

UTILIZAÇÃO DE MODELOS PARA SIMULAÇÃO E PREVISÃO DE SÉRIES
HIDROLÓGICAS MENSAIS

por

M.E.P.Maceira^{1,2}, J.Kelman^{2,3} e J.M.Damázio^{2,3}

RESUMO -- Este artigo apresenta um conjunto de metodologias para identificação, estimação e verificação de modelos PAR. A seguir descreve-se uma comparação de desempenho entre o modelo PAR e modelos atualmente em uso no setor elétrico brasileiro.

INTRODUÇÃO

Quando em um estudo de sistemas de recursos hídricos o intervalo de tempo considerado é menor que um ano a modelagem deve levar em conta o comportamento sazonal das propriedades probabilísticas das afluições naturais. Este é o caso do modelo pioneiro de Thomas e Fiering (1962), desenvolvido para vazões mensais, onde utiliza-se o processo autoregressivo de ordem um com parâmetro variável mês a mês.

Recentemente alguns pesquisadores passaram a investigar o uso dos chamados modelos autoregressivos periódicos (modelo PAR) numa tentativa de melhor modelar a variação sazonal da estrutura de autocorrelação de séries de vazões mensais.

Este artigo apresenta uma descrição de um conjunto de metodologias para identificação, estimação e verificação de modelos PAR. Apresenta-se ainda comparações entre modelos PAR e modelos atualmente em uso no setor elétrico brasileiro.

DESCRIÇÃO DE MODELO PAR(p)

O modelo PAR(p_1, p_2, \dots, p_s) pode ser descrito por:

$$\phi^m(B) \cdot \left(\frac{z_t - \mu_m}{\sigma_m} \right) = a_t \quad (1)$$

¹ Aluno da Coordenação dos Programas de Pós Graduação em Engenharia da Universidade Federal do Rio de Janeiro (COPPE/UFRJ).

² Pesquisador do Centro de Pesquisas de Energia Elétrica, CEPEL, C.P. 2754, CEP-20001, Rio de Janeiro, RJ.

³ Professor da COPPE/UFRJ.

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 da estação m ($m = 1, 2, \dots, s$),
- s é o número de estações ($s = 12$ para dados mensais),
- N é o número de anos,
- μ_m é a média sazonal de período s ,
- σ_m é o desvio-padrão sazonal de período s ,
- $\phi^m(B)$ é o operador autoregressivo de ordem p_m ,
- $\phi^m(B) = (1 - \phi_1^m B - \phi_2^m B^2 - \dots - \phi_{p_m}^m B^{p_m})$,
- p_m é a ordem do operador autoregressivo da estação m ,
- a_t são ruídos independentes e com distribuição normal média zero e variância $\sigma_a^2(m)$.

A função de autocorrelação a cada estação m do modelo, $\rho^m(k)$, é dada pela solução de (Maceira, 1987):

$$\rho^m(k) = \phi_1^m \rho^{m-1}(k-1) + \dots + \phi_{p_m}^m \rho^{m-p_m}(k-p_m) \quad (2)$$

A equação (2) é uma equação diferencial cuja solução pode ser expressa por uma combinação de ondas senoidais e decaimentos exponenciais o que faz com que $\rho^m(k)$ tenda suavemente a zero à medida que k cresce.

A função de autocorrelação parcial a cada estação do modelo, ϕ_{kk}^m , é dada pela solução de (Maceira, 1987):

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

As funções ϕ_{kk}^m caem abruptamente a zero a partir de $k=p_m + 1$

Identificação do modelo

A identificação consiste em determinar as ordens p_m dos operadores autoregressivos de cada mês $m=1, s$. Esta determinação é feita obtendo-se estimativas $\hat{\phi}_{kk}^m$, $k=1, N/4$, substituindo-se em (3) ρ_j^m , $j=1, k$, pelos valores amostrais (equação 6). Sabe-se que (Maceira, 1987) se a ordem do operador de um mês qualquer m é p então $\hat{\phi}_{kk}^m$ para $k > p$ é aproximadamente normal com média zero e variância N^{-1} . Para cada mês a ordem do operador é escolhida procurando-se o maior valor de p_m tal que todas estimativas $\hat{\phi}_{kk}^m$ para $k > p_m$ não sejam mais significativas.

Estimação do modelo

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

As médias e os desvios padrões sazonais são estimados por:

$$\hat{\mu}_m = N^{-1} \sum_{i=1}^N z_{(i-1)s+m} \quad m=1, 2, \dots, s \quad (4)$$

$$(\hat{\sigma}_m)^2 = N^{-1} \sum_{i=1}^N (z_{(i-1)s+m} - \hat{\mu}_m)^2 \quad m=1, 2, \dots, s \quad (5)$$

Os parâmetros ϕ são estimados substituindo-se em (2) ρ por suas estimativas dadas por:

$$\hat{\rho}_j^m = \frac{N^{-1} \sum_{i=1}^N (z_{(i-1)s+m} - \hat{\mu}_m) (z_{(i-1)s+m-j} - \hat{\mu}_{m-j})}{\hat{\sigma}_m \cdot \hat{\sigma}_{m-j}} \quad (6)$$

Os valores de $\sigma_a^{2(m)}$ são estimados pelas variâncias sazonais dos resíduos obtidos substituindo-se na equação do modelo, (1), todos os parâmetros por suas estimativas.

Verificação do modelo

Estimados os parâmetros é necessário testar a adequação do modelo, isto é, verificar se as hipóteses assumidas estão satisfeitas.

Hipótese de normalidade dos ruídos -- Esta hipótese é testada calculando-se as assimetrias sazonais

$$\hat{\gamma}_a^m = \frac{N^{-1} \sum_{i=1}^N \hat{a}^3_{(i-1)s+m}}{(\hat{\sigma}^{(m)})^3} \quad (7)$$

Segundo Mcleod (1974) $\hat{\gamma}_a^m$ tem aproximadamente distribuição normal com média zero e variância $6N^{-1}$. Pode-se garantir a obtenção de ruídos normais através da transformação de Box-Cox previamente aplicada aos dados. O modelo passa a ser escrito:

$$z_t^{\lambda_m} = \frac{(z_t + \text{const})^{\lambda_m}}{\lambda_m} \quad \lambda_m \neq 0 \quad (8b)$$

$$z_t^{\lambda_m} = \ln z_t \quad \lambda_m = 0 \quad (8c)$$

onde λ_m é o expoente da transformação Box-Cox da estação m. Note que μ_m e σ_m passam a representar momentos da série transformada.

Hipótese de independência dos ruídos -- Esta hipótese é testada calculando-se as autocorrelações sazonais dos ruídos por:

$$r_a^{(m)}(k) = \frac{N^{-1} \sum_{i=1}^N \hat{a}_{(i-1)s+m} \hat{a}_{(i-1)s+m-k}}{\hat{\sigma}_a^{(m)} \hat{\sigma}_a^{(m-k)}} \quad (9)$$

Mcleod e Hipel (1983) mostraram que se o modelo é adequado, $r_a^{(m)}(k)$ tem aproximadamente distribuição normal com média zero e variância menor que N^{-1} e que as estatísticas de portmanteau, dadas por:

$$Q_{m,L} = N \sum_{j=1}^L (\hat{r}_a^{(m)}(j))^2 + \frac{L(L+1)}{2N} \quad m = 1, 2, \dots, s \quad (10)$$

são assintoticamente independentes e tem distribuição quiquadrado com $(L-p_m)$ graus de liberdade. Recomenda-se o valor de $N/4$ para L ,

Um valor significativamente alto para $Q_{m,L}$ indica que a ordem do operador da estação m não está adequada. Deve-se variar esta ordem até se obter ruídos independentes. Por vezes é necessário variar a ordem dos operadores de meses adjacentes.

O modelo deve também ser testado em conjunto usando-se a estatística agregada:

$$Q_L = \sum_{m=1}^s Q_{m,L} \quad (11)$$

tendo Q_L distribuição quiquadrado com $\sum_{m=1}^s (L-p_m)$ graus de liberdade.

ESTUDOS COMPARATIVOS DE DESEMPENHO DO MODELO PAR

Nos estudos descritos a seguir procurou-se comparar o desempenho dos modelos PAR(p) com o desempenho de modelos atualmente em uso no setor elétrico brasileiro. Essas comparações de desempenho foram feitas para duas aplicações de modelos estocásticas mensais: programação dinâmica e geração de séries sintéticas (simulação).

Na aplicação de programação dinâmica o modelo PAR(p) foi comparado com o modelo de Thomas-Fiering. Este modelo é atualmente usado na determinação das estratégias de operação mensal do sistema elétrico brasileiro, através da programação dinâmica estocástica (Fortunato et al, 1985), porque permite uma solução de compromisso entre viabilidade computacional (nº pequeno de estados) e representação da estrutura de correlação das afluências. Recentemente Stedinger, 1984, propôs usar na programação dinâmica a previsão da afluência do próximo mês como variável de estado avaliando-se as decisões através da distribuição da próxima afluência condicionada à previsão. Isto permite incorporar qualquer modelo estocástico sem aumentar o esforço computacional. Os melhores modelos neste caso são aqueles que fornecem distribuições para a próxima afluência condicionada à previsão com menor variância. Em outras palavras, deve-se procurar um modelo que forneça o menor erro de previsão um passo à frente.

Na aplicação de simulação o modelo PAR(p) foi comparado com o modelo de Thomas-Fiering e com o modelo de desagregação desenvolvido no CEPEL (Kelman et al, 1978). O último modelo é usado nos estudos de planejamento da expansão do sistema elétrico brasileiro (Ventura et al, 1986). Para estes estudos um bom modelo deve gerar séries sintéticas que reproduzam os períodos secos do histórico.

Para efeito desses estudos os modelos foram ajustados à série de vazões mensais afluentes a um reservatório hipotético representativo do sistema hidroelétrico da região sudeste do Brasil entre os anos de 1931 e 1977. Esta série está expressa em Mwh. Terry et al, 1985, apresenta detalhes de como transformar um histórico multivariado de vazões afluentes a um sistema de reservatórios num histórico univariado de energias afluentes.

A Tabela 1 apresenta os parâmetros ajustados de energias e as estatísticas de portmanteau obtidas onde pode-se ver que hipótese de independência é aceita em todos os meses ao nível de confiança de 5%. No teste em conjunto do modelo obteve-se para Q_{12} , 118,21 que não é significativa a 5% ($\chi_{98,0.05}^2 = 122,11$).

Tabela 1 - "Ajuste do modelo PAR(p) a série de afluições mensais
ao reservatório equivalente - região sudeste"

MÊS (m)	1	2	3	4	5	6	7	8	9	10	11	12
λ^m	0,66	0,89	0,22	0,36	0,61	1,52	0,41	0,41	-0,55	-0,52	-0,19	1,00
p_m	9	8	11	1	2	1	3	1	1	4	4	1
ϕ_1^m	0,682	0,640	0,584	0,79	0,633	0,870	0,681	0,912	0,816	0,349	0,676	0,579
ϕ_2^m	-0,002	-0,078	0,224	-	0,275	-	0,020	-	-	0,074	-0,335	-
ϕ_3^m	-0,218	-0,046	-0,254	-	-	-	0,271	-	-	0,762	-0,141	-
ϕ_4^m	-0,223	-0,114	0,005	-	-	-	-	-	-	-0,605	0,371	-
ϕ_5^m	-0,018	-0,307	0,182	-	-	-	-	-	-	-	-	-
ϕ_6^m	0,480	0,439	-0,289	-	-	-	-	-	-	-	-	-
ϕ_7^m	-0,197	0,174	-0,060	-	-	-	-	-	-	-	-	-
ϕ_8^m	0,576	-0,341	0,189	-	-	-	-	-	-	-	-	-
ϕ_9^m	-0,409	-	0,035	-	-	-	-	-	-	-	-	-
ϕ_{10}^m	-	-	-0,175	-	-	-	-	-	-	-	-	-
ϕ_{11}^m	-	-	0,371	-	-	-	-	-	-	-	-	-
$\hat{\mu}_m$	1126	8921	36	90	473	703185	93	84	1,8	1,9	4,4	17031
$\hat{\sigma}_m$	235	2744	3,2	9,8	69,7	225618	8,6	7,7	0,002	0,003	0,1	4702
$\hat{\sigma}_a^{2(m)}$	0,421	0,488	0,425	0,376	0,249	0,243	0,160	0,169	0,334	0,559	0,573	0,664
$Q_{12,m}$	3,01 (3)	3,78 (4)	3,82 (1)	8,62 (11)	11,48 (10)	14,00 (11)	15,67 (9)	16,41 (11)	18,70 (11)	7,57 (8)	3,84 (8)	11,31 (11)

OBS.: Os valores entre parênteses na linha correspondem às estatísticas de portmanteau referem-se aos graus de liberdade da distribuição.

Aplicação na programação dinâmica

Os desempenhos dos modelos PAR(p) e Thomas-Fiering foram avaliados calculando-se os erros de previsão em um mês à frente para os anos de 1978 a 1980. Os erros de previsão um mês à frente foram calculados por:

$$e_t = \hat{z}_t - z_t \quad (12)$$

onde \hat{z}_t são os valores previstos pelo modelo calculados por

$$\hat{z}_t^{\lambda_m} = \hat{\mu}_m + \hat{\sigma}_m \left(\hat{\phi}_1^m \frac{(z_{t-1}^{\lambda_m} - \hat{\mu}_{m-1})}{\hat{\sigma}_{m-1}} + \dots + \hat{\phi}_p^m \frac{(z_{t-p}^{\lambda_m} - \hat{\mu}_{m-p})}{\hat{\sigma}_{m-p}} \right) \quad (13)$$

se $\lambda_m \neq 1$ é necessário um tratamento a z_t .

se $\lambda_m = 0$, tem-se:

$$\hat{z}_t = \exp [\hat{z}_t^{\lambda_m} + \sigma_a^2(m)] \quad (14)$$

se $\lambda_m \neq 0$ a previsão deve ser obtida por expansão em série de Taylor (ver Maceira, 1987) de (8b).

Os modelos foram comparados em termos de erro médio quadrático, erro percentual absoluto médio, erro percentual absoluto mediano e erro absoluto médio de previsão.

A Tabela 2 apresenta os resultados obtidos, onde pode-se verificar que os erros de previsão do modelo PAR(p) tendem a ser menores que os do modelo de Thomas-Fiering. No entanto aplicando-se o teste de soma dos ranks de Wilconxon (Bradley, 1968) aos quadrados dos erros de previsão de cada modelo não se pode rejeitar a hipótese de que os erros de cada modelo pertençam a mesma população, mesmo ao nível de significância de 10%.

Tabela 2 - "Erros de Previsão"

CRITÉRIO	MAPE	MEDIAN APE	MAD	MSE
PAR(p)	0,2646	0,1791	0,3974E+04	0,3354E+08
PAR(p)	0,2551	0,1789	0,3822E+04	0,2764E+08

Aplicação na geração de séries sintéticas

A fim de aplicar o modelo PAR(p) na geração de séries sintéticas, o modelo (1) deve ser colocado da seguinte forma:

$$z_t^{\lambda_m} = \hat{\mu}_m + \hat{\phi}_1 \frac{\hat{\sigma}_m}{\hat{\sigma}_{m-1}} (z_{t-1}^{\lambda_{m-1}} - \hat{\mu}_{m-1}) + \dots + \hat{\phi}_p \frac{\hat{\sigma}_m}{\hat{\sigma}_{m-p}} (z_{t-p}^{\lambda_{m-p}} - \hat{\mu}_{m-p}) + \hat{\sigma}_m a_t$$

Obtida a série Z_t aplica-se o inverso da transformação (8b) ou (8c) para obter-se z_t .

As séries sintéticas do modelo PAR foram obtidas a partir da condição inicial

$$z_{-j}^{\lambda_m} = \hat{\mu}_{-j} \quad \forall j \geq 0$$

tendo-se desprezado os primeiros 50 anos gerados para não condicionar a análise. Maceira (1987) apresenta detalhes da geração de séries pelos modelos Thomas-Fiering e de desagregação.

Os desempenhos dos modelos foram avaliados com base em estatísticas gerais (médias, desvios-padrões, correlogramas, etc) e relativas a períodos secos (comprimento de sequência-negativa, déficits, etc).

Estatísticas gerais -- A Tabela 3 compara os principais momentos das séries históricas e geradas pelos três modelos. Em geral, os momentos das séries geradas pelos modelos estão próximos dos correspondentes momentos históricos à exceção da correlação lag um anual do modelo Thomas-Fiering que é bem menor que a histórica. A figura 1 mostra os correlogramas anuais históricos e gerados onde pode-se notar que o problema de não representação da correlação a nível anual do modelo Thomas-Fiering ocorre apenas no lag um. A figura 2 mostra os correlogramas histórico e gerados para o mês de julho onde pode-se notar que o modelo Thomas-Fiering não representa bem a estrutura de correlação do histórico.

Estatísticas relativas a períodos secos Para a análise do comportamento dos modelos em relação à estas estatísticas as séries sintéticas foram divididas em segmentos de comprimento igual ao do histórico, e calculada a frequência de estatísticas sintéticas menores que o valor histórico. A tabela 4 apresenta os resultados para quatro destas estatísticas. Pode-se notar a dificuldade que os três modelos apresentam para gerar séries sintéticas com máximo comprimento de soma de sequências iguais ou equivalentes ao ocorrido no histórico. Neste particular o modelo PAR(p) apresenta ligeira vantagem. Os três modelos tiveram maior sucesso no caso do máximo deficit.

Tabela 3 - "Principais momentos das séries históricas e geradas"

	JAN	FEV	MAR	ABR	MAI	JUN	JUL	AGO	SET	OUT	NOV	DEZ	ANUAL
MÉDIA (Mw.mês)													
HIST	22551	24060	22696	15939	11146	9030	7535	6211	5959	7463	10355	16634	13303
TF	22539	23784	22672	15951	11144	9064	7549	6261	5975	7442	10302	16530	13268
PAR(p)	22742	24204	23114	16181	11255	9036	7573	6251	5959	7462	10357	16832	13414
DESAG	21994	23667	22223	15726	10939	8906	7414	6095	5878	7344	10210	16385	13065
DESVIO PADRÃO (Mw.mês)													
HIST	6964	8212	3056	4552	2656	1960	1629	1349	1809	2129	3039	4701	2751
TF	7105	8357	5253	4510	2653	1995	1638	1332	1746	2107	2987	4717	2637
PAR(p)	7160	8562	8868	4746	2787	2100	1703	1412	1719	2266	3075	4685	2898
DESAG	6549	8064	3324	4743	2776	1990	1668	1358	1785	2142	3044	4764	2789
CORRELAÇÃO (t, t-1)													
HIST	0,66	0,57	0,66	0,77	0,85	0,87	0,89	0,91	0,80	0,54	0,62	0,57	0,25
TF	0,66	0,60	0,67	0,78	0,84	0,90	0,89	0,91	0,78	0,57	0,59	0,55	0,11
PAR(p)	0,69	0,59	0,69	0,81	0,85	0,87	0,88	0,92	0,79	0,55	0,62	0,56	0,29
DESAG	0,60	0,49	0,69	0,80	0,87	0,88	0,90	0,91	0,80	0,54	0,62	0,57	0,26

Tabela 4 - "Estatísticas Relativas a Períodos Secos".

OBS.: Para os modelos apresenta-se a menor e a maior estatística obtida em 21 segmentos e entre parentesis a frequência de valores menores que o valor histórico. Para a definição de sequências, usou-se como nível a média sazonal do histórico. α é o nível de regularização.

	Máxima Soma de sequência (10 ⁴ Mw.mês)	Máx.Comprimento de sequência (meses)	Máx.Deficit(10 ⁴ Mw.mês)	
			$\alpha = 0,90$	$\alpha = 0,70$
Histórico	19,7	44	15,6	3,8
PAR	7,3-29,7(0,91)	18-62(0,95)	7,2-33,3(0,62)	2,7-6,4(0,38)
TF	4,1-17,2(1,00)	14-40(1,00)	6,2-22,4(0,86)	2,2-6,9(0,48)
DESAG	5,2-16,6(0,95)	14-33(1,00)	8,7-21,9(0,48)	2,9-6,8(0,33)

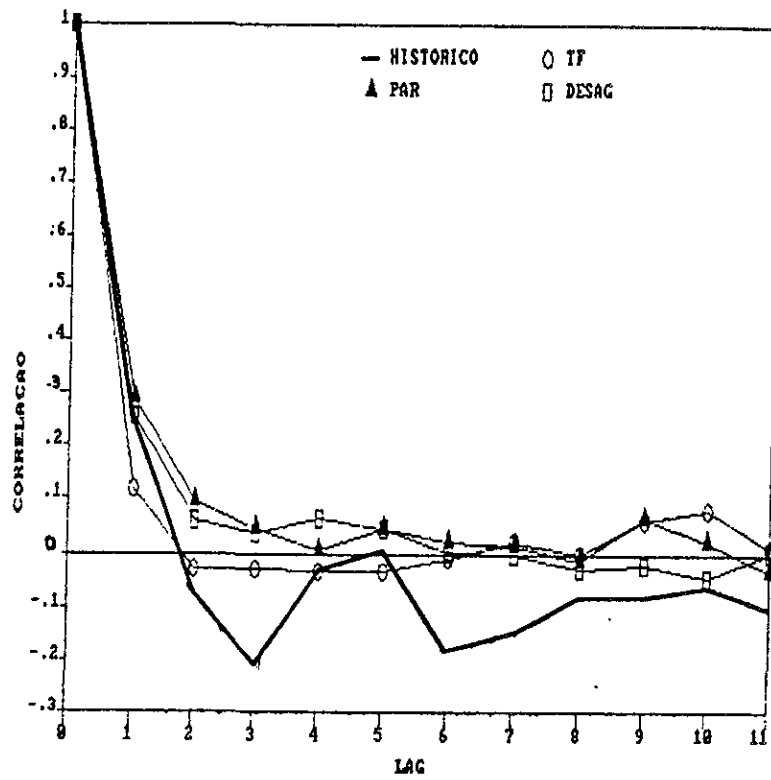


Figura 1 - "Correlogramas anuais"

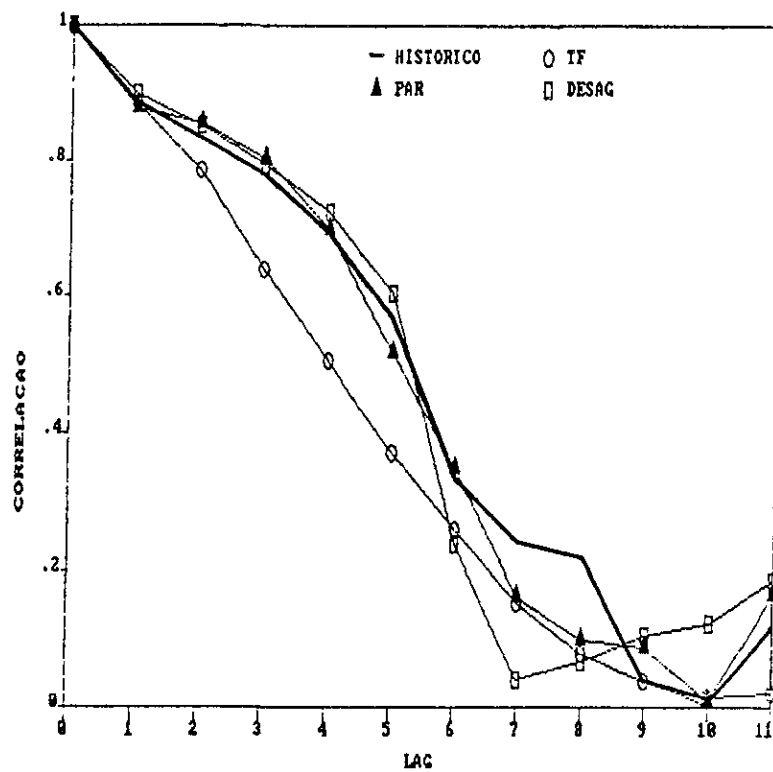


Figura 2 - "Correlogramas do mês de julho"

CONCLUSÕES

O modelo PAR é uma alternativa atraente para a modelagem de séries hidrológicas mensais podendo-se usar na identificação, estimação e verificação as metodologias apresentadas neste artigo.

AGRADECIMENTOS

Os autores agradecem aos engenheiros Mário Veiga Ferraz Pereira e Gerson Couto de Oliveira do Departamento de Sistemas do CEPEL, pelas diversas sugestões.

REFERÊNCIAS

- BRADLEY, J. V. (1968), "Distribution-Free Statistical Tests", Prentice-Hall, Inc, New Jersey.
- FORTUNATO, L.A.M., ARARIPE NETO, T. A., BARRETO, L.A.L., e FERREIRA, C., (1985), "Operation Planning Studies of the Brazilian Generation System", IFAC Electric Energy Systems, Rio de Janeiro, Brasil.
- KELMAN, J., OLIVEIRA, G.C., PEREIRA, M.V.F. e Costa, C.C.G., (1978), "Modelo de Séries Hidrológicas", Relatório Técnico CEPEL.
- MACEIRA, M.E.P., (1987), "Aplicação de modelos PARMA em séries hidrológicas mensais", tese de mestrado a ser submetido a COPPE-UFRJ, Rio de Janeiro, Brasil.
- MCLEOD, A.I., (1974), "Contributions to Applied Time Series", tese de Mestrado, Universidade de Waterloo, Waterloo, Ontario, Canada.
- MCLEOD, A.I., e HIPEL, K.W., (1983), "Periodic Autoregression Modelling", Dept. of Statistical and Actuarial Sciences, The University of Western Ontario, London, Ontario, Canada.
- TERRY, L.A., PEREIRA, M.V.F., ARARIPE NETO, T.A., SILVA, L.F.C.A. e SALES, P.R.H., (1986), "Coordinating the Energy Generation of the Brazilian National Hydrothermal Electrical Generation Systems Interfaces, volume 16, Number 1.
- THOMAS, H.A., FIERING, M.B., (1962), Mathematical Synthesis of Streamflow sequences for the analysis of river basins by simulation, A. Maas et al, "Design of Water Resource Systems", chap.12, Cambridge, Mass., Harvard University Press.
- VENTURA, A., ALBUQUERQUE, J.C.R., ROSENBLATT, J., LIMA, W.M., (1986), "Short Term Expansion Planning Studies in the Brazilian Generation System-Interrelationship-GCPS-GCOI", I International Workshop on Energy Operation Planning of the Brazilian Interconnected Power System, Rio de Janeiro, Brasil.

USE OF PAR MODELS IN THE SIMULATION AND
FORECASTING OF MONTHLY HYDROLOGICAL SERIES

by

M.E.P.Maceira, J.Kelman and J.M.Damázio

ABSTRACT -- A set of methods that can be used in the identification, estimation and diagnostic check stages of PAR models development are explained. Also, an assessment of PAR model's performance as compared with models which are actually in use by the Brazilian electric sector is presented.