

4 Modelo de Geração de Séries Sintéticas de Energia

4.1. Considerações Iniciais

A adoção de critérios probabilísticos em diversas atividades do planejamento e operação de sistemas hidrotérmicos interligados criou a necessidade da modelagem probabilística de afluições 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 (seqüências) de afluições 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). Na Figura 14 é mostrado que um sistema hidroelétrico pode suprir a demanda e que está associada ao risco p , previamente selecionado.

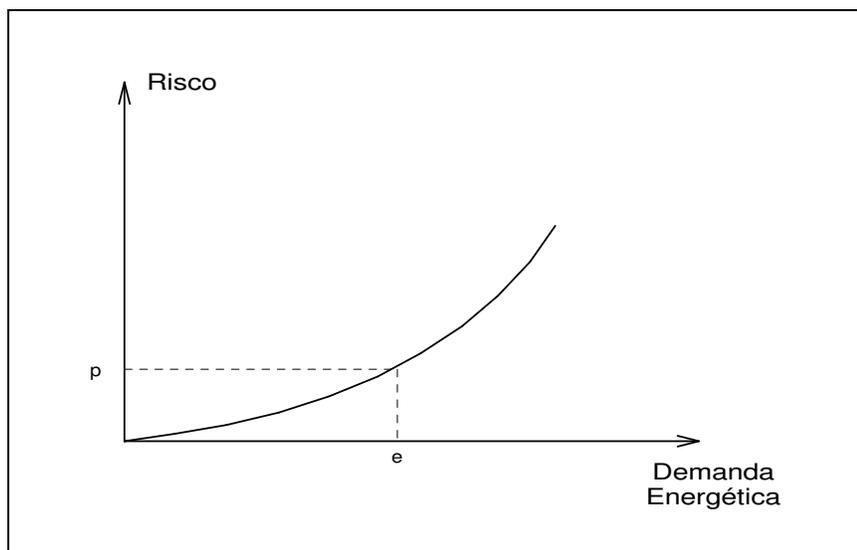


Figura 14: Relação Demanda x Risco em um Ano Qualquer

O único cenário disponível na prática, o registro de afluições 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 aflúncias, diferentes da série histórica mas igualmente possíveis. 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.

Nesta seção será descrito o modelo autorregressivo periódico, $PAR(p)$, proposto para ser utilizado no Modelo Estratégico de Geração Hidrotérmica a Subsistemas Equivalentes Interligados – NEWAVE, no Modelo de Simulação a Usinas Individualizadas para Subsistemas Hidrotérmicos Interligados – SUISHIO e no Modelo de Determinação da Coordenação da Operação a Médio Prazo – DECOMP.

4.2. Processos Estocásticos e Séries Temporais

Medindo-se, por exemplo, a vazão afluente mensal a um dado local por um período de cinco anos, pode-se obter a curva 1 da Figura 15. Realizando-se as mesmas medições para outro segmento de cinco anos, obtém-se outra curva, que é em geral diferente da primeira.

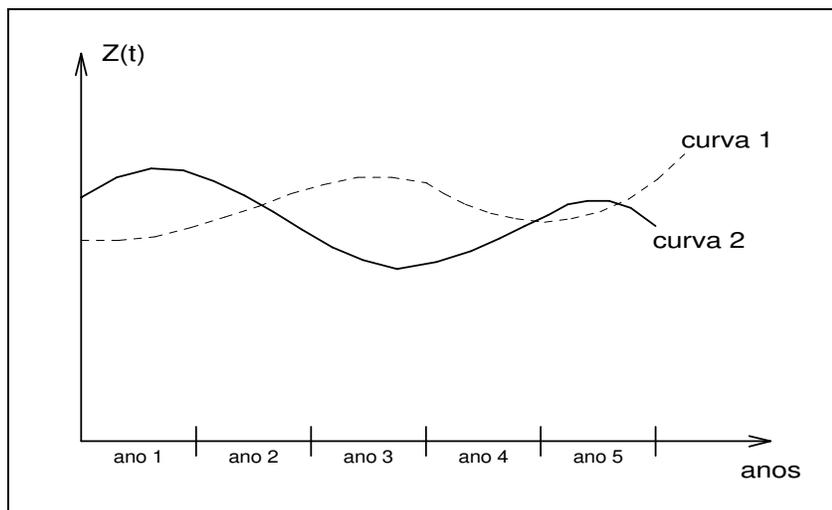


Figura 15: Vazão Afluente Anual x Anos

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 com as respectivas probabilidades. Cada trajetória é também chamada de série temporal.

Na prática, só está disponível uma realização do processo estocástico, a série histórica. Assim, 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 à variável aleatória do processo estocástico no instante t .

Um processo estocástico é totalmente descrito pelo conjunto de todas as séries temporais que o compõem, com suas respectivas probabilidades, ou pela distribuição de probabilidade conjunta de todas as variáveis aleatórias envolvidas. Como na realidade não estão disponíveis o conjunto de todas as séries temporais e a distribuição de probabilidade conjunta de todas as variáveis aleatórias, 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 considerar 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, sazonalidades (características que se repetem dentro de um ano). Quanto ao último fator, o processo de afluências anuais é considerado estacionário e o processo de afluências mensais é considerado não estacionário.

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

- média amostral:

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

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

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

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

$$\hat{\gamma}(1) = \frac{1}{N} \sum_{i=2}^N (z_i - \hat{\mu})(z_{i-1} - \hat{\mu}) \quad (4.3)$$

A auto-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 pelo desvio-padrão das variáveis envolvidas, resultando na correlação de ordem um, também conhecida por dependência serial anual.

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

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

Pode-se medir também a dependência espacial entre as vazões afluentes a duas usinas hidroelétricas, isto é, se a série histórica de uma usina é linearmente dependente da série histórica de outra, por meio 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:

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

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

A média amostral de cada mês é dada por:

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

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

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

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

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

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

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

4.3. O Modelo Autorregressivo Periódico

4.3.1. Descrição do Modelo

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

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

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

ou

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

onde:

Z_t é uma série sazonal de período s

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

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

N é o número de anos

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

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

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

p_m é a ordem do operador autorregressivo do período m

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

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

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

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

tomando o valor esperado obtemos para cada período:

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

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

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

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

Fixando-se m e variando k de 1 a p_m em (4.12) obtém-se, para cada período, um conjunto de equações comumente denominado de equações de Yule-Walker. Para um período m qualquer:

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

Chamando 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:

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

O conjunto de valores ϕ_{kk}^m , $k = 1, 2, \dots$, é chamado de função de autocorrelação parcial do período m . O conjunto de funções ϕ_{kk}^m , $m = 1, 2, \dots, s$, 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}^m 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 (4.12) fica:

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

Multiplicando ambos os lados da equação (4.10a) por a_t e tomando o valor esperado:

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

Substituindo (4.17) em (4.16):

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

Esta expressão é válida para qualquer período m .

4.3.2. Ajuste do Modelo

Box e Jenkins (Box & Jenkins, 1976) 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, com base em estimativas das funções $\rho^m(k)$ e ϕ_{kk}^m obtidas a partir da série amostral. Na modelagem autorregressiva periódica isto 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 consideradas 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.

4.3.2.1. 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 (4.15) as autocorrelações pelos respectivos valores amostrais (4.8) e (4.9). Segundo Maceira (1989), se a ordem do operador autorregressivo de um

período qualquer m é p_m , então $\hat{\phi}_{kk}^m$ para $k > p_m$ tem distribuição aproximadamente Normal com média zero e variância N^{-1} . Para cada período m procura-se a maior ordem i tal que todas as estimativas $\hat{\phi}_{kk}^m$ para $k > i$ não sejam mais significativas

4.3.2.2. 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 (Harvey, 1981).

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

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

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

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 (4.18).

4.3.2.3. Verificação

Na etapa de verificação são utilizados procedimentos estatísticos que permitem verificar se o modelo estimado é adequado e portanto, capaz de gerar as séries sintéticas. A grande maioria dos testes se baseia 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 $k > 0$.

4.3.3. Geração de Séries Sintéticas com o Modelo PAR(p)

Quando se dispõe de uma série temporal, por exemplo, 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 uma outra série diferente da série histórica, mas igualmente possível (Kelman, 1992).

Utiliza-se o modelo PAR 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 (4.10a):

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

De posse da expressão (4.18) pode-se escrever (4.10a) na forma:

$$Z_t = \mu_m + \phi_1^m \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{p_m}^m \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \sigma_m \sigma_a^{(m)} E_t \quad (4.19a)$$

ou

$$Z_t = \mu_m + \phi_1^m \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{p_m}^m \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \quad (4.19b)$$

$$\sigma_m \sqrt{1 - \phi_1^m \rho^m(1) - \phi_2^m \rho^m(2) - \dots - \phi_{p_m}^m \rho^m(p_m)} E_t$$

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

A expressão (4.19b) é particularmente conveniente na geração de séries sintéticas, pois permite um esquema de geração seqüencial 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 necessário 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.

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

A estrutura de geração das séries sintéticas mostrada na Figura 16 é usualmente conhecida como geração em paralelo ou pente. Esta é a estrutura de geração utilizada pelo modelo NEWAVE.

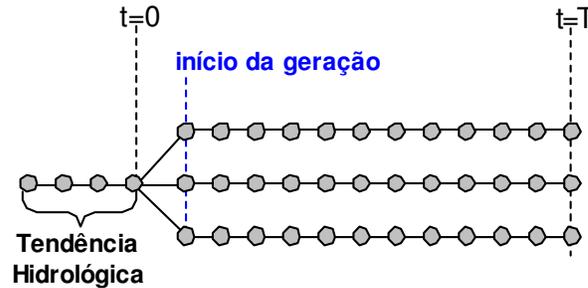


Figura 16: Tendência Hidrológica

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 devem-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 17. 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 18. 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.

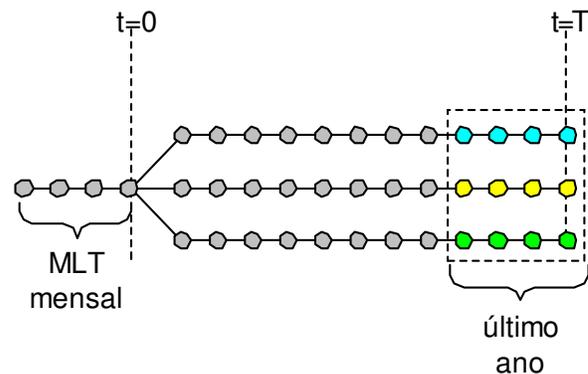


Figura 17: Geração não condicionada – Primeira Etapa

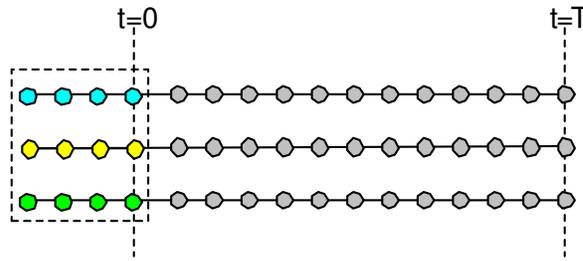


Figura 18: Geração não condicionada – Segunda Etapa

No modelo NEWAVE as séries sintéticas de energias naturais afluentes são necessárias no módulo de cálculo da política de operação hidrotérmica e no módulo de simulação da operação. A política de operação é calculada através de um processo iterativo onde a cada iteração são realizadas uma simulação *backward* e uma simulação *forward*. De maneira simplificada, a simulação da operação pode ser considerada como uma simulação *forward*.

As séries sintéticas utilizadas nas simulações *forward* e *backward* são geradas a partir do mesmo modelo matemático, porém o procedimento para o cálculo destas séries é ligeiramente diferente.

4.3.3.1. Simulação Forward

A geração das séries sintéticas para a simulação *forward* é feita de forma seqüencial, isto é, a cada instante t , Z_t é obtido como uma função dos p_m valores passados e de um sorteio E_t . Assim, para cada instante t é sorteado um conjunto diferente de ruídos. Na Figura 19 é apresentado o esquema de geração para a simulação *forward*, supondo três séries sintéticas.

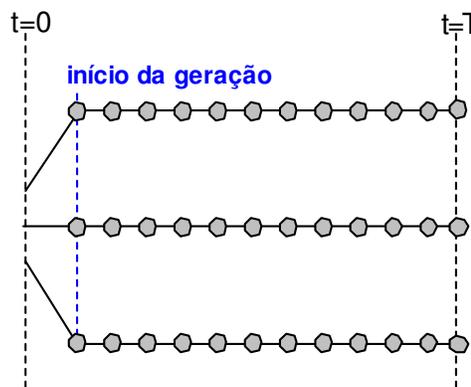


Figura 19: Geração em Paralelo (Pente) para a simulação Forward

Durante a simulação *forward* apenas um valor de afluência é utilizado a cada período e série. Já na simulação *backward*, um número maior de valores de afluência é utilizado a cada período e série. Cada valor de afluência que é utilizado em um determinado período e série é denominado de ‘abertura’.

A geração das afluências para as aberturas é feita de forma seqüencial, isto é, a cada instante t e série i , $Z_{t,i}$ é obtido como uma função dos p_m valores passados da série i e de um sorteio E_t . Os valores passados da série i são aqueles gerados para a simulação *forward*, Figura 19.

A cada instante t é sorteado um conjunto diferente de n_{leq} ruídos E_t , onde n_{leq} é o número de aberturas utilizadas no cálculo da política. O conjunto de ruídos sorteado em um instante t é utilizado para gerar as n_{leq} aberturas de todas as séries sintéticas.

Na Figura 20 é mostrada a seqüência de geração para as aberturas da simulação *backward*, supondo três séries sintéticas com três aberturas.

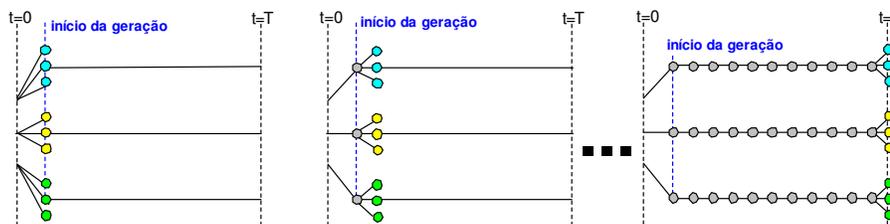


Figura 20: Seqüência de geração para a simulação Backward

Na Figura 21 é apresentado o esquema de geração para a simulação *backward*.

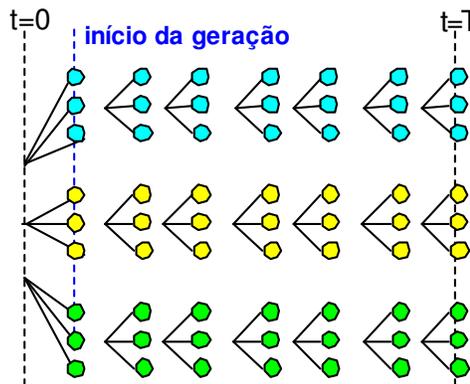


Figura 21: Geração em Paralelo (Pente) para a simulação Backward

4.3.4. Vazões Incrementais Negativas

Um problema freqüente 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:

$$Z_t = \mu_m + \phi_1^m \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{p_m}^m \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \sigma_m a_t > 0 \quad (4.20a)$$

ou

$$a_t > - \left(\frac{\mu_m}{\sigma_m} \right) - \phi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) - \dots - \phi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \quad (4.20b)$$

Muitos pesquisadores consideram que os resíduos a_t apresentam distribuição Normal e uma possível não-normalidade pode ser corrigida pela transformação Box-Cox (Maceira, 1989). Como as séries sintéticas produzidas serão utilizadas em modelos que calculam as estratégias ótimas de operação de um sistema multi-reservatórios, 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 (Maceira & Bezerra, 1997). Conseqüentemente, a variável ξ_t tem distribuição Normal com média zero e desvio-padrão igual a $\sigma_\xi^{2(m)}$:

$$\xi_t = \ln(a_t - \Delta) \quad (4.21)$$

Os parâmetros Δ e $\sigma_\xi^{2(m)}$ são estimados de tal forma a preservar os momentos dos resíduos, através das seguintes relações (Charbeneau, 1978):

$$\sigma_\xi^{2(m)} = \ln(\theta) \quad (4.22)$$

$$\Delta = \frac{-\sigma_a^{(m)}}{\sqrt{(\theta-1)}} \quad (4.23)$$

Fazendo Δ igual à expressão (4.20b) pode-se escrever:

$$\theta = 1 + \frac{\sigma_a^{2(m)}}{(-\Delta)^2} \quad (4.24)$$

$\sigma_\xi^{2(m)}$ pode ser diretamente obtido a partir das expressões (4.22) e (4.24).

4.3.5. Correlação Espacial

A fim de se gerar vazões mensais multivariadas considera-se que os resíduos espacialmente não correlacionados, ξ_t , podem ser transformados em resíduos espacialmente correlacionados, W_t , da seguinte forma:

$$W_t = D\xi_t \quad (4.25)$$

onde D é uma matriz quadrada de dimensão igual ao número de subsistemas equivalentes.

A matriz D pode ser estimada por:

$$DD^T = \hat{U} \quad (4.26)$$

onde \hat{U} pode ser, em princípio, a estimativa da matriz de covariâncias $[E(\hat{\xi}_t, \hat{\xi}_t^T)]$. Por exemplo, o elemento u_{ab} é a covariância espacial lag zero entre os resíduos das usinas a e b . 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 (Maceira & Jardim, 2000). A matriz D pode ser estimada adotando-se D como uma matriz triangular inferior ou por decomposição espectral.

4.4. Avaliação do Desempenho do Modelo

4.4.1. Geração em Paralelo

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 e operação de um sistema de geração de energia elétrica, 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 (Kelman & Pereira, 1977).

Em geral, aceita-se que as variáveis aleatórias selecionadas devem estar relacionadas com o conceito de seqüência: uma seqüência 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.

Na Figura 22 é ilustrado o conceito de seqüência e das variáveis associadas. A linha contínua representa aflúências e a linha pontilhada um limite pré-estabelecido. Os intervalos $(t_1 - t_2)$ e $(t_3 - t_4)$ correspondem às seqüências negativas, isto é, as vazões estão abaixo dos limites.

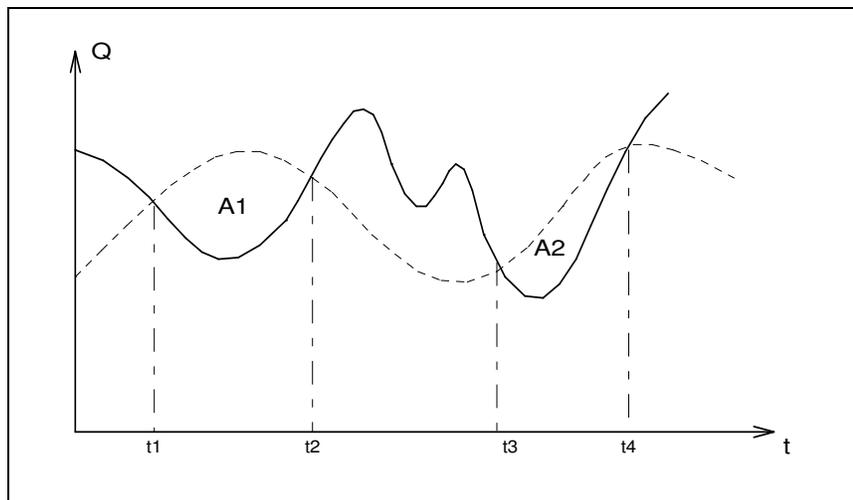


Figura 22: Seqüência Negativa

As variáveis associadas são:

Comprimento de seqüência: corresponde ao comprimento dos intervalos $(t_2 - t_1)$, $(t_4 - t_3)$, na Figura 22 e é calculado com em (4.27).

$$C = t_2 - t_1 \quad (4.27)$$

Soma de seqüência: corresponde à área abaixo do limite durante a seqüência, equação (4.28). Na Figura 22, as somas de seqüências são as áreas A1 e A2

$$S = \sum_{i=t_1}^{t_2} (Z_i - \bar{\mu}_i) \quad (4.28)$$

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

$$I = \frac{S}{C} \quad (4.29)$$

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

- Soma de seqüência negativa
- Comprimento de seqüência negativa
- Intensidade de seqüência 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 seqüência) utilizamos o teste de igualdade de duas distribuições multinomiais. Uma descrição detalhada deste teste é fornecida por Mood et al (1974).

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

$$H_0: p_{1,j} = p_{2,j} \quad j = 1, \dots, k+1 \quad (4.30)$$

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:

$$\chi_k^2 = \sum_{i=1}^2 \sum_{j=1}^{k+1} \frac{\left(N_{ij} - n_i \frac{(N_{1j} + N_{2j})}{(n_1 + n_2)} \right)^2}{n_i \frac{(N_{1j} + N_{2j})}{(n_1 + n_2)}} \quad (4.31)$$

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 e n_i é o número de observações da i -ésima variável aleatória.

Caso as variáveis aleatórias em estudo sejam contínuas, por exemplo, a soma e intensidade de seqüência, o teste de Kolmogorov-Smirnov (Stephens, 1974) 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 freqüências 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:

$$W_{\text{crítico}} = 1,358 \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \quad (4.32)$$

Quando $W > W_{\text{crítico}}$ a hipótese de igualdade entre as distribuições é rejeitada. Descrição detalhada é fornecida por Bradley (1968).

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

$$S_t = \sum_{j=1}^t (Z_j - \beta \bar{\mu}) \quad (4.33)$$

onde β é chamado de nível de regularização.

Neste caso a seqüência de somas parciais é dada por:

$$S_t = \sum_{j=1}^t Z_j - \beta t \bar{\mu} \quad (4.34)$$

e está ilustrada na Figura 23.

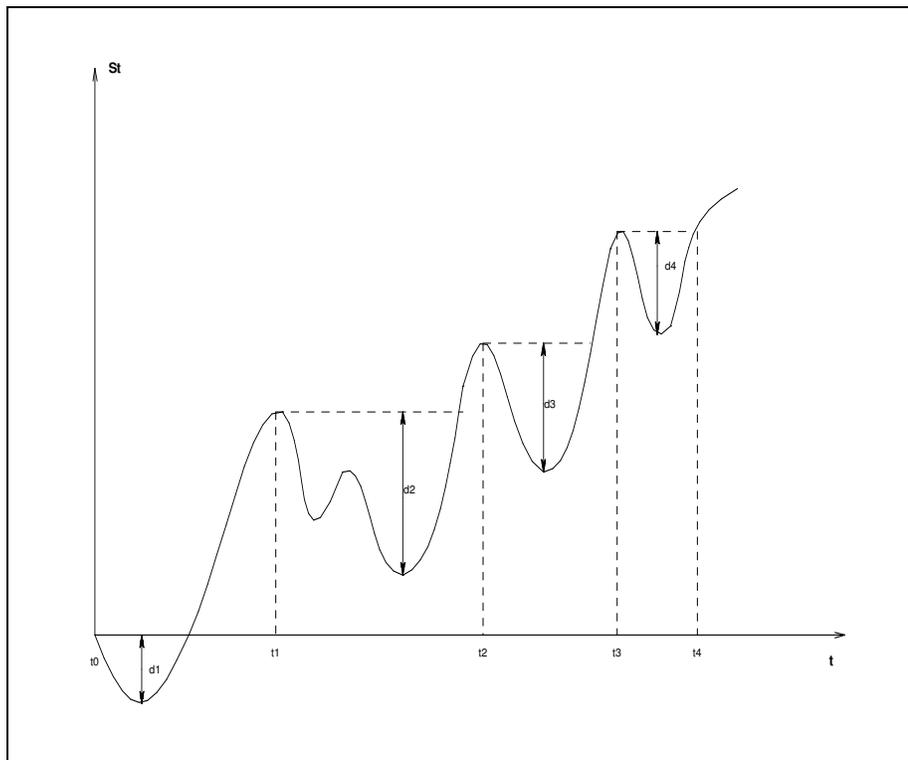


Figura 23: Volume de regularização (Déficit)

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 necessário 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, têm grande interesse nos estudos de regularização das afluições.

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 (1883) 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 freqüências.

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 as seguintes estatísticas:

- Máximo comprimento de seqüência
- Máxima soma de seqüência
- Máxima intensidade de seqüência
- Máximo déficit para um determinado nível de regularização
- Comprimento do período crítico
- Vazão média de período crítico

4.4.2. Geração em Árvore

Alguns testes foram implementados para verificar se os cenários gerados estão coerentes com o modelo ajustado. Para tanto foram avaliados os momentos mensais condicionados de primeira e segunda ordem para os valores gerados.

4.4.2.1. 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 aos subsistemas equivalentes. 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.

Seja o modelo PAR(p) para uma usina hidrelétrica expresso como:

$$\left(\frac{Z_t - \mu_m}{\sigma_m} \right) = \phi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + a_t \quad (4.35)$$

onde a_t é uma série de ruídos independentes com média zero e variância $\sigma_a^{2(m)}$, Z_t é a série de energias naturais afluentes a um subsistema, e os parâmetros μ_m e σ_m são a média e o desvio padrão históricos, respectivamente, do subsistema 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:

$$E(Z_t | Z_{t-}) = \left[\phi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \phi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \right] \sigma_m + \mu_m \quad (4.36)$$

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

$$\text{VAR}(Z_t | Z_{t-}) = \sigma_m^2 \sigma_a^{2(m)} = \sigma_m^2 [1 - \phi_1^m \rho^m(1) - \phi_2^m \rho^m(2) - \dots - \phi_{p_m}^m \rho^m(p_m)] \quad (4.37)$$

E o desvio padrão condicionado é dado por:

$$\sigma_{Z_t|Z_{t-}} = \sigma_m \sigma_a^{(m)} = \sigma_m \sqrt{1 - \phi_1^m \rho^m(1) - \phi_2^m \rho^m(2) - \dots - \phi_{p_m}^m \rho^m(p_m)} \quad (4.38)$$

4.4.2.2.

Parâmetros das distribuições amostrais

Em Yevjevich (1972) 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.

a) Média

O valor esperado da média amostral, $E(\bar{x})$, é igual à média da população (4.39) e a variância da média amostral, $\text{VAR}(\bar{x})$, é igual à variância da população dividida pelo número de elementos da amostra (4.40).

$$E(\bar{x}) = \mu = E(Z_t | Z_{t-}) \quad (4.39)$$

$$\text{VAR}(\bar{x}) = \frac{\mu_2}{N} = \frac{\text{VAR}(Z_t | Z_{t-})}{N} \quad (4.40)$$

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 assintoticamente Normal. Então o intervalo de confiança de 95% calculado para a média amostral estimada a partir dos N cenários hidrológicos gerados, para um dado subsistema, é dado por:

$$IC_\mu = E(Z_t | Z_{t-}) \pm 1,96 \frac{\sigma_{Z_t|Z_{t-}}}{\sqrt{N}} \quad (4.41)$$

b) Desvio Padrão

O valor esperado do desvio padrão amostral é igual ao desvio padrão da população (4.42).

$$E(s) = \sigma = \sigma_{z_t|z_{t-}} \quad (4.42)$$

A variância do desvio padrão amostral, $VAR(s)$, é definida em (4.43a). Substituindo o momento de quarta ordem em (4.43a) obtém-se (4.43b). A definição do momento central de quarta ordem de população correspondente à distribuição Lognormal pode ser encontrada em Yevjevich (1972).

$$VAR(s) = \frac{\mu_4 - \mu_2^2}{4\mu_2 N} \quad (4.43a)$$

$$VAR(s) = \frac{\mu_4 - \mu_2^2}{4\mu_2 N} = \frac{\mu_2(\eta^8 + 6\eta^6 + 15\eta^4 + 16\eta^2 + 2)}{4N} \quad (4.43b)$$

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 afluências.

Com o intuito de estabelecer um intervalo de confiança de 95% (4.44) para a estimativa do desvio padrão, foi calculada uma tabela com valores-padrão, VP (Jardim, 2002). A Tabela 1 foi construída através da simulação de 100.000 amostras com 1000 cenários, geradas por distribuições Lognormal com σ_n^2 variável e μ_n igual a 1.

$$IC_\sigma = \sigma_{z_t|z_{t-}} + VP \frac{\sigma_{z_t|z_{t-}} \sqrt{\eta^8 + 6\eta^6 + 15\eta^4 + 16\eta^2 + 2}}{\sqrt{4N}} \quad (4.44)$$

Tabela 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

Valores-Padrão (VP)		
σ_n	Limite Inferior	Limite Superior
0.30	-1.9311	2.0045
0.50	-1.8305	2.0673
0.70	-1.6777	2.1220
0.80	-1.5493	2.0849
0.90	-1.3965	2.0005
1.00	-1.1970	1.8652
1.10	-0.9889	1.6254

4.5. Resumo

Neste capítulo foi apresentada a formulação do modelo autorregressivo periódico e todas as etapas que envolvem o seu ajuste. Também foi mostrada a estratégia adotada para evitar a geração de cenários de energia negativos e para geração de cenários multivariados. Finalmente foram apresentados os testes utilizados para aferir o modelo ajustado.

No próximo capítulo serão apresentadas técnicas de agregação de cenários e técnicas de amostragem que serão aplicadas ao modelo PAR(p) para definição da árvore de cenários a ser utilizada no problema de planejamento da operação energética.