DEPARTAMENTO DE ENGENHARIA ELÉTRICA

Internal Research Reports

ISSN

Number 4 | December 2009

Uma nova abordagem para identificação das ordens "*p*" em modelos auto-regressivos periódicos – PAR (p)

Reinaldo Castro Souza

Fernando Luiz Cyrino Oliveira

PUC RIO

Internal Research Reports

Number 4 | December 2009

Uma nova abordagem para identificação das ordens "*p*" em modelos auto-regressivos periódicos – PAR (p)

Reinaldo Castro Souza

Fernando Luiz Cyrino Oliveira

CREDITS

Publisher:
MAXWELL / LAMBDA-DEE
Sistema Maxwell / Laboratório de Automação de Museus, Bibliotecas Digitais e Arquivos
http://www.maxwell.vrac.puc-rio.br/

Organizers:
Alexandre Street de Aguiar
Delberis Araújo Lima

Cover:
Ana Cristina Costa Ribeiro

Uma nova abordagem para identificação das ordens *"p"* em modelos auto-regressivos periódicos – PAR (p)

*Reinaldo Castro Souza – reinaldo@ele.puc-rio.br
Departamento de Engenharia Elétrica
Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)

*Fernando Luiz Cyrino Oliveira – fcyrino@ele.puc-rio.br
Departamento de Engenharia Elétrica
Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio)

Resumo: O modelo auto-regressivo periódico da família Box & Jenkins, PAR (p), é empregado na modelagem das séries de vazões hidrológicas e/ou de energias naturais afluentes utilizadas no planejamento da operação energética no Brasil pelo Sistema Newave. Recentemente alguns aspectos da modelagem começaram a ser questionados e pesquisas diversas vêm sendo realizadas. Este trabalho visa estudar a fase de identificação das ordens *p* dos modelos.

Atualmente a identificação é feita com base na avaliação da significância dos coeficientes da função de autocorrelação parcial (FACP), baseados na aproximação assintótica de Quenouille. A proposta deste estudo é a aplicação da técnica de computação intensiva *Bootstrap* para estimar a significância dos coeficientes da FACP.

Os resultados obtidos mostram que a identificação via *Bootstrap* é sensivelmente mais parcimoniosa e aproxima-se da proposta de STEDINGER (2001), numa análise em que o autor critica a forma atual de identificação implementada no Newave.

Palavras-Chave: PAR (p), Ordem p, Função de Autocorrelação Parcial, Bootstrap.

Abstract: The periodic autoregressive model for the Box & Jenkins family, PAR (p), is employed to model the series of hydrological flows used for planning the energy operation in Brazil by the Newave System. Recently, some aspects of modeling began to be questioned and several researches have been carried out. This paper focus on the identification phase of the orders p models.

Nowadays, the identification is based on evaluating the significance of the coefficients of the partial autocorrelation function (ACF), based on the asymptotics of Quenouille. The purpose of this study is on the application of the computer-intensive Bootstrap technique to estimate the significance of the coefficients of the FACP.

The results inform that the identification by Bootstrap is considerably more parsimonious and is closed to the study proposed by STEDINGER (2001), in an analysis in which the author criticizes the current form of identification implemented in Newave.

Keywords: PAR (p), Order p, Partial Autocorrelation Function, Bootstrap.

1 INTRODUÇÃO

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 (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, HARVEY (1981), um dos pressupostos para a aplicação da metodologia Box & Jenkins).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). De maneira similar, um PARMA (p,q) consiste num modelo ARMA (p,q) para cada período em estudo. 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 auto-regressivos. 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. 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 de intervalo de tempo menor que o ano, tais como séries mensais, têm como característica o comportamento periódico das suas propriedades probabilísticas, como por exemplo, a média, a variância, a assimetria e a estrutura de autocorrelação. A análise deste tipo de séries pode ser feita pelo uso de formulações auto-regressivas cujos parâmetros apresentam um comportamento periódico. A esta classe, denomina-se modelos auto-regressivos periódicos. Os mesmos são referenciados por PAR(p), onde p é a ordem do modelo, ou seja, o número de termos auto-regressivos do modelo. Em geral, p é um vetor, $p = (p_1, p_2, ..., p_{12})$, onde cada elemento fornece a ordem de cada período (mês, no caso de séries mensais).

O modelo PAR (p) é descrito matematicamente por:

$$\left(\frac{Z_t - \mu_m}{\sigma_m}\right) = \phi_1^m \left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right) + \phi_2^m \left(\frac{Z_{t-2} - \mu_{m-2}}{\sigma_{m-2}}\right) + \ldots + \phi_{p_m}^m \left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right) + a_t \qquad (1)$$

Z_t Série sazonal de período s.

s Número de períodos (s = 12 para séries mensais).

t Índice do tempo, t = 1, 2, ..., sN, função do ano T (T = 1, 2, ..., N) e do período m (m = 1,2, ..., 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 auto-regressivo do período m.

p_m Ordem do operador auto-regressivo 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 auto-regressivo de ordem *p* para cada um dos períodos (meses) das séries hidrológicas históricas de vazões e/ou ENAs (Energia Natural Afluente) de cada um dos subsistemas brasileiros (Sudeste/Centro-Oeste, Sul, Nordeste e Norte). Além dos vínculos hidráulicos, os subsistemas são conectados eletricamente através de grandes troncos de interligação, constituindo desta forma um sistema interligado.

Neste contexto, o modelo utilizado no planejamento da operação de médio prazo do sistema elétrico brasileiro é Sistema Newave, em que o problema de planejamento da operação energética de médio prazo é representado por um problema de programação linear multiestágio, cujo objetivo é a minimização do custo total de operação (custo de combustível das unidades térmicas mais a penalidade de atendimento à demanda) ao longo do horizonte de planejamento. O modelo é baseado na técnica de PDDE (Programação Dinâmica Dual Estocástica) e considerando uma representação agregada do parque hidrelétrico, o modelo Newave pode considerar vários subsistemas interligados, permite a representação estática ou dinâmica da configuração do sistema, discretização da carga própria em até três patamares (por exemplo, pesada, média e leve), representação dos cortes no suprimento do mercado de energia elétrica em até quatro patamares de déficit, além da consideração de diversos cenários de energias afluentes, obtidos através de modelos auto-regressivos periódicos, PAR (p).

O objetivo deste trabalho é a investigação do método de identificação das ordens de cada um dos modelos ajustados – fase que tem sido alvo de críticas por alguns autores, dado que ordens mal escolhidas podem gerar resultados não confiáveis e cenários sintéticos inconsistentes – e proposição de uma nova abordagem para a mesma tarefa através da técnica de computação intensiva *Bootstrap*.

2 IDENTIFICAÇÃO DAS ORDENS "*p*"

Segundo MACEIRA (1989), a identificação tradicional das ordens p dos modelos PAR (p) é feita através 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:

$$\rho_k^m = E\left[\left(\frac{Z_t - \mu_m}{\sigma_m}\right)\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] \tag{2}$$

O conjunto de funções de autocorrelação ρ_k^m dos períodos m = 1, ..., s, descrevem a estrutura de dependência temporal da série. Estas funções são dadas por:

$$\begin{aligned}\rho_k^m &= E\left[\left(\frac{Z_t - \mu_m}{\sigma_m}\right)\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] = \varphi_1^m E\left[\left(\frac{Z_{t-1} - \mu_{m-1}}{\sigma_{m-1}}\right)\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] + \ldots \\ &+ \varphi_{p_m}^m E\left[\left(\frac{Z_{t-p_m} - \mu_{m-p_m}}{\sigma_{m-p_m}}\right)\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right] + E\left[a_t\left(\frac{Z_{t-k} - \mu_{m-k}}{\sigma_{m-k}}\right)\right]\end{aligned} \tag{3}$$

Conhecidos os parâmetros de um modelo PAR(p) as funções ρ_k^m são dadas pela solução de (3) 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 (3) obtemos para cada período um conjunto de equações comumente denominado de equações de Yule-Walker. Para um período m qualquer:

$$\begin{bmatrix} 1 & \rho_1^{m-1} & \rho_2^{m-1} & \ldots & \rho_{p_m-1}^{m-1} \\ \rho_1^{m-1} & 1 & \rho_1^{m-2} & \ldots & \rho_{p_m-2}^{m-2} \\ \rho_2^{m-1} & \rho_1^{m-2} & 1 & \ldots & \rho_{p_m-3}^{m-3} \\ \vdots & \vdots & \vdots & \ldots & \vdots \\ \rho_{p_m-1}^{m-1} & \rho_{p_m-2}^{m-2} & \rho_{p_m-3}^{m-3} & \ldots & 1 \end{bmatrix} \begin{bmatrix} \varphi_1^m \\ \varphi_2^m \\ \varphi_3^m \\ \vdots \\ \varphi_{p_m}^m \end{bmatrix} = \begin{bmatrix} \rho_1^m \\ \rho_2^m \\ \rho_3^m \\ \vdots \\ \rho_{p_m}^m \end{bmatrix} \tag{4}$$

Chamando φ_{kj}^m o *j*-ésimo parâmetro auto-regressivo de um processo de ordem k, φ_{kk}^m é o último parâmetro deste processo. As equações de Yule-Walker para cada período m podem ser reescritas da seguinte forma, (SOUZA, 2004):

$$\begin{bmatrix} 1 & \rho_1^{m-1} & \rho_2^{m-1} & \cdots & \rho_{k-1}^{m-1} \\ \rho_1^{m-1} & 1 & \rho_1^{m-2} & \cdots & \rho_{k-2}^{m-2} \\ \rho_2^{m-1} & \rho_1^{m-2} & 1 & \cdots & \rho_{k-3}^{m-3} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \rho_{k-1}^{m-1} & \rho_{k-2}^{m-2} & \rho_{k-3}^{m-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} \varphi_{k1}^m \\ \varphi_2^m \\ \varphi_3^m \\ \vdots \\ \varphi_{kk}^m \end{bmatrix} = \begin{bmatrix} \rho_1^m \\ \rho_{k2}^m \\ \rho_{k3}^m \\ \vdots \\ \rho_{kk}^m \end{bmatrix} \tag{5}$$

Omitindo a notação do período *m* no intuito de facilitar as notações, ao conjunto de valores φ_{kk}, k = 1, 2,..., 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 auto-regressivo 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 auto-regressivos de cada período p_m, m = 1, ..., s. Isto pode ser feito obtendo-se estimativas φ_{kk}, k = 1, ..., N/4 e substituindo em (5) as autocorrelações pelos respectivos valores amostrais. Se a ordem do operador auto-regressivo 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 (1958)). 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 e a ordem máxima admitida é seis, haja vista que estudos mostram que ordens elevadas apresentam maiores chances de conter coeficientes auto-regressivos negativos que, eventualmente, poderão produzir coeficientes positivos indesejáveis nos cortes de *Benders* (fase de otimização), MACEIRA (2004).

3 RESULTADOS ASSINTÓTICOS PARA A DISTRIBUIÇÃO DE PROBABILIDADES DE AUTOCORRELAÇÃO E AUTOCORRELAÇÃO PARCIAL

As distribuições de probabilidade destas variáveis aleatórias são extremamente 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:

$$V(\rho_k) \cong \frac{1}{N} \sum_{i=-\infty}^{\infty} \left(\rho_i^2 + \rho_{i+k}\rho_{i-k} - 4\rho_k \rho_i \rho_{i-k} + 2\rho_k \rho_i \rho_{i-k} + 2\rho_k^2 \rho_i^2 \right) \tag{6}$$

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

$$V(\rho_k) \cong \frac{1}{N}\left[1+\sum_{i=1}^{q}\left(\rho_i^2\right)\right] \tag{7}$$

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

$$V(\varphi_{kk}) \cong \frac{1}{N} \tag{8}$$

Logo, os procedimentos para teste de hipóteses são:

$H_{01} : \rho_k = 0$ $\qquad\qquad\qquad\qquad$ $H_{02} : \varphi_{kk} = 0$

E a partir dos resultados mostrados anteriormente, é possível construir os intervalos de confiança ao nível (1- α) para os parâmetros de interesse:

$$I_{\rho_k} = \left[\rho_k \pm Z_{(1-\alpha/2)}\sqrt{V(\rho_k)}\right] \tag{9}$$

$$I_{\varphi_{kk}} = \left[\varphi_{kk} \pm Z_{(1-\alpha/2)}\sqrt{\frac{1}{N}}\right] \tag{10}$$

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 comportamento teórico esperado das mesmas.

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 idéia de que a forma clássica de identificação das ordens *p* pode não ser a mais adequada, haja vista os resultados mostrados anteriormente, STEDINGER (2001) propõe 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. A ordem máxima considerada é seis.

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*).

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*.

4 *BOOTSTRAP* NA INDENTIFICAÇÃO DAS ORDENS "p"

O termo *Bootstrap* tem origem na expressão de língua inglesa "*lift oneself by pulling his/her bootstrap*", ou seja, alguém levantar-se puxando seu próprio cadarço de bota. O que está por trás desta expressão é o fato de que, através do método, é possível obter-se propriedades de grandes amostras a partir de um número reduzido de observações.

O *Bootstrap*, introduzido por EFRON (1979), é uma técnica estatística não paramétrica computacionalmente intensiva que permite a avaliação da variabilidade de estimadores com base nos dados de uma única amostra existente.

A técnica é indicada para problemas nos quais os procedimentos estatísticos convencionais sejam de difícil aplicação. Em geral apresenta vantagens se usado em situações de amostras pequenas ou grandes, desde que forneça resultados próximos aos obtidos por meios assintóticos usuais em grandes amostras ou superior a amostras reduzidas.

Operacionalmente, a técnica consiste de um sorteio com reposição dos elementos de uma amostra aleatória, gerando uma "amostra *Bootstrap*", de tamanho igual à original. Extrai-se um número suficiente de amostras a fim de se obter a "distribuição *Bootstrap*" de qualquer estatística de interesse do pesquisador. Desta forma, o conjunto de observações *Bootstrap* corresponde a uma estimativa da verdadeira distribuição amostral da estatística em questão. Como mostrado em EFRON (1993), à medida que o tamanho da amostra tende ao infinito, a distribuição *Bootstrap* converge para a distribuição verdadeira da estatística.

Considere $X = (x_1, x_2, ..., x_{n-1}, x_n)$ a amostra original finita de tamanho *n* obtida de um modelo probabilístico desconhecido descrito por sua função de distribuição acumulada F e $\theta = S(X)$ a estatística de interesse. Denota-se X_i^*, i = 1, 2,..., B como sendo a *i*-ésima amostra *Bootstrap* de tamanho *n* obtida da amostra original X através de sorteios com reposição. Para cada amostra tem-se a correspondente estimativa *Bootstrap* da estatística de interesse, isto é, $\theta_i^* = S(X_i^*)$.

Define-se a média, variância e erro padrão do estimador *Bootstrap* de θ por, respectivamente:

$$\theta^* = \frac{\sum_{i=1}^{B} \theta_i^*}{B} \tag{11}$$

$$Var(\theta_i^*) = \frac{\sum_{i=1}^{B} (\theta_i^* - \theta^*)^2}{B-1} \tag{12}$$

$$SE_{boot} = \sqrt{Var(\theta_i^*)} \tag{13}$$

Sem maiores detalhes matemáticos, é possível provar, ver EFRON (1993), que:

$$Var(SE_{boot}) \cong \frac{C_1}{n^2} + \frac{C_2}{nB} \tag{14}$$

C1 e C2 são constantes que dependem da distribuição populacional F, mas não dependem de *n* e B. Portanto, a incerteza associada ao estimador *Bootstrap* dependerá, em última análise, do tamanho da amostra original *n*, isto é, mesmo que se gere uma infinidade de amostras *Bootstrap*, a incerteza do estimador não vai à zero.

4.1 A distribuição Bootstrap de ρ_k e φ_{kk}

Para construção da distribuição *Bootstrap* de ρ_k e φ_{kk} é necessário um algoritmo que preserve a estrutura de autocorrelação da série. De posse da série histórica de dados, obtêm-se ρ_k^{*l}, a estimativa de ρ_k na *l*-ésima replicação *Bootstrap* tomando-se, com reposição, n – k pares da amostra original de pares e construindo-se com eles a amostra *Bootstrap*. A seguir, calcula-se a estimativa pelo método usual. Repetindo-se o processo B vezes, tem-se o estimador *Bootstrap* de ρ_k:

$$\rho_k^* = \sum_{l=1}^{\mathbf{B}} \rho_k^{*l} \Big/ B \tag{15}$$

O conjunto de pares de observações com defasagem k fornece as informações para a estimativa do parâmetro ρ_k. A estimativa desta autocorrelação, obtida deste conjunto, é um elemento da distribuição amostral do estimador clássico. Deste conjunto pode-se obter uma amostra *Bootstrap*, de índice *l*, e a respectiva estimativa do parâmetro ρ_k. Um conjunto muito grande de estimativas ρ_k^{*l} constitui uma aproximação da distribuição amostral de ρ_k. Pela Lei dos Grandes Números, tem-se que:

$$\rho_k^* = \left[\sum_{l=1}^{\mathbf{B}} \rho_k^{*l} \Big/ B \right] \to E[\rho_k] \tag{16}$$

Dispondo-se da distribuição *Bootstrap* do estimador ρ_k é possível construir a distribuição *Bootstrap* de φ_{kk} em função da autocorrelação *Bootstrap* de defasagem k e anteriores, da maneira usual.

Neste momento, após execução do algoritmo, são obtidas as distribuições *Bootstrap* de ρ_k e φ_{kk}. Uma medida de precisão destas estatísticas pode ser calculada com base nestes conjuntos. Os erros padrão *Bootstrap* de ρ_k e φ_{kk} são, respectivamente:

$$s^*(\rho_k) = \sqrt{\frac{\left[\sum_{l=1}^{B}(\rho_k^{*l} - \rho_k^*)^2\right]}{B}} \tag{17}$$

$$s^*(\varphi_{kk}) = \sqrt{\frac{\left[\sum_{l=1}^{B}(\varphi_{kk}^{*l} - \varphi_{kk}^*)^2\right]}{B}} \tag{18}$$

Desta forma, intervalos de confiança *Bootstrap* podem ser construídos sem hipótese de normalidade, tanto para ρ_k quanto para φ_{kk}. Neste caso, tem-se disponível, segundo NETO (1991):

$$I_{\rho_k} = \left[\rho_k \pm t_{gl,(1-\alpha)} s^*(\rho_k)\right] \tag{19}$$

$$I_{\varphi_{kk}} = \left[\varphi_{kk} \pm t_{gl,(1-\alpha)} s^*(\varphi_{kk})\right] \tag{20}$$

Na próxima sessão são mostrados os resultados obtidos através da utilização da técnica no caso específico da modelagem periódica.

5 RESULTADOS OBTIDOS

Por meio da aplicação do *Bootstrap* foram obtidas as estimativas para os coeficientes da FAC e FACP para cada um dos períodos (janeiro a dezembro) e para cada um dos *lags* selecionados de todos os subsistemas, respeitando a restrição de p_m máximo igual a seis. Neste trabalho, as séries utilizadas foram as ENA com histórico de janeiro/1931 a dezembro/2005, disponibilizadas pelo Operador Nacional do Sistema Elétrico, ONS.

Como forma de garantir a convergência (através da Lei dos Grandes Números) dos estimadores *Bootstrap*, foram realizadas 10.000 simulações para cada um dos ρ_k e φ_{kk} de cada um dos períodos. Evidentemente, uma melhor ou pior aproximação depende fundamentalmente da qualidade da amostra, o que, independentemente do emprego do *Bootstrap*, configura-se uma dificuldade em estatística.

A seguir os resultados obtidos para as ordens de cada um dos períodos dos subsistemas brasileiros. Para a forma clássica de identificação, são mostrados os valores de p encontrados segundo os critérios estabelecidos. Para a proposta via *Bootstrap* são considerados apenas as ordens referentes ao Critério 2 de identificação, haja vista que este foi considerado mais razoável do ponto de vista matemático e pela razoabilidade hidrológica, proposta por STEDINGER (2001). O objetivo é, portanto, comparar os resultados obtidos por meio das duas modelagens e verificar a aplicabilidade da técnica proposta. Vale ressaltar que os critérios considerados são:

Critério 1: para cada período m procura-se a maior ordem i tal que todas as estimativas φ_{kk} para k > i não sejam mais significativas, com ordem máxima igual a seis (seguindo a orientação usada pelo Newave).

Critério 2: para cada período m procura-se a maior ordem i tal que todas as estimativas φ_{kk} para k < i sejam significativas. Este é o critério proposto por STEDINGER (2001) e não admite *lags* intermediários não significantes. A ordem máxima considerada é seis.

	SUDESTE/CENTRO-OESTE		
	Clássica		Bootstrap
	Critério 1	Critério 2	Critério 2
Jan	5	1	1
Fev	6	1	1
Mar	1	1	1
Abr	2	2	2
Mai	3	1	1
Jun	1	1	1
Jul	3	3	3
Ago	1	1	1
Set	1	1	1
Out	3	3	2
Nov	1	1	1
Dez	4	1	1

Figura 1: Resultados da estimação das ordens para o subsistema Sudeste/Centro-Oeste

	SUL		
	Clássica		Bootstrap
	Critério 1	Critério 2	Critério 2
Jan	1	1	1
Fev	1	1	1
Mar	4	1	1
Abr	5	1	1
Mai	1	1	1
Jun	5	1	1
Jul	4	2	1
Ago	1	1	1
Set	1	1	1
Out	1	1	1
Nov	6	1	1
Dez	1	1	1

Figura 2: Resultados da estimação das ordens para o subsistema Sul

	NORDESTE		
	Clássica		Bootstrap
	Critério 1	Critério 2	Critério 2
Jan	1	1	1
Fev	4	2	1
Mar	1	1	1
Abr	1	1	1
Mai	1	1	1
Jun	1	1	3
Jul	1	1	1
Ago	1	1	1
Set	1	1	1
Out	3	1	1
Nov	2	2	1
Dez	5	1	1

Figura 3: Resultados da estimação das ordens para o subsistema Nordeste

	NORTE		
	Clássica		Bootstrap
	Critério 1	Critério 2	Critério 2
Jan	1	1	1
Fev	4	2	1
Mar	1	1	1
Abr	1	1	1
Mai	1	1	2
Jun	1	1	1
Jul	3	3	4
Ago	1	1	2
Set	5	1	2
Out	2	2	1
Nov	5	1	1
Dez	1	1	1

Figura 4: Resultados da estimação das ordens para o subsistema Norte

6 CONCLUSÕES E CONSIDERAÇÕES FINAIS

A utilização do PAR (p) para modelar de séries de vazões hidrológicas e/ou de energias naturais afluentes vem sendo aplicada no planejamento da operação energética no Brasil há muitos anos e recentemente alguns aspectos da modelagem começaram a ser questionados e investigações vêm surgindo à luz de projetos de pesquisa realizados em alguns dos centros de referência no país.

O emprego do *Bootstrap* para estimar os coeficientes das funções de autocorrelação e autocorrelação parcial levou à identificação de ordens inferiores na maioria (92% dos casos) dos períodos de todos os subsistemas, tornando os modelos substancialmente mais parcimoniosos, um dos principais fundamentos da metodologia proposta por Box & Jenkins (1976), baseados na Teoria Geral de Sistemas Lineares.

Ainda, a utilização do Critério 2 para seleção das ordens mostrou-se realmente mais adequada, confirmando o julgamento de STEDINGER (2001), numa análise em que o autor critica a forma clássica de identificação e afirma que a solução para o desafio enfrentado na modelagem 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. Em geral, as ordens identificadas pelo Critério 2 da identificação clássica e via *Bootstrap* foram idênticos, salvo as poucas exceções que, no caso extremo, diferiu em duas unidades (mês de junho no subsistema Nordeste).

Isto posto, o *Bootstrap* apresentou-se eficiente na identificação das ordens dos modelos de cada período, aproximando-se do Critério 2 do método clássico, já proposto por STEDINGER (2001) como, provavelmente, a maneira mais adequada. Desta forma, recomenda-se como alternativa mais acertada a utilização da técnica de computação intensiva na fase crítica de identificação das ordens *p* ou, como mudança inicial no Sistema Newave, a utilização do Critério 2 em substituição ao Critério 1, implantado atualmente.

7 REFERENCIAS BIBLIOGRÁFICAS

1 Box, G. E. P. & Jenkins, G. M. & Reinsel, G. C. (1994). Time Series Analysis: Forecasting and Control , Prentice Hall, Englewood Clifls, New Jersey.

2 Efron, B. & Tibshirani, R. J. (1993). An Introduction to the Bootstrap. Chapman & Hall, New York.

3 Harvey, A. C. (1981). Time Series Models. Philip Allan. London.

4 Hipel, K. W. & McLeod, A. I. (1994). Time Series Modelling of Water Resources and Environmental Systems. Elsevier. Amsterdam.

5 Hosking, J. R. M. (1984). Modelling Persistence in Hydrological Time Series Using Fractional Differencing. Water Resource Research, 20, 1898-1908.

6 Maceira, M. E. P. (1989). Operação Ótima de Reservatórios com Previsão de Afluências. Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, Brasil.

7 Maceira, M. E. P. & Damázio, J. M. (2004). The use of PAR (p) model in the stochastic dual dynamic programming optimization scheme used in the operation planning of the Brazilian hydropower system. 8th International Conference on Probabilistic Methods Applied to Power Systems, Iowa State University, Ames, Iowa.

8 Maceira, M. E. P. & Bezerra, C. V. (1997). Stochastic Streamflow Model for Hydroeletric Systems. Proceedings of 5th International Conference on Probabilistic Methods Applied to Power Systems, pp. 305-310, Vancouver, Canadá.

9 Morettin, P. A. & Toloi, C. M. C. (1987). Previsão de Séries Temporais. Atual Editora, São Paulo.

10 Neto, A. C. (1991). Bootstrap em séries temporais. Tese de Doutorado, DEE, PUC-Rio, Brasil.

11 Neto, A. C. & Souza, R. C. (1996). A Bootstrap Simulation Study in ARMA (p,q) Structures. Journal of Forecasting, 15, 343-353.

12 ONS. Operador Nacional do Sistema Elétrico. Disponível em <http://www.ons.com.br/home/>. Acesso em 2009.

13 Penna, D. D. J. (2009). Definição da árvore de cenários de afluências para o planejamento da operação energética de médio prazo. Tese de Doutorado, DEE, PUC-Rio, Brasil, 2009.

14 Rasmussen, R. F. & Salas, J. D. & Fagherazzi, L. & Rassam, J. C. & Bobee, R. (1996). Estimation and validation of contemporaneous PARMA models for streaflow simulation. Water Resource Research, 32, 3151-3160.

15 Salas, J. D & Obeysekera, J. T. B. (1982). ARMA model identification of hydrologic time series. Water Resource Research, 18, 1011-1021.

16 Souza, R. C. & Camargo, M. E. Análise e Previsão de Séries Temporais: os modelos ARIMA. Sedigraf, Ijuí.

17 Stedinger, J. R. (2001) Report on the evaluation of CEPEL's PAR models. Cornell University. School of Civil and Environmental Engineering. Ithaca, New York.

18 Stedinger, J. R. & Taylor, M. R. (1982). Synthetic Streamflow Generation: 1. Model Verification and Validation. Water Resource Research, 18, 909-918.