

4. Representação da Incerteza Hidrológica no Planejamento da Operação

Neste capítulo será apresentada a modelagem da incerteza hidrológica no planejamento da operação energética de médio prazo, com foco na construção da árvore de cenários e no modelo estocástico gerador empregado nas simulações. Ao final, após a contextualização do problema, serão apresentados os pontos passíveis de melhorias metodológicas.

4.1 A Construção da Árvore de Cenários

Conforme descrito, por meio da representação de um problema de programação estocástica multiestágios, resolvido pela técnica de PDDE, o planejamento da operação de médio prazo é realizado pelo modelo NEWAVE. No contexto do planejamento energético, esta metodologia foi proposta no Brasil por (PEREIRA & PINTO, 1991). A mesma técnica pode ser usada na gestão de portfólios de finanças, como em (CONSIGLI & DEMPSTER, 1998).

A representação dos vários estágios e da estocasticidade do problema é realizada por meio da construção de uma árvore de cenários considerando as variáveis de estado do problema. Cada caminho da árvore é chamado de "cenário" e cada nó representa a realização da variável estocástica considerada no problema. A cardinalidade da árvore é função do número de nós e ramos. Do ponto de vista computacional, a resolução de problemas de otimização para cada um dos nós percorrendo todos os caminhos possíveis da árvore é extremamente custoso e o tempo de processamento não é factível. Em (CHIRALAKSANAKUL, 2003), o autor propõe alguns métodos para atacar situações deste tipo, cuja solução computacional é complexa.

Historicamente, o primeiro método de decomposição do problema de programação linear estocástico foi empregado para um modelo de dois estágios e ficou conhecido na literatura por *L-shaped* (VAN SLYKE & WETS, 1969). Em

problemas reais, a cardinalidade da árvore é elevada e a solução encontrada para a representação da mesma foi a utilização de técnicas de amostragem para seleção de sub árvores, com número de caminhos reduzido, mas que representam de forma fiel às características originais do modelo, caso a árvore fosse percorrida em sua totalidade. Em (PENNA, 2009), uma revisão dos métodos de seleção está disponível. Nos trabalhos de (da COSTA, et al., 2006) e (da COSTA, 2007), os autores propõem uma metodologia para seleção de cenários multivariados, gerando uma árvore reduzida, porém representativa, por meio de técnicas de Agregação (nas quais a variável ENA compõem os eixos de agregação) e de Componentes Principais, em que há redução de dimensionalidade diretamente do espaço de estado de vazões.

No contexto do modelo NEWAVE, cada caminho da árvore é chamado de "cenário hidrológico" ou "cenário de afluência", embora a variável estocástica em questão seja a ENA. As realizações desta variável seguem um modelo estocástico autorregressivo ajustado a partir dos dados históricos.

No modelo vigente, para a construção da árvore são realizados sorteios aleatórios de ruídos a partir de uma distribuição Lognormal multivariada, conforme será apresentado na próxima seção deste capítulo, e incorporação a um modelo autorregressivo. De posse dos ruídos selecionados e a partir de sucessivas operações matemáticas, é possível transformar uma árvore de cenários de variáveis aleatórias independentes em uma árvore de cenários de afluências. É este o processo empregado no modelo NEWAVE.

Para iniciar o processo de ilustração da construção da árvore de cenários, a Figura 4.1 apresenta uma árvore de três estágios com um conjunto de três nós sucessores para cada estágio de tempo. O número de possíveis nós sucessores é chamado de "abertura". Portanto, no exemplo a seguir, tem-se uma árvore de três estágios e três aberturas.

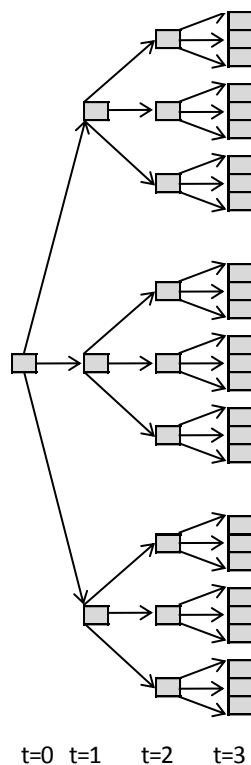


Figura 4.1 - Exemplo de uma Árvore de Cenários Completa

A solução do problema de planejamento associado à árvore de cenários completa, em princípio, exigiria a solução de um problema para cada nó. Supondo a árvore completa da Figura 4.1, a cardinalidade da árvore de cenários completa é igual a 3^3 . No caso do problema real, considera-se 20 aberturas, por exemplo, e 60 períodos de simulações. Pode-se perceber que mesmo uma representação extremamente simples da incerteza hidrológica torna o percurso de toda a árvore de cenários inviável do ponto de vista computacional. Portanto, para solução deste problema deve-se procurar aplicar técnicas que contornem o problema da dimensionalidade da árvore de cenários, evitando-se percorrê-la em sua totalidade. Atualmente, o modelo NEWAVE seleciona a sub árvore de forma aleatória, embora outras formas de seleção tenham sido propostas em (PENNA, 2009).

A partir deste ponto, a construção da árvore será apresentada de forma detalhada. Conforme apresentado no Capítulo 3, a PDDE emprega, para resolução do problema estocástico multiestágios, a chamada "Decomposição de Benders" para a aproximação da FCF, que consiste em diversos estágios, no qual cada um seja dependente do subsequente, não sendo necessário a discretização de todos os

espaços de estados do sistema em questão (como ocorre na PDDD). Com isso, informações obtidas no período atual são repassadas para períodos posteriores permitindo que haja influências e interferências entre estes períodos. No método de Decomposição de Benders são realizados dois processos: *Forward e Backward*.

O passo *Forward* abrange o processo iniciado no estágio 1 e finalizado no último estágio, T . Neste processo ocorre uma recursão direta, não incluindo informações a respeito do custo futuro da operação na primeira iteração, e é resolvido através de Programação Linear. Dessa forma, a tendência do modelo é a utilização, em cada estágio, da capacidade máxima das usinas hidrelétricas, o que conseqüentemente levará à utilização de usinas termelétricas no futuro e a um possível déficit.

Por sua vez, o passo *Backward* segue o caminho contrário do anterior, iniciando no último estágio e finalizando no primeiro. A cada novo estágio cria-se uma restrição com relação ao estágio anterior, que trará informações importantes à formação das FCF e, conseqüentemente, à otimização da operação do sistema hidrotérmico.

Desta forma, as recursões *Forward e Backward* têm por objetivo estimar, de forma iterativa, o valor ótimo (associado a um intervalo de confiança) do valor esperado do custo total de operação do sistema. A *Forward* determina o chamado Z_{SUP} , que é uma estimativa para o limite superior deste valor esperado. Por sua vez, no passo *Backward* é determinado o chamado valor de Z_{INF} , a estimativa do limite inferior do valor esperado do custo total de operação. O algoritmo de solução é dito convergido quando a diferença entre Z_{SUP} e Z_{INF} for menor que uma dada tolerância, considerando o intervalo de confiança de ambos.

O primeiro passo para a construção da árvore de cenários é a definição do número de aberturas que farão parte do passo *Backward*, isto é, o número de sorteios referentes aos nós que poderão ser realizados para cada estágio do período de planejamento. No caso do exemplo da Figura 4.1, verifica-se que para cada estágio são sorteados três novos possíveis valores aleatórios, e estes serão considerados no processo de otimização da simulação recursiva, criação dos cortes, e na formação dos caminhos *Forward* de simulação. No modelo

NEWAVE estes sorteios são realizados uma única vez para todos os meses do período de estudo.

Em seguida, define-se o número de caminhos *Forward* que serão considerados durante o processo de otimização. Esta definição é importante pois determina o número de possíveis cenários a serem visitados na árvore. Desta forma, para cada simulação *Forward* deve-se sortear a transição entre os possíveis nós para todos os estágios de operação.

A Figura 4.2 apresenta quatro possíveis caminhos *Forward* para uma árvore formada a partir de três aberturas e três estágios.

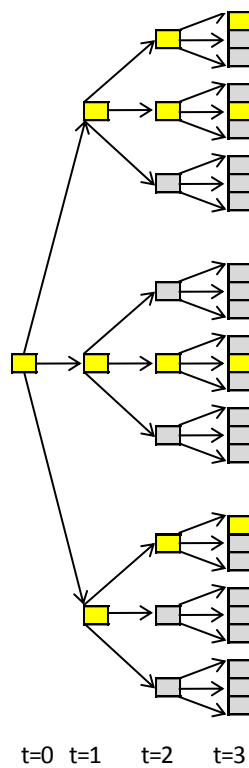


Figura 4.2 - Sub árvore Passo *Forward*

No passo *Backward* todas as ramificações da sub árvore selecionada são consideradas na construção da aproximação da FCF. Portanto, no exemplo da Figura 4.2, a sub árvore do passo *Backward* é mostrada a seguir:

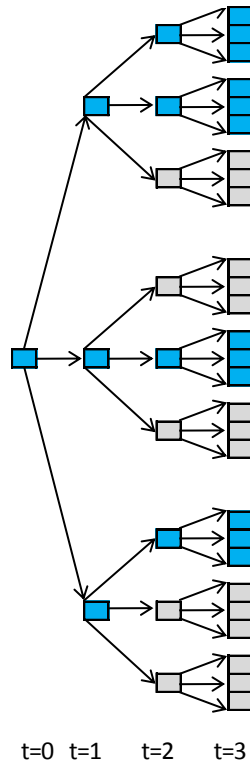


Figura 4.3 - Sub árvore Passo *Backward* (estágio 3)

Para o detalhamento da construção da árvore de cenários, a partir deste ponto, com vistas à simplificação da visualização e facilidade de entendimento, serão exibidas nas próximas figuras somente nove das 27 possibilidades do terceiro estágio. Por conseguinte, a partir de um conjunto de variáveis aleatórias que seguem uma distribuição de probabilidade definida, considere a seguinte árvore:

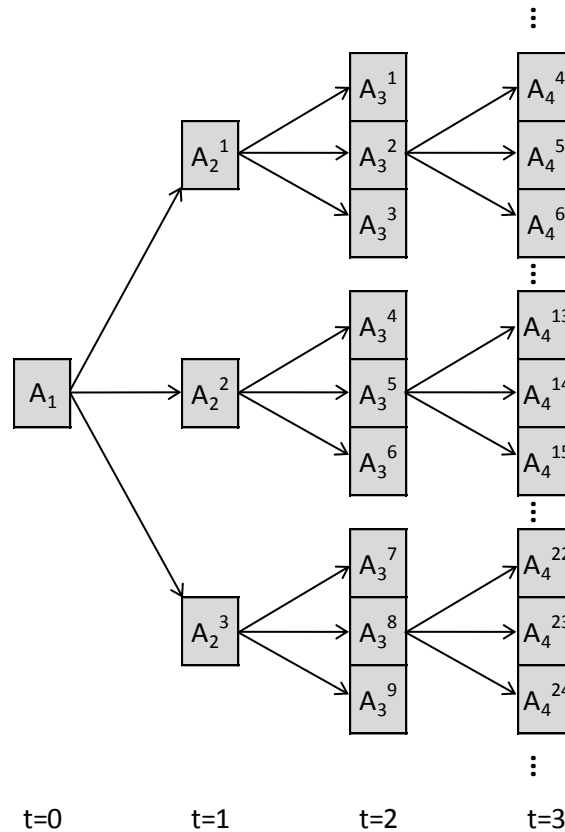


Figura 4.4 - Ruídos para Construção da Árvore de Cenários

Na Figura 4.4, o espaço amostral para um estágio t é definido como Ω_t , e um cenário nesse espaço pode ser definido como ω_t . A realização desse cenário é representada pelos elementos $A_t^{\omega_t}$. Segundo (MATOS, 2008), o primeiro estágio é considerado determinístico, tendo valor previamente conhecido. No caso do modelo NEWAVE, é usada a tendência hidrológica ou a média de longo termo (MLT), (PENNA, 2009). Cada nó do estágio $t \geq 1$, $A_t^{\omega_t}$, possui apenas um nó antecessor, denominado $a(\omega_t)$. Por sua vez, cada nó do estágio $t < T$, em que T é o número total de estágios, tem um conjunto de nós sucessores definido como $\Delta(\omega_t)$. Dessa forma, os nós formam uma árvore de cenários, conforme ilustrado pela Figura 4.4.

Como descrito anteriormente, as realizações que formam os nós sucessores, $\Delta(\omega_t)$, são as mesmas para qualquer cenário do estágio t , isto é, por exemplo,

$A_3^1 = A_3^4 = A_3^7 = A_3^a$, $A_3^2 = A_3^5 = A_3^8 = A_3^b$ e $A_3^3 = A_3^6 = A_3^9 = A_3^c$. Isto posto, a árvore de cenários do exemplo pode ser simplificada.

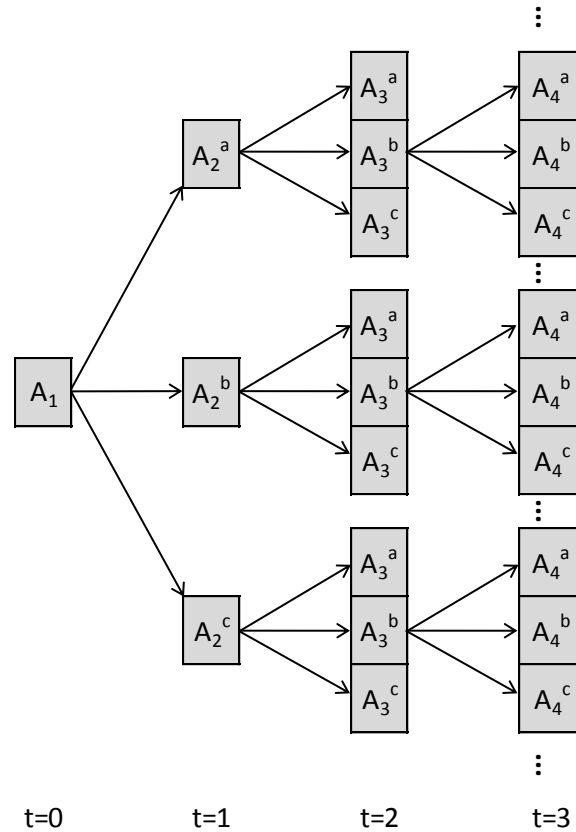


Figura 4.5 - Ruídos Repetidos para Construção da Árvore de Cenários

Assim, o elemento 9 do segundo estágio, ilustrado na Figura 4.5, é constituído pela realização A_3^c quando os nós antecessores forem $[A_1, A_2^c]$. Pode-se, então, definir $\omega_2 = 9 = [A_1, A_2^c, A_3^c]$. O nó antecessor a $\omega_2 = 9$, $a(\omega_2)$, é dado pelo último nó do cenário $\omega_1 = 3$, e o conjunto de nós sucessores, $\Delta(\omega_2)$, formam os elementos $\omega_2 = 25, 26, 27$.

Em geral, se dois cenários têm a mesma sequência de nós entre o primeiro e o k -ésimo estágios, esses têm a mesma parcela de informações durante estes períodos. Consequentemente, decisões associadas a tais cenários devem ser idênticas até o k -ésimo período. Este requerimento é chamado de Condição de Implementabilidade, definida por (OLIVEIRA, 2007).

Portanto, a partir da definição do número de aberturas *Backward*, do número e sequência de caminhos *Forward* e do sorteio de ruídos que podem compor a árvore de cenários, um modelo autorregressivo periódico ajustado ao histórico de ENA é aplicado nestes ruídos e então são gerados valores sintéticos da variável ENA para o horizonte de interesse. Vale lembrar que para o estudo do planejamento da operação energética de médio prazo, o modelo NEWAVE utiliza 20 aberturas *Backward*, 200 caminhos *Forward* e 60 (ou 120) estágios. A amostragem da árvore de cenários é realizada de forma aleatória.

No próximo capítulo será apresentado o modelo estatístico utilizado para gerar a árvore de cenários hidrológicos utilizada na solução do problema de planejamento da operação energética.

4.2 Modelos Estocásticos em Hidrologia

De acordo com (SALAS, et al., 1982), os processos estocásticos naturais são, em geral, estacionários em sentido lato, isto é, os momentos de primeira e segunda ordem da distribuição de probabilidades não são afetados por variações devido à escolha da origem dos tempos, um dos pressupostos para a aplicação da metodologia Box & Jenkins (BOX, et al., 2008).

Os processos periódicos surgiram a partir do trabalho de (GLADYSEV, 1961) e a partir de então inúmeros trabalhos nas áreas de hidrologia e econometria surgiram.

Na classe de modelos periódicos, dois modelos se destacam: PAR (*periodic autoregressive*) e PARMA (*periodic ARMA*). O modelo PAR(p) ajusta para cada período da série um modelo autorregressivo. Em hidrologia, a modelagem PAR(p) foi originalmente proposta por (THOMAS & FIERING, 1962) e analisados em (PAGANO, 1978), apresentando as propriedades assintóticas destes modelos assumindo normalidade do processo estocástico gerador. No trabalho de (TROUTMAN, 1979), o autor apresenta as condições necessárias e suficientes para estacionariedade periódica. Em (REIS, 2013), o autor propõe uma abordagem Bayesiana para a previsão e simulação de vazões no contexto do SEB.

De maneira similar, um PARMA(p,q) consiste num modelo ARMA(p,q) para cada período em estudo. Esses modelos foram desenvolvidos em

(VECCHIA, 1985a) e estendidos em (ULA, 1993), (BOSHNAKOV, 1996) e (LUND, 2000). As propriedades assintóticas dos modelos PARMA são apresentadas em (ANDERSON & VECCHIA, 1993) e (BASAWA & LUND, 2001), assumindo normalidade do processo. Entretanto, algumas dificuldades de uso desses modelos são destacadas em (MCLEOD, 1994) e em (RASMUSSEN, et al., 1996). Os autores argumentam que a extrapolação dos modelos PAR para os modelos PARMA não é uma tarefa trivial em função das dificuldades no processo de estimação paramétrica.

Uma série de outros trabalhos utilizando modelos periódicos para dados hidrológicos está disponível na literatura. Algumas referências, em ordem cronológica, são os trabalhos de (PARZEN & PAGANO, 1979), (OBEYSEKERA & SALAS, 1982), (VECCHIA, 1985b), (BARTOLINI, et al., 1988), (SHAHJAJAN MONDAL, 2006), (LUND, et al., 2006), (OLIVEIRA & SOUZA, 2011) e (SOUZA, et al., 2012).

O foco deste trabalho é o modelo PAR, empregado na geração de séries artificiais de ENA no modelo de planejamento da operação energética de médio prazo, o NEWAVE. Na próxima seção o modelo PAR será detalhado em todas as suas fases: identificação, estimação e simulação.

4.3

O Modelo Autorregressivo Periódico

Considere a partir de agora uma série temporal periódica Z_t , $t = 1, \dots, N$, sendo N o número de observações da série com período S . Para simplificar, assumamos que $N/S = n$ é um número inteiro. Em outras palavras, n é o número de anos disponível na base de dados. O índice t pode ser escrito como $t = t(r, m) = (r-1)S + m$, em que $r = 1, \dots, n$ e $m = 1, \dots, S$. No caso de séries mensais, $S = 12$ e r e m denotam o ano e mês, respectivamente. Nos casos em que:

$$\mu_m = E[Z_{t(r,m)}] \quad (4.1)$$

$$\gamma_{k,m} = \text{Cov}(Z_{t(r,m)}, Z_{t(r,m)-k}) \quad (4.2)$$

existem e dependem somente de k e m , Z_t é dita periodicamente correlacionada ou estacionária periódica (GLADYSEV, 1961).

Deste modo, o modelo PAR de ordem (p_1, p_2, \dots, p_s) é dado, na sua forma padronizada, por:

$$\left(\frac{Z_{t(r,m)} - \mu_m}{\sigma_m} \right) = \sum_{i=1}^{p_m} \varphi_{i,m} \left(\frac{Z_{t(r,m)-i} - \mu_{m-i}}{\sigma_{m-i}} \right) + a_{t(r,m)} \quad (4.3)$$

com $p_m < S$ e $a_{t(r,m)}$ é um processo $NID(0, \sigma_{m(a)}^2)$ (McLEOD, 1994). Haja vista a padronização do modelo, a expressão teórica da variância dos resíduos σ_m^2 é mostrada a seguir.

$$\sigma_m^2 = 1 - \sum_{k=1}^{p_m} \varphi_{k,m} \rho_{k,m} \quad (4.4)$$

em que $\rho_{k,m}$ é a correlação entre $Z_{t(r,m)}$ e $Z_{t(r,m)-k}$.

A introdução dos modelos PAR na literatura foi feita por (THOMAS & FIERING, 1962) para modelagem e simulação de vazões de rios. No caso do modelos NEWAVE e MDDH, as séries modeladas (de cada subsistema que compõem o SIN) são de Energia Natural Afluyente, conforme mencionado anteriormente.

O modelo PAR, conforme mostrado em (TROUTMAN, 1979), pode ser escrito na forma de um modelo MA (médias móveis) da seguinte maneira:

$$\left(\frac{Z_{t(r,m)} - \mu_m}{\sigma_m} \right) = \sum_{i=1}^{\infty} \psi_{i,m} a_{t(r,m)-i} \quad (4.5)$$

em que os $\psi_{i,m}$ são calculados recursivamente usando a expressão a seguir.

$$\psi_{i,m} = \sum_{j=1}^{p_m} \phi_{i,m} \psi_{i-j,m-j}, \quad i \geq 1 \quad (4.6)$$

$\psi_{0,m} = 1$ e $\psi_{i,m} = 0$ se $i < 0$. A condição necessária e suficiente para estacionariedade periódica é, segundo (TROUTMAN, 1979):

$$\sum_{i=0}^{\infty} \psi_{i,m}^2 < \infty, \quad m = 1, \dots, S \quad (4.7)$$

O modelo PARMA pode ser encarado como um extensão do modelo PAR, incluindo a componente de médias móveis na equação do modelo. Para detalhes, ver (VECCHIA, 1985a) e (VECCHIA, 1985b).

A seguir serão detalhadas as fases da modelagem PAR.

4.3.1 Identificação Estrutural

A primeira fase da estruturação de modelos da família Box & Jenkins é a identificação estrutural. Esta fase é comumente realizada por meio da análise das Funções de Autocorrelação (FAC) e Autocorrelação Parcial (FACP), nos casos dos polinômios autorregressivos e médias móveis “puros”. No caso de modelos ARMA, é comum a utilização de funções penalizadoras, já que os comportamentos das FAC e FACP não são bem definidos. É possível ainda o emprego de outras formas alternativas de identificação, como as Funções de Autocorrelação Inversa e Estendida (SOUZA, et al., 2012). Os Critérios de Informação, como o AIC (*Akaike Information Criterion*), BIC (*Bayesian Information Criterion*), HQC (*Hannan-Quinn Information Criterion*) são funções penalizadoras para seleção de modelos e medidas relativas da qualidade do ajuste, com base no número de parâmetros e na função de verossimilhança (SCHWARZ, 1978).

Sob a hipótese de normalidade, um estimador assintótico e eficiente de $\rho_{k,m}$, para a construção da Função de Autocorrelação Periódica - FACPPe - é dado por (TROUTMAN, 1979):

$$c_{k,m} = \frac{1}{n} \sum_{r=1}^n \left(\frac{Z_{t(r,m)} - \hat{\mu}_m}{\hat{\sigma}_m} \right) \left(\frac{Z_{t(r,m)-k} - \hat{\mu}_{m-k}}{\hat{\sigma}_{m-k}} \right) \quad (4.8)$$

em que $\hat{\mu}_m = \left(\sum_{r=1}^n Z_{t(r,m)} / n \right)$ e $\hat{\sigma}_m = \left(\sqrt{(1/n-1) \sum_{r=1}^n (Z_{t(r,m)} - \hat{\mu}_m)^2} \right)$ são os estimadores de média e desvio padrão para o período m .

O cálculo dos coeficientes da Função de Autocorrelação Parcial Periódica, FACPPe é determinado pela resolução do sistemas de equações de Yule-Walker, dados, para cada período m , da seguinte forma:

$$\begin{bmatrix} 1 & \rho_{1,m-1} & \rho_{2,m-1} & \cdots & \rho_{k-1,m-1} \\ \rho_{1,m-1} & 1 & \rho_{1,m-2} & \cdots & \rho_{k-2,m-2} \\ \rho_{2,m-1} & \rho_{1,m-2} & 1 & \cdots & \rho_{k-3,m-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1,m-1} & \rho_{k-2,m-2} & \rho_{k-3,m-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k1,m} \\ \varphi_{k2,m} \\ \varphi_{k3,m} \\ \vdots \\ \varphi_{kk,m} \end{bmatrix} = \begin{bmatrix} \rho_{1,m} \\ \rho_{2,m} \\ \rho_{3,m} \\ \vdots \\ \rho_{k,m} \end{bmatrix} \quad (4.9)$$

Ao conjunto de valores $\varphi_{kk,m}$, chamamos de FACPPe, do período m . Este conjunto é outra forma de representar a estrutura de dependência do processo estocástico ao longo do tempo. Em um processo autorregressivo de ordem p_m , a FACPPe será diferente de zero para k menor ou igual a p_m e zero para k maior que p_m .

No contexto do NEWAVE, a identificação da estrutura dos modelos PAR é realizada com base na análise da significância dos coeficientes da FACPPe para cada um dos m períodos do histórico. Uma descrição mais detalhada sobre a identificação de modelos PAR, e um a proposta alternativa, está disponível em (OLIVEIRA & SOUZA, 2011).

Sob a hipótese de que $\varphi_{kk,m} = 0$, faz-se a identificação da estrutura autorregressiva do modelo PAR com base no *lags* k significantes, haja vista que a aproximação assintótica para a variância da FACPPe é dada por (BARTLETT, 1946):

$$\text{Var}(\varphi_{kk,m}) \cong \frac{1}{n} \quad (4.10)$$

De acordo com (CEPEL, 2001), no modelo NEWAVE, a identificação da estrutura dos modelos PAR é dada de acordo com o critério: "para cada período m , procura-se a maior ordem i tal que todas as estimativas φ_{kk}^m para $k > i$ não sejam mais significativas".

Uma extensa discussão acerca da validade do critério de identificação do modelo NEWAVE é feita em (SOUZA, et al., 2012), em que os autores propõem formas alternativas de identificação, obtendo, em geral, modelos mais parcimoniosos.

4.3.2 Estimação Paramétrica

Considere $\beta_m = [\varphi_{1,m}, \dots, \varphi_{p_m,m}]$ o vetor de parâmetros autorregressivos do período m . De acordo com (PAGANO, 1978), pode-se obter, por meio do sistema de equações de Yule-Walker, o estimador assintoticamente eficiente $\hat{\beta}_m$ da seguinte forma:

$$\sum_{i=1}^{p_m} \hat{\varphi}_{i,m} c_{k-i,m-i} = c_{k,m} \quad , k = 1, \dots, p_m \quad (4.11)$$

Conforme a equação (4.4), a variância estimada dos resíduos do modelo pode ser obtida por:

$$\hat{\sigma}_m^2 = 1 - \sum_{k=1}^{p_m} \hat{\varphi}_{k,m} c_{k,m} \quad (4.12)$$

Em (PAGANO, 1978), é mostrado que $\sqrt{n}(\hat{\beta}_m - \beta_m)$ é assintoticamente normal com média zero e matriz de variâncias e covariâncias dada por $\frac{1}{n} I_m^{-1}$:

$$I_m = \frac{1}{\sigma_m^2} (\gamma_{i-j,m-j}) \quad (4.13)$$

Na prática, o estimador \hat{I}_m é obtido substituindo os valores de γ por c .

No mesmo trabalho, (PAGANO, 1978) mostrou que as estimativas dos parâmetros para diferentes períodos m são assintoticamente descorrelatadas, o que significa que a matriz de informação conjunta de $\beta_1 \dots \beta_s$ é bloco diagonal. Destarte, o modelo autorregressivo de cada período pode ser estimado de forma independente dos demais.

No caso de modelos PARMA, a estimação paramétrica de todos os períodos deve ser feita de forma simultânea, incluindo a variância dos resíduos e através de técnicas de otimização não linear para avaliação da convergência da função de máxima verossimilhança (VECCHIA, 1985a), (LI & HUI, 1988).

4.3.3 Simulação de Cenários

O histórico de ENAs disponível para o ajuste dos modelos PAR é uma série temporal que, na verdade, consiste em apenas uma das possíveis realizações do processo estocástico gerador. O objetivo é, portanto, dado um modelo autorregressivo ajustado à série original, aproximar o comportamento estocástico e, sinteticamente, gerar tantas novas séries temporais quanto se queira, diferentes do histórico original, mas igualmente possíveis do ponto de vista estatístico (OLIVEIRA, 2010).

Nesta sessão são apresentados os principais aspectos da modelagem PAR para a simulação de cenários. Maiores detalhes podem ser conferidos em (MACEIRA, 1989), (PENNA, 2009) e (OLIVEIRA, 2010).

De acordo com (CEPEL, 2001), uma vez que as séries sintéticas que serão geradas pelo modelo PAR serão utilizadas em modelos que calculam as estratégias ótimas de operação de sistemas multi reservatório, baseados na PDDE, o modelo de geração de cenários deve ser aplicado diretamente aos dados de ENA originais, isto é, sem quaisquer transformações estatísticas, e deve ser capaz de lidar com resíduos que apresentam forte coeficiente de assimetria.

Por conseguinte, a solução adotada pelo CEPEL para lidar com o problema citado anteriormente foi ajustar uma distribuição Lognormal com três parâmetros aos resíduos mensais. Essa distribuição, aplicada a dados hidrológicos, foi originalmente proposta por (CHARBENEAU, 1978). De acordo com o autor, a distribuição Lognormal com três parâmetros só é aplicável a distribuições com

coeficiente de assimetria positivo. Assim, a variável θ é a única raiz real (sempre ≥ 1) da equação (4.14), em que γ representa o coeficiente de assimetria dos resíduos gerados pelo modelo.

$$\theta = \left[\left(1 + \frac{\gamma^2}{2} \right) + \left(\gamma^2 + \frac{\gamma^4}{4} \right)^{\frac{1}{2}} \right]^{\frac{1}{3}} + \left[\left(1 + \frac{\gamma^2}{2} \right) - \left(\gamma^2 + \frac{\gamma^4}{4} \right)^{\frac{1}{2}} \right]^{\frac{1}{3}} - 1 \quad (4.14)$$

A estrutura de uma variável aleatória A_t , em um instante t qualquer, que segue uma distribuição Lognormal com três parâmetros é a seguinte:

$$A_t = e^{\xi_t} + \Delta_t \quad (4.15)$$

em que A_t é Lognormal e ξ_t é uma Normal com média μ_{ξ} e desvio padrão σ_{ξ} . A variável Δ é o deslocamento da Lognormal, e equivale ao terceiro parâmetro. Segundo (CHARBENEAU, 1978) os momentos da variável aleatória são:

$$\sigma_{\xi_t} = \sqrt{\ln(\theta)} \quad (4.16)$$

$$\mu_{\xi_t} = \frac{1}{2} * \ln \left(\frac{\sigma_m}{\theta(\theta-1)} \right) \quad (4.17)$$

A variável θ também pode ser obtida por:

$$\theta = 1 + \frac{\sigma_m^2}{\Delta^2} \quad (4.18)$$

É importante ressaltar que além da correlação temporal definida pelo modelo PAR, existe uma correlação espacial entre os diferentes reservatórios equivalentes de energia do sistema. Isto ocorre uma vez que as usinas hidrelétricas que estão geograficamente próximas tendem a ser atingidas por períodos úmidos e secos ao mesmo tempo. Nesse sentido, modifica-se, no problema do planejamento da operação, o resíduo que será utilizado para a geração das séries sintéticas para

considerar a correlação espacial. Embora os resíduos não sejam espacialmente correlacionados, esta modificação é feita com o intuito de preservar a correlação espacial das energias afluentes (CEPEL, 2001). O procedimento de incorporação da correlação entre os subsistemas na geração de séries sintéticas é mostrado a seguir.

Uma vez que os ruídos dos quatro subsistemas devem ser gerados de forma conjunta, define-se o vetor aleatório multivariado $b_t \sim N_4(0, I)$. O vetor b_t tem dimensão (4×1) e é gerado aleatoriamente com base em uma distribuição gaussiana padrão $(N(0,1))$. Para obter resíduos espacialmente correlacionados, entre os subsistemas R e S , $R \neq S$, para um mês m , utiliza-se a chamada "matriz de carga", $D_m^{R \times S}$, de dimensão (4×4) . Os detalhes para o cálculo desta matriz estão disponíveis em (OLIVEIRA, 2010). Portanto, os resíduos espacialmente correlacionados W_t são:

$$W_t = D_m^{R \times S} b_t \quad (4.19)$$

Retomando a equação (4.15), considere, a partir deste momento, a substituição da variável aleatória ξ_t , Normal com média μ_ξ e desvio padrão σ_ξ , pela variável W_t . Finalmente, a estrutura dos resíduos empregados na geração de cada um período per do cenário cen , com $per = 1, \dots, Nper$ e $cen = 1, \dots, Ncen$, é dada por, de acordo com (OLIVEIRA, 2010):

$$A_{t(per, cen)} = e^{W_t(per, cen)\sigma_{\xi_t(per, cen)} + \mu_{\xi_t(per, cen)}} + \Delta_{t(per, cen)} \quad (4.20)$$

No contexto da geração de cenários de ENA, uma vez que, por definição, esta variável não pode assumir valores negativos, o deslocamento da distribuição Lognormal é obtido por:

$$\Delta_{t(per, cen)} = - \left(\frac{\mu_m}{\sigma_m} \right) - \sum_{i=1}^{p_m} \varphi_{i,m} \left(\frac{C_{t(per, cen)-i} - \mu_{m-i}}{\sigma_{m-i}} \right) \quad (4.21)$$

$C_{t(per,cen)}$ representa o valor simulado para o período per do cenário cen . Vale ressaltar que no início da simulação, $C_{t(per,cen)}$ é substituído pela tendência hidrológica, explicitado em detalhes em (PENNA, 2009).

Finalmente, o valor simulado de ENA para um dado período per do cenário cen é obtido a partir de:

$$C_{t(per,cen)} = \mu_m + \sigma_m \left[\sum_{i=1}^{p_m} \varphi_{i,m} \left(\frac{C_{t(per,cen)-i} - \mu_{m-i}}{\sigma_{m-i}} \right) + \left(e^{W_{t(per,cen)} \sigma_{\xi_{t(per,cen)}} + \mu_{\xi_{t(per,cen)}}} + \Delta_{t(per,cen)} \right) \right] \quad (4.22)$$

Em (OLIVEIRA, 2010), mostrou-se que, substituindo a equação (4.21) em (4.22), tem-se a seguinte estrutura final para o modelo gerador de cenários:

$$C_{t(per,cen)} = \sigma_m \left[e^{W_{t(per,cen)} \sigma_{\xi_{t(per,cen)}} + \mu_{\xi_{t(per,cen)}}} \right] \quad (4.23)$$

Conforme descrito no Capítulo 3, a FCF a ser minimizada pelo processo da PDDE é um poliedro convexo que pode ser aproximado por uma função linear por partes, através dos Cortes de Benders. A forma da equação linear por partes, que representa a aproximação do valor esperado da FCF, depende da hipótese assumida sobre a estocasticidade das afluições ao sistema de reservatórios equivalentes. Esta hipótese é de que a variável aleatória segue um modelo baseado em uma estrutura autorregressiva linear.

Conforme apresentado pelas equações (3.18), (3.20), (3.25) e (3.27) e em (3.29), a estrutura do modelo estocástico definido influi diretamente na solução do problema de planejamento, afetando diretamente os resultados e as decisões da política de operação obtida como resposta da PDDE.

No modelo apresentado em (4.23), nota-se estrutura não linear, que pode gerar não convexidades no poliedro que compõe a FCF. Na equação (3.29), o termo $\varphi_{p,i,t}$ representa o resultado do cálculo da derivada $\frac{\partial ENA_{i,t}}{\partial ENA_{i,t-p}}$ quando

aplicada a Regra da Cadeia em (3.15) ou, equivalentemente, em (4.3). Considerando o modelo obtido em (4.23), a expressão da derivada analítica da

variável de estado em relação ao estágio anterior, $t-1$, é apresentada em detalhes em (OLIVEIRA & SOUZA, 2011) e fica:

$$\frac{\partial ENA_{i,t}}{\partial ENA_{i,t-1}} = \sigma_m \left(e^{W_{t(per.cen)} \sigma_{\varepsilon_t(per.cen)} + \mu_{\varepsilon_t(per.cen)}} \right) \left(W_{t(per.cen)} \frac{2\varphi_1 \sigma_m^2}{2\theta \Delta^3 \ln(\theta)} + \frac{2\varphi_1 \sigma_m (2\theta - 1)}{\Delta^3 (\theta^2 - \theta)} \right) \quad (4.24)$$

Desta forma, o impacto da estrutura de simulação que emprega a distribuição Lognormal com três parâmetros na PDDE está comprovada. Isto posto, este trabalho insere-se neste contexto com vistas à proposição de um modelo alternativo para geração das realizações estocásticas da variável de estado que compõem a árvore de cenários utilizada nos modelos de otimização.

A próxima seção apresenta a ideia da proposta geral desta Tese, que será pormenorizada no Capítulo 5.

4.4

Proposta de Trabalho

Em (OLIVEIRA, 2010) foi proposto um método baseado na técnica não paramétrica de computação intensiva *Bootstrap*, (EFRON & TIBSHIRANI, 1993), para geração dos ruídos do modelo gerador de séries sintéticas em substituição ao uso da variável aleatória Lognormal. O método propunha que os próprios resíduos do modelo PAR(p) ajustado ao histórico de ENA fossem sorteados de forma aleatória – com reposição - e adicionados à equação de geração dos cenários em substituição à estrutura não linear da equação (4.20).

Contudo, esta abordagem não garante a geração de valores simulados para a construção da árvore de cenários não negativos e a estratégia utilizada foi o descarte do nó negativo e novos sorteios eram realizados até que somente valores positivos fossem gerados. Nesta abordagem, a amostra de ruídos para cada período m tem dimensão igual ao número de anos do histórico de dados, o que torna relativamente elevada a chance de que realizações negativas da variável de estado sejam geradas. Contudo, mesmo com esta dificuldade, o método obteve cenários que cumpriram de maneira satisfatória os requisitos de "qualidade" das séries, no que tange à manutenção das características do histórico, de acordo com uma série de testes estatísticos. Este modelo foi aplicado por (CASTRO, 2012) na

PDDE e também se apresentou satisfatório quanto aos resultados das decisões da política da operação de planejamento.

Entretanto, no referido trabalho só foram consideradas simulações *Forward*. Posteriormente, quando as simulações do passo *Backward* foram introduzidas, em função das aberturas consideradas e do aumento do número de possibilidades de simulações, haja vista que todas as ramificações das sub árvores selecionadas são visitadas e devem garantir valores de ENA positivos, o modelo proposto por (OLIVEIRA, 2010) não se mostrou eficiente, uma vez que o mesmo não garante a obtenção de valores simulados não negativos em todos os nós da árvore. No trabalho de (CASTRO, 2012), optou-se simplesmente por não considerar, descartando, os valores negativos no processo de otimização, o que será tratado de outra forma nesta Tese.

Assim, o objetivo deste trabalho é a proposição de um novo tratamento para construção da árvore de cenários hidrológicos e de ENAs que atenda de maneira satisfatória os pressupostos necessários para o sucesso das simulações *Forward* e *Backward*, evitando a não convexidade da FCF. Esta proposta mantém o uso do modelo PAR, contudo toda a fase de simulação de cenários é revista, empregando técnicas não paramétricas, ajuste de distribuições de probabilidade e de simulação estocástica.

O modelo proposto, assim como o arcabouço teórico empregado, são apresentados no próximo capítulo. É mister ressaltar que, embora a notação utilizada ao longo deste capítulo esteja "simplificada", o modelo PAR empregado utiliza as configurações do Programa Mensal de Operação (PMO), disponibilizado mensalmente pelos órgãos oficiais do SEB. Uma descrição mais detalhada sobre esta modelagem dinâmica está apresentada no Apêndice A e pode ser encontrada também em (FERREIRA, 2013).