

2

O modelo autorregressivo periódico – PAR (p)

Neste capítulo é apresentado ao leitor o modelo autorregressivo periódico, PAR (p), foco deste trabalho. No intuito de introduzir e contextualizar matematicamente o problema, são mostrados, brevemente, os principais conceitos e abordagens da metodologia proposta por Box & Jenkins, a formulação básica dos principais modelos e os estágios fundamentais da modelagem. Conceitos matemáticos e estatísticos mais detalhados podem ser encontrados nas referências bibliográficas utilizadas.

2.1

A metodologia Box & Jenkins

Suponha, *a priori*, que exista um sistema que atue como um filtro, estimulado por uma série de ruídos brancos (sequência de variáveis aleatórias independentes e identicamente distribuídas, em geral seguindo uma distribuição gaussiana de probabilidades com média nula e variância constante), resultantes de um processo de geração de números aleatórios, e que com esse estímulo seja gerada pelo sistema uma sequência de valores observados seguindo um padrão, que corresponde à série temporal Z_t .

Evidentemente, em situações reais tem-se o caminho contrário, isto é, conhece-se o conjunto de observações sequenciais Z_t geradas pelo sistema em questão, ao qual se busca associar um modelo que corresponda aos processos internos ao sistema que as gerou. Uma vez que se estabeleça um modelo operacional para essa representação, a série aleatória a_t de valores em torno de zero corresponde à sequência de valores (resíduos) que resulta ao extrair de Z_t os valores obtidos com o modelo ajustado a essa mesma série de valores observados Z_t .

De acordo com HAMILTON (1984), a metodologia Box & Jenkins para a previsão se baseia no ajuste de modelos sequenciais denominados ARIMA (autorregressivos integrados e de médias móveis) a séries temporais de valores

observados de forma que a diferença entre os valores gerados pelos modelos e os valores observados resulte numa sequência de ruídos brancos.

Os modelos ARIMA são matematicamente muito robustos e capazes de descrever os processos de geração de uma variedade de séries temporais para os previsores (que correspondem aos filtros) sem a necessidade, *a priori*, de outras séries explicativas do fenômeno em questão.

Pode-se associar o conceito inicial de um filtro estimulado por uma série aleatória do tipo ruído branco à metodologia de Box & Jenkins conforme Figura 2.1, em que é representado um conjunto de sucessivos filtros aos quais se associam os parâmetros dos modelos ARIMA (p,d,q) que representam os sistemas estimulados pela série a_t que geraram a série temporal Z_t : o filtro de médias móveis (parâmetro q), o filtro autorregressivo estacionário (parâmetro p) e o filtro de integração não-estacionário (parâmetro d).

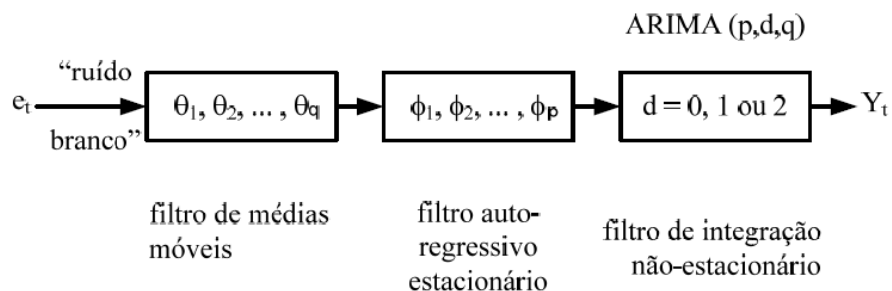


Figura 2.1 – Os filtros do modelo ARIMA (p,d,q)

Fonte: adaptado de SOUZA (2004)

A diferença básica entre a regressão clássica e os modelos de séries temporais é que nos ARMA (p,q) ou ARIMA (p,d,q) não se pode assumir independência entre observações. Ao contrário, os modelos autorregressivos e de médias móveis se propõem a modelar o grau de autocorrelação entre desvios e observações defasadas.

De forma geral, quando se faz referência a modelos ARMA (p,q), esses modelos estão sendo ajustados à série original. Já ao fazer-se referência a modelos ARIMA (p,d,q), considera-se que a série foi diferenciada. Supondo Z_t a série já diferenciada, os modelos ARMA (p,q) são representados em sua forma geral da seguinte maneira:

$$Z_t = \alpha + \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \dots + \varphi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (2.1)$$

O melhor modelo deve ser parcimonioso, utilizando o menor conjunto de parâmetros possível para ajuste à série de dados observados. Os parâmetros p e q representam o número de parâmetros relativos aos comprimentos de defasagem em que se observam valores significativos das autocorrelações e que correspondem a particularidades do sistema de geração das séries que devem ser explicadas pelo modelo (pois correspondem a um padrão de geração). O processo gerador dos dados da série é dito aleatório linear se o modelo ajustado Z_t pode ser descrito como uma combinação linear de valores defasados de Z_t e a_t .

Esses modelos podem, para facilitar a compreensão, ser separados em dois modelos complementares: os modelos de médias móveis e os modelos autorregressivos. Os primeiros, MA (q), correspondem aos processos de médias móveis de ordem q em que cada observação Z_t é gerada por uma média ponderada dos erros aleatórios de q períodos passados. O modelo autorregressivo genérico AR (p) modela uma autorregressão da variável Z_t com ela mesma defasada ($Z_{t-1}, Z_{t-2}, \dots, Z_{t-p}$), para os p períodos de defasagem em que a autocorrelação parcial entre as variáveis é significativa.

Existe um princípio de dualidade entre os modelos do tipo MA e AR de forma que haja a seguinte correspondência entre eles: um MA (q) de ordem finita corresponde a um AR (p) de ordem infinita e um AR (p) de ordem finita a MA (q) de ordem infinita, BOX, JENKINS & REINSEL (1994).

A metodologia de Box & Jenkins segue três estágios principais, conforme mostrado na Figura 2.2:

- Identificação de possíveis modelos e de seus parâmetros (p, d, q);
- Estimação dos parâmetros (p, q) e testes de diagnóstico;
- Previsão, simulação ou controle do sistema de geração dos valores observados Z_t .

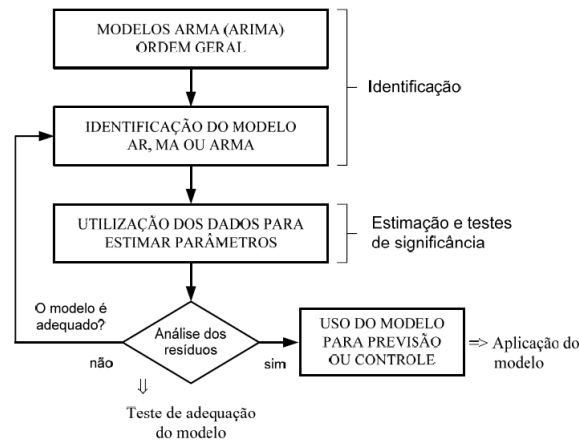


Figura 2.2 – Estágios da metodologia Box & Jenkins

Fonte: adaptado de SOUZA (2004)

Portanto, como tarefa inicial é preciso determinar p e q para a identificação de possíveis modelos. Para isso procede-se ao exame dos coeficientes das funções de autocorrelação e autocorrelação parcial, que permitem medir a força relativa de interação entre as variáveis Z_t defasadas.

A combinação dos termos ponderados por esses dois coeficientes, na ausência de aleatoriedade, poderiam revelar o modelo exato ARMA (p,q). Contudo, a aleatoriedade está presente na amostra do processo, o que leva a desvios dos verdadeiros valores observados. Logo, pode haver enganos na identificação dos coeficientes de autocorrelação com base na análise de dados amostrais. Esses enganos são revelados no teste de adequação do modelo ou análise dos resíduos.

2.2 O modelo PAR (p)

Conforme mostrado em HIPEL & McLEOD (1994), algumas séries históricas, dentre estas as hidrológicas sazonais, exibem uma estrutura de autocorrelação que depende não somente do intervalo de tempo entre as observações, mas também do período observado. Ainda, segundo SALAS, BOES & SMITH (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 HARVEY (1981).

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 AR (p). Em hidrologia, a modelagem PAR (p) surgiu a partir das pesquisas de THOMAS & FIERING (1962), de acordo com HIPEL & McLEOD (1994).

De maneira similar, um PARMA (p,q) consiste num modelo ARMA (p,q) para cada período em estudo. De acordo com RASMUSSEN (1996), a extrapolação dos modelos PAR (p) para os modelos PARMA (p,q) não é uma tarefa trivial e pode não ser justificável dado o bom desempenho dos modelos autorregressivos. Ainda, conforme mostrado em HOSKING (1984), estão descritos na literatura procedimentos para modelagem de séries hidrológicas que apresentam longa dependência e possuem o parâmetro d do modelo ARIMA (grau de diferenciação) assumindo valores fracionários. Estes modelos são conhecidos como ARFIMA, TREVISAN, SOUZA & SOUZA (2000). A estimação de d em geral é baseada na função periodograma e periodograma suavizado. Esses modelos não serão abordados neste trabalho.

De acordo com MACEIRA (1989), séries hidrológicas com periodicidade 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. A esta classe, denomina-se modelos autorregressivos periódicos, 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 (mês, no caso de séries mensais).

O modelo PAR (p) é descrito matematicamente por:

$$\left(\frac{Z_t - \mu_m}{\sigma_m} \right) = \varphi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \varphi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + a_t \quad (2.2)$$

Z_t	Série sazonal de período s .
s	Número de períodos ($s = 12$ para séries mensais).
t	Índice do tempo, $t = 1, 2, \dots, sN$, função do ano T ($T = 1, 2, \dots, N$) e do período m ($m = 1, 2, \dots, s$).
N	Número de anos.
μ_m	Média sazonal do período m .
σ_m	Desvio-padrão sazonal do período m .
φ_m^i	i -ésimo coeficiente autorregressivo do período m .
p_m	Ordem do operador autorregressivo do período m .
a_t	Série de ruídos independentes com média zero e variância $\sigma_a^{2(m)}$. A fim de não sobrecarregar a notação, os ruídos $a_t^{(m)}$ serão tratados apenas como a_t .

A metodologia ajusta, portanto, um modelo autorregressivo de ordem p para cada um dos períodos (meses) das séries hidrológicas históricas. No caso do Newave, estes modelos são ajustados às séries de vazões e/ou ENAs (Energia Natural Afluyente) de cada um dos subsistemas brasileiros (Sudeste/Centro-Oeste, Sul, Nordeste e Norte). Além dos vínculos hidráulicos, os subsistemas são conectados eletricamente por meio de grandes troncos de interligação, constituindo desta forma um sistema interligado.

A Energia Natural Afluyente é calculada a partir das vazões naturais e das produtibilidades equivalentes ao armazenamento de 65% do volume útil dos reservatórios dos aproveitamentos hidroelétricos. A ENA pode ser calculada em base diária, semanal, mensal ou anual e também por bacias e por subsistemas, de acordo com os sistemas de aproveitamentos hidroelétricos existentes nas configurações de bacias hidrográficas e de subsistemas elétricos, MARCATO (2002).

Um dos objetivos deste trabalho é a investigação do método de identificação das ordens de cada um dos modelos ajustados, propondo uma nova abordagem para a mesma tarefa por meio da técnica de computação intensiva *Bootstrap*.

2.2.1 Identificação das ordens “p”

Segundo MACEIRA (1989), a identificação tradicional das ordens p dos modelos PAR (p) é feita por meio da análise das funções de autocorrelação (FAC) e autocorrelação parcial (FACP).

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

$$\rho_k^m = E \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] \quad (2.3)$$

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 dadas por:

$$\begin{aligned} \rho_k^m &= E \left[\left(\frac{Z_t - \mu_m}{\sigma_m} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] = \\ &\varphi_1^m E \left[\left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] + \dots \\ &\dots + \varphi_{p_m}^m E \left[\left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] + E \left[a_t \left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}} \right) \right] \end{aligned} \quad (2.4)$$

Conhecidos os parâmetros de um modelo PAR(p), as funções ρ_k^m são dadas pela solução da equação (2.4) e podem ser expressas por uma combinação de decaimentos exponenciais e/ou ondas senoidais (para detalhes, ver BOX, JENKINS & REINSEL (1994)), 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 2.4, obtemos para cada período um conjunto de equações comumente denominado de equações de Yule-Walker. Para um período m qualquer:

$$\begin{bmatrix} 1 & \rho_1^{m-1} & \rho_2^{m-1} & \dots & \rho_{p_m-1}^{m-1} \\ \rho_1^{m-1} & 1 & \rho_1^{m-2} & \dots & \rho_{p_m-2}^{m-2} \\ \rho_2^{m-1} & \rho_1^{m-2} & 1 & \dots & \rho_{p_m-3}^{m-3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{p_m-1}^{m-1} & \rho_{p_m-2}^{m-2} & \rho_{p_m-3}^{m-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \varphi_1^m \\ \varphi_2^m \\ \varphi_3^m \\ \vdots \\ \varphi_{p_m}^m \end{bmatrix} = \begin{bmatrix} \rho_1^m \\ \rho_2^m \\ \rho_3^m \\ \vdots \\ \rho_{p_m}^m \end{bmatrix} \quad (2.5)$$

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:

$$\begin{bmatrix} 1 & \rho_1^{m-1} & \rho_2^{m-1} & \dots & \rho_{k-1}^{m-1} \\ \rho_1^{m-1} & 1 & \rho_1^{m-2} & \dots & \rho_{k-2}^{m-2} \\ \rho_2^{m-1} & \rho_1^{m-2} & 1 & \dots & \rho_{k-3}^{m-3} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \rho_{k-1}^{m-1} & \rho_{k-2}^{m-2} & \rho_{k-3}^{m-3} & \dots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k1}^m \\ \varphi_{k2}^m \\ \varphi_{k3}^m \\ \vdots \\ \varphi_{kk}^m \end{bmatrix} = \begin{bmatrix} \rho_1^m \\ \rho_{k2}^m \\ \rho_{k3}^m \\ \vdots \\ \rho_{kk}^m \end{bmatrix} \quad (2.6)$$

Omitindo a notação do período m no intuito de facilitar as notações, ao conjunto de valores φ_{kk} , $k = 1, 2, \dots$, chamamos de função autocorrelação parcial 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 função de autocorrelação parcial φ_{kk} será diferente de zero para k menor ou igual a p_m e zero para k maior que p_m .

Portanto, a identificação *clássica* 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 φ_{kk} , $k = 1, \dots, N/4$ e substituindo em (2.6) as autocorrelações pelos respectivos valores amostrais. Se a ordem do operador autorregressivo de um período qualquer m é p_m , então φ_{kk} para $k > p_m$ tem distribuição aproximadamente Gaussiana com média zero e variância $1/N$ (aproximação de Quenouille). Para cada período m procura-se a maior ordem i tal que todas as estimativas φ_{kk} para $k > i$ não sejam mais significativas. Esta forma de identificação é a utilizada no Newave, CEPEL (2000), e a ordem máxima admitida é seis, já que estudos mostram que ordens elevadas apresentam maiores chances de conter coeficientes autorregressivos negativos que, eventualmente, poderão produzir coeficientes positivos indesejáveis nos cortes de *Benders* (fase de otimização), MACEIRA (2004).

Após a etapa de identificação é necessário obter estimativas para os diversos parâmetros do modelo. Segundo HIPEL & McLEOD (1994), para os modelos autorregressivos, os estimadores de momento são em geral tão eficientes quanto os obtidos pelo método da máxima verossimilhança. Os parâmetros φ_i^m , $i = 1, \dots$,

p_m , são estimados substituindo-se na equação (2.5) ρ_k^{m-k} , $j = 0, \dots, (p_m - 1)$, $k = 1, \dots, p_m$, por suas estimativas amostrais, MACEIRA, 1989.

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 a seguir, em que os ρ_i^m são substituídos por seus estimadores, PENNA (2009):

$$\sigma_a^{2(m)} = 1 - \varphi_1^m \rho_1^m - \varphi_2^m \rho_2^m - \dots - \varphi_{p_m}^m \rho_{p_m}^m \quad (2.7)$$

A etapa final, antes da geração dos cenários sintéticos, consiste em testar a adequação dos modelos, verificando se as hipóteses assumidas foram satisfeitas. Uma série de estatísticas e critérios pode ser empregada nessa fase a fim de escolher-se o modelo mais adequado; a grande maioria baseia-se na verificação da hipótese de os resíduos ajustados a_t serem *ruídos brancos*, isto é, sequências de variáveis aleatórias independentes, identicamente distribuídas e, não necessariamente, mas usualmente, com distribuição Gaussiana de média zero, variância constante e que não apresente qualquer estrutura de dependência serial (MORETTIN, 1987).

2.2.2

Resultados assintóticos para a distribuição de probabilidades de autocorrelação e autocorrelação parcial

De acordo com NETO (1991), as distribuições de probabilidade da FAC e FACP são bastante complexas. Em ANDERSON (1942) apud NETO (1991) mostra-se que se o parâmetro estimado, ρ_k , é nulo e o tamanho da série é relativamente grande, então o estimador de ρ_k tem distribuição gaussiana: $\rho_k \sim N[0, V(\rho_k)]$. BARTLETT (1946) apud NETO (1991) propôs uma expressão aproximada para a variância de ρ_k :

$$V(\rho_k) \cong \frac{1}{N} \sum_{i=-\infty}^{\infty} (\rho_i^2 + \rho_{i+k} \rho_{i-k} - 4\rho_k \rho_i \rho_{i-k} + 2\rho_k \rho_i \rho_{i-k} + 2\rho_k^2 \rho_i^2) \quad (2.8)$$

Para processos nos quais $\rho_k = 0$ para $k > q$, pode-se usar a aproximação de Bartlett, dada por:

$$V(\rho_k) \cong \frac{1}{N} \left[1 + 2 \sum_{i=1}^q (\rho_i^2) \right] \quad (2.9)$$

Para as autocorrelações parciais amostrais existe o resultado provado por QUENOUILLE (1958) apud NETO (1991), mostrando que sob a hipótese de um processo autorregressivo de ordem p , a variância aproximada de φ_{kk} é dada por (para N grande, pode-se supor normalidade):

$$V(\varphi_{kk}) \cong \frac{1}{N} \quad (2.10)$$

Logo, os procedimentos para testar as hipóteses:

$$H_{01} : \rho_k = 0$$

$$H_{02} : \varphi_{kk} = 0$$

são bastante grosseiros e, em suma, se $\rho_k = 0$ e $\varphi_{kk} = 0$ e N suficientemente grande, é possível afirmar que $\rho_k \sim N[0, V(\rho_k)]$ e $\varphi_{kk} \sim N[0, 1/n]$.

Assim, tem-se, segundo NETO (1991), os pivôs $(\rho_k - 0)/V(\rho_k)$ e $\sqrt{n}(\varphi_{kk} - 0)$. A partir deste resultado, para verificar se os parâmetros de interesse são estatisticamente nulos, constrói-se os intervalos de confiança ao nível $(1 - \alpha)$:

$$I_{\rho_k} = \left[\pm Z_{(1-\alpha/2)} \sqrt{V(\rho_k)} \right] \quad (2.11)$$

$$I_{\varphi_{kk}} = \left[\pm Z_{(1-\alpha/2)} \sqrt{\frac{1}{N}} \right] \quad (2.12)$$

Portanto, são utilizados a FAC, FACP e os intervalos anteriores como ferramentas para identificação da mais provável estrutura de modelagem da série, comparando com o seu comportamento teórico esperado.

Contudo, existem modelos em todas as estruturas cujos valores da FAC e FACP, embora não nulos, são aceitos como pertencentes, respectivamente, aos intervalos de confiança mostrados.

Assim, a técnica tradicional aponta para a aceitação de processos como sendo ruídos brancos, quando, na verdade, trata-se de modelos com parâmetros

situados em determinadas regiões do espaço paramétrico que possuem baixos valores de ρ_k e φ_{kk} . Desta forma, a definição clara de processos desta natureza (conhecidos como “Quase Ruído Branco”, QRB) é fundamental para um estudo rigoroso em séries temporais. O conceito de QRB foi introduzido na literatura por NETO (1991).

Corroborando a ideia de que a forma clássica de identificação das ordens p pode não ser a mais adequada, como mostram os resultados anteriores, STEDINGER (2001) sugere algumas reflexões acerca desta fase e propõe o seguinte critério para a seleção das defasagens (*lags*) significantes: para cada período m procura-se a maior ordem i tal que todas as estimativas φ_{kk} para $k < i$ sejam significativas. Este critério não admite *lags* intermediários não-significantes, o que ocorre na forma clássica.

STEDINGER (2001) ainda afirma que uma modelagem que considera individualmente cada mês não pode produzir o melhor ou um conjunto adequado de modelos. Sem maiores detalhes matemáticos e dada a natureza dos fenômenos naturais envolvidos, o autor questiona que não faria sentido considerar, por exemplo, que os fluxos de fevereiro dependam daqueles de janeiro, dezembro, novembro, outubro, setembro e agosto ($p = 6$), e para as previsões de janeiro seja necessário somente o histórico de dezembro ($p = 1$). O mesmo raciocínio é mostrado em MATALAS (1967), em que o autor afirma que um sistema de planejamento adequado deve considerar a interação entre várias variáveis do sistema hidrológico.

STEDINGER (2001) finaliza afirmando que a solução para o desafio da modelagem enfrentado seria a adoção de uma visão mais ampla do modelo de decisões, que deve ser guiado pela razoabilidade hidrológica dos modelos selecionados com o reconhecimento da intrínseca ligação entre os modelos escolhidos para diferentes períodos.

Portanto, dadas as observações e análises críticas à maneira clássica de identificação das ordens p , este trabalho propõe o emprego do *Bootstrap* como forma alternativa à estimação da distribuição de probabilidades dos coeficientes da FAC e FACP e posterior análise da significância dos *lags* significativos para identificação das ordens p .

2.2.3

Geração de cenários hidrológicos sintéticos

O histórico de vazões e/ou ENAs disponível para o ajuste dos modelos 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.

Nesta sessão são apresentados todos os conceitos e detalhes matemáticos e estatísticos utilizados para geração de cenários hidrológicos sintéticos via métodos tradicionais, da forma como estão implementados no modelo Newave. Conforme já descrito, os modelos utilizados para modelagem e geração cenários de aflúncias são referenciados como PAR(p), onde p é a ordem do modelo, ou seja, o número de termos autorregressivos. Em geral, p é um vetor, $p = [p_1, p_2, \dots, p_{12}]$, onde cada elemento fornece a ordem do respectivo período. Matematicamente tem-se:

$$\left(\frac{Z_t - \mu_m}{\sigma_m} \right) = \varphi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \varphi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + a_t \quad (2.13)$$

Manipulando a equação do modelo para isolar o valor das vazões e/ou ENAs, têm-se as seguintes expressões:

$$\frac{Z_t}{\sigma_m} = \frac{\mu_m}{\sigma_m} + \varphi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \varphi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + a_t \quad (2.14)$$

$$Z_t = \frac{\mu_m}{\sigma_m} \sigma_m + \varphi_1^m \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \varphi_{p_m}^m \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \sigma_m a_t \quad (2.15)$$

$$Z_t = \mu_m + \varphi_1^m \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \varphi_{p_m}^m \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \sigma_m a_t \quad (2.16)$$

Um dos problemas frequentes em sistemas com reservatórios em cascata, onde as vazões incrementais podem ser muito pequenas, é a geração de vazões mensais negativas. Para obter Z_t positivo é necessário que ele seja maior que zero:

$$Z_t = \mu_m + \varphi_1^m \sigma_m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \varphi_{p_m}^m \sigma_m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + \sigma_m a_t > 0 \quad (2.17)$$

Colocando a inequação 2.17 em função de a_t :

$$a_t > - \left(\frac{\mu_m}{\sigma_m} \right) - \varphi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) - \dots - \varphi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \quad (2.18)$$

Chamando o lado direito da inequação 2.18 de Δ , tem-se:

$$a_t > \Delta \quad (2.19)$$

Portanto, a variável Δ é função apenas dos momentos (média e variância) do período m e dos coeficientes autorregressivos e é dada por:

$$\Delta = - \left(\frac{\mu_m}{\sigma_m} \right) - \varphi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) - \dots - \varphi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) \quad (2.20)$$

Muitos pesquisadores consideram que os resíduos a_t apresentam distribuição Normal e uma possível não-normalidade pode ser corrigida por transformações não-lineares, BOX & COX (1964) apud SHAHJAJAN MONDAL (1996). Entretanto, 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, segundo a literatura, foi ajustar uma distribuição Lognormal com três parâmetros aos resíduos mensais a_t , MACEIRA & MERCIO (1997).

A seguir são mostrados em detalhes os passos para modelagem da estrutura dos ruídos que serão gerados com base em simulações Monte Carlo – geração

pseudoaleatória de observações de alguma distribuição de probabilidades e uso da amostra obtida para aproximar a função de interesse.

Definindo as seguintes variáveis:

$$\xi_t \sim N(\mu_\xi, \sigma_\xi^2) \quad (2.21)$$

$$a_t = e^{\xi_t} + \Delta \quad (2.22)$$

Então, a variável a_t é Lognormal, como segue:

$$a_t \sim LNormal(\mu_\xi, \sigma_\xi^2, \Delta) \quad (2.23)$$

Visto que os ruídos dos quatro subsistemas são gerados de forma conjunta, define-se o vetor aleatório multivariado $b_t \sim N_4(0,1)$. Logo, segue que:

$$\frac{\xi_t - \mu_\xi}{\sigma_\xi} = b_t \quad (2.24)$$

$$\xi_t = b_t \sigma_\xi + \mu_\xi \quad (2.25)$$

$$a_t = e^{\xi_t} + \Delta \quad (2.26)$$

Logo, é possível reescrever da seguinte forma a_t com a seguinte estrutura:

$$a_t = e^{b_t \sigma_{\xi_t} + \mu_{\xi_t}} + \Delta \quad (2.27)$$

O vetor b_t tem dimensão (4×1) e é gerado aleatoriamente com base em uma distribuição gaussiana padrão $(N(0,1))$. Os parâmetros Δ e σ_{ξ_t} são estimados de tal forma a preservar os momentos dos resíduos, conforme CHARBENEAU (1978) e reproduzido em PEREIRA et al. (1984).

$$\sigma_{\xi_t} = \sqrt{\ln(\theta)} \quad (2.28)$$

$$\theta = 1 + \frac{\sigma_{a(m)}^2}{\Delta^2} \quad (2.29)$$

$$\mu_{\xi_t} = 0,5 * \ln\left(\frac{\sigma_{a(m)}}{\theta(\theta-1)}\right) \quad (2.30)$$

As definições anteriores são dadas no intuito de preservar os momentos dos resíduos. De acordo com CHARBENEAU (1978) 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 a seguir, em que γ representa o coeficiente de assimetria dos resíduos gerados pelo modelo.

$$\theta^3 + 3\theta^2 - (4 + \gamma^2) = 0 \quad (2.31)$$

Para se gerar vazões mensais multivariadas, assume-se que os resíduos espacialmente não correlacionados, b_t , podem ser transformados em resíduos espacialmente correlacionados, W_t , da seguinte forma:

$$W_t = Db_t \quad (2.32)$$

D é uma matriz quadrada de dimensão igual ao número de subsistemas (4x4). A matriz D pode ser estimada por:

$$DD^T = U \quad (2.33)$$

Segundo PENNA (2009), no intuito de preservar dependências espaciais entre os subsistemas, faz-se U sendo a estimativa da matriz correlação espacial mensal entre as ENAs dos pares de subsistemas. A matriz D pode ser estimada pelo método da Decomposição Espectral, CASELLA (2001).

De acordo com MENDES (1990), para estimar D , adotam-se os seguintes procedimentos matemáticos:

A equação matricial $DD^T = U$ é resolvida por meio da obtenção dos autovalores e autovetores da matriz U , assim:

$$PL^{1/2}L^{1/2}P^T = U \quad (2.34)$$

$$DD^T = U \quad (2.35)$$

Portanto, $D = PL^{1/2}$. Em que P e L são as matrizes ($k \times k$) de autovetores e autovalores de U , respectivamente. Assim, U é decomposta em P e L , obtendo-se posteriormente a matriz de carga, D . Tem-se, portanto, m matrizes D , $m = 12$ no caso de séries mensais.

A partir dos cenários gerados, o modelo Newave busca calcular estratégias ótimas de operação para diversos cenários hidrológicos. A dimensão do espaço de estados é dada pelo número de reservatórios somado à ordem máxima do modelo PAR(p).

Na fase de otimização, o modelo estocástico “linear” PAR (p) será utilizado na estrutura de derivadas durante a fase de recursões do algoritmo de PDDE, responsável pela construção iterativa da função de custo futuro, MARCATO (2002).

Contudo, dada a necessidade de garantir vazões e/ou ENAs não-negativas e a partir de algumas manipulações matemáticas, é possível observar uma alteração na equação linear autorregressiva do modelo, assumido uma estrutura não-linear, o que se torna um agravante na PDDE. No próximo item as manipulações matemáticas e o impacto causado são detalhados e avaliados.

2.2.4 A estrutura não-linear do PAR (p)

A partir deste ponto são realizadas algumas manipulações matemáticas na estrutura da equação do PAR (p) e é avaliado o impacto da equação final obtida na fase da construção da função de custo futuro.

Assim, manipulando matematicamente a estrutura do modelo PAR (p) e definindo algumas variáveis auxiliares, tem-se:

$$\left(\frac{Z_t - \mu_m}{\sigma_m} \right) = \varphi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}} \right) + \dots + \varphi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}} \right) + a_t \quad (2.36)$$

Seguindo a proposta de FINARDI et al. (2009), definindo as variáveis Φ_t^m e β_t , pode-se reescrever Z_t :

$$\Phi_t^m = \varphi_t^m \frac{\sigma_m}{\sigma_{m-p_m}} \quad (2.37)$$

$$\beta_t = \mu_m - \Phi_1^m \mu_{m-1} - \dots - \Phi_{p_m}^m \mu_{m-p_m} \quad (2.38)$$

$$Z_t = \Phi_1^m Z_{t-1} + \dots + \Phi_{p_m}^m Z_{t-p_m} + \beta_t + a_t \sigma_m \quad (2.39)$$

Assumindo a premissa de que o as vazões e/ou ENAs devem assumir valores não-negativos, tem-se:

$$\Phi_1^m Z_{t-1} + \dots + \Phi_{p_m}^m Z_{t-p_m} + \beta_t + a_t \sigma_m > 0 \quad (2.40)$$

Isolando o resíduo a_t :

$$a_t > \frac{(-\Phi_1^m Z_{t-1} - \dots - \Phi_{p_m}^m Z_{t-p_m} - \beta_t)}{\sigma_m} \quad (2.41)$$

Colocando a_t na estrutura mostrada anteriormente, seguindo uma distribuição Lognormal de três parâmetros e explicitando a variável Δ :

$$a_t = e^{(b_t \sigma_{\xi_t} + \mu_{\xi_t})} + \Delta \quad (2.42)$$

$$\Delta = \frac{(-\Phi_1^m Z_{t-1} - \dots - \Phi_{p_m}^m Z_{t-p_m} - \beta_t)}{\sigma_m} \quad (2.43)$$

Voltando à equação inicial do modelo e fazendo substituições recursivas, o modelo PAR (p), a princípio autorregressivo linear, assume a seguinte estrutura:

$$Z_t = \Phi_1^m Z_{t-1} + \dots + \Phi_{p_m}^m Z_{t-p_m} + \beta_t + a_t \sigma_m \quad (2.44)$$

$$Z_t = \Phi_1^m Z_{t-1} + \dots + \Phi_{p_m}^m Z_{t-p_m} + \beta_t + (e^{(b_t \sigma_{\xi_t} + \mu_{\xi_t})} + \Delta) \sigma_m \quad (2.45)$$

$$Z_t = \Phi_1^m Z_{t-1} + \dots + \Phi_{p_m}^m Z_{t-p_m} + \beta_t + (e^{(b_t \sigma_{\xi_t} + \mu_{\xi_t})}) \sigma_m + \Delta \sigma_m \quad (2.46)$$

$$Z_t = \Phi_1^m Z_{t-1} + \dots + \Phi_{p_m}^m Z_{t-p_m} + \beta_t + (e^{(b_t \sigma_{\xi_t} + \mu_{\xi_t})}) \sigma_m + \frac{(-\Phi_1^m Z_{t-1} - \dots - \Phi_{p_m}^m Z_{t-p_m} - \beta_t)}{\sigma_m} \sigma_m \quad (2.47)$$

$$Z_t = (e^{(b_t \sigma_{\xi_t} + \mu_{\xi_t})}) \sigma_m \quad (2.48)$$

Incorporando a correlação entre os subsistemas, tem-se:

$$Z_t = (e^{(W_t \sigma_{\xi_t} + \mu_{\xi_t})}) \sigma_m \quad (2.49)$$

Desta maneira, em função das manipulações matemáticas para garantir que $Z_t > 0$, é possível notar a alteração na estrutura do modelo PAR (p), tornando-o não-linear. Na fase de otimização, a estratégia de solução é válida apenas em problemas lineares. Portanto, ao se ter um modelo não-linear que é aplicado direto ao problema linear, este último perde uma das principais características, a linearidade.

Sabe-se que durante as recursões do algoritmo de otimização, são calculadas derivadas e estas incluem as do modelo PAR (p). Em uma estrutura autorregressiva, $\partial Z_t / \partial Z_{t-i}$ é, em princípio, apenas o próprio coeficiente estimado pelo modelo. Entretanto, na nova equação do modelo, a partir das manipulações para garantir $Z_t > 0$, perde-se essa referência.

Apenas de forma ilustrativa são mostrados os cálculos a seguir, segundo FINARDI et al. (2009). Para a estrutura não-linear, derivando em relação à primeira defasagem (análogo para as demais até a ordem do período m), tem-se:

$$\frac{\partial Z_t}{\partial Z_{t-1}} = \lambda_{z_{t1}} = \frac{\partial((e^{(W_t \sigma_{\xi_t} + \mu_{\xi_t})}) \sigma_m)}{\partial Z_{t-1}} \quad (2.50)$$

$$\frac{\partial \Delta}{\partial Z_{t-1}} = \Delta' = -\varphi_1 \quad (2.51)$$

$$\frac{\partial \theta}{\partial Z_{t-1}} = \theta' = \sigma_a^2 \left(\frac{-2}{\Delta^3} \right) \Delta' = \frac{2\varphi_1 \sigma_a^2}{\Delta^3} \quad (2.52)$$

$$\frac{\partial \sigma_{\xi_t}}{\partial Z_{t-1}} = \sigma_{\xi_t}' = \frac{1}{2 \ln(\theta)} \frac{1}{\theta} \theta' = \frac{2\varphi_1 \sigma_a^2}{2\theta \Delta^3 \ln(\theta)} \quad (2.53)$$

$$\frac{\partial \mu_{\xi_t}}{\partial Z_{t-1}} = \mu_{\xi_t}' = \frac{(\theta^2 - \theta)^{0.5}}{\sigma_a} \frac{1}{2(\theta^2 - \theta)^{1.5}} (2\theta\theta' - \theta') = \frac{2\varphi_1 \sigma_a (2\theta - 1)}{\Delta^3 (\theta^2 - \theta)} \quad (2.54)$$

$$\lambda_{Z_{t1}} = \sigma_m \frac{\partial (e^{(W_t \sigma_{\xi_t}) + \mu_{\xi_t}})}{\partial Z_{t-1}} = \sigma_m (e^{(W_t \sigma_{\xi_t}) + \mu_{\xi_t}}) (\xi_t \sigma_{\xi_t} + \mu_{\xi_t}) \quad (2.55)$$

$$\lambda_{Z_{t1}} = \sigma_m (e^{(W_t \sigma_{\xi_t}) + \mu_{\xi_t}}) \left(\xi_t \frac{2\varphi_1 \sigma_a^2}{2\theta \Delta^3 \ln(\theta)} + \frac{2\varphi_1 \sigma_a (2\theta - 1)}{\Delta^3 (\theta^2 - \theta)} \right) \quad (2.56)$$

Nota-se, portanto, o impacto da não-linearidade assumida pelo PAR (p) no cálculo das derivadas mostradas anteriormente. Assim, a proposta deste trabalho é sugerir uma nova abordagem para a geração e cenários hidrológicos de forma a garantir a linearidade do modelo autorregressivo a fim de que a PDDE possa ser realizada a partir da abordagem proposta neste trabalho.