gerados de vazão lateral são somados aos cenários de vazões totais das usinas de montante defasadas por uma regra de defasagem linear. Esta regra leva em consideração: (a) o tempo de viagem entre a usina e as usinas imediatamente a montante, (b) as vazões naturais totais geradas para as usinas de montante, e (c) a vazão lateral gerada para a usina em questão, conforme 4.3.3.1.

$$ZT(t) = \left[\beta * \sum_{i=1}^{nmont} ZT_{mont_i}(t-1) \right] + \left[(1-\beta) * \sum_{i=1}^{nmont} ZT_{mont_i}(t) \right] + Z_{lat}(t) \quad (4.21)$$

Onde:

$$\beta = \frac{ndias}{ndmes};$$

ndias é o tempo de viagem em dias;

ndmes é o número de dias do mês;

ZT_{mont_i} é a vazão total da usina de montante i .

Maiores detalhes sobre a consideração do tempo de viagem no cálculo da vazão incremental podem ser consultados na seção 5.1.10.2 do Manual de Referência do modelo DECOMP.

4.3.4 Cálculo de vazões para Postos Artificiais

A geração de cenários sintéticos é realizada somente para usinas com postos de vazão natural. As vazões naturais históricas são obtidas retirando-se o efeito de ações antrópicas (usos consuntivos, bombeamento, etc.) e incorporando-se as vazões relativas à evaporação líquida dos reservatórios. Porém, em algumas representações se faz necessário o uso de registros de vazões artificiais. Postos artificiais correspondem à vazão natural, levando-se em conta o efeito de qualquer regra de operação no local, ou à montante do local considerado, que possa contribuir na alteração do regime natural. Em geral, os postos artificiais são dotados de uma regra particular para o cálculo de sua vazão.

A partir destas regras e dos cenários de vazões naturais totais gerados pelo GEVAZP é possível calcular as vazões para os postos artificiais. O cálculo dos cenários de vazões para estes postos ocorre após a geração de cenários para postos principais e o cálculo de cenários de vazão total para os postos satélites. O fluxograma abaixo, Figura 4.4, descreve as etapas do processo de cálculo de vazão para os postos artificiais.

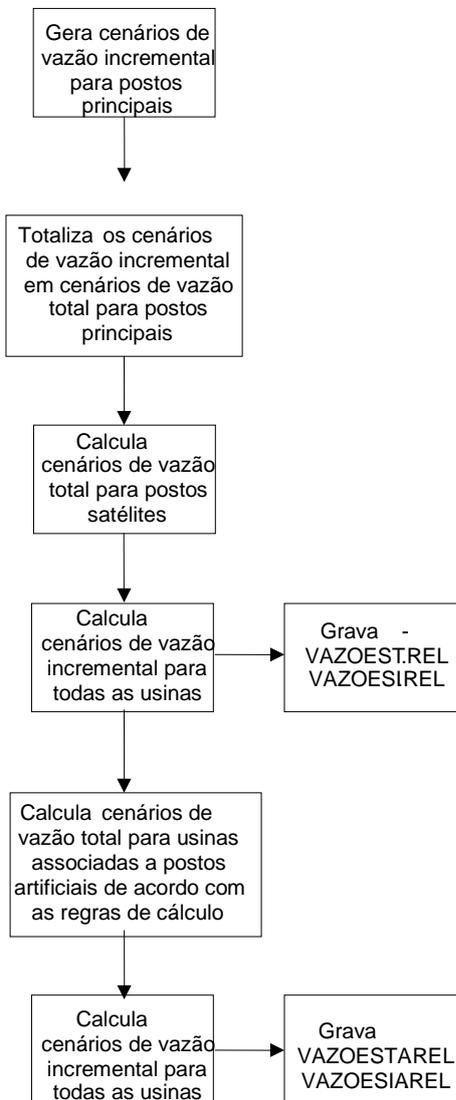


Figura 4.4 – Fluxograma da geração de cenários de vazão para postos artificiais.

5 APLICAÇÃO DAS TÉCNICAS DE AGREGAÇÃO DE CENÁRIOS AO MODELO GEVAZP

Devido à restrição de tempo computacional, no modelo DECOMP é interessante trabalhar com o menor número de cenários de vazões possível. Entretanto, um número reduzido de séries não é suficiente para caracterizar bem um processo estocástico com tal variabilidade. Para resolver este problema, a técnica empregada é a de agregação de cenários.

Agregação de cenários é uma técnica estatística capaz de elaborar critérios que possibilitam agrupar cenários similares em determinados grupos, de tal forma que a árvore reduzida preserve as características do processo estocástico do qual as vazões são geradas. Esta técnica multivariada é um caso particular da Análise de Conglomerados (Cluster Analysis).

5.1 Agregação de Cenários

No modelo de planejamento da operação de curto prazo (DECOMP), a incerteza acerca das vazões afluentes às diversas usinas hidroelétricas do sistema é representada por cenários hidrológicos através de uma árvore de afluências, ilustrada na Figura 5.1, com probabilidades de ocorrência associadas a cada ramo. Admite-se que as probabilidades de ocorrência são as mesmas para cada ramo, isto é, os cenários são equiprováveis.

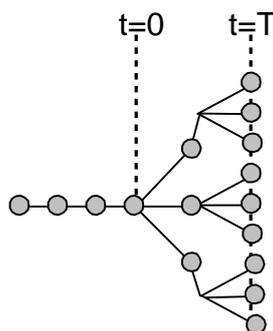


Figura 5.1 – Estrutura de Árvore

Dizer que os cenários são equiprováveis não significa dizer que as vazões tenham a mesma probabilidade de ocorrência. Os cenários são sorteados por um modelo de geração de vazões sob condições idênticas, logo todos têm a mesma possibilidade de ocorrência. Na prática é usual chamar estes cenários igualmente possíveis de cenários equiprováveis.

O modelo de geração de séries sintéticas de vazões (GEVAZP) é o responsável pela geração dos cenários hidrológicos necessários ao modelo DECOMP. Para que o processo estocástico de afluências seja bem representado, a árvore de afluências deve conter muitos ramos, pois poucos cenários hidrológicos não conseguem descrever adequadamente as características do processo estocástico de afluências. Porém a árvore não pode conter muitos ramos uma vez que isto tornaria a solução do problema de planejamento de curto prazo inviável por necessitar de muito esforço computacional.

A restrição de tempo computacional pode ser contornada utilizando-se Análise de Conglomerados. O principal objetivo da aplicação das técnicas de agregação ao modelo GEVAZP é a redução do número de cenários hidrológicos através da escolha de um conjunto representativo de cenários hidrológicos.

Os cenários que fazem parte deste conjunto representativo devem conter toda a informação necessária para representar o processo estocástico de afluências. Eles são obtidos através de agrupamento de cenários semelhantes, possuem características similares aos demais componentes do

grupo em que estão localizados e têm pesos diferentes, de acordo com o número de cenários contidos no grupo.

Este capítulo mostra como a Análise de Conglomerados é incorporada ao modelo de geração de séries sintéticas de vazões, em particular, como os objetos e variáveis utilizados são definidos, o tratamento feito nos dados de entrada, os testes implementados para a validação dos resultados obtidos e outras definições utilizadas durante o processo de agregação.

5.1.1 Definições

Para esta aplicação os objetos a serem agrupados são vetores de vazões naturais afluentes e as variáveis que os descrevem são as vazões naturais a cada usina hidroelétrica considerada na configuração do sistema hidroelétrico brasileiro. Estes vetores de vazões são gerados de maneira multivariada conforme descrito no Capítulo 4, e tendem a preservar as características do registro histórico, tais como a média, desvio padrão e correlação entre as vazões das usinas hidroelétricas.

$$Z^T_i = [z_{i1} \ z_{i2} \ z_{i3} \ \Lambda \ z_{ir}] \quad i = 1, \dots, N \quad e \quad r = 1, \dots, P \quad (5.1)$$

onde N é número de objetos a serem agrupados, isto é, o número total de cenários hidrológicos gerados por período e por ramo da árvore de afluências, e P é o número de variáveis utilizadas para descrever cada objeto, ou seja, o número de usinas hidroelétricas que participam do processo de geração de cenários sintéticos.

Antes de inicializar o processo de agregação, as variáveis devem ser padronizadas para que vazões com diferentes ordens de grandeza não prejudiquem o agrupamento dos vetores de vazões. Após a padronização o vetor de vazões é definido como:

$$Z^{*T}_i = [z^*_{i1} \ z^*_{i2} \ z^*_{i3} \ \Lambda \ z^*_{ir}] \quad i = 1, \dots, N \quad e \quad r = 1, \dots, P \quad (5.2)$$

onde cada componente z^*_{ir} é a variável z_{ir} padronizada, calculada como:

$$z^*_{ir} = \frac{z_{ir} - \mu_r}{\sigma_r} \quad (5.3)$$

A média e o desvio padrão, das vazões de cada usina hidroelétrica que compõem o vetor de vazões Z_i , de um determinado estágio para os N cenários, são representados por μ_r e σ_r , respectivamente.

$$\mu_r = \frac{1}{N} \sum_{i=1}^N z_{ir} \quad (5.4)$$

$$\sigma_r = \sqrt{\frac{1}{N} \sum_{i=1}^N (z_{ir} - \mu_r)^2} \quad (5.5)$$

A cada estágio do horizonte de estudo e a cada ramo da árvore de afluências são gerados 1000 cenários hidrológicos distintos (N=1000) e equiprováveis para o processo de agregação. O número de aberturas em cada estágio é o número de agrupamentos obtidos a partir dos N cenários, onde N é um dado de entrada.

Supondo a não existência de sazonalidade no processo estocástico de afluências, aconselha-se que o número definido de aberturas em um estágio seja no mínimo igual ao definido no estágio anterior. Isto porque a incerteza com relação às afluências futuras aumenta ao longo do horizonte de estudo. Logo, não faz sentido usar um número menor de cenários para representar uma incerteza maior quanto às afluências futuras.

5.1.2 Método de Agrupamento Utilizado

O método escolhido foi o método de agrupamento não hierárquico K-Means, pois o número de cenários hidrológicos que são fornecidos 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 algoritmo do método K-Means é acoplado ao modelo GEVAZP após a rotina de geração dos N cenários hidrológicos multivariados equiprováveis. Os cenários resultantes do processo de agregação são os cenários efetivamente representados na estrutura da árvore de aflúncias.

5.1.3 Medida de Distância

A métrica utilizada para calcular a similaridade entre os objetos (cenários hidrológicos multivariados) e os centróides dos grupos é a distância Mahalanobis. Esta métrica é definida como em (5.6).

$$d_{i,j} = \sqrt{(Z_i - Z_j)' \Sigma^{-1} (Z_i - Z_j)} \quad (5.6)$$

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

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

Utilizando esta métrica pretende-se preservar além da média e do desvio padrão, a correlação espacial entre as usinas hidroelétricas.

As variáveis que descrevem os objetos são padronizadas pela média e desvio padrão mensal, de acordo com (5.3).

Antes da padronização dos dados, é aplicada a transformação logarítmica nos vetores de vazões (5.7) para que a dispersão das variáveis possa ser reduzida. Como a transformação logarítmica é monotônica, a ordem original das distâncias entre os objetos não é afetada.

$$x_{ir} = \log(z_{ir}) \quad (5.7)$$

Geralmente os objetos mais dispersos são representantes de cenários com vazões elevadas e, quando não se aplica a transformação logarítmica nos dados, estes objetos tendem a se isolar em um grupo. Por outro lado, cenários com baixas aflúncias têm pouca dispersão entre si e tendem a se localizar em um mesmo grupo. Como a transformação logarítmica diminui a dispersão entre os objetos, a formação de alguns grupos com poucos cenários e outros com muitos é evitada. Desta maneira, tantos os cenários favoráveis (úmidos) quanto os cenários críticos (secos) são representados de forma mais equilibrada.

A Figura 5.2 mostra cenários agregados resultantes para a usina de Itutinga, com e sem a aplicação da transformação logarítmica nos dados durante o processo de agregação.

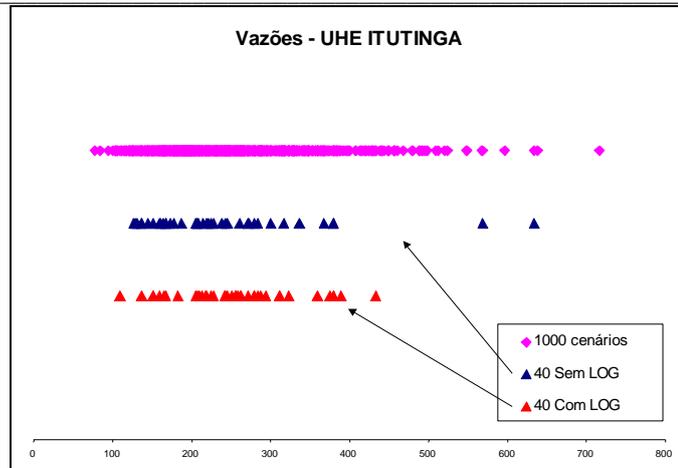


Figura 5.2 - Aplicação da Transformação Logarítmica

5.1.4 Determinação dos Cenários Representativos

O processo de agregação é inicializado através do sorteio de sementes aleatórias para representar os centroides dos grupos. Estes pontos iniciais são objetos do conjunto de entrada escolhidos aleatoriamente. Logo, são cenários hidrológicos do conjunto original gerado sinteticamente. Desta maneira, pode-se garantir que nenhum grupo ficará vazio.

Nos passos seguintes até a convergência do processo de agregação, o centroide dos grupos (C_k) será o ponto médio destes grupos, isto é:

$$C_k = \frac{1}{NO_k} \sum_{i=1}^{NO_k} Z_i \tag{5.8}$$

onde NO_k é o número de cenários hidrológicos pertencentes ao grupo k . O cenário hidrológico Z_i é um vetor de vazões multivariadas.

Após a convergência do processo, o centroide dos grupos será o objeto mais próximo do ponto médio deste grupo. A Figura 5.3 ilustra como é escolhido o representante de cada grupo formado.

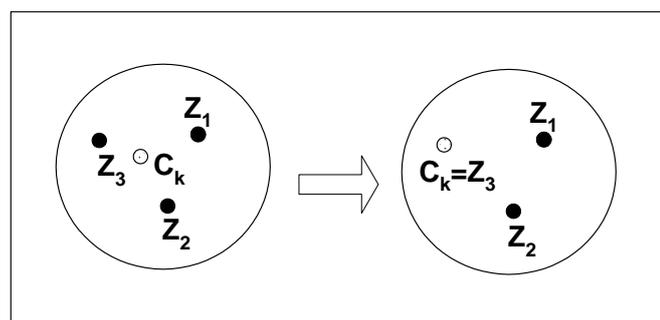


Figura 5.3 – Escolha do Cenário Representativo

A escolha do cenário representante é realizada desta forma para garantir que este sempre será um dos cenários gerados pelo modelo ajustado. Logo, os cenários resultantes do processo de agregação preservam as características captadas pelo modelo autorregressivo periódico, tal como a estrutura da correlação espacial.

Tomando a Figura 5.3 como exemplo, supondo que após a convergência do processo de agregação os cenários hidrológicos Z_1 , Z_2 e Z_3 , cada um composto por vazões afluentes à todas as usinas

hidroelétricas, formem um grupo cujo ponto médio é representado pelo vetor C_k . O cenário hidrológico Z_3 é escolhido para representar o grupo, pois é o mais próximo do ponto médio deste.

Cada cenário representante tem associado a ele uma probabilidade de ocorrência. A probabilidade P_k associada ao grupo k é calculada como:

$$P_k = \frac{NO_k}{N} \quad (5.9)$$

onde N é número de cenários equiprováveis de vazões multivariadas gerados antes do processo de agregação.

5.1.5 Correlação Temporal e Espacial

Após o processo de agregação, permanecem apenas K cenários de vazões multivariadas não equiprováveis. Deste processo não resultam novos valores de vazões multivariadas. Sendo assim, as correlações temporal e espacial são preservadas.

A correlação temporal ou serial não é alterada, pois a cada estágio (mês) é gerado novamente todo o conjunto de cenários necessários para descrever o processo estocástico (N cenários). A geração deste conjunto de cenários é realizada para cada nó do estágio corrente, e isto é feito levando-se em conta o passado destes nós.

Cada cenário hidrológico considerado no processo de agregação é composto por vazões afluentes a cada usina hidroelétrica pertencente à configuração do sistema hidroelétrico brasileiro, e estes cenários são gerados de forma a preservar a correlação espacial histórica. Desde que os cenários hidrológicos selecionados como representativos da amostra pertencem ao conjunto original de cenários e, como não há troca de vazões entre cenários e entre usinas hidroelétricas, a correlação espacial entre as vazões às diversas usinas hidroelétricas não é prejudicada.

5.1.6 Estrutura do Novo processo de Geração de Cenários Sintéticos em Árvore

Supondo-se, por exemplo, que a cada estágio, um nó dê origem a três novos nós na estrutura da árvore, como ilustrado na Figura 5.4.

No primeiro estágio serão gerados N ($N=1000$, por exemplo) cenários equiprováveis de vazões multivariadas a partir de uma única sequência de vazões multivariadas conhecidas (Z_3, Z_2, Z_1). A seguir, o processo de agregação é aplicado resultando em três cenários não equiprováveis de vazões multivariadas (Z^1_4, Z^2_4, Z^3_4).

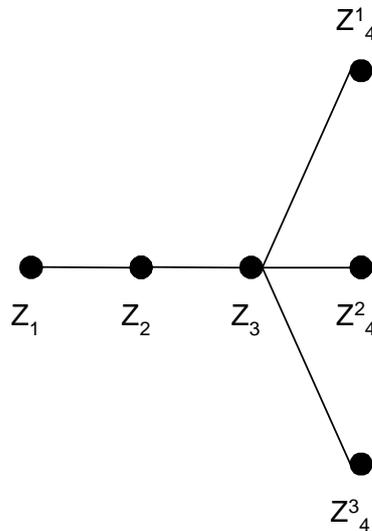


Figura 5.4 – Caso Exemplo (1º per. Estocástico)

No segundo estágio, para a sequência (Z^1_4, Z_3, Z_2) , são gerados N cenários equiprováveis de vazões multivariadas. Do processo de agregação três cenários não equiprováveis de vazões multivariadas são produzidos $(Z^{1}_{5,1}, Z^{2}_{5,1}, Z^{3}_{5,1})$. Ainda neste estágio, a partir da sequência (Z^2_4, Z_3, Z_2) são gerados N cenários equiprováveis de vazões multivariadas. Da mesma forma, o processo de agregação é aplicado e, três cenários não equiprováveis de vazões multivariadas são obtidos $(Z^{1}_{5,2}, Z^{2}_{5,2}, Z^{3}_{5,2})$. Por último, a partir da sequência (Z^3_4, Z_3, Z_2) , são gerados N cenários de vazões multivariadas. O processo de agregação é novamente aplicado resultando em três cenários não equiprováveis de vazões multivariadas $(Z^{1}_{5,3}, Z^{2}_{5,3}, Z^{3}_{5,3})$, Figura 5.5.

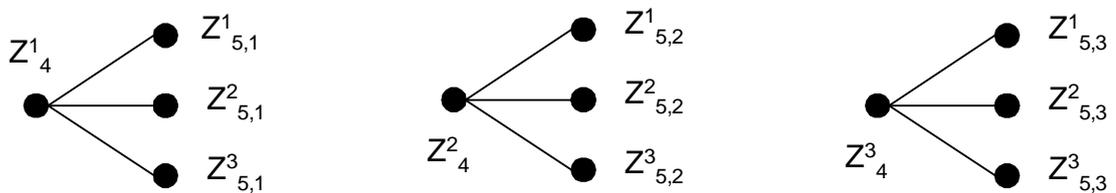


Figura 5.5 - Caso Exemplo (2º per. Estocástico)

Este procedimento é repetido para cada um dos estágios do horizonte. A árvore de afluências gerada é mostrada na Figura 5.6.

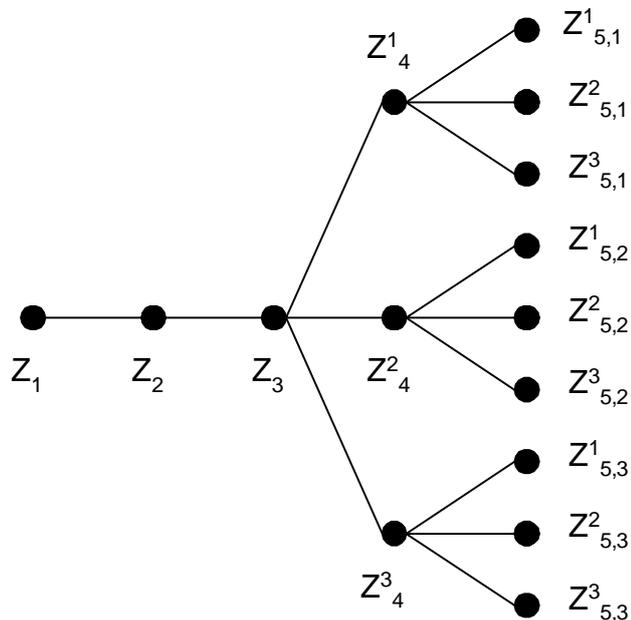


Figura 5.6 – Caso Exemplo (Árvore Completa)

A Figura 5.7 ilustra o novo processo de geração de séries sintéticas a ser utilizado pelo modelo DECOMP. Note que a cada estágio do horizonte de estudo, o processo de agregação é incorporado à geração de cenários hidrológicos sintéticos.

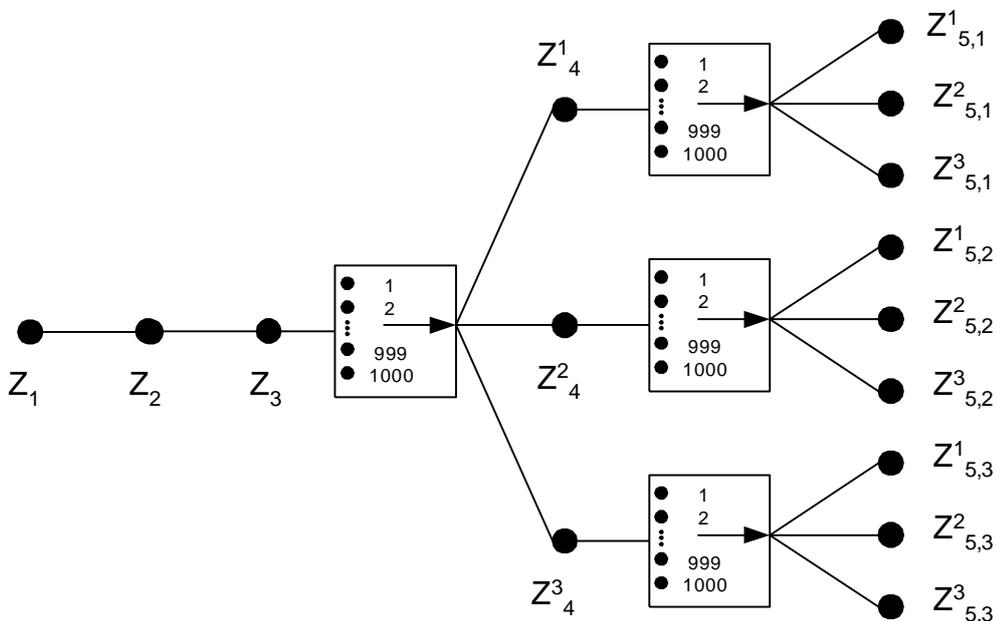


Figura 5.7 – Geração de Séries Sintéticas com Agregação

6 CRITÉRIOS PARA TESTAR A ADEQUAÇÃO DO MODELO

6.1 Testes para Séries Geradas em Paralelo (Pente)

6.1.1 Verificação das Estatísticas Anuais e Mensais

Qualquer característica do processo que tenha sido usada na definição do modelo estocástico de afluições não pode ser utilizada para testar a adequação do modelo. Por exemplo, se as médias sazonais são parâmetros do modelo que está sendo analisado, é comum comparar as médias históricas e geradas: já se sabe "a priori" que elas serão preservadas. Na prática, os momentos são verificados na tentativa de achar inconsistências ou desvios.

6.1.2 Avaliação do Desempenho do Modelo

Um modelo de geração de séries sintéticas deve preservar as principais características da série observada. Em termos práticos, a utilidade de um modelo pode ser aferida por sua capacidade de reproduzir distribuições de probabilidade de variáveis aleatórias relevantes ao processo. Estas variáveis devem naturalmente refletir os requisitos da aplicação proposta. No caso do planejamento da expansão de um sistema de geração, a representação de períodos críticos, as correlações serial e espacial entre as vazões afluentes aos diversos reservatórios, curvas de regularização para diversos reservatórios, etc., devem de alguma forma ser levados em conta nesta avaliação [17].

Em geral, aceita-se que as variáveis aleatórias selecionadas devem estar relacionadas com o conceito de **sequência**: uma sequência negativa é o período de tempo em que as vazões afluentes estão continuamente abaixo de valores pré-determinados, por exemplo, as médias mensais, precedidos e sucedidos por valores acima destes limites.

A Figura 6.1 ilustra o conceito de sequência e das variáveis associadas. A linha contínua representa afluições e a linha pontilhada um limite pré-estabelecido. Os intervalos $(t_1 - t_2)$ e $(t_3 - t_4)$ correspondem às sequências negativas, isto é, as vazões estão abaixo dos limites.

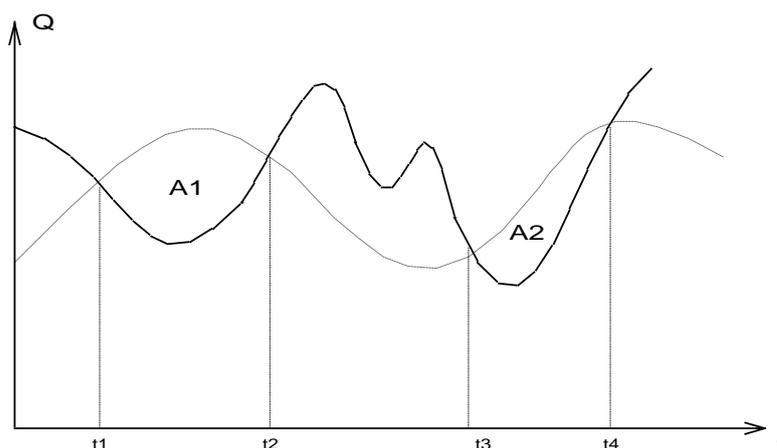


Figura 6.1 - Sequência Negativa

As variáveis associadas são:

Comprimento de sequência corresponde ao comprimento dos intervalos $(t_2 - t_1)$, $(t_4 - t_3)$, na Figura 6.1.

$$C = t_2 - t_1 \tag{6.1}$$

Soma de sequência corresponde à área abaixo do limite durante a sequência. Na Figura 6.1, as somas de sequências são as áreas A_1 e A_2 .

$$S = \sum_{i=t_1}^{t_2} (Z_i - \mu_i) \tag{6.2}$$

Intensidade de sequência corresponde ao valor médio abaixo do limite, isto é, a soma de sequência dividida pelo respectivo comprimento de sequência.

$$I = \frac{S}{C} \tag{6.3}$$

Repetindo estes cálculos para cada sequência negativa é possível obter da série histórica amostras de três variáveis aleatórias:

- a) soma de sequência negativa
- b) comprimento de sequência negativa
- c) intensidade de sequência negativa

O mesmo procedimento é realizado para a série sintética, onde o limite pré-determinado é o mesmo usado na série histórica, e obtemos outras três amostras. Podemos, então, testar a hipótese de que as duas amostras são provenientes de distribuições idênticas.

Seja n_1 o número de observações da variável aleatória X , obtida a partir da série histórica e n_2 o número de observações da variável aleatória Y , obtida a partir da série gerada. Caso essas variáveis aleatórias sejam discretas (ex.: comprimento de sequência) utilizamos o teste de igualdade de duas distribuições multinomiais.

O teste se baseia na divisão do espaço amostral em $(k+1)$ subconjuntos e a hipótese nula estabelece que:

$$H_0: p_{1,j} = p_{2,j} \quad j = 1, \dots, k+1 \tag{6.4}$$

onde $p_{1,j}$ é a probabilidade de que uma observação da primeira população pertença ao j -ésimo subconjunto e $p_{2,j}$ é a correspondente probabilidade para a segunda população.

Demonstra-se que:

$$\chi_k^2 = \sum_{i=1}^2 \sum_{j=1}^{k+1} \frac{\left(N_{ij} - n_i \frac{(N_{1j} + N_{2j})}{(n_1 + n_2)} \right)^2}{n_i \frac{(N_{1j} + N_{2j})}{(n_1 + n_2)}} \tag{6.5}$$

tem assintoticamente distribuição Qui-quadrado com k graus de liberdade, onde N_{ij} é o número de observações do j -ésimo subconjunto da i -ésima população.

Uma descrição detalhada deste teste é fornecida por Mood et al [18].

Caso as variáveis aleatórias em estudo sejam contínuas, por exemplo, a soma e intensidade de sequência, o teste de Smirnov para duas amostras é bastante adequado.

Sob a hipótese nula da igualdade entre duas distribuições, a variável aleatória $W = \text{Max}_x |S_1(x) - S_2(x)|$ tem distribuição conhecida. $S_1(x)$ é a distribuição de frequências acumulada relativa à série histórica e $S_2(x)$ é a correspondente distribuição relativa à série gerada. Para grandes valores de n_1 e n_2 , o percentual de 95% de W é aproximadamente dado por:

$$W_{\text{crítico}} = 1,358 \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \tag{6.6}$$

Quando $W > W_{\text{crítico}}$ a hipótese de igualdade entre as distribuições é rejeitada.

Descrição detalhada é fornecida por Bradley [19].

Outra variável aleatória que pode ser considerada na análise de desempenho é a de **somas parciais** que pode ser definida como a sequência $\{S_t\}$, $t = 1, 2, \dots, 12N$, onde:

$$S_t = \sum_{j=1}^t (Z_j - \beta\mu) \tag{6.7}$$

onde β é chamado de nível de regularização.

Neste caso a sequência de somas parciais é dada por:

$$S_t = \sum_{j=1}^t Z_j - \beta t \mu \tag{6.8}$$

e está ilustrada na Figura 6.2.

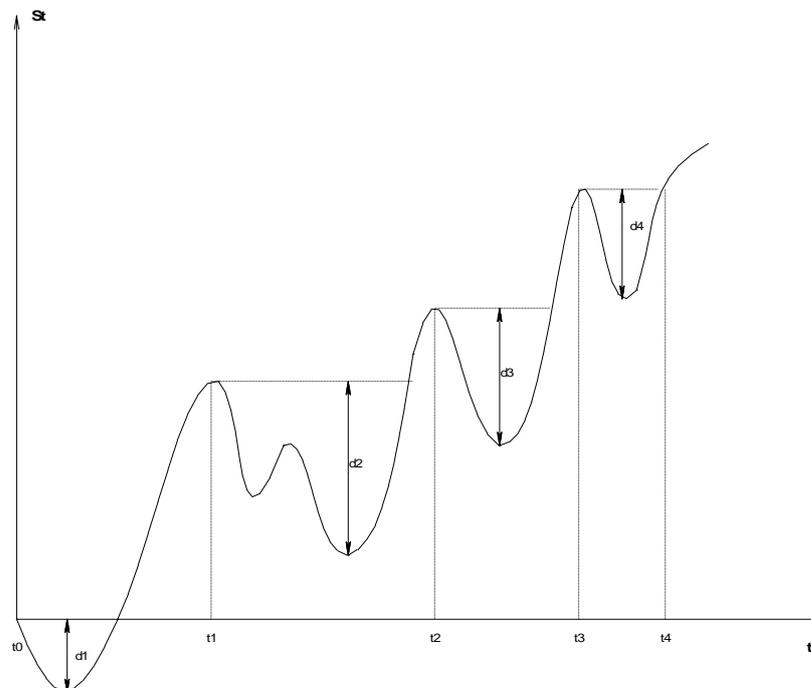


Figura 6.2 - Volume de regularização (Déficit)

O período t_0 a t_1 só transcorrerá sem ocorrência de déficit caso esteja disponível uma capacidade de armazenamento igual ou maior que $d_1 \cdot \Delta t$, onde Δt é o intervalo de discretização. Analogamente para o intervalo (t_1, t_2) é necessária uma capacidade de armazenamento no mínimo igual a $d_2 \cdot \Delta t$ e assim sucessivamente. Portanto, as grandezas d_1, d_2, \dots , que são observações de uma variável aleatória chamada déficit, tem grande interesse nos estudos de regularização das aflúências.

Um modelo orientado para aplicação no setor energético deve estar apto a produzir séries sintéticas que revelem uma distribuição amostral para a variável aleatória **déficit** estatisticamente indistinguível da distribuição amostral obtida a partir da série histórica.

Verificou-se, entretanto, que os índices mais relevantes para o planejamento estavam em geral associados a valores extremos das distribuições. O período crítico, por exemplo, corresponde à pior situação hidrológica em todo o histórico. Esta constatação é consequência da existência de reservatórios plurianuais, pouco sensíveis a flutuações de curto prazo. O primeiro índice do tipo "máximo" a ser definido foi o chamado máximo déficit, que corresponde ao volume do reservatório capaz de regularizar uma vazão especificada.

Rippl propôs utilizar o **máximo déficit**, obtido a partir da série histórica, como valor de projeto para a capacidade de um reservatório.

Para estudo das variáveis do tipo "máximo", o procedimento é o seguinte:

- divide-se a série gerada em M segmentos de comprimento igual ao da série histórica;
- calcula-se o valor do índice para cada segmento, obtendo-se assim uma distribuição de frequências.

Como a série histórica possui apenas um valor, não faz sentido falar de aderências de distribuições, mas sim da "tipicidade" do valor histórico em relação à distribuição dos valores gerados. Em outras palavras, deseja-se saber a probabilidade da amostra histórica ser sorteada dado que o modelo de geração escolhido é verdadeiro.

Em termos univariados, o desempenho do modelo pode ser medido pela proporção de índices gerados maiores ou menores do que o índice histórico. Se esta proporção for muito pequena, isto é uma indicação de que a observação histórica é atípica para o modelo considerado.

Esta análise pode ser realizada para os seguintes índices:

- máximo comprimento de sequência
- máxima soma de sequência
- máxima intensidade de sequência
- máximo déficit para um determinado nível de regularização
- comprimento do período crítico
- vazão média de período crítico

Maiores detalhes podem ser encontrados em [6].

6.2 Testes para Séries Geradas em Árvore

Alguns testes foram implementados para verificar se os cenários hidrológicos resultantes do processo de agregação conseguem manter as mesmas características do conjunto de cenários gerados originalmente e, se os cenários gerados estão coerentes com o modelo ajustado. Foram realizados testes para avaliar os momentos mensais condicionados de primeira e segunda ordem para os valores gerados, os valores de média, desvio padrão e correlação espacial entre as usinas hidroelétricas

resultantes dos cenários gerados e agregados e a aderência entre as distribuições marginais das vazões geradas e agregadas.

6.2.1 Teste de Momentos Condicionados de Ordem Um e Dois

Os valores teóricos correspondentes à média e ao desvio-padrão condicionados podem ser derivados dos modelos estocásticos mensais ajustados às usinas hidroelétricas. Estes valores devem ser utilizados como referência de comparação para avaliar os valores de média e desvio-padrão estimados a partir dos N cenários hidrológicos gerados. Através deste teste, será possível verificar se os cenários gerados são coerentes com o modelo estocástico usado para gerar a árvore.

Conforme apresentado no Capítulo 4, para uma usina hidroelétrica seja o modelo PAR(p) expresso como:

$$\left(\frac{Z_t - \mu_m}{\sigma_m}\right) = \phi_1^m \cdot \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right) + \dots + \phi_{p_m}^m \cdot \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right) + a_t \tag{6.9}$$

onde a_t é uma série de ruídos independentes com média zero e variância $\sigma_a^{2(m)}$, Z_t é a série de vazões afluentes a uma usina hidroelétrica e, os parâmetros μ_m e σ_m são a média e o desvio padrão históricos, respectivamente, da usina considerada.

A média de Z_t condicionada aos valores passados $Z_{t-} = \{Z_{t-1}, Z_{t-2}, \dots, Z_{t-p_m}\}$ é dada por:

$$E(Z_t | Z_{t-}) = \mu_m + \left[\phi_1^m \cdot \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right) + \dots + \phi_{p_m}^m \cdot \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right) \right] \cdot \sigma_m \tag{6.10}$$

A variância de Z_t condicionada aos valores passados Z_{t-} é dada por:

$$VAR(Z_t | Z_{t-}) = \sigma_m^2 \sigma_a^{2(m)} = \sigma_m^2 \left[1 - \phi_1^m \rho_1^m - \phi_2^m \rho_2^m - \dots - \phi_{p_m}^m \rho_{p_m}^m \right] \tag{6.11}$$

E o desvio padrão condicionado é dado por:

$$\sigma(Z_t | Z_{t-}) = \sigma_m \sigma_a^{(m)} = \sigma_m \sqrt{1 - \phi_1^m \rho_1^m - \phi_2^m \rho_2^m - \dots - \phi_{p_m}^m \rho_{p_m}^m} \tag{6.12}$$

6.2.2 Parâmetros das distribuições amostrais

Em [21] são apresentados os parâmetros da distribuição amostral para a média e o desvio-padrão. Estes parâmetros são necessários para verificar se os momentos condicionados de primeira e segunda ordem obtidos a partir dos N cenários gerados são condizentes com os valores teóricos calculados para estes momentos.

6.2.2.1 Média

O valor esperado da média amostral, $E(\bar{x})$, é igual à média da população (6.13) e a variância da média amostral, $VAR(\bar{x})$, é igual à variância da população dividida pelo número de elementos da amostra (6.14).

$$E(\bar{x}) = \mu = E(Z_t | Z_{t-}) \tag{6.13}$$

$$VAR(\bar{x}) = \frac{\mu_2}{N} = \frac{VAR(Z_t | Z_{t-})}{N} \tag{6.14}$$

onde μ e μ_2 são a média e a variância da população.

Pelo teorema do limite central pode-se dizer que a média tem distribuição Normal. Então o intervalo de confiança de 95% calculado para a média amostral estimada a partir dos N cenários hidrológicos gerados, para uma dada usina hidroelétrica, é dado por:

$$IC_{\mu} = E(Z_t | Z_{t-}) \pm 1,96 \frac{\sigma_{Z_t|Z_{t-}}}{\sqrt{N}} \tag{6.15}$$

6.2.2.2 Desvio Padrão

O valor esperado do desvio padrão amostral, $E(s)$, é igual ao desvio padrão da população (6.16).

$$E(s) = \sigma = \sigma_{Z_t|Z_{t-}} \tag{6.16}$$

A variância do desvio padrão amostral, $VAR(s)$, é definida em (6.17a). Substituindo o momento de quarta ordem em (6.17a) obtêm-se (6.17b). A definição do momento central de quarta ordem de população correspondente à distribuição Lognormal pode ser encontrada em [20].

$$VAR(s) = \frac{\mu_4 - \mu_2^2}{4\mu_2 N} \tag{6.17a}$$

$$VAR(s) = \frac{\mu_4 - \mu_2^2}{4\mu_2 N} = \frac{\mu_2(\eta^8 + 6\eta^6 + 15\eta^4 + 16\eta^2 + 2)}{4N} \tag{6.17b}$$

onde:

- μ_2 e μ_4 são os momentos centrados de segunda e quarta ordem da população
- $\eta = \sqrt{e^{\sigma_n^2} - 1}$ é o coeficiente de variação da população e σ_n^2 é o parâmetro de forma da Distribuição Lognormal 3 parâmetros ajustada para a geração dos cenários de aflúências.

Para estabelecer um intervalo de confiança de 95%, para a estimativa do desvio padrão, foi calculada uma tabela com valores-padrão, VP. A Tabela 6.1 foi construída através da simulação de 100.000 amostras com 1000 cenários, geradas por distribuições Lognormal com σ_n variável e μ_n igual a 1.

$$IC_{\sigma} = \sigma_{Z_t|Z_{t-}} + VP \frac{\sigma_{Z_t|Z_{t-}} \sqrt{\eta^8 + 6\eta^6 + 15\eta^4 + 16\eta^2 + 2}}{\sqrt{4N}} \tag{6.18}$$

Tabela 6.1 – Valores Padrão para IC 95%

| Valores-Padrão (VP) | | |
|---------------------|-----------------|-----------------|
| σ_n | Limite Inferior | Limite Superior |
| 0.05 | -1.9661 | 1.9570 |
| 0.07 | -1.9650 | 1.9648 |
| 0.10 | -1.9439 | 1.9850 |
| 0.30 | -1.9311 | 2.0045 |
| 0.50 | -1.8305 | 2.0673 |
| 0.70 | -1.6777 | 2.1220 |

| Valores-Padrão (VP) | | |
|---------------------|-----------------|-----------------|
| σ_n | Limite Inferior | Limite Superior |
| 0.80 | -1.5493 | 2.0849 |
| 0.90 | -1.3965 | 2.0005 |
| 1.00 | -1.1970 | 1.8652 |
| 1.10 | -0.9889 | 1.6254 |

6.2.3 Testes de Aderência

Testes de aderência, também conhecidos como testes de adequação de ajuste ou testes de adaptabilidade, têm como objetivo determinar se certa distribuição postulada é razoável na presença de dados coletados (amostra), isto é, diz a respeito ao grau de concordância entre a distribuição da amostra e da população da qual esta foi extraída. Estes testes pressupõem que as amostras sejam independentes, o que não ocorre com os cenários gerados e agregados. Estas amostras são dependentes, pois os cenários agregados são obtidos a partir dos cenários gerados. A utilização destes testes, para comparar as distribuições de probabilidades obtidas através dos cenários gerados e agregados, não é feita de maneira formal, porém fornece uma indicação sobre o grau de adequação entre estas duas distribuições.

6.2.3.1 Teste de Kolmogorov-Smirnov

O teste de Kolmogorov-Smirnov determina se os valores de uma amostra podem razoavelmente ser considerados como provenientes de uma população com uma determinada distribuição (distribuição teórica). O teste procura especificar a distribuição de freqüência acumulada que ocorreria sob a distribuição teórica, e a comparação desta com a distribuição de freqüência acumulada observada, então se determina o ponto em que essas duas distribuições acusam maior divergência.

Seja $F_0(X)$ uma distribuição de freqüência acumulada completamente especificada, isto é, a distribuição teórica. Para qualquer valor de X , o valor de $F_0(X)$ é a proporção de casos esperados com escores iguais a, ou menores do que X . E seja $S_N(X)$ a distribuição de freqüência acumulada de uma amostra aleatória de N observações.

Pela hipótese de nulidade, de que a amostra tenha sido extraída da distribuição teórica especificada, espera-se que, para cada valor de X , $S_N(X)$ esteja suficientemente próximo de $F_0(X)$. O teste de Kolmogorov-Smirnov visa a maior das diferenças entre $F_0(X)$ e $S_N(X)$. O maior valor de $F_0(X) - S_N(X)$ é chamado desvio máximo, D :

$$D = \text{máximo}(|F_0(X) - S_N(X)|) \tag{6.19}$$

A distribuição amostral de D , sob H_0 , é conhecida e alguns valores críticos de dessa distribuição são apresentados na Tabela 6.2. Note que a significância de um valor D depende de N .

Tabela 6.2 - Valores Críticos D (Kolmogorov-Smirnov)

| Tamanho amostra (N) | Nível de significância (α) | | |
|---------------------|-------------------------------------|------|------|
| | 0.1 | 0.05 | 0.01 |
| 5 | 0.51 | 0.56 | 0.67 |
| 10 | 0.37 | 0.41 | 0.49 |
| 15 | 0.3 | 0.34 | 0.4 |

| Tamanho amostra (N) | Nível de significância (α) | | |
|---------------------|----------------------------|-------------------------|-------------------------|
| | 0.1 | 0.05 | 0.01 |
| 20 | 0.26 | 0.29 | 0.36 |
| 25 | 0.24 | 0.27 | 0.32 |
| 30 | 0.22 | 0.24 | 0.29 |
| 35 | 0.2 | 0.23 | 0.27 |
| 40 | 0.19 | 0.21 | 0.25 |
| 45 | 0.18 | 0.2 | 0.24 |
| 50 | 0.17 | 0.19 | 0.23 |
| >50 | $\frac{1.22}{\sqrt{N}}$ | $\frac{1.36}{\sqrt{N}}$ | $\frac{1.63}{\sqrt{N}}$ |

6.2.3.2 Teste de Cramér-von Mises

O teste de Cramér-von Mises tem o mesmo objetivo que o teste de Kolmogorov-Smirnov, isto é, verificar se a amostra é proveniente de uma população com determinada distribuição teórica. Este teste pertence a uma classe quadrática de testes de aderência, que contabilizam a diferença quadrática entre $F_0(X)$ e $S_N(X)$. Segundo [21], este teste apresenta uma maior potência com relação ao teste apresentado em 6.2.2.1.

O teste de Cramér-von Mises (W^2) possui a seguinte formulação:

$$W^2 = \sum_{i=1}^N \left[F_0(x_i) - \frac{2i-1}{2N} \right]^2 + \left(\frac{1}{12N} \right) \tag{6.20}$$

Partindo-se de um conjunto com N observações, $\{x_1, x_2, \dots, x_n\}$, e considerando que tais observações devem estar ordenadas em ordem crescente, considera-se que $F_0(x_i)$ é o valor da distribuição acumulada teórica para o valor x_i e que o valor de $S_N(x_i)$ é dado por $\left(\frac{2i-1}{2N} \right)$ para cada valor observado x_i .

O valor final a ser comparado aos valores críticos tabelados na Tabela 6.3 é:

$$T = \left(W^2 - \frac{0.4}{N} + \frac{0.6}{N^2} \right) \left(1 + \frac{1}{N} \right) \tag{6.21}$$

Tabela 6.3 – Valores Críticos T (Cramér-von Mises)

| | | | |
|---|-------------------------------------|------|------|
| | Nível de significância (α) | | |
| | 0.1 | 0.05 | 0.01 |
| T | 0.35 | 0.46 | 0.74 |

Outros tipos de testes de aderência não-paramétricos podem ser encontrados em [22] e em [21].

Com estes testes pretende-se verificar a aderência da distribuição condicionada univariada de vazões agregadas de cada usina hidroelétrica com uma distribuição empírica obtida a partir de uma amostra de N realizações geradas para a estrutura de árvore.

Na Figura 6.3 é apresentado um exemplo de distribuições marginais univariadas de vazões, considerando 1000 cenários equiprováveis e 40 cenários não equiprováveis resultante do processo de agregação do conjunto inicial de cenários.

Distribuição Marginal Univariada - Teste de Aderência KS

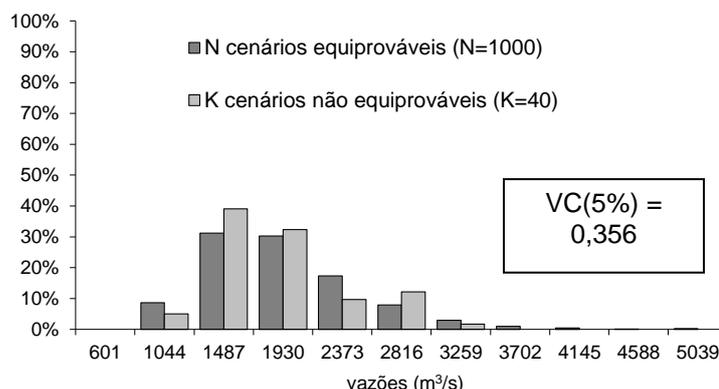


Figura 6.3 – Teste de Aderência KS

Para este exemplo, o teste de Kolmogorov-Smirnov mostra que as duas distribuições são aderentes. O valor crítico do teste, com um nível de significância de 5%, é de 0.356 e a maior diferença encontrada entre as duas distribuições acumuladas é de 0.063.

6.2.4 Comparação das Estatísticas Geradas e Agregadas

O objetivo desta comparação é testar se as estatísticas média, desvio padrão e correlação espacial obtidas para cada usina hidroelétrica a partir dos cenários agregados estão de acordo com as mesmas estatísticas obtidas a partir dos N cenários gerados. Testes de hipóteses formais não podem ser usados neste caso, pois o conjunto dos cenários agregados advém dos cenários gerados, o que faz com que estas amostras sejam dependentes.

Este teste é baseado no ajuste de uma reta aos valores das estatísticas de cada usina hidroelétrica, dos cenários gerados e agregados, através de um modelo de regressão linear de equação única, (6.22). A reta ajustada deve sempre passar pela origem dos eixos.

$$X_{AG} = \beta X_{GER} \tag{6.22}$$

onde $X^t_{AG} = [x^1_{AG} \ x^2_{AG} \ x^3_{AG} \ \Lambda \ x^p_{AG}]$ $X^t_{GER} = [x^1_{GER} \ x^2_{GER} \ x^3_{GER} \ \Lambda \ x^p_{GER}]$ são as estatísticas (média ou desvio padrão ou correlação cruzada) obtidas a partir dos cenários hidrológicos agregados e gerados, respectivamente, para cada uma das P usinas hidroelétricas consideradas na configuração do sistema.

De acordo com estas estatísticas é traçado um diagrama de dispersão e então, a reta de regressão é obtida. Se os cenários agregados estão representando satisfatoriamente os cenários gerados, espera-se que o coeficiente angular β da reta ajustada esteja próximo de 1 e que o coeficiente de determinação R^2 , que mede o grau de ajuste da reta, esteja próximo de 100%.

A Figura 6.4 ilustra uma comparação feita entre a média dos valores de vazões gerados e agregados. Cada ponto representa um par ordenado (média valores gerados, média valores agregados) de uma usina hidroelétrica.

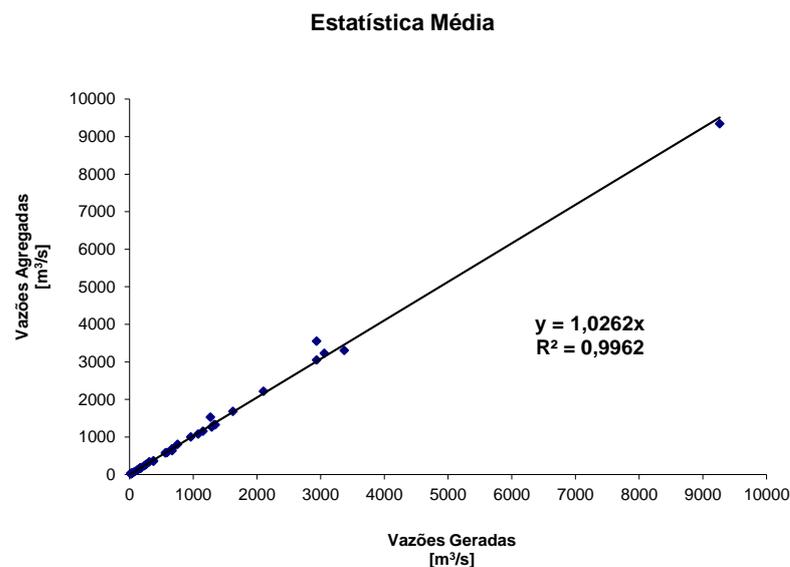


Figura 6.4 - Médias - Valores Gerados X Valores Agregados

Pode-se observar através deste exemplo que a reta de regressão foi bem ajustada ($R^2 = 99.62\%$) e que as médias obtidas a partir dos cenários agregados são ligeiramente maiores do que para os cenários gerados, para a maioria das usinas analisadas ($\beta = 1.0262$).

6.3 Verificação dos Coeficientes Autorregressivos Periódicos

Um coeficiente autorregressivo negativo na equação da afluência de um mês específico pode provocar a presença de um coeficiente positivo associado à afluência prevista para um mês anterior, no corte de Benders que descreve a função de custo futuro deste mês [1][3]. Um corte de Benders deste tipo leva a uma inconsistência ao se atualizar a previsão da afluência para o mês em curso. O custo marginal de operação associado a uma previsão de afluência, por exemplo, superior a média do mês é maior que o custo marginal de operação associado a uma previsão de afluência inferior a média do mês.

Coefficientes positivos nos cortes de Benders têm origem nos coeficientes autorregressivos negativos do modelo estocástico de afluências. Entretanto, a presença de coeficientes autorregressivos negativos não necessariamente implica em coeficientes positivos nos cortes de Benders.

A possibilidade de um coeficiente autorregressivo negativo gerar coeficiente positivo associado à afluência prevista, para o mês em curso no corte de Benders, pode ser avaliada a partir do exame da equação de regressão da afluência de um mês futuro qualquer em função da afluência prevista para o mês em curso e das afluências passadas.

Por exemplo, supondo o seguinte modelo autorregressivo de aflúências para o mês de fevereiro do ano N:

$$A_{FEV} = 0,641 \cdot A_{JAN} - 0,336 \cdot A_{DEZ} + 0,624 \cdot A_{NOV} - 0,421 \cdot A_{OUT} - 1,09 \cdot A_{SET} + 1,83 \cdot A_{AGO} + W_{FEV}$$

onde $A_{FEV} = \left(\frac{Z_{FEV} - \mu_{FEV}}{\sigma_{FEV}} \right)$, $A_{JAN} = \left(\frac{Z_{JAN} - \mu_{JAN}}{\sigma_{JAN}} \right)$, etc.

A contribuição da aflúência do mês de janeiro em fevereiro é dada pelo coeficiente 0,641, positivo.

Supondo que a aflúência de jan/N é dada por $A_{JAN} = 0,934 \cdot A_{DEZ} + W_{JAN}$, substituindo na equação anterior e omitindo-se as parcelas referentes aos resíduos, pode-se escrever:

$$A_{FEV} = 0,641 \cdot 0,934 \cdot A_{DEZ} - 0,336 \cdot A_{DEZ} + 0,624 \cdot A_{NOV} - 0,421 \cdot A_{OUT} - 1,09 \cdot A_{SET} + 1,83 \cdot A_{AGO}$$

Reagrupando, obtém-se a regressão da aflúência de fev/N em função da aflúência prevista em dez/N-1 e aflúências passadas:

$$A_{FEV} = 0,2627 \cdot A_{DEZ} + 0,624 \cdot A_{NOV} - 0,421 \cdot A_{OUT} - 1,09 \cdot A_{SET} + 1,83 \cdot A_{AGO}$$

Logo, a contribuição da aflúência do mês de dezembro em fevereiro é dada pelo coeficiente 0,2627, positivo.

Da mesma forma, a aflúência de dez/N-1 é dada por:

$$A_{DEZ} = 0,971 \cdot A_{NOV} + W_{DEZ}$$

Substituindo na equação anterior e omitindo-se as parcelas referentes aos resíduos, obtém-se:

$$A_{FEV} = 0,2627 \cdot 0,971 \cdot A_{NOV} + 0,624 \cdot A_{NOV} - 0,421 \cdot A_{OUT} - 1,09 \cdot A_{SET} + 1,83 \cdot A_{AGO}$$

Reagrupando, obtém-se a regressão da aflúência de fev/N em função da aflúência prevista em nov/N-1 e aflúências passadas:

$$A_{FEV} = 0,8791 \cdot A_{NOV} - 0,421 \cdot A_{OUT} - 1,09 \cdot A_{SET} + 1,83 \cdot A_{AGO}$$

Novamente, a contribuição da aflúência do mês de novembro em fevereiro é dada pelo coeficiente 0,8791, positivo.

Por outro lado, a aflúência de nov/N-1 é representada por: $A_{NOV} = 0,816 \cdot A_{OUT} + W_{NOV}$. Substituindo-se na equação anterior e omitindo-se as parcelas referentes aos resíduos, escreve-se:

$$A_{FEV} = 0,8791 \cdot 0,816 \cdot A_{OUT} - 0,421 \cdot A_{OUT} - 1,09 \cdot A_{SET} + 1,83 \cdot A_{AGO}$$

Reagrupando, obtém-se a regressão da aflúência de fev/N em função da aflúência prevista em out/N-1 e aflúências passadas:

$$A_{FEV} = 0,2963 \cdot A_{OUT} - 1,09 \cdot A_{SET} + 1,83 \cdot A_{AGO}$$

A contribuição da aflúência do mês de outubro em fevereiro é dada pelo coeficiente 0,2963.

Já a aflúência de out/N-1 é dada por:

$$A_{OUT} = 0,470 \cdot A_{SET} - 0,151 \cdot A_{AGO} + 0,581 \cdot A_{JUL} + W_{OUT}$$

Substituindo na equação anterior e omitindo-se as parcelas referentes aos resíduos, obtém-se:

$$A_{FEV} = 0,2963 \cdot (0,470 \cdot A_{SET} - 0,151 \cdot A_{AGO} + 0,581 \cdot A_{JUL}) - 1,09 \cdot A_{SET} + 1,83 \cdot A_{AGO}$$

Finalmente, reagrupando, obtém-se a regressão da aflúência de fev/N em função da aflúência prevista em set/N-1 e aflúências passadas.

$$A_{FEV} = -0,9507 \cdot A_{SET} + 1,7853 \cdot A_{AGO} + 0,1722 \cdot A_{JUL}$$

O coeficiente negativo na afluência prevista para o mês em curso, set/N, confirma a influência do coeficiente autorregressivo negativo associado à afluência de setembro no modelo estocástico de fev/N. Nota-se que os demais coeficientes autorregressivos negativos do modelo estocástico de fev/N (dez e out/N-1) não resultaram em coeficientes negativos na regressão da afluência de fev/N correspondente. Em decorrência, as funções de custo futuro desses meses não apresentarão coeficientes positivos associados às afluências previstas para estes meses.

Em geral, modelos autorregressivos com ordem elevada apresentam maiores chances de conter coeficientes autorregressivos negativos, que eventualmente, poderão produzir coeficientes positivos indesejáveis nos cortes de Benders. Neste caso, poder-se-ia procurar reduzir a ordem dos modelos autorregressivos. Por outro lado, a incorporação de um número de afluências passadas, variável e maior que um, no modelo de geração de séries sintéticas teve o objetivo bem sucedido de melhorar a representação de secas longas, tais como a verificada no período 1952 a 1956 na região Sudeste. Isto não era bem representado com o modelo autorregressivo fixo de ordem 1 que vinha sendo utilizado anteriormente nos modelos do setor. O modelo PAR(p) foi capaz de gerar séries sintéticas que reproduziram com sucesso os valores médios e dispersões mensais e anuais, assim como as estatísticas dos períodos secos (soma, comprimento e intensidade de sequência negativa, etc.). Com esse mesmo objetivo, o modelo PAR(p) foi empregado com sucesso em outros países, tais como, Estados Unidos e Canadá.

No sentido de conciliar os objetivos originais do modelo PAR(p) com a sua incorporação nos modelos de otimização de forma explícita, deve-se buscar uma ordem suficientemente grande, mas que não produza coeficientes positivos nos cortes de Benders associados à afluência prevista para o mês em curso.

Uma solução para o problema, quando ele ocorre, é analisar novamente a função de autocorrelação parcial do mês em questão (no exemplo anterior corresponde ao mês de Fevereiro) de $k = p_{original}$ até $k = 1$, procurando o próximo coeficiente ϕ_{kk} significativo. A ordem deste novo coeficiente será o novo valor de p. A seguir, novos coeficientes são estimados e uma análise dos coeficientes, como a que foi reproduzida acima, deve realizar-se. Este procedimento deve ser repetido até que o conjunto de parâmetros de todos os meses não produza contribuição negativa em qualquer mês futuro.

7 REFERÊNCIAS

- [1] Maceira, M.E.P.; Costa, A .P.C., Marcato, A .L.M. – “Newave – Especificação Funcional”, Relatório Técnico CEPEL nº DPP/PEL – 349/99.
- [2] Maceira, M.E.P.; Xavier, L.N.; Marcato, A .L.M. – “Modelo de Referência do Modelo Suishi-o”, Relatório Técnico CEPEL nº DP/DEA – 53041/06.
- [3] Pimentel, A.L.G; Lima, A.L.D.S.; Serra, F.C; Xavier, L.N – “Modelo DECOMP - Manual de Referência Versão 12.0”, Relatório Técnico CEPEL nº DP/DEA– 46491/05.
- [4] Maceira, M.E.P.; Jardim, D.L.D.D. – “Manual de Referência de Análise de Conglomerados”, Relatório Técnico CEPEL nº DPP/PEN – 109/2001.
- [5] Salas, J.D., Delleur, J.W., Yevjevich, V., Lane, W.L., "Applied Modeling of Hydrologic Time Series", *Water Resources Publications*, 1980.
- [6] Maceira, M.E.P., "Operação Ótima de Reservatórios com Previsão de Afluências", Tese de Mestrado, COPPE/UFRJ, março 1989.
- [7] Box, G.E.P., Jenkins, G.M., "Time Series Analysis - Forecasting and Control", Holden-Day, 1970.
- [8] Kelman, J., Oliveira, G.C., Maceira, M.E.P., "Modelo de Séries Hidrológicas (MSH) - Manual de Metodologia", Relatório Técnico CEPEL nº 242/92.
- [9] Oliveira, G.C., Costa, J.P., Damázio, J.M., Kelman, J., "Multivariate Weekly Streamflow Forecasting Model", *IFAC Conference on Systems Analysis Applied to Water and Related Land Resources*, Lisbon, Portugal, 1985.
- [10] Noakes, D.J., McLeod, A.I., Hipel, K.W., "Forecasting Seasonal Hydrological Time Series", Technical Report, Dept. of Statistical and Actuarial Sciences, University of Waterloo, Waterloo, Ontario, Canada, 1983.
- [11] Noakes, D.J., McLeod, A.I., Hipel, K.W., "Forecasting Monthly Riverflow Time Series", *International Journal of Forecasting*, 1, pp. 179-190, 1985
- [12] Harvey, A.C., "Time Series Models", P. Allan, 1981
- [13] G.E.P.Box, D.R.Cox, "An Analysis of Transformations", *Journal of The Royal Statistical Society*, A127, pp. 211-252, 1964.
- [14] Maceira, M.E.P., Mercio, C.M.V.B., "Stochastic Sreamflow Model for Hydroelectric Systems", *5th International Conference PMAPS - Probabilistic Methods Applied to Power Systems*, Vancouver, Canada, 1997.
- [15] R.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.
- [16] Press, W.H.; Flannery, B.P.; Teukolosjy, S.A.; Vetterling, W.T. (1992) “Numerical Recipes in FORTRAN 77: The Art of Scientific Computing”. Second Edition, Hardcover.

-
- [17] Kelman, J., Pereira, M.V.F., "Critérios de Avaliação para Modelos de Séries Hidrológicas", *IV SNPTEE*, 1977.
- [18] Mood, A.M., Graybill, F.A., Boes, D.C., "Introduction to the Theory of Statistics", McGraw-Hill Book Company, Third Edition, 1974.
- [19] Bradley, J., "Distribution Free Statistical Tests", Prentice Hall Inc., 1968.
- [20] Yevjevich, V., Probability and Statistics in Hydrology, 1972, Fort Collins, Water Resources Publications.
- [21] Stephens, M.A. 1974, "EDF Statistics for Goodness of Fit and Some Comparisons", *Journal of the American Statistical Association*, v.69, n.347 (Sep), pp. 730-737
- [22] Siegel, S., 1975, *Estatística Não Paramétrica*, São Paulo, Mc Graw-Hill.