

ABE - ASSOCIAÇÃO BRASILEIRA DE ESTATÍSTICA  
SBE - SOCIEDADE BRASILEIRA DE ECONOMETRIA

# 3<sup>ª</sup> ESCOLA DE SÉRIES TEMPORAIS E ECONOMETRIA

$$= \theta(B)a_t$$

Identificação e estimação  
recursiva de modelos  
de séries temporais

PAULO R. DE HOLANDA SALES  
BASÍLIO DE BRAGANÇA PEREIRA  
AILTON MESQUITA VIEIRA

## PREFÁCIO

A série de artigos que constituem o objeto deste mini-curso tratam de procedimentos estatísticos iterativos para identificação e estimação de modelos de séries temporais lineares multivariadas.

Baseados em regressões autorregressivas recursivas de algoritmo

de identificação e estimação recursiva de modelos de séries temporais

de séries temporais, como

substantivo, reduzem

em procedimentos lineares.

## IDENTIFICAÇÃO E ESTIMAÇÃO RECURSIVA

### DE MODELOS DE SÉRIES TEMPORAIS

A metodologia (inicialmente chamada *Iterative Identification and Estimation*) é apresentada

de forma ilustrada com aplicações nos casos de modelos ARMA univariados, alguns

dos classes de modelos multivariados, tais como SIMVAR, e de funções de transferência

de sistemas de controle. O procedimento iterativo é aplicado ao caso

de identificação e estimação de modelos de séries temporais multivariadas.

PAULO ROBERTO DE HOLANDA SALES

ELETRÓBRÁS E UFRJ

BASILIO DE BRAGANÇA PEREIRA

IM E COPPE/UFRJ

AILTOM MESQUITA VIEIRA

ELETRÓBRÁS

Finalmente, as pesquisas realizadas neste mini-curso são

de natureza básica e de interesse acadêmico, sendo que alguns dos

resultados foram publicados em artigos científicos e em livros.

Este mini-curso foi desenvolvido no âmbito do projeto de pesquisa de desenvolvimento de

tecnologias de identificação e estimação de modelos de séries temporais multivariadas, financiado

## PREFÁCIO

## Prefácio

A série de artigos que constituem o objeto deste mini-curso tratam de procedimentos automáticos lineares para identificação e estimação de modelos de séries temporais uni e multivariadas.

Baseados em regressões-autorregressões recursivas os algoritmos apresentados além de não exigir conhecimentos profundos de identificação de modelos, como em outras metodologias (ex: Box-Jenkins) apresentam ainda substancial redução nos tempos computacionais em virtude de serem baseadas em procedimentos lineares.

A metodologia (inicialmente chamada Hannan-Rissanen) é apresentada e ilustrada com aplicações nos casos de modelos **ARMA** univariados, algumas classes de modelos multivariados, tais como **SURARMA**, o de Função de Transferência ou **ARMAX**, o multivariado **MARMA** e indica-se como proceder no caso Função de Transferência Multivariado ou **MARMAX**.

A teoria destes algoritmos é baseada em inúmeros trabalhos teóricos do Professor E.J.Hannan e seus colaboradores, desenvolvidos a partir de 1980 e apresentados em seu recente livro com o Professor M.Deistler e corroboram a imagem de que, em sua longa carreira, o professor Hannan tem, invariavelmente, indicado as direções nas quais o campo da Análise de Séries Temporais, irá se mover, mais adiante.

Finalmente, as pesquisas resumidas neste minicurso são desenvolvimentos da tese de doutoramento de um dos outros (PRHS) e com alguns desenvolvimentos computacionais do projeto de tese de doutoramento de AMV, ambos sob a orientação de BBP.

Resumo

- Introdução

- Identificação e estimação linear

- Aplicações

- Conclusões

- Referências

## CONTEÚDO

### Prefácio

#### I - Procedimento linear de identificação e estimação de modelos **ARMA** para séries hidrológicas temporais

- Resumo
- Introdução
- Estimação não linear/linear
- Aplicação
- Conclusões
- Referências

#### II - Procedimento linear de identificação e estimação de função de transferência para séries hidrológicas temporais

- Resumo
- Introdução
- Formas alternativas de representação de sistemas dinâmicos li neares
- Procedimentos lineares para o Modelo **ARMAX**
- Aplicação
- Conclusões
- Referências

#### III - Procedimento linear de identificação e estimação de modelos **SURARMA** aplicados a séries hidrológicas temporais multivariadas

- Resumo
- Introdução
- Identificação e estimação linear
- Aplicação
- Conclusões
- Referências

IV - Procedimento linear de identificação e estimação de modelos para séries hidrológicas temporais multivariadas

- Resumo
- Introdução
- Identificação e estimação linear
- Aplicação
- Conclusões
- Referências

PROCEDIMENTOS LINEARES DE IDENTIFICAÇÃO E ESTIMAÇÃO  
DE MODELOS ARMA PARA SÉRIES HIDROLÓGICAS TEMPORAIS

POR

P. R. HOLANDA SALES<sup>1</sup>, B. B. PEREIRA<sup>2</sup> e A. M. VIEIRA<sup>1</sup>

**Resumo** - - É apresentado um método para identificação da ordem dos operadores autorregressivos médias móveis e estimação dos parâmetros desconhecidos de uma sequência  $\{z_t\}$ , ARMA  $(p_0, q_0)$ . O método é uma versão recursiva do algoritmo "off-line" de três estágios, baseado em mínimos quadrados padrão, apresentado inicialmente por Hannan & Rissanen (1982) e que incorpora as modificações indicadas por Hannan & Kavalieris (1984) e Hannan, Kavalieris & Mackisack (1986). O atrativo maior do algoritmo se prende ao fato de que um dos problemas mais cruciais no ajuste de um modelo ARMA  $(p_0, q_0)$  a uma série temporal é a identificação correta das ordens  $p_0$  e  $q_0$ , a partir das observações  $z_1, z_2, \dots, z_t$ . Os procedimentos de estimação são igualmente importantes e a maioria dos que têm sido propostos na literatura, são baseados no princípio da máxima verossimilhança. Muito embora sejam desejáveis estimativas exatas de máxima verossimilhança, estimativas simplificadas são, também, atrativas pela razão de um menor tempo computacional e avaliações teóricas muito mais simples. Então, o algoritmo apresentado no presente artigo pode ser descrito como se segue. Em primeiro lugar é ajustado a uma série temporal de comprimento  $T$ , um modelo autorregressivo de longa extensão. A seguir, os "innovations" obtidos do modelo selecionado são utilizados em uma segunda etapa que consiste na estimação preliminar dos parâmetros do modelo ARMA  $(p, q)$ . Finalmente, a partir de uma segunda estimativa dos resíduos, obtida no estágio anterior, e pela utilização de critérios tais como AIC  $(p, q)$  e BIC  $(p, q)$ , uma estimativa fortemente consistente é obtida para os parâmetros  $(p_0, q_0)$ . Como caso de aplicação, foi utilizada a série hidrológica temporal de vazões naturais médias mensais ao reservatório de Furnas, na bacia do Rio Grande, no período 1931-1984, obtida do banco de dados hidrometeorológicos (BDH) das Centrais Elétricas Brasileiras S.A. (ELETROBRÁS).

INTRODUÇÃO

Seja  $\{z_t\}$  uma série de tempo discreta, estacionária e ergódica gerada por um modelo ARMA  $(p_0, q_0)$ ,

$$\sum_{j=0}^{p_0} \phi_j z_{t-j} = \sum_{j=0}^{q_0} \theta_j a_{t-j} \quad (1)$$

<sup>1</sup> Engenheiro, Departamento de Operação Energética, ELETROBRÁS, Rio de Janeiro - RJ

<sup>2</sup> Professor, COPPE/Universidade Federal do Rio de Janeiro, Rio de Janeiro RJ

onde,  $\phi_0 = \theta_0 = 1$   
 $\phi_{p_0} \neq 0, \theta_{q_0} \neq 0$   
 $\langle a_t \rangle = 0, \langle a_t a_s \rangle = \delta_{s,t} \sigma^2, \langle \cdot \rangle = \text{valor esperado}$

A hipótese do valor esperado de  $a_t$  ser zero implica, na prática, que os dados devam possuir, também, média zero. Por simplicidade e, por não fazer a menor diferença na obtenção de resultados assintóticos, a correção em média será ignorada a seguir. O interesse maior reside na estimação de  $p_0, q_0$ , bem como dos parâmetros do sistema,  $\phi_j$  e  $\theta_j$  a partir das observações  $z_1, z_2, \dots, z_t$ .

Pela introdução das funções de transferência,

$$\phi_{p_0}(B) = \sum_{j=0}^{p_0} \phi_j B^j \quad e, \quad (2)$$

$$\theta_{q_0}(B) = \sum_{j=0}^{q_0} \theta_j B^j \quad (3)$$

onde,  $B$  é o operador atraso, isto é,  $Bz_t = z_{t-1}$ , a equação (1) pode ser reescrita como,

$$\phi_{p_0}(B) z_t = \theta_{q_0}(B) a_t \quad (4)$$

As condições de estacionaridade e invertibilidade impõem que os polinômios em  $B$ , dos graus  $p_0$  e  $q_0$ ,  $\phi_{p_0}(B)$  e  $\theta_{q_0}(B)$ , sejam primos, isto é, não possuam fatores comuns. Além disso,  $\phi_{p_0}(B) \neq 0$  e  $\theta_{q_0}(B) \neq 0$  para  $|B| < 1$ . Então, a sequência  $\{a_t\}$ , com média zero e variância  $\sigma^2$  é chamada de "innovations".

Por conveniência, seja  $\beta$  um vetor cujos  $k = p + q$  elementos são os parâmetros autorregressivos e médias móveis,  $\phi$  e  $\theta$ . Então o conjunto completo de  $(k + 1)$  parâmetros do processo ARMA pode ser escrito como  $\underline{\beta}, \sigma_a^2$  ou  $\underline{\beta}, \sigma_a^2$ .

Como dito anteriormente,  $\{z_t\}$  é uma série de tempo discreta, estacionária e ergódica. Entretanto, para situações práticas mais gerais,  $\{z_t\}$  pode ser obtida a partir de alguma transformação, como por exemplo, do tipo Box e Cox e/ou uma dessazonalização da série original (Holanda Sales e Vieira, 1986),  $\{y_t\}$ . Neste caso, para efeitos de previsão e/ou geração cuidados especiais devem ser tomados quando da obtenção de transformações inversas.

De acordo com Box e Jenkins (1976) se as observações são geradas por um modelo ARIMA, a log-verossimilhança não condicional é dada por,

$$L(\underline{\beta}, \sigma_a^2) = f_t(\underline{\beta}) - T \ln \sigma_a^2 - \frac{S(\underline{\beta})}{2 \sigma_a^2} \quad (5)$$

onde,  $f_t(\beta)$  é uma função dos parâmetros  $\phi$  e  $\theta$ . Então, a função soma dos quadrados não condicional é dada por,

$$S(\beta) = \sum_{t=-\infty}^T \langle a_t | \beta, z_t \rangle^2 \quad (6)$$

onde,  $\langle a_t | \beta, z_t \rangle$  é o valor esperado condicional de  $\{a_t\}$  dados  $\beta$  e  $z_t$ . Em termos práticos, a soma infinita pode ser substituída pela soma finita,

$$\sum_{t=1-K}^T \langle a_t | \beta, z_t \rangle^2 \quad (7)$$

onde K tal que o efeito dos transientes iniciais é desprezível.

Geralmente,  $f_t(\beta)$  só é importante para T relativamente pequeno. Para valores moderados e grandes de T, a expressão (5) é dominada por  $S(\beta)/2\sigma^2$ . Então, a função soma dos quadrados, no espaço dos parâmetros ( $\beta$ ) é muito próxima da função de verossimilhança ou da função log-verossimilhança. Segue, então, em particular, que as estimativas dos parâmetros obtidas pela minimização da soma dos quadrados de (6), isto é, estimativas de mínimos quadrados, são aproximações das estimativas de máxima verossimilhança.

Entretanto, simplificações consideráveis ocorrem na minimização com respeito a  $\beta$ , da soma dos quadrados,

$$S(\beta) = \sum_{t=1-K}^T \langle a_t | \beta, z_t \rangle^2 \quad (8)$$

se cada uma das parcelas é uma função linear dos parâmetros  $\beta$ . Por outro lado, a linearidade de  $\langle a_t | \beta, z_t \rangle$  é de alguma forma diferente para os parâmetros autorregressivos,  $\phi$  e para os parâmetros médias móveis,  $\theta$ .

Para um processo autorregressivo puro, de ordem p,

$$\langle a_t | \beta, z_t \rangle = \phi_p(B) \langle z_t \rangle \quad (9)$$

e,

$$\frac{\partial \langle a_t | \beta, z_t \rangle}{\partial \phi_i} = - \langle z_{t-i} \rangle + \phi_p(B) \frac{\partial \langle z_t \rangle}{\partial \phi_i} \quad (10)$$

Para  $j > 0$ ,  $\langle z_j \rangle = z_j$  e  $\partial \langle z_j \rangle / \partial \phi_i = 0$ , enquanto para  $j < 0$ ,  $\langle z_j \rangle$  e  $\partial \langle z_j \rangle / \partial \phi_i$  são funções de  $\phi_j$ . Então, exceto pelo efeito dos valores iniciais,  $\langle a_t | \beta, z_t \rangle$  é uma função linear dos  $\phi$ 's.

Considerando agora um processo médias móveis de ordem q, tem-se:



$$\langle a_t | \underline{\beta}, z_t \rangle = \theta_q^{-1} (B) \langle z_t \rangle \quad (11)$$

e,

$$\frac{\partial \langle a_t | \underline{\beta}, z_t \rangle}{\partial \theta_i} = \theta_q^{-2} (B) \langle z_{t-i} \rangle + \theta_q^{-1} (B) \frac{\partial \langle z_t \rangle}{\partial \theta_i} \quad (12)$$

ou seja, a sequência  $\langle a_t | \underline{\beta}, z_t \rangle$  é sempre uma função não linear dos parâmetros.

Por outro lado, expandindo  $\langle a_t | \underline{\beta}, z_t \rangle$  em série de Taylor no entorno de  $\underline{\beta}'_0 = (\beta_{1,0}, \beta_{2,0}, \dots, \beta_{k,0})$ , tem-se, aproximadamente,

$$\langle a_t | \underline{\beta}, z_t \rangle = \langle a_{t,0} | \underline{\beta}_0, z_t \rangle - \sum_{i=1}^k (\beta_i - \beta_{i,0}) v_{i,t} \quad (13)$$

onde,

$$v_{i,t} = - \left. \frac{\partial \langle a_t | \underline{\beta}, z_t \rangle}{\partial \beta_i} \right|_{\underline{\beta} = \underline{\beta}_0} \quad (14)$$

Então, se  $V$  é uma matriz  $(T + K) \times k$ ,  $\{v_{i,t}\}$ , o sistema de equações (13) pode ser escrito como,

$$\langle \underline{a}_0 \rangle = V(\underline{\beta} - \underline{\beta}_0) + \langle \underline{a} \rangle \quad (15)$$

onde,  $\langle \underline{a}_0 \rangle$  e  $\langle \underline{a} \rangle$  são vetores colunas de  $(T + K)$  elementos.

Os ajustamentos  $(\underline{\beta} - \underline{\beta}_0)$ , os quais minimizam  $S(\underline{\beta}) = \langle \underline{a} \rangle' \langle \underline{a} \rangle$ , podem ser obtidos por mínimos quadrados lineares, através da regressão dos  $\langle a_0 | \beta_0, z_t \rangle$  nos  $v$ 's. Dado que os  $\langle a_t | \underline{\beta}, z_t \rangle$  não são exatamente lineares nos parâmetros  $\underline{\beta}$ , um simples ajustamento não produz, em geral, estimativas de mínimos quadrados. Entretanto, a convergência é rápida se estimativas obtidas no estágio de identificação são utilizadas inicialmente.

### ESTIMAÇÃO NÃO LINEAR/LINEAR

Seja um modelo ARMA  $(p, q)$  como expresso em (4) e ponha-se como objetivo, na fase de estimação a sua separação em componentes linear e não linear. Assim,

$$\phi_p (B) z_t = \theta_q (B) a_t \quad (16)$$

então,

$$a_t = \phi_p (B) \theta_q^{-1} (B) z_t \quad (17)$$

ou,

$$a_t = \phi_p(B) (\varepsilon_t | \underline{\theta}) \quad (18)$$

onde,

$$(\varepsilon_t | \underline{\theta}) = \theta_q^{-1} (B) z_t \quad (19)$$

ou seja,

$$z_t = \theta_q (B) (\varepsilon_t | \underline{\theta}) \quad (20)$$

Para um conjunto qualquer conhecido de  $\underline{\theta}$ ,  $\{\varepsilon_t\}$  pode ser obtida recursivamente a partir de (20), isto é,

$$\varepsilon_t = z_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (21)$$

A recursão pode ser iniciada fazendo-se os  $\varepsilon_t$  não conhecidos iguais a zero. Obtida a série  $\{\varepsilon_t\}$ , as estimativas condicionais  $(\hat{\phi} | \hat{\theta})$  podem ser obtidas rapidamente. Elas serão nada mais nada menos que as estimativas dos parâmetros autorregressivos no modelo linear expresso em (17), o qual pode ser escrito como,

$$a_t = \varepsilon_t - \phi_1 \varepsilon_{t-1} - \phi_2 \varepsilon_{t-2} - \dots - \phi_p \varepsilon_{t-p} \quad (22)$$

Entretanto, ser for adotado a aproximação na qual os valores iniciais são zero, isto é, iguais as suas esperanças não condicionais, para o modelo expresso em (17) e, se uma estimativa inicial para o vetor de parâmetros  $\underline{\beta} = (\underline{\phi}, \underline{\theta})$  é dada por  $\underline{\beta}_0 = (\underline{\phi}_0, \underline{\theta}_0)$ , então,

$$a_{t,0} = \phi_{p,0} (B) \theta_{q,0}^{-1} (B) z_t \quad (23)$$

e,

$$-\frac{\partial a_t}{\partial \phi_i} | \underline{\beta}_0 = u_{i,t} = u_{t-i} \quad (24)$$

$$-\frac{\partial a_t}{\partial \theta_i} | \underline{\beta}_0 = v_{i,t} = v_{t-i} \quad (25)$$

onde,

$$u_t = \theta_{q,0}^{-1} (B) z_t = \phi_{p,0}^{-1} (B) a_{t,0} \quad (26)$$

$$v_t = -\theta_{q,0}^{-2} (B) \phi_{p,0} (B) z_t = -\theta_{q,0}^{-1} (B) a_{t,0} \quad (27)$$

A série  $\{a_t\}$ , bem como as auxiliares  $\{u_t\}$  e  $\{v_t\}$ , podem ser obtidas recursivamente, fazendo-se os valores iniciais iguais a

zero, isto é,

$$a_{t,0} = z_t - \phi_{1,0} z_{t-1} - \dots - \phi_{p,0} z_{t-p} + \theta_{1,0} a_{t-1} + \dots + \theta_{q-0} a_{t-q,0} \quad (28)$$

$$u_t = \theta_{1,0} u_{t-1} + \dots + \theta_{q,0} u_{t-q} + z_t \quad (29)$$

$$= \phi_{1,0} u_{t-1} + \dots + \phi_{p,0} u_{t-p} + a_{t,0} \quad (30)$$

$$v_t = \theta_{1,0} v_{t-1} + \dots + \theta_{q,0} v_{t-q} - a_{t,0} \quad (31)$$

Correspondendo a (13), a equação de regressão linear aproximada é dada por,

$$a_{t,0} = \sum_{i=1}^p (\phi_i - \phi_{i,0}) u_{t-i} + \sum_{i=1}^q (\theta_i - \theta_{i,0}) v_{t-i} + a_t \quad (32)$$

Os ajustamentos são, então, os coeficientes da regressão de  $a_{t,0}$  em  $u_{t-i}$  e  $v_{t-i}$ . Pela adição dos valores ajustados as primeiras estimativas  $\hat{a}_0$ , um conjunto de segundas estimativas é formado e o processo continua até que a convergência seja obtida. Entretanto, corroborando o já dito no último parágrafo do item anterior, a convergência é rápida se estimativas iniciais obtidas no estágio de identificação são utilizadas.

Alternativamente, as séries auxiliares,  $\{u_t\}$  e  $\{v_t\}$ , podem ser vistas como contemporaneamente relacionadas. Logo, a técnica SURE (Sistema de equações de regressão aparentemente não relacionadas) poderia ser utilizada na obtenção de estimativas lineares consistentes e assintoticamente eficientes (Harvey, 1982) para os  $(k+1)$  parâmetros do modelo ARMA  $(p_0, q_0)$ .

Então, um algoritmo linear de três estágios para identificação e estimação dos parâmetros de uma sequência  $\{z_t\}$ , ARMA  $(p_0, q_0)$  pode ser sumarizado como se segue:

#### Estágio I - Preliminar

Ajusta-se aos dados um modelo autorregressivo de ordem  $n$ . As estimativas de  $\{a_t\}$  no modelo,

$$\phi_n(B) z_t = a_t \quad (33)$$

são dados por,

$$\hat{a}_t = \sum_{i=0}^n \hat{\phi}_i z_{t-i} \quad (34)$$

onde,

$$\hat{\phi}_0 = 1 \text{ e } z_t = 0 \quad \forall t \leq 0 \quad (35)$$

Neste caso,  $n$  é selecionado de maneira a minimizar,

$$\text{BIC}(v) = \ln \hat{\sigma}_v^2 + v \frac{\ln T}{T} \quad (36)$$

onde,  $\hat{\sigma}_v^2$  é a estimativa de  $\sigma^2$  a partir da autorregressão de ordem  $v$ . É suficiente que  $v \leq (\ln T)^{1+\delta}$ ,  $\delta > 0$ .

### Estágio II - Identificação e Estimação Inicial

Com as séries  $\{z_t\}$  e  $\{\hat{a}_t\}$  e, a partir da regressão de  $z_t$  em  $z_{t-i}$  ( $i = 1, 2, \dots, p$ ) e  $\hat{a}_t$  em  $\hat{a}_{t-i}$  ( $i = 1, \dots, q$ ), obtêm-se as estimativas  $\tilde{p}$ ,  $\tilde{q}$ ,  $\hat{\sigma}_{p,q}^2$  das ordens  $p_0$ ,  $q_0$  e da variância dos resíduos,  $\hat{\sigma}_a^2$ . São obtidas, também, estimativas iniciais,  $\beta_0 = (\phi_0, \theta_0)$ , do vetor de coeficientes,  $\beta = (\phi, \theta)$ , do modelo,

$$\phi_p(B) z_t = \theta_q(B) a_t \quad (37)$$

Neste caso, as estimativas são tais que minimizam,

$$\text{BIC}(p,q) = \ln \sigma_{p,q}^2 + (p+q) \frac{\ln T}{T} \quad (38)$$

### Estágio III - Estimação Final

Com  $(\tilde{p}, \tilde{q})$  e as estimativas iniciais,  $\beta_0 = (\phi_0, \theta_0)$ , obtidas no estágio de identificação anterior, as equações (28), (29) ou (30) e (31) são empregadas na obtenção da série  $\{a_{t,0}\}$  e das séries auxiliares  $\{u_t\}$  e  $\{v_t\}$ . A partir daí, a equação (32) é utilizada na obtenção de uma segunda estimativa para os coeficientes. O processo continua até que a convergência seja obtida.

Por outro lado, é importante que se observe que a equação (32) pode, alternativamente, ser escrita como,

$$a_{t,0} + \sum_{i=1}^{\tilde{p}} \phi_{i,0} u_{t-i} + \sum_{i=1}^{\tilde{q}} \theta_{i,0} v_{t-i} = \sum_{i=1}^{\tilde{p}} \phi_i u_{t-i} + \sum_{i=1}^{\tilde{q}} \theta_i v_{t-i} + a_t \quad (39)$$

ou, com auxílio das equações (30) e (31),

$$a_{t,0} + (u_t - a_{t,0}) + (v_t + a_{t,0}) = \sum_{i=1}^{\tilde{p}} \phi_i u_{t-i} + \sum_{i=1}^{\tilde{q}} \theta_i v_{t-i} + a_t \quad (40)$$

ou ainda,

$$a_{t,0} + u_t + v_t = \sum_{i=1}^{\bar{p}} \phi_i u_{t-i} + \sum_{i=1}^{\bar{q}} \theta_i v_{t-i} + a_t \quad (41)$$

Em outras palavras, basta que se forme, a cada iteração, a regressão da variável  $(a_{t,0} + u_t + v_t)$  nas variáveis auxiliares  $u_{i,t} (i=1, \bar{p})$  e  $v_{i,t} (i=1, \bar{q})$ .

### APLICAÇÃO

Como exemplo de aplicação do algoritmo proposto foi selecionada a série de vazões naturais médias mensais de Furnas. Este local de aproveitamento está situado no rio Grande, no município de Alpinópolis em Minas Gerais. Sua área de drenagem é de 50.464 km<sup>2</sup>. O volume total do reservatório é de 22,950 km<sup>3</sup>, sendo o volume útil igual a 17,217 km<sup>3</sup>. Com relação à usina, sua potência efetiva total é de 1312 MW com oito unidades geradoras de 164 MW. O tipo de turbina é Francis e o engolimento total da usina é de 1810 m<sup>3</sup>/s.

O período de observação da série é de janeiro de 1931 a dezembro de 1984. A tabela 1 resume as principais estatísticas mensais da série analisada.

TABELA 1. Estatísticas Correspondentes a Série de Vazões Naturais Médias Mensais do Aproveitamento de Furnas (1931-1984)

MÊS	MÉDIA (m <sup>3</sup> /s)	MAXIMO (m <sup>3</sup> /s)	MÍNIMO (m <sup>3</sup> /s)	D.PADRAO (m <sup>3</sup> /s)	COEFICIENTE ASSIMETRIA
JAN	1714	3529	571	711	0,659
FEV	1679	3264	355	690	0,372
MAR	1520	3650	463	665	1,090
ABR	1028	2367	379	369	1,043
MAI	747	1679	288	249	1,109
JUN	630	2392	310	291	4,271
JUL	520	1368	275	172	2,302
AGO	431	1082	200	141	1,924
SET	449	1989	196	264	4,048
OUT	543	1934	276	252	3,407
NOV	768	2018	324	321	1,389
DEZ	1285	3117	252	509	0,957

A transformação do Box e Cox selecionada indicou  $\lambda_1 = 0,0$  e  $\lambda_2 = -179,0$ , isto é, uma transformação do tipo logorítmo natural.

Como as estatísticas médias e variâncias mensais da série transformada eram periódicas, foi feita uma padronização mensal da mesma no sentido de se obter estacionaridade de segunda ordem. Da série resultante, obteve-se uma outra com média zero e variância unitária. A seguir, procedeu-se a aplicação do algoritmo proposto.

No estágio I, foram ajustados modelos autorregressivos para  $\delta = 0,0$  e  $\delta = 0,5$ . O critério BIC (p) indicou p=2 como mostrado

na tabela 2, a seguir.

TABELA 2. Modelos Autorregressivos Ajustados a Série Transformada da Dependente Padronizada do Aproveitamento de Furnas  
Transformação:  $z = \ln(x + (-179.))$

MODELO AJUSTADO AR(P)	VARIÂNCIA DOS RESÍDUOS (VR)	CRITÉRIO DE AKAIKE BIC(P) = $\ln(VR) + P \ln(T)/T$
AR (1)	0,44936	- 0,790
AR (2)	0,43712	- 0,808
AR (3)	0,43327	- 0,806
AR (4)	0,43342	- 0,796
AR (5)	0,43421	- 0,784
AR (6)	0,43496	- 0,773
AR (7)	0,43632	- 0,759
AR (8)	0,43576	- 0,751
AR (9)	0,43657	- 0,739
AR (10)	0,43677	- 0,728
AR (11)	0,43646	- 0,719
AR (12)	0,43772	- 0,706
AR (13)	0,43888	- 0,694
AR (14)	0,43951	- 0,682
AR (15)	0,44020	- 0,671
AR (16)	0,44193	- 0,657

A seguir, passou-se para o estágio II, onde foram ajustadas "regressões" ARMA (p,q) à série transformada e à série de resíduos obtida do ajustamento de uma AR(2) à primeira. A tabela 3 apresenta os principais resultados onde, entre parêntesis, estão indicadas as variâncias dos resíduos. O critério BIC(p,q) indicou p=q=1.

TABELA 3. Valores de BIC(p,q) a Partir do Ajustamento de Regressões Lineares à Série Transformada Dependente Padronizada e à Série de Resíduos Obtida no Estágio I.

p \ q	q					
	0	1	2	3	4	5
0	-	- 0,169 (0,83600)	- 0,276 (0,74410)	- 0,401 (0,64962)	- 0,496 (0,58502)	- 0,594 (0,52505)
1	- 0,790 (0,44948)	- 0,807 (0,43746)	- 0,802 (0,43503)	- 0,792 (0,43529)	- 0,781 (0,43572)	- 0,771 (0,43570)
2	- 0,801 (0,44016)	- 0,801 (0,43542)	- 0,792 (0,43503)	- 0,782 (0,43529)	- 0,771 (0,43572)	- 0,761 (0,43570)
3	- 0,801 (0,43545)	- 0,791 (0,43544)	- 0,781 (0,43543)	- 0,772 (0,43520)	- 0,761 (0,43570)	- 0,752 (0,43503)
4	- 0,791 (0,43545)	- 0,781 (0,43545)	- 0,771 (0,43543)	- 0,761 (0,43556)	- 0,751 (0,43566)	- 0,742 (0,43503)
5	- 0,781 (0,43556)	- 0,771 (0,43556)	- 0,761 (0,43555)	- 0,751 (0,43555)	- 0,741 (0,43565)	- 0,731 (0,43571)

A partir das estimativas iniciais dos parâmetros do modelo ARMA (1,1) identificado no estágio anterior,  $\hat{\phi}_1=0,84098$ ,  $\hat{\sigma}_1=0,22658$  e  $\hat{\sigma}_a^2=0,43746$ , passou-se ao

estágio III do algoritmo proposto. As estimativas finais foram obtidas após quatro iterações, como sumarizado na tabela 4. Entre parêntesis estão relacionados os erros padrão das estimativas.

TABELA 4. Estimativas Finais dos Parâmetros  $\phi_1, \theta_1$  e  $\sigma_a^2$  do Modelo ARMA (1,1) ajustado à Série Transformada Dependente Padronizada do Aproveitamento de Furnas

ITERAÇÃO	$\phi_1$	$\theta_1$	$\sigma_a^2$
1	0,71888 (0,03530)	0,02496 (0,03796)	0,44920
2	0,85067 (0,02805)	0,25400 (0,05049)	0,43438
3	0,85441 (0,02691)	0,26374 (0,04995)	0,43428
4	0,85443 (0,02690)	0,26383 (0,04994)	0,43427

Como exemplo para validação do algoritmo proposto, recorreu-se a identificação e estimação dos parâmetros do modelo ARMA (p, q), ajustado à série transformada dependente padronizada de Furnas, através da utilização da metodologia de Box e Jenkins (1976). Estes resultados constam em um dos relatórios do Grupo Coordenador para Operação do Sistema Interligado - GCOI (1987) e são apresentados, de forma resumida, na tabela 5.

TABELA 5. Estimativas de Máxima Verossimilhança para os Parâmetros do Modelo ARMA (1,1) Ajustado à Componentes Estocástica Dependente Padronizada do Aproveitamento de Furnas

PARÂMETROS	ESTIMATIVA	ERRO PADRAO
$\phi_1$	0,8492	0,0273
$\theta_1$	0,2510	0,0502
$\sigma_a^2$	0,4244	-

#### CONCLUSÕES

Tendo-se em vistas os resultados apresentados, as seguintes principais conclusões podem ser tiradas:

- i) O algoritmo utilizado mostrou-se amplamente viável sob o ponto de vista computacional. Isto porque os estágios I, II e III são nada mais nada menos do que aplicações de regressões e autotregressões lineares;
- ii) a série temporal analisada foi considerada consistente e homogênea, isto é, isenta de erros grosseiros e/ou sistemáticos. Não foi observado, também, a existência da componente de tendência a longo termo;

- iii) o modelo identificado no estágio II do algoritmo, ARMA (1,1), coincidiu com o obtido pela metodologia de Box e Jenkins. Entretanto, enquanto o primeiro é um método de identificação automática o segundo, não, isto é, a atuação e habilidade do analista são fundamentais na identificação correta do particular modelo;
- iv) em vista do item anterior, o método proposto é muito mais simples, onde praticamente é eliminada a retroação após o estágio de estimação;
- v) as estimativas finais para os parâmetros do modelo identificado foram obtidas após quatro iterações no estágio III do algoritmo. Entretanto, como realçado na introdução do trabalho, uma segunda iteração no estágio II seria suficiente para obtenção de melhores estimativas preliminares, visando com isso a diminuição do número final de iterações no último estágio.

#### REFERÊNCIAS

- BOX, G.E.P. & JENKINS, G.M. - Time Series Analysis - Forecasting and Control, Holden - Day, San Francisco, 1976.
- GCOI - Sistema Prevaz - Seleção de Modelos de Previsão de Vazões Médias Mensais para os Aproveitamentos Considerados no Programa de Operação, Relatório GCOI-SCEN-GTEH-01/87, 1987.
- HANNAN, E. J. & RISSANEN, J. - Recursive Estimation of Mixed Autorregressive - Moving Average Order, Biometrika, v.69, pp. 81-94, 1982. Correction, Biometrika, v.70, pp.303, 1983.
- HANNAN, E.J.; KAVALIERIS, L.; MACKISACK, M. - Recursive Estimation of Linear Systems, Biometrika, v.73, pp.1986.
- HARVEY, A.C. - The Econometric Analysis of Time Series, Philip Allan, Oxford, 1982.
- HOLANDA SALES, P.R. & VIEIRA, A.M. - Previsão de Vazões no Planejamento da Operação do Sistema Hidroelétrico Brasileiro, II Simpósio Luso-Brasileiro sobre Hidráulica e Recursos Hídricos, Lisboa, 1986.



UM PROCEDIMENTO LINEAR DE IDENTIFICAÇÃO PARA  
 MODELOS DE FUNÇÃO DE TRANSFERÊNCIA PARA  
 SÉRIES HIDROLÓGICAS TEMPORAIS

POR

P.R. HOLANDA SALES<sup>(1)</sup>, B.B. PEREIRA<sup>(2)</sup> e A.M. VIEIRA<sup>(1)</sup>

Resumo - - A aplicação da metodologia ARIMA (autorregressivo - integrado - médias móveis) de Box & Jenkins (1976) vem sendo, cada mais, utilizada em Hidrologia Estocástica para previsão e controle de séries hidrológicas temporais. Entretanto, situações há em que a série a ser prevista e/ou controlada pode ser dinamicamente relacionada a outras, chamadas de séries "exógenas". Nestes casos, o desempenho de um particular modelo pode ser melhorado pela inclusão de uma ou mais séries exógenas. A classe de modelos chamada de Função de Transferência é particularmente útil para formulação deste tipo de problema. Muito embora os modelos de Função de Transferência sejam, em geral, mais potentes que os modelos ARIMA, eles não têm sido largamente utilizados. Um dos maiores obstáculos reside no estágio de identificação do modelo. Os procedimentos tradicionais ou são empíricos e, necessitam de uma certa habilidade e conhecimento do fenômeno físico, ou se baseiam quase que integralmente, em um pré-branqueamento das séries de entrada e/ou saída e no estudo das correlações cruzadas (Box & Jenkins, 1976). Este último procedimento, apesar de ser simples para o caso de uma única série de entrada, é muito difícil de ser generalizado para múltiplas entradas.

Então, é apresentado neste artigo um algoritmo para identificação de Funções de Transferência, baseado, principalmente, no procedimento de estimação de mínimos quadrados lineares das séries originais ou filtradas. Como caso de aplicação foram utilizadas as séries de vazões naturais médias mensais em Três Marias e Sobradinho e a série de vazões incrementais médias mensais a este último reservatório. Os dados vão de janeiro de 1931 a dezembro de 1984 e foram obtidos do Banco de Dados Hidrometeorológicos (BDH) das Centrais Elétricas Brasileiras S.A. (ELETROBRÁS).

### INTRODUÇÃO

Seja uma série temporal discreta, representada por  $z_1, z_2, \dots$ ,

- 
1. Engenheiros, Departamento de Operação Energética, ELETROBRÁS, Rio de Janeiro, RJ
  2. Professor Titular, COPPE/Universidade Federal do Rio de Janeiro, RJ

$z_t, \dots, z_T$ , suposta não sazonal e estacionária. A família de modelos ARMA constitui um conjunto bastante flexível e de fácil compreensão na modelagem de tal série e, pode ser escrita como,

$$\phi(B)(z_t - \mu) = \theta(B)a_t \quad (1)$$

onde,

$\mu$  é a média teórica da série  $z_t$  ;  
 $B$  é o operador atraso, isto é,  $B^k z = z_{t-k}$  ;  
 $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  é o operador autorregressivo de ordem  $p$  (AR( $p$ )), tal que as raízes da equação característica  $\phi(B) = 0$  caíam fora do círculo unitário para assegurar a estacionaridade e  $\phi_i = 1, 2, \dots, p$  são os parâmetros autorregressivos ;

$\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  é o operador médias móveis de ordem  $q$  (MA( $q$ )), tal que as raízes da equação característica  $\theta(B) = 0$  caíam fora do círculo unitário para garantir a invertibilidade e  $\theta_i = 1, 2, \dots, q$  são os parâmetros médias móveis ;

$a_t$  é a série de resíduos, suposta ruído branco, com média zero e variância constante,  $\sigma_a^2$ , e é frequentemente considerada normalmente distribuída.

A notação ARMA ( $p, q$ ) é utilizada para representar um modelo ARMA com  $p$  parâmetros autorregressivos e  $q$  parâmetros médias móveis. Convém observar que este tipo de modelo é não identificável quando os operadores AR e MA possuem zeros em comum.

Na determinação de um particular modelo ARMA para uma dada série é recomendável que se siga os estágios de identificação, estimação e testes de verificação. Nas fases de identificação e estimação podem ser utilizados procedimentos lineares como sugerido por Holanda Sales et al. (1987). Nos testes de verificação as hipóteses de independência, normalidade e homocedasticidade dos resíduos devem ser verificadas. A hipótese de independência é crucial e sua violação pode causar drásticas consequências. Entretanto, heterocedasticidade e não normalidade dos resíduos podem, frequentemente, ser corrigidas através de uma adequada transformação da série original, tal como a proposta por Box & Cox (1964).

Para o caso específico de séries hidrológicas temporais, as quais apresentam, em geral, uma forte componente sazonal, procedimentos de dessazonalização e transformação como sugerido por Holanda Sales & Vieira (1986) podem ser utilizados a priori.

Considere agora uma situação onde, a partir do entendimento físico do fenômeno,  $N$  séries temporais de comprimento  $n$ ,  $x_{t,j}$ ,  $t = 1, 2, \dots, n$ ,  $j = 1, 2, \dots, N$  são supostas exógenas e a série  $y_t$ ,  $t = 1, 2, \dots, n$  é suposta endógena. O modelo de função de transferência-ruído que relaciona as séries  $x_{t,j}$  e  $y_t$  pode ser escrito como,

$$y_t = \mu_y = \sum_{j=1}^N v_j(B) (x_{t,j} - \mu_j) + \epsilon_t \quad (2)$$

onde,

$$v_j(B) = \frac{w_j(B)}{\delta_j(B)} B^{b_j} = v_{0,j} + v_{1,j} B + v_{2,j} B^2 + \dots \text{ é a}$$

função de transferência da  $j$ -ésima série com os pesos  $v_{0,j}, v_{1,j}, \dots$ , os quais são chamados de função impulso resposta;

$w_j(B) = w_{0,j} - w_{1,j} B - w_{2,j} B^2 - \dots - w_{s,j} B^s$  é o operador do numerador da  $j$ -ésima função de transferência, tal que as raízes de  $w_j(B)=0$  caíam fora do círculo unitário e,  $w_{i,j}, i=0,1,2,\dots,s$ , são os parâmetros de  $w_j(B)$ ;

$\delta_j(B) = 1 - \delta_{1,j} B - \delta_{2,j} B^2 - \dots - \delta_{r,j} B^r$  é o operador do denominador da função de transferência e, para garantir a estabilidade do sistema, as raízes de  $\delta_j(B)=0$  devem cair fora do círculo unitário e,  $\delta_{i,j}, i=1,2,\dots,r$  são os parâmetros de  $\delta_j(B)$ ;

$\mu_j$  é a média da  $j$ -ésima série de entrada;

$b_j$  é o atraso no tempo para  $x_{t,j}$  afetar  $y_t$ ;

$\mu_y$  é a média da série  $y_t$ ;

$\epsilon_t = \frac{\Theta(B)}{\Phi(B)} a_t$  é o ruído ARMA onde,  $a_t$  é uma sequência ruído branco, com média zero e variância  $\sigma_a^2$  e,  $\Theta(B)$  e  $\Phi(B)$  são, respectivamente, os operadores médios móveis e autorregressivo, como definido na equação (1)

Aqui, também, o modelo é não identificável se os operadores AR e MA possuírem zeros em comum e/ou o numerador e o denominador da função de transferência possuírem raízes comuns.

Como no caso de modelos ARMA, quando do ajustamento de um particular modelo de função de transferência-ruído para um dado conjunto de observações, é recomendável que se siga os estágios de identificação, estimação e testes de verificação do modelo. Entretanto, o maior problema na identificação de modelos de função de transferência reside em se encontrar estimativas apropriadas para a função de transferência,  $v_j(B)$ . Desde que  $v_j(B)$  é obtida, pode-se expressá-la de maneira simples na forma racional  $w_j(B)/\delta_j(B)$ .

#### FORMAS ALTERNATIVAS DE REPRESENTAÇÃO DE SISTEMAS DINÂMICOS LINEARES

Pela observação da equação (2), representativa de um modelo de Função de Transferência-Ruído (MFT), nota-se que alternativamente poder-se-ia pensar em um sistema dinâmico linear exógeno-ARMAX. Como realçado por Hannan (1987) qualquer Modelo de Função de Transferência pode ser escrito na forma de um Modelo ARMAX e

vice-versa. A forma preferível de representação de um sistema depende da particular aplicação que se venha a ter na prática. Assim é que, a equação (2) poderia ser encarada como representativa de uma situação em que se considera "erro nas variáveis" do modelo. Entretanto, pela consideração de "erro nas equações", ter-se-ia o Modelo ARMAX ou derivações deste.

As diversas formas de representação de um sistema dinâmico linear podem ser melhor compreendidas se se considera uma entrada e uma saída e, ainda, sendo estas estacionárias. Para tanto, sejam  $A(B)$ ,  $B(B)$ ,  $C(B)$  e  $D(B)$ , polinômios em  $B$  dos graus  $p$ ,  $r$ ,  $s$ , e  $q$ , respectivamente;  $\{x_t\}$ , a entrada determinística do sistema ou "ruído-livre" e,  $\{y_t\}$ , o ruído observado de saída. Seja, também,  $\{a_t\}$ , uma sequência de variáveis aleatórias independentes com média zero e variância  $\sigma_a^2$ , chamada de inovações.

Então, as equações a seguir, tentam melhor explicar a idéia exposta:

(i) Modelo de Função de Transferência - MFT

$$y_t = \frac{B(B)}{A(B)} x_t + \frac{D(B)}{C(B)} a_t \quad (3)$$

(ii) Modelo Autorregressivo - Médias Móveis Exógeno - ARMAX

$$A(B) y_t = B(B) x_t + D(B) a_t \quad (4)$$

(iii) Ajustamento Dinâmico com Ruído Autorregressivo

$$A(B) y_t = B(B) x_t + \frac{1}{C(B)} a_t \quad (5)$$

(iv) Ajustamento Dinâmico com Ruído ARMA

$$A(B) y_t = B(B) x_t + \frac{D(B)}{C(B)} a_t \quad (6)$$

A equação (3) - MFT - poderia, alternativamente, ser representada, sob forma de equações relacionadas, como a seguir:

$$A(B) u_t = B(B) x_t \quad (7)$$

$$C(B) v_t = D(B) a_t \quad (8)$$

$$y_t = u_t + v_t \quad (9)$$

onde,  $\{u_t\}$  é a saída hipotética do sistema, chamada de "ruído-livre" e  $\{v_t\}$  é o ruído aditivo, o qual é "gerado" a partir do ruído branco,  $\{a_t\}$ , por um modelo ARMA ( $p, q$ ).

Neste caso, torna-se mais natural considerar que o ruído ou erro  $\{v_t\}$  é adicionado à variável determinística de saída,  $\{u_t\}$ , que à equação do sistema. Por esta razão, o modelo expresso pela equação (3) é chamado de modelo com "erro nas variáveis".

Por outro lado, os modelos expressos pelas equações (4), (5) e (6) são nada mais nada menos que representações naturais de "erros nas equações". Isto porque, o ruído ou erro,  $\{v_t\}$ , afeta diretamente, de uma forma aditiva, as equações do sistema, isto é,

$$A(B) y_t = B(B) x_t + v_t \quad (10)$$

onde,  $\{v_t\}$ , é definido como um ruído autorregressivo (AR), médias móveis (MA) ou autorregressivo-médias móveis (ARMA), conforme o caso.

Em termos prático, pode-se dizer que quando a variável endógena, ao instante de tempo  $t$ , é afetada por estimativas (previsões) das variáveis exógenas, nos instantes  $t, t+1, \dots$ , então é preferível adotar-se um modelo de representação do tipo Função de Transferência-Ruído, ou seja, um modelo com erro nas variáveis. Caso contrário, deve-se adotar um modelo com erro nas equações.

No caso específico de séries físicas, como as séries hidrológicas temporais de vazões naturais médias mensais, é preferível uma representação do tipo ARMAX, ou seja, de erro nas equações. Isto porque, considerando o objetivo específico de previsões, trabalha-se, usualmente, com vazões já observadas na seleção de um particular modelo, isto é, são com estas medições que as previsões são comparadas. Então, o que apenas se admite é que as duas fontes de erro - inadequabilidade do modelo e erro de medição - são computadas juntas na equação de observação.

#### PROCEDIMENTO LINEAR PARA MODELOS ARMAX

Sejam,  $\{z_t\}$ ,  $\{x_t\}$  e  $\{y_t\}$ , respectivamente, as séries endógena e exógenas. Admita-se, também, que elas possam ser relacionadas através do seguinte modelo ARMAX

$$C_p(B) z_t = D_r(B) x_t + E_s(B) y_t + F_q(B) a_t \quad (11)$$

onde,

$$\begin{aligned} C(B) &= 1 - c_1 B - \dots - c_p B^p \\ D(B) &= d_0 - d_1 B - \dots - d_r B^r \\ E(B) &= e_0 - e_1 B - \dots - e_s B^s \\ F(B) &= 1 - f_1 B - \dots - f_q B^q \end{aligned}$$

Ponha-se agora como objetivo a estimação das ordens  $p, r, s$  e  $q$ , bem como do vetor de coeficiente do sistema,  $\underline{\beta} = (\underline{c}, \underline{d}, \underline{e}, \underline{f})$  e da variância,  $\sigma_a^2$ , dos resíduos.

A série de resíduos,  $\{a_t\}$ , pode ser obtida a partir da equação (11), através de

$$a_{t,o} = F_{q,o}^{-1}(B) [C_{p,o}(B) z_t - D_{r,o}(B) x_t - E_{s,o}(B) y_t] \quad (12)$$

e, as séries auxiliares,  $\{m_t\}$ ,  $\{n_t\}$ ,  $\{u_t\}$  e  $\{v_t\}$  são, então, da

das por:

$$m_t = -\frac{\partial a_t}{\partial c_i} \beta_o = F_{q,o}^{-1} (B) z_t = m_{t-1} \quad (13)$$

$$n_t = -\frac{\partial a_t}{\partial d_i} \beta_o = F_{q,o}^{-1} (B) x_t = n_{t-1} \quad (14)$$

$$u_t = -\frac{\partial a_t}{\partial e_i} \beta_o = F_{q,o}^{-1} (B) y_t = u_{t-1} \quad (15)$$

$$v_t = -\frac{\partial a_t}{\partial f_i} \beta_o = F_{q,o}^{-1} (B) a_t \quad (16)$$

onde,  $\beta_o = (c_o, d_o, e_o, f_o)$  é uma estimativa inicial para o vetor dos parâmetros.

A série  $\{a_t\}$ , bem como as auxiliares  $\{m_t\}$ ,  $\{n_t\}$ ,  $\{u_t\}$  e  $\{v_t\}$  podem ser obtidas recursivamente, fazendo-se os valores iniciais iguais as suas esperanças não condicionais, isto é, iguais a zero, da seguinte forma,

$$a_{t,o} = z_t - c_{1,o} z_{t-1} - \dots - c_{p,o} z_{t-p} - d_{o,o} x_t + \dots + d_{r,o} x_{t-r} + e_{o,o} y_t + e_{1,o} y_{t-1} + \dots + e_{s,o} y_{t-s} + f_{1,o} a_{t-1} + \dots + f_{q,o} a_{t-q} \quad (17)$$

$$m_t = f_{1,o} m_{t-1} + \dots + f_{q,o} m_{t-q} + z_t \quad (18)$$

$$n_t = f_{1,o} n_{t-1} + \dots + f_{q,o} n_{t-q} - x_t \quad (19)$$

$$u_t = f_{1,o} u_{t-1} + \dots + f_{q,o} u_{t-q} - y_t \quad (20)$$

$$v_t = f_{1,o} v_{t-1} + \dots + f_{q,o} v_{t-q} - a_t \quad (21)$$

Por outro lado, seja  $\langle a_t | \zeta, z_t, y_t \rangle$  o valor esperado condicional de  $\{a_t\}$  dados o vetor de parâmetros  $\zeta$  e  $z_t, x_t$  e  $y_t$ . Expandindo  $\langle a_t | \zeta, z_t, x_t, y_t \rangle$  em série de Taylor no entorno  $t$  de  $\zeta'_o = (\zeta_{1,o}, \zeta_{2,o}, \dots, \zeta_{k,o})$ , tem-se, aproximadamente,

$$\langle a_t | \zeta, z_t, x_t, y_t \rangle = \langle a_{t,o} | \zeta_o, z_t, x_t, y_t \rangle - \sum_{i=1}^k (\zeta_i - \zeta_{i,o}) \eta_{i,t} \quad (22)$$

onde,

$$\eta_{i,t} = \frac{\delta \langle a_t | \zeta, z_t, x_t, y_t \rangle}{\delta \zeta_i} \quad \zeta = \zeta_o$$

Correspondendo a (22), a equação de regressão linear aproximada é dada por,

$$a_{t,0} = \sum_{i=1}^p (c_i - c_{i,0}) m_{t-i} + \sum_{i=1}^r (d_i - d_{i,0}) n_{t-i} + \sum_{i=1}^s (e_i - e_{i,0}) u_{t-i} + \sum_{i=1}^q (f_i - f_{i,0}) v_{t-i} + a_t \quad (23)$$

Logo, os ajustamentos são os coeficientes da regressão de  $a_{t,0}$  em  $m_{t-i}$ ,  $n_{t-i}$ ,  $u_{t-i}$  e  $v_{t-i}$ .

Pela adição dos valores ajustados às primeiras estimativas,  $\beta_0 = (c_0, d_0, e_0, f_0)$ , um segundo conjunto de estimativas é formado. O processo deve, então, continuar até que a convergência final seja obtida.

Assim, um algoritmo linear para identificação das ordens e estimação dos parâmetros do modelo ARMAX, definido pela equação (11) pode ser sumarizado como se segue:

#### Estágio I - Preliminar

Ajuste, por mínimos quadrados, um modelo de regressão/autorregressão do tipo

$$\sum_{i=0}^h \hat{c}_{1i} z_{t-i} = \sum_{i=1}^h \hat{d}_{1i} x_{t-i} + \sum_{i=1}^h \hat{e}_{1i} y_{t-i} + \hat{a}_{1t} \quad (24)$$

com,

$$\sigma_h^2 = \frac{1}{T} \sum_{t=1}^T a_{1t}^2 \quad (25)$$

Selecione  $h$ , tal que

$$\text{BIC}(h) = \ln \sigma_h^2 + 3h \frac{\ln T}{T} \quad (26)$$

seja mínimo.

#### Estágio II - Identificação e Estimação Preliminar

Para cada  $p$ ,  $r$ ,  $s$  e  $q$ , obtenha a regressão de  $z_t$  em

$$- z_{t-i} \quad (i = \overline{1, p})$$

$$x_{t-i} \quad (i = \overline{1, r})$$

$$y_{t-i} \quad (i = \overline{1, s})$$

$$\hat{a}_{1_{t-i}} \quad (i = \overline{1, q})$$

Seja,  $\underline{\beta}$  um vetor cujos  $k = p + r + s + q$  elementos são os parâmetros do modelo, selecione, então,  $k$  de tal maneira que minimize a expressão,

$$BIC(k) = \ln \sigma_k^2 + k \frac{\ln T}{T} \quad (27)$$

### Estágio III - Estimação Final

Com a segunda estimativa dos resíduos, obtida no estágio anterior, isto é,

$$\hat{a}_{2_t} = \sum_{i=1}^p \hat{c}_{2_i} z_{t-i} - \sum_{i=1}^r \hat{d}_{2_i} x_{t-i} - \sum_{i=1}^s \hat{e}_{2_i} y_{t-i} - \sum_{i=1}^q \hat{f}_{2_i} \hat{a}_{2_t} \quad (28)$$

onde,

$$\hat{a}_{2_t} = z_t = x_t = y_t = 0 \quad (t \leq 0)$$

e, com as séries auxiliares dadas por (18), (19), (20) e (21), aplique a equação (23) no sentido de se obter uma nova estimativa para os parâmetros.

Este procedimento deve ser repetido, até que a convergência final seja obtida. Como ressaltado por Hannan et al. (1986), as estimativas finais são assintoticamente eficientes, ou seja, os estimadores possuem as mesmas distribuições assintóticas que as dos estimadores de máxima verossimilhança, quando as restrições normais são impostas.

### APLICAÇÃO

Como exemplo de aplicação do algoritmo proposto foi utilizado séries de vazões naturais médias mensais da bacia do rio São Francisco. Esta bacia tem uma área de drenagem de 641.000 km<sup>2</sup> e um estirão ao longo do curso principal de 2.700 km. O sistema estudado neste trabalho compreende as séries de vazões naturais médias mensais dos aproveitamentos de Três Marias e Sobradinho. Devido ao grande estirão fluvial sem controle existente entre estes dois aproveitamentos e considerando que a bacia intermediária é o segmento que recebe os maiores afluentes, a série de vazões naturais médias mensais observadas neste trecho da bacia, foi também considerada. O esquema da figura 1 caracteriza o sistema estudado. Os valores entre parenteses que aparecem na figura 1, são as distâncias em km até a foz.



As principais características dos aproveitamentos considerados podem ser vistas na tabela 1.

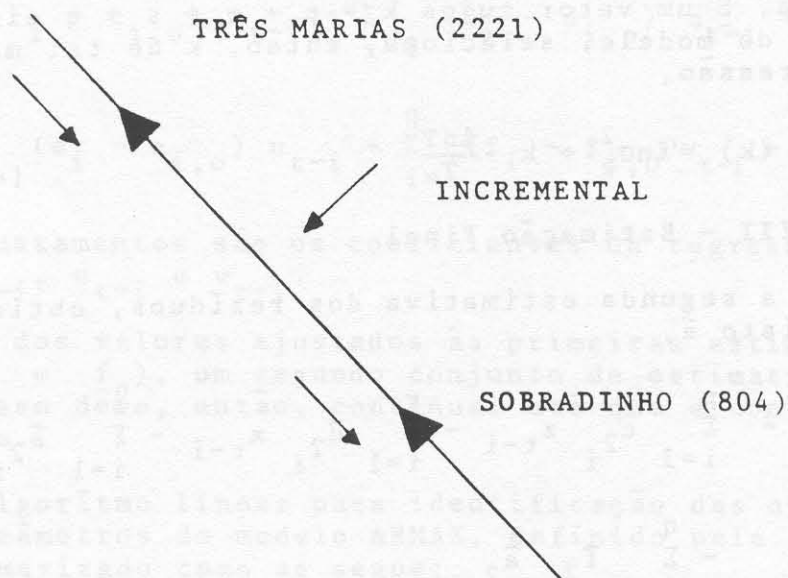


Figura 1. Esquema da Bacia do Rio São Francisco

Tabela 1. Principais Características dos Reservatórios

CARACTERÍSTICAS	TRÊS MARIAS	SOBRADINHO
Área de Drenagem (km <sup>2</sup> )	50.560	498.425
Nível Máximo Normal (m)	568,30	392,50
Nível Mínimo Operativo (m)	545,00	380,50
Nível Máximo Maximorum (m)	569,20	393,50
Nível Máximo de Segurança (m)	-	397,30
Volume Morto (km <sup>3</sup> )	4,250	5,447
Volume Útil (km <sup>3</sup> )	15,278	28,669
Capacidade do Vertedor de Superfície (m <sup>3</sup> /s)	8.700	6.300
Capacidade de Vertedor de Fundo (m <sup>3</sup> /s)	-	16.260
Descarga de Restrição para Cálculo de Volume de Espera (m <sup>3</sup> /s)	2.500	8.000

Na aplicação então do algoritmo à bacia do rio São Francisco, a série de vazões naturais médias mensais em Sobradinho é a série endógena e as séries de vazões naturais médias mensais em Três Marias e na bacia incremental são as séries exógenas.

O período de observação das séries é de janeiro de 1931 a dezembro de 1984. A tabela 2 resume as principais estatísticas mensais das séries analisadas.

Tabela 2. Estatísticas Correspondentes às Séries de Vazões Naturais Médias Mensais dos Aproveitamentos de Três Marias e Sobradinho e da Bacia Intermediária.

LOCAL	MÊS	MÉDIA (m <sup>3</sup> /s)	MÁXIMO (m <sup>3</sup> /s)	MÍNIMO (m <sup>3</sup> /s)	DESVIO PADRÃO (m <sup>3</sup> /s)	COEFICIENTE DE ASSIMETRIA
TRÊS MARIAS	JAN	1333	3249	259	623	1,060
	FEV	1275	3859	199	727	1,620
	MAR	1054	2949	330	553	1,768
	ABR	704	1649	201	338	1,182
	MAI	423	817	129	157	0,862
	JUN	313	580	157	105	0,907
	JUL	257	454	115	83,1	0,852
	AGO	207	348	92	66,0	0,847
	SET	194	368	92	65,3	1,097
	OUT	257	586	103	85,9	1,466
	NOV	535	1119	234	221	1,126
	DEZ	1028	2289	137	454	0,915
SOBRADINHO	JAN	4353	7846	1722	1288	0,682
	FEV	4798	10073	1467	1946	1,183
	MAR	4632	15364	1693	2488	2,452
	ABR	3687	7959	1432	1704	1,135
	MAI	2337	8760	1092	1233	3,198
	JUN	1576	3936	950	541	2,185
	JUL	1298	2588	851	368	1,667
	AGO	1112	2075	747	294	1,625
	SET	972	1700	671	258	1,616
	OUT	1078	1975	640	277	1,127
	NOV	1759	3264	876	595	0,943
	DEZ	3222	6257	1127	1107	0,726
INCREMENTAL	JAN	3148	5918	1406	974	0,842
	FEV	3460	8311	1238	1518	1,370
	MAR	3446	12514	1320	2088	2,614
	ABR	2797	6626	1165	1355	1,206
	MAI	1771	7639	760	1054	3,723
	JUN	1208	3236	806	440	2,478
	JUL	1013	2069	623	292	1,791
	AGO	877	1673	629	235	1,749
	SET	768	1379	572	208	1,714
	OUT	847	1645	504	230	1,312
	NOV	1354	2616	631	517	1,063
	DEZ	2422	4837	926	879	0,703

A transformação de Box & Cox (1964) foi aplicada a todas as séries, podendo ser vista na tabela 3 as transformações selecionadas. Nesta tabela pode-se notar que em todas as séries a transformação indicada foi do tipo logaritmo natural.

Tabela 3. Transformações de Box & Cox Seleccionadas para as Séries de Vazões Naturais Médias Mensais dos Aproveitamentos de Três Marias e Sobradinho e da Bacia Intermediária.

SÉRIE	$\lambda_1$	$\lambda_2$
TRÊS MARIAS	0	-90
SOBRADINHO	0	-529
INCREMENTAL	0	-468

Como as médias e variâncias mensais das séries transformadas eram periódicas, foi feita uma padronização mensal das mesmas, no sentido de se obter estacionaridade de segunda ordem. Das séries resultantes, obteve-se outras com média zero e variância unitária. A seguir, procedeu-se a aplicação do algoritmo proposto, onde  $\{z_t\}$  identifica a série endógena (Sobradinho),  $\{x_t\}$  e  $\{y_t\}$  as séries exógenas, respectivamente, Três Marias e Incremental.

No estágio I foram ajustados vários modelos de regressão/autorregressão a partir da aplicação das equações (24) e (25), considerando diversos valores para  $h$ . O critério BIC ( $h$ ), como mostrado na equação (26), indicou  $h=2$ . A tabela 4, a seguir, sumariza os principais resultados do estágio I.

Tabela 4. Modelos de "Regressão/Autorregressão" Ajustados as Séries Dependentes Padronizadas Transformadas.

VARIÁVEL DEPENDENTE	$h$	VARIÁVEIS NA EQUAÇÃO	VARIÂNCIA DOS RESÍDUOS	ÍNDICE DE AKAIKE BIC ( $h$ )
$z_t$	1	$z_{t-1}, x_{t-1}, y_{t-1}$	0,32158	-1,10454
$z_t$	2	$z_{t-1}, z_{t-2}, x_{t-1}, x_{t-2}, y_{t-1}, y_{t-2}$	0,29349	-1,16597
$z_t$	3	$z_{t-1}, z_{t-2}, z_{t-3}, x_{t-1}, x_{t-2}, x_{t-3}, y_{t-1}, y_{t-2}, y_{t-3}$	0,29142	-1,14307

A seguir, passou-se ao estágio II, onde foram ajustadas "Regressões/Autorregressões" às séries  $\{z_t\}$ ,  $\{x_t\}$ ,  $\{y_t\}$  e à série de resíduos  $\{a_t\}$  obtida no estágio anterior. A tabela 5 apresenta os valores do critério BIC ( $k$ ), bem como as variâncias dos resíduos (valores entre parênteses) das "regressões/autorregressões" considerados no estágio II.

Tabela 5. Valores de BIC (k),  $k = p+r+s+q$ , a partir do ajustamento de "Regressões/Autorregressões" lineares às Séries Dependentes Padronizadas Transformadas e à Série de resíduos obtida no estágio I.

p \ q		0			1			2				
		r \ s	0	1	2	r \ s	0	1	2	r \ s	0	1
0	0		-0.6675 (0.5079)	-0.6708 (0.5012)		-0.2058 (0.8059)	-0.6577 (0.5078)	-0.6939 (0.4849)		-0.3147 (0.7156)	-0.6463 (0.5085)	-0.688 (0.4829)
	1	-0.9308 (0.3903)	-1.1122 (0.3223)	-1.1115 (0.3194)		-0.9913 (0.3637)	-1.1051 (0.3214)	-1.1418 (0.3067)		-1.0258 (0.3479)	-1.0951 (0.3214)	-1.1307 (0.3071)
	2	-0.9195 (0.3909)	-1.1442 (0.3091)	-1.1710 (0.2979)		-0.9938 (0.3592)	-1.1596 (0.3013)	-1.1600 (0.2982)		-1.0183 (0.3471)	-1.1486 (0.3016)	-1.1505 (0.2981)
1	0	-0.8599 (0.4190)	-0.9260 (0.3883)	-0.9159 (0.3883)		-0.8522 (0.4180)	-0.9525 (0.3744)	-1.0061 (0.3513)		-0.8442 (0.4172)	-0.9412 (0.3749)	-0.9976 (0.3508)
	1	-1.0940 (0.3283)	-1.1045 (0.3216)	-1.1081 (0.3173)		-1.0954 (0.3245)	-1.0949 (0.3215)	-1.1304 (0.3072)		-1.0864 (0.3242)	-1.0850 (0.3214)	-1.1631 (0.2943)
	2	-1.1557 (0.3055)	-1.1448 (0.3058)	-1.1723 (0.2946)		-1.1819 (0.2947)	-1.1709 (0.2950)	-1.1649 (0.2938)		-1.1740 (0.2941)	-1.1631 (0.2943)	-1.1585 (0.2928)
2	0	-0.8485 (0.4196)	-0.9155 (0.3885)	-1.0045 (0.3519)		-0.8498 (0.4149)	-0.9730 (0.3632)	-1.0355 (0.3377)		-0.8437 (0.4133)	-0.9730 (0.3632)	-1.0567 (0.3774)
	1	-1.0831 (0.3286)	-1.0980 (0.3205)	-1.1446 (0.3028)		-1.1137 (0.3155)	-1.1074 (0.3143)	-1.1454 (0.2996)		-1.1030 (0.3157)	-1.0965 (0.3146)	-1.1484 (0.2957)
	2	-1.1762 (0.2964)	-1.1654 (0.2966)	-1.1660 (0.2935)		-1.1732 (0.2943)	-1.1623 (0.2946)	-1.1570 (0.2932)		-1.1652 (0.2937)	-1.1545 (0.2939)	-1.1580 (0.2899)

No estágio II ficou identificado o modelo ARMAX com as seguintes ordens:  $p=1$ ,  $r=2$ ,  $s=0$  e  $q=1$ . As estimativas iniciais dos parâmetros obtidas neste estágio indicaram  $\hat{c}_1 = 0,80752$ ,  $\hat{d}_1 = 0,62279$ ,  $\hat{d}_2 = -0,52025$ ,  $\hat{f}_1 = -0,41308$  e  $\hat{\sigma}_a^2 = 0,29469$ .

A partir das estimativas iniciais do estágio II do algoritmo, passou-se ao estágio III. As estimativas finais foram obtidas após duas iterações, como sumarizado na tabela 6. Entre parênteses estão assinalados os erros padrão das estimativas.

Tabela 6. Estimativas Finais dos Parâmetros  $\hat{c}_1$ ,  $\hat{d}_1$ ,  $\hat{d}_2$ ,  $\hat{f}_1$  e  $\hat{\sigma}_a^2$  do Modelo ARMAX ajustado às Séries Dependentes Padronizadas Transformadas.

ITERAÇÃO	$\hat{c}_1$	$\hat{d}_1$	$\hat{d}_2$	$\hat{f}_1$	$\hat{\sigma}_a^2$
1	0,81914 (0,06693)	0,63267 (0,03821)	-0,53639 (0,06394)	-0,43626 (0,08272)	0,29326
2	0,82612 (0,06704)	0,63235 (0,03806)	-0,54163 (0,06382)	-0,44517 (0,08266)	0,29274

#### CONCLUSÕES

A partir dos resultados apresentados no caso de aplicação, as seguintes conclusões podem ser tiradas:

- i) O algoritmo proposto, por ser calcado em regressões/autor regressões lineares, mostrou-se viável sob o ponto de vista computacional;
- ii) As séries temporais analisadas foram consideradas consistentes e homogêneas, ou seja, isentas de erros grosseiros e/ou sistemáticos;
- iii) A partir da dimensão das séries: 54a x 12m. Não foi possível observar, em nenhuma delas, a existência de componente de tendência a longo termo e/ou mudança de nível;
- iv) As séries analisadas foram consideradas estacionárias de 2ª ordem após uma transformação de Box & Cox, do tipo logaritmo natural, seguida de uma padronização das mesmas;
- v) Tendo-se em vista que o critério de Akaike,  $BIC(k)$ ,  $k=p+r+s+q$ , engloba, em seu bojo, o critério da parcimônia, a série exógena  $\{y_t\}$ , correspondente às vazões naturais incrementais à Sobradinho, foi considerada não relevante para pertencer a estrutura final do modelo ARMAX ajustado;
- vi) O fato de se ter, no modelo final,  $p=1$ ,  $r=2$  e  $q=1$ , aliado ao tempo de viagem, entre Três Marias e Sobradinho, ratifica a conclusão anterior;
- vii) O método proposto parece ser muito mais simples, pois prati

camente elimina a retroação após o estágio de estimação.

- viii) Em vista do item anterior, a intervenção do analista ao longo do processo é bastante minimizada;
- ix) Em outras palavras, o algoritmo proposto fornece, também, um método seguro de identificação para modelos ARMAX;
- x) A convergência final das estimativas no último estágio do algoritmo, foi obtida após somente duas iterações. Tal fato corrobora, sobretudo, as afirmações anteriores acerca da eficiência do método proposto.

#### REFERÊNCIAS

- BOX, G.E.P. & COX, D.R. - An Analysis of Transformations, Journal of Royal Statistical Society, Series B, v.26, 1964.
- BOX, G.E.P. & JENKINS, G.M. - Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, 1976.
- HANNAN, E.J.; KAVALIENS, L. and MACKISACK, M. - Recursive Estimation in Linear Systems. Biometrika, v.73, pp.119-133, 1986.
- HANNAN, E.J. - Approximation of Linear Systems, In Time Series and Econometric Modelling, Ed.I.B.MacNeill, pp.1-12, 1987.
- HOLANDA SALES, P.R. & VIEIRA, A.M. - Previsão de Vazões no Planejamento da Operação do Sistema Hidroelétrico Brasileiro. II Simpósio Luso-Brasileiro sobre Hidráulica e Recursos Hídricos, Lisboa, Fevereiro/1986.
- HOLANDA SALES, P.R., PEREIRA, B.B., VIEIRA, A.M. - Procedimentos Lineares de Identificação e Estimação de Modelos ARMA para Séries Hidrológicas Temporais. VII Simpósio Brasileiro de Hidrologia e Recursos Hídricos e III Simpósio Luso-Brasileiro de Hidráulica e Recursos Hídricos, Salvador, Novembro/1987.

UM PROCEDIMENTO LINEAR DE IDENTIFICAÇÃO E ESTIMAÇÃO  
PARA MODELOS SURARMA APLICADOS A SÉRIES HIDROLÓGICAS  
TEMPORAIS MULTIVARIADAS

POR

P.R. HOLANDA SALES<sup>(1)</sup>, B.B. PEREIRA<sup>(2)</sup> e A.M. VIEIRA<sup>(1)</sup>

**RESUMO** - - Como já realçado por diversos autores, cada vez mais vêm sendo empregados, na prática, modelos multivariados para séries temporais. Basicamente, isto é devido ao grande sucesso de técnicas univariadas as quais encorajam, por sua vez, a extensão da metodologia de forma a incluir no modelo outras séries relacionadas, bem como o avanço tecnológico a partir do surgimento de computadores cada vez mais potentes. A sub-família dos modelos multivariados ARMA, chamada de SURARMA, ou seja, modelos ARMA para séries temporais aparentemente não relacionadas, vem despertando particular interesse. Isto se prende ao fato de que dado  $m$  séries univariadas, os "innovations"  $a_{i_t}$  ( $1 \leq i \leq m$ ) em  $\phi_i(B)z_{i_t} = \theta_i(B)a_{i_t}$ , podem ser contemporaneamente relacionados.

Então, o presente artigo visa o desenvolvimento de um algoritmo linear para identificação e estimação dos parâmetros da subclasse de modelos SURARMA para séries hidrológicas multivariadas temporais. Basicamente, em um primeiro estágio, são ajustados, através de procedimentos lineares, modelos univariados ARMA a cada série temporal componente do sistema. As estimativas conjuntas de máxima verossimilhança, condicionada a matriz de variância - covariância dos resíduos é, como se sabe, um problema de mínimos quadrados não linear generalizado. Entretanto, pela expansão dos resíduos em séries de Taylor, as quais utilizam séries auxiliares obtidas de derivadas parciais dos resíduos em relação aos parâmetros e, pela utilização de mínimos quadrados lineares generalizados um procedimento iterativos é utilizado a partir das estimativas iniciais univariadas. O emprego de mínimos quadrados generalizado prende-se ao fato de que os resíduos são contemporaneamente correlacionados. Como caso de aplicação foram utilizadas as séries de vazões naturais médias mensais afluentes dos reservatórios de Furnas, Itumbiara e Ilha Solteira. Os dados compreendem o período de janeiro de 1931 a dezembro de 1984.

### INTRODUÇÃO

Seja um conjunto de  $m$  séries temporais, representadas pelo vetor,

$$\tilde{z}_t = (z_{1_t}, z_{2_t}, \dots, z_{m_t})'$$

- 
1. Engenheiros, Departamento de Operação Energética, ELETROBRÁS, Rio de Janeiro, RJ
  2. Professor Adjunto, COPPE/Universidade Federal do Rio de Janeiro, RJ

seja, também,

$$\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_m)$$

o vetor das médias de  $z_t$ .

Admita-se, agora, que as  $m$  séries de tempo possam ser modeladas a partir de uma classe especial de modelos para séries temporais multivariadas na qual os resíduos  $a_{i_t}$  dos  $m$  modelos ARMA, sejam correlacionados contemporaneamente mas não serialmente, isto é,

$$\phi_i(B) (z_{i_t} - \mu_i) = \theta_i(B) a_{i_t}, \quad i = \overline{1, m} \quad (1)$$

onde,

$\phi_i(B) = 1 - \phi_{i_1} B - \dots - \phi_{i_{p_i}} B^{p_i}$  é o operador autorregressivo, de ordem  $p_i$ , da série  $z_i$ ;

$\theta_i(B) = 1 - \theta_{i_1} B - \dots - \theta_{i_{q_i}} B^{q_i}$  é o operador médias móveis, de ordem  $q_i$ , da série  $z_i$ ;

$\mu_i$  = média de série  $z_i$ ;

$\underline{a}_t = (a_{1_t}, \dots, a_{m_t})' \sim \text{NID}(0, \Omega)$

$\Omega = \sigma_{g_i}$  é a matriz de variância-covariância contemporânea de  $\underline{a}_t$  que neste caso é definida como  $\langle a_t, a'_t \rangle = \Omega$ , se  $t = \tau$  e igual a zero, caso contrário, com  $\langle : \rangle$  representando valor esperado.

O modelo definido pela equação (1) é chamado de contemporâneo ARMA ( $p, q$ ) ou CARMA ( $p, q$ ) onde,  $p = \max(p_1, p_2, \dots, p_m)$  e  $q = \max(q_1, q_2, \dots, q_m)$ .

Admite-se, ainda, neste caso, que as raízes das equações características  $\phi_i(B) = 0$  e  $\theta_i(B) = 0$ ,  $i = 1, 2, \dots, m$  caiam fora do círculo unitário para que sejam garantidas as condições de estacionaridade e invertibilidade dos modelos.

O modelo CARMA pode, por outro lado, ser encarado como um conjunto de  $m$  modelos univariados ARMA onde os "innovations" são relacionados contemporaneamente. Entretanto, para o caso em que  $\sigma_{g_i} = 0$ , para  $g \neq i$ , o modelo se reduz a um conjunto de  $m$  modelos univariados ARMA ( $p_i, q_i$ ) como definidos por Box & Jenkins (3). Por outro lado, as matrizes dos parâmetros do modelo CARMA ( $p, q$ ) são diagonais, isto é, são do tipo  $\phi(B) = \text{diag}\{\phi_1(B), \dots, \phi_m(B)\}$  e  $\theta(B) = \text{diag}\{\theta_1(B), \dots, \theta_m(B)\}$ . Caso contrário, se elas fossem triangulares, ou mesmo completas, então, os modelos apropriados seriam o de Função de Transferência, no primeiro caso, ou



o multivariado, no segundo. Por estas razões, a classe de modelos CARMA pode ser vista como uma extensão natural do caso univariado e como um caso particular do multivariado de acordo com Hipel (1985), Camacho et al. (1985) e Salas et al. (1985). Uma outra interpretação para o modelo CARMA seria a de um conjunto de  $m$  modelos ARMA aparentemente não relacionados. Daí, também, a denominação, na literatura especializada, de modelos SURARMA, segundo Moriarty & Salamon (1980), Harvey (1982) e Umashankar & Ledolter (1983).

No caso específico do modelo SURARMA ser considerado como uma classe particular de um modelo geral multivariado, a equação (1) pode ser escrita como,

$$\begin{aligned} \tilde{z}_t &= \begin{bmatrix} \phi_{11} & & & \\ & \phi_{21} & & \\ & & \ddots & \\ & & & \phi_{m1} \end{bmatrix} \tilde{z}_{t-1} - \dots - \begin{bmatrix} \phi_{1p} & & & \\ & \phi_{2p} & & \\ & & \ddots & \\ & & & \phi_{mp} \end{bmatrix} \tilde{z}_{t-p} = \\ &= \underline{a}_t - \begin{bmatrix} \theta_{11} & & & \\ & \theta_{21} & & \\ & & \ddots & \\ & & & \theta_{m1} \end{bmatrix} \underline{a}_{t-1} - \dots - \begin{bmatrix} \theta_{1q} & & & \\ & \theta_{2q} & & \\ & & \ddots & \\ & & & \theta_{mq} \end{bmatrix} \underline{a}_{t-q} \quad (2) \end{aligned}$$

onde,

$$\tilde{z}_t = z_t - \mu$$

Nos modelos contemporâneos, a diagonalização das matrizes dos parâmetros permite que o modelo seja decomposto em suas equações componentes. Então, os parâmetros do modelo não precisam ser estimados conjuntamente e, por este motivo, procedimentos univariados podem ser empregados segundo Harvey (1982). Uma vez estimados os parâmetros do modelo, os resíduos de cada componente podem ser recuperados de tal forma que se obtenha "innovations" contemporaneamente correlacionadas. A partir daí, a matriz de variância-covariância pode ser estimada e parte-se para a obtenção de novos resíduos não correlacionados, ou seja, ruído branco. Estes últimos devem ser utilizados, então, na verificação do modelo.

IDENTIFICAÇÃO E ESTIMAÇÃO LINEAR

Considere o sistema de  $m$  equações ARMA ( $p_i, q_i$ ), como

definido pela equação (1), isto é,

$$\phi_i(B) \bar{z}_{i_t} = \theta_i(B) a_{i_t}, \quad 1 \leq i \leq m \quad (3)$$

onde,

$$\bar{z}_{i_t} = z_{i_t} - \mu_i$$

Seja  $\beta_i$  um vetor cujos  $k_i = p_i + q_i$  elementos são os parâmetros autorregressivos e média móvels,  $\phi_i$  e  $\theta_i$ , ou seja,  $\beta_i = (\phi_i, \theta_i)$ . Seja, também,  $\beta_{i_0} = (\phi_{i_0}, \theta_{i_0})$  uma estimativa inicial para  $\beta_i$ .

Então, para cada equação  $i$  ( $1 \leq i \leq m$ ) do sistema, tem-se:

$$a_{i_{t,0}} = \phi_{i_{p_{i,0}}}^{-1} (B) \theta_{i_{q_{i,0}}} (B) \bar{z}_{i_t} \quad (4)$$

e,

$$-\frac{\partial a_{i_t}}{\partial \phi_j} \Big|_{\beta_{i_0}} = u_{i_{t-j}} \quad (5)$$

$$-\frac{\partial a_{i_t}}{\partial \theta_j} \Big|_{\beta_{i_0}} = v_{i_{t-j}} \quad (6)$$

onde,

$$u_{i_t} = \theta_{i_{q_{i,0}}}^{-1} (B) \bar{z}_{i_t} = \phi_{i_{p_{i,0}}}^{-1} (B) a_{i_{t,0}} \quad (7)$$

$$v_{i_t} = -\theta_{i_{q_{i,0}}}^{-2} (B) \phi_{i_{p_{i,0}}} (B) \bar{z}_{i_t} = -\theta_{i_{q_{i,0}}}^{-1} (B) a_{i_{t,0}} \quad (8)$$

A série  $\{a_{i_t}\}$ , bem como as auxiliares  $\{u_{i_t}\}$  e  $\{v_{i_t}\}$ , podem ser obtidas recursivamente, fazendo-se os valores iniciais iguais a zero, isto é,

$$a_{i,t,0} = \bar{z}_{i,t} - \phi_{i,1,0} \bar{z}_{i,t-1} - \dots - \phi_{i,p_i,0} \bar{z}_{i,t-p_i} + \theta_{i,1,0} a_{i,t-1,0} + \dots + \theta_{i,q_i,0} a_{i,t-q_i,0} \quad (9)$$

$$u_{i,t} = \theta_{i,1,0} u_{i,t-1} + \dots + \theta_{i,q_i,0} u_{i,t-q_i} + \bar{z}_{i,t} \quad (10)$$

$$= \phi_{i,1,0} u_{i,t-1} + \dots + \phi_{i,p_i,0} u_{i,t-p} + a_{i,t,0} \quad (11)$$

$$v_{i,t} = \theta_{i,1,0} v_{i,t-1} + \dots + \theta_{i,q_i,0} v_{i,t-q_i} - a_{i,t,0} \quad (12)$$

Expandindo-se  $\{a_{i,t}\}$  em s\u00e9rie de Taylor no entorno de  $\underline{\beta}_{i_0} = \{\underline{\phi}_{i_0}, \underline{\theta}_{i_0}\}$ , tem-se aproximadamente,

$$a_{i,t,0} = \sum_{j=1}^{p_i} (\phi_{i,j} - \phi_{i,j,0}) u_{i,t-j} + \sum_{j=1}^{q_i} (\theta_{i,j} - \theta_{i,j,0}) v_{i,t-j} + a_{i,t} \quad (13)$$

Os ajustamentos s\u00e3o, ent\u00e3o, para cada equa\u00e7\u00e3o  $i$  ( $1 \leq i \leq m$ ) do sistema, os coeficientes da regress\u00e3o linear simples de  $a_{i,t,0}$  em  $u_{i,t-j}$  e  $v_{i,t-j}$ . Pela adi\u00e7\u00e3o dos valores ajustados as primeiras estimativas  $\underline{\beta}_{i_0}$ , um conjunto de segundas estimativas \u00e9 formado e, o processo continua at\u00e9 que a converg\u00eancia seja obtida.

O sistema de equa\u00e7\u00f5es dado por (13) pode, compactamente, ser escrito de acordo com notaa\u00e7\u00e3o vetorial, como

$$a_{i,t,0} = \sum_{j=1}^{k_i} (\underline{\beta}_{i,j} - \underline{\beta}_{i,j,0}) X_{i,t-j} + a_{i,t} \quad (14)$$

onde,

$$X_{i,t-j} = - \frac{\partial a_{i,t}}{\partial \beta_{i,j}} \Big|_{\beta_i = \beta_{i_0}} \quad (15)$$

ou, ainda, sob forma matricial, como

$$a_{i_0} = X_i \beta_i + a_i, \quad i = 1, \dots, m \quad (16)$$

onde, cada  $a_{i_0}$  é um vetor de  $T \times 1$ , o número de variáveis independentes, para um dado  $i$ , é dado por  $k_i$  e a soma,  $k_1 + k_2 + \dots + k_m$  é igual a  $k$ .

Nenhuma das variáveis ou parâmetros nas  $m$  equações precisam ser relacionadas; a conexão entre as equações reside, somente, nos termos relativos aos resíduos, os quais são contemporaneamente relacionados nas diferentes equações.

Entretanto, as  $m$  equações (16) podem ser escritas juntas, como

$$a_{t,0} = X_t \beta + a_t, \quad t = 1, 2, \dots, T \quad (17)$$

onde,  $a_{t,0}$  é um vetor  $m \times 1$  e  $X_t$  é uma matriz  $m \times k$ , dada por,

$$X_t = \begin{bmatrix} X'_{1t} & 0 & \dots & 0 \\ 0 & X'_{2t} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & X'_{mt} \end{bmatrix} \quad (18)$$

onde,  $X_{i_t}$ , dado pela equação (15), é um vetor  $k_i \times 1$  das variáveis independentes da  $i$ -ésima equação. O vetor  $\beta$  é  $k \times 1$  e é definido exatamente como em (16) e os  $a_t$ 's são serialmente não correlacionados com média zero e matriz de covariância,  $\langle a_t a_t' \rangle = \Omega$ .

As  $m$  equações (17), possuem uma matriz de covariância dos resíduos  $a^* = (a_1', \dots, a_m')'$ , dada por,

$$\langle a^* a^{*'} \rangle = \begin{bmatrix} \Omega & 0 & \dots & 0 \\ 0 & \Omega & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \Omega \end{bmatrix} \quad (19)$$

a qual pode ser escrita de forma mais consisa como,

$$\langle a^* a^{*'} \rangle = I \otimes \Omega \quad (20)$$

Entretanto, não é necessário a utilização dos produtos de Kronecker no sentido de se mostrar que o estimador GLS de  $\beta$ , é dado por

$$\hat{\beta} = \left( \sum_{t=1}^T X_t' \Omega^{-1} X_t \right)^{-1} \sum_{t=1}^T X_t' \Omega^{-1} a_{t,0} \quad (21)$$

Por outro lado, é importante observar que, em geral,  $\Omega$  é desconhecida e, por isso, (21) não pode ser utilizada diretamente. Entretanto, ela pode ser estimada pela aplicação de mínimos quadrados ordinários ("OLS") separadamente a cada uma das  $m$  equações definidas em (16). Após a convergência do processo individual se,  $a_i$  representa o vetor  $T \times 1$  dos resíduos OLS na  $i$ -ésima equação, então, o estimador do elemento  $ij$  de  $\Omega$  é dado por,

$$\hat{\omega}_{ij} = \frac{a_i' a_j}{T}, \quad i, j = 1, \dots, m \quad (22)$$

Pela substituição de  $\hat{\Omega}$  em (21) e, pela utilização das estimativas iniciais de  $\beta_i$ , ao final do processo de convergência das  $m$  equações iniciais, obtêm-se um estimador SURE factível.

O estimador da matriz de variância-covariância assintótica do estimador SURE é, então, dada por,

$$\text{VAR}(\hat{\beta}) = \left( \sum_{t=1}^T X_t' \hat{\Omega}^{-1} X_t \right)^{-1} \quad (23)$$

Além disso, como  $\hat{\beta}$  é um estimador linear, a sua distribuição de probabilidade será multivariada normal, quando os resíduos são normais. Para tanto, tudo o que se necessita é de uma transformação do tipo Box & Cox (1964) aplicada aos dados originais.

Finalmente, um algoritmo linear para estimação e identificação de um modelo multivariado SURARMA ( $p, q$ ), pode ser sumariado como se segue:

### Estágio I - Preliminar

A cada uma das  $m$  equações de um modelo SURARMA como de finido pela equação (3), aplique o algoritmo de identificação e estimação linear como proposto por Holanda Sales et al. (1987). A fase final do algoritmo corresponde à aplicação das equações (9), (10) ou (11), (12) e (13).

### Estágio II - Identificação

Com as  $m$  séries de resíduos  $\{a_{i_t}\}$ , obtidas no estágio anterior, obtenha a matriz de correlação cruzada às séries  $\{a_{i_t}\}$  e  $\{a_{j_t}\}$ ,  $i, j = 1, \dots, m$ .

Se o modelo SURARMA for adequado, somente as correlações

de ordem zero devem ser significativamente diferentes de zero. Sob a hipótese de que o modelo é adequado, os limites de um intervalo de confiança de aproximadamente 95% são dados por  $\pm 2/\sqrt{T-1}$ . O teste de significância para as correlações cruzadas pode ser obtido a partir do gráfico das correlações cruzadas dos resíduos,  $R(k)$ ,  $k = 0, +1, +2, \dots, +n$ ,  $i > j$ , com os seus respectivos intervalos de confiança ou, ainda, através de uma matriz resumo das correlações cruzadas,  $R(k) = R_{a_i a_j}(k)$ , onde "+" indica um valor superior a  $2/\sqrt{T-1}$ , "-", um valor inferior a  $-2/\sqrt{T-1}$  e "." um valor entre  $-2/\sqrt{T-1}$  e  $2/\sqrt{T-1}$ . Um outro teste que poderia ser aplicado é o de Portmanteau, o qual considera as  $n$  correlações cruzadas como um todo.

Esta técnica de pré-branqueamento é similar aos procedimentos de identificação propostos por Box & Jenkins (1976) para modelos ARMA multivariados. Entretanto, em algumas situações, nenhum modelo é rigorosamente apropriado. Este é o caso dos modelos SURARMA onde os resíduos para diferentes períodos de tempo, são não correlacionados. No caso da existência de correlações significativas para ordens diferentes de zero, outros modelos como Função de Transferência ou Multivariado Completo, devem ser pesquisados. Além disso, o possível ganho na eficiência dos estimadores, no caso específico dos modelos SURARMA só ocorre quando as correlações de ordem zero são significativas. Se estas correlações forem fracas, é preferível a utilização de técnicas de modelagem univariadas.

### Estágio III - Estimação Final

Aplique a equação (22) aos resíduos univariados obtidos no Estágio I no sentido de se obter uma estimativa da matriz de variância-covariância  $\Omega$ . A partir daí, utilize as equações (21) e (23) para a obtenção das estimativas de mínimos quadrados lineares generalizados, bem como a matriz de variância-covariância das estimativas e as novas séries de resíduos. O processo continua pela aplicação das equações (15), (18), (21), (22) e (23), até que a convergência final seja obtida.

### APLICAÇÃO

Como caso de aplicação do algoritmo proposto foram utilizadas as séries de vazões naturais médias mensais aos reservatórios de Furnas, no Rio Grande, Itumbiara, no Rio Paranaíba e Ilha Solteira, no Rio Paraná. Os dados compreendem o período de janeiro de 1931 a dezembro de 1984 e foram obtidos no Departamento de Operação Energética (DEOP) das Centrais Elétricas Brasileiras S.A. - ELETROBRÁS.

A tabela 1, a seguir, resume as principais características dos aproveitamentos considerados.

**TABELA 1 - Principais Características dos Reservatórios de Furnas, Itumbiara e Ilha Solteira**

CARACTERÍSTICAS	FURNAS	ITUMBIARA	I. SOLTEIRA
ÁREA DE DRENAGEM (km <sup>2</sup> )	50.464	95.000	375.460
NÍVEL MÁXIMO NORMAL (m)	768,00	520,00	328,00
NÍVEL MÁXIMO OPERATIVO (m)	768,00	520,00	328,00
NÍVEL MÁXIMO MAXIMORUM (m)	769,30	521,20	329,00
NÍVEL MÁXIMO DE SEGURANÇA (m)	769,30	521,20	329,00
VOLUME MORTO (km <sup>3</sup> )	5,733	4,573	8,232
VOLUME ÚTIL (km <sup>3</sup> )	17,217	12,454	12,828
CAPACIDADE DO VERTEDOR DE SUPERFÍCIE (m <sup>3</sup> /s)	13.000	16.270	37.905
CAPACIDADE DO VERTEDOR DE FUNDO (m <sup>3</sup> /s)	13.000	16.270	37.905
DESCARGA DE RESTRIÇÃO PARA CÁLCULO DE VOLUME DE ESPERA (m <sup>3</sup> /s)	4.000	7.000	2.000

As principais estatísticas mensais das séries analisadas estão listadas nas tabelas 2, 3 e 4.

**TABELA 2 - Estatísticas Mensais das Séries de Vazões Naturais Médias Mensais do Aproveitamento de Furnas**

LOCAL	MÊS	MÉDIA (M <sup>3</sup> /S)	MAXIMO (M <sup>3</sup> /S)	MÍNIMO (M <sup>3</sup> /S)	DESVIO PADRAO (M <sup>3</sup> /S)	COEFICIENTE DE ASSIMETRIA
FURNAS	JAN	1512	3215	571	639	1,242
	FEV	1535	3114	355	654	0,802
	MAR	1368	3650	463	596	1,626
	ABR	915	1690	379	306	0,972
	MAI	663	1257	288	205	1,038
	JUN	543	921	310	149	0,769
	JUL	463	808	275	123	1,010
	AGO	383	721	200	108	1,139
	SET	388	1048	196	156	1,966
	OUT	471	927	276	155	1,625
	NOV	680	1431	324	260	1,216
	DEZ	1120	2370	252	401	0,773

**TABELA 3 - Estatísticas Mensais das Séries de Vazões Naturais Médias Mensais do Aproveitamento de Itumbiara**

LOCAL	MÊS	MÉDIA (M <sup>3</sup> /S)	MAXIMO (M <sup>3</sup> /S)	MÍNIMO (M <sup>3</sup> /S)	DESVIO PADRAO (M <sup>3</sup> /S)	COEFICIENTE DE ASSIMETRIA
ITUMBIARA	JAN	2364	4350	680	876	0,806
	FEV	2617	5160	725	1145	0,817
	MAR	2485	5320	916	1070	1,113
	ABR	1844	4210	798	709	1,602
	MAI	1193	2320	494	378	1,259
	JUN	918	1750	423	282	1,069
	JUL	723	1250	310	212	0,847
	AGO	561	1030	271	174	0,925
	SET	477	1010	263	152	1,510
	OUT	620	1140	254	203	0,972
	NOV	1023	2240	483	386	1,357
	DEZ	1728	3570	554	660	1,003

TABELA 4 - Estatísticas Mensais das Séries de Vazões Naturais Médias Mensais do Aproveitamento de Ilha Solteira

LOCAL	MÊS	MÉDIA (M3/S)	MÁXIMO (M3/S)	MÍNIMO (M3/S)	DESVIO PADRAO ( M3/S)	COEFICIENTE DE ASSIMETRIA
I. SOLTEIRA	JAN	7617	13862	2933	2568	0,732
	FEV	8266	15293	2534	3043	0,534
	MAR	7925	15273	3532	2892	0,925
	ABR	5852	11200	2843	1890	1,274
	MAI	3990	6651	2240	1059	1,143
	JUN	3185	5010	2085	770	0,863
	JUL	2643	4113	1664	590	0,797
	AGO	2193	3485	1399	470	0,899
	SET	2045	3908	1280	577	1,749
	OUT	2408	4403	1430	591	1,576
	NOV	3420	6693	2059	1016	1,388
	DEZ	5452	10048	1856	1727	0,939

A aplicação da transformação de Box & Cox (1964) às séries originais conduziu aos resultados mostrados na tabela 5. Nesta tabela pode-se notar que, em todas as séries, a transformação selecionada foi do tipo logaritmo natural.

TABELA 5 - Transformações de Box & Cox Selecionadas para as Séries de Vazões Naturais Médias Mensais dos Aproveitamentos de Furnas, Itumbiara e Ilha Solteira

SÉRIE	$\lambda_1$	$\lambda_2$
FURNAS	0	- 179
ITUMBIARA	0	- 179
I. SOLTEIRA	0	-1238

As médias e variâncias mensais das séries transformadas eram periódicas. Procedeu-se, então, uma padronização mensal de todas as séries no sentido de se obter estacionaridade de segunda ordem das mesmas. O algoritmo proposto foi, por conseguinte, aplicado as séries resultantes.

O estágio I do algoritmo consistiu em se identificar modelos ARMA ( $p_i, q_i$ ) univariados a cada uma das  $i = 1, 3$  séries estudadas. Neste estágio foram obtidas, também, estimativas dos parâmetros dos modelos selecionados. Cabe ressaltar que estas estimativas são as iniciais do processo iterativo do estágio III e foram obtidas, basicamente, pela aplicação do algoritmo proposto por Holanda Sales et al (1987) a cada uma das séries individualmente.

A tabela 6, a seguir, sumariza os modelos selecionados no estágio I do algoritmo para cada uma das séries analisadas. Entre parêntesis são apresentados os erros das estimativas.



TABELA 6 - Modelos ARMA ( $p_i, q_i$ ) Seleccionados para as Séries de Furnas, Itumbiara e Ilha Solteira

SÉRIE	(p, q)	ESTIMATIVAS		
FURNAS	(1, 1)	$\hat{\phi}_1=0,8421(0,0237)$	$\hat{\theta}_1=-0,2398(0,0426)$	$\hat{\sigma}_a^2=0,4343$
ITUMBIARA	(1, 0)	$\hat{\phi}_1=0,7351(0,0267)$		$\hat{\sigma}_a^2=0,4595$
I. SOLTEIRA	(1, 1)	$\hat{\phi}_1=0,8229(0,0258)$	$\hat{\theta}_1=-0,1620(0,0449)$	$\hat{\sigma}_a^2=0,4104$

A seguir, no estágio II, com os resíduos univariados estimados no estágio I, obteve-se as Funções de Covariância e Correlação Cruzadas Amostrais. Os gráficos 1, 2 e 3 corroboram a hipótese de dependência contemporânea entre as séries. Há, portanto, evidências de que um modelo multivariado como o SURARMA deve conduzir a estimativas para os parâmetros mais eficientes que técnicas univariadas.

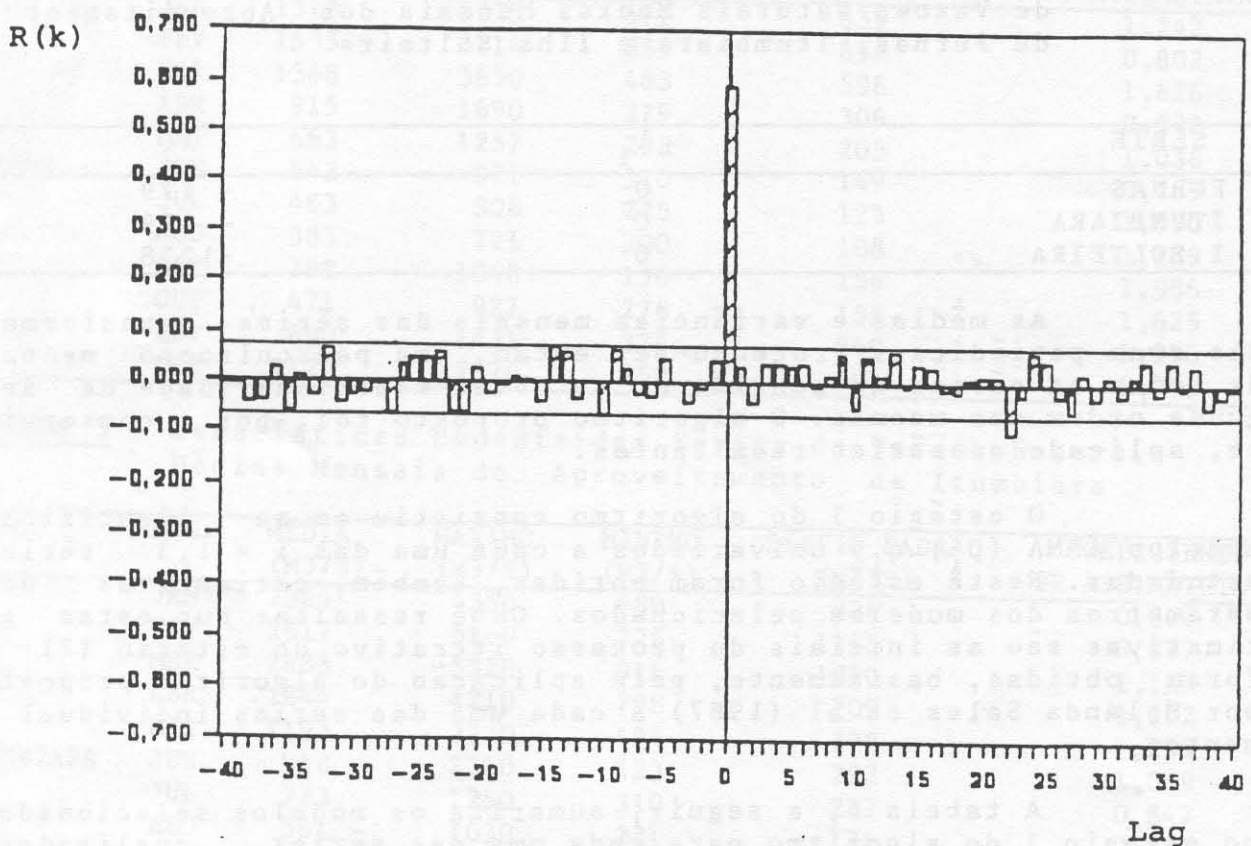


Figura 1. FUNÇÃO DE CORRELAÇÃO CRUZADA AMOSTRAL FURNAS x ITUMBIARA

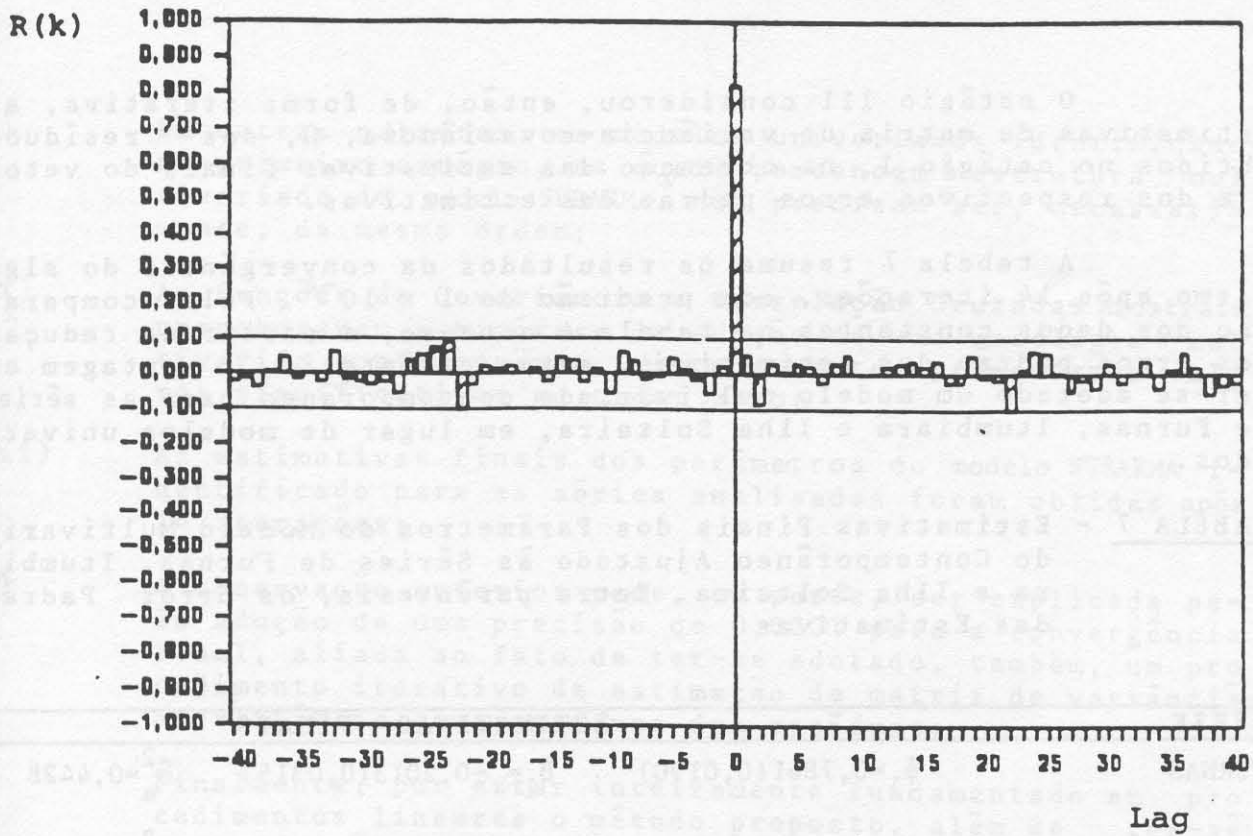


Figura 2. FUNÇÃO DE CORRELAÇÃO CRUZADA AMOSTRAL  
FURNAS x ILHA SOLTEIRA

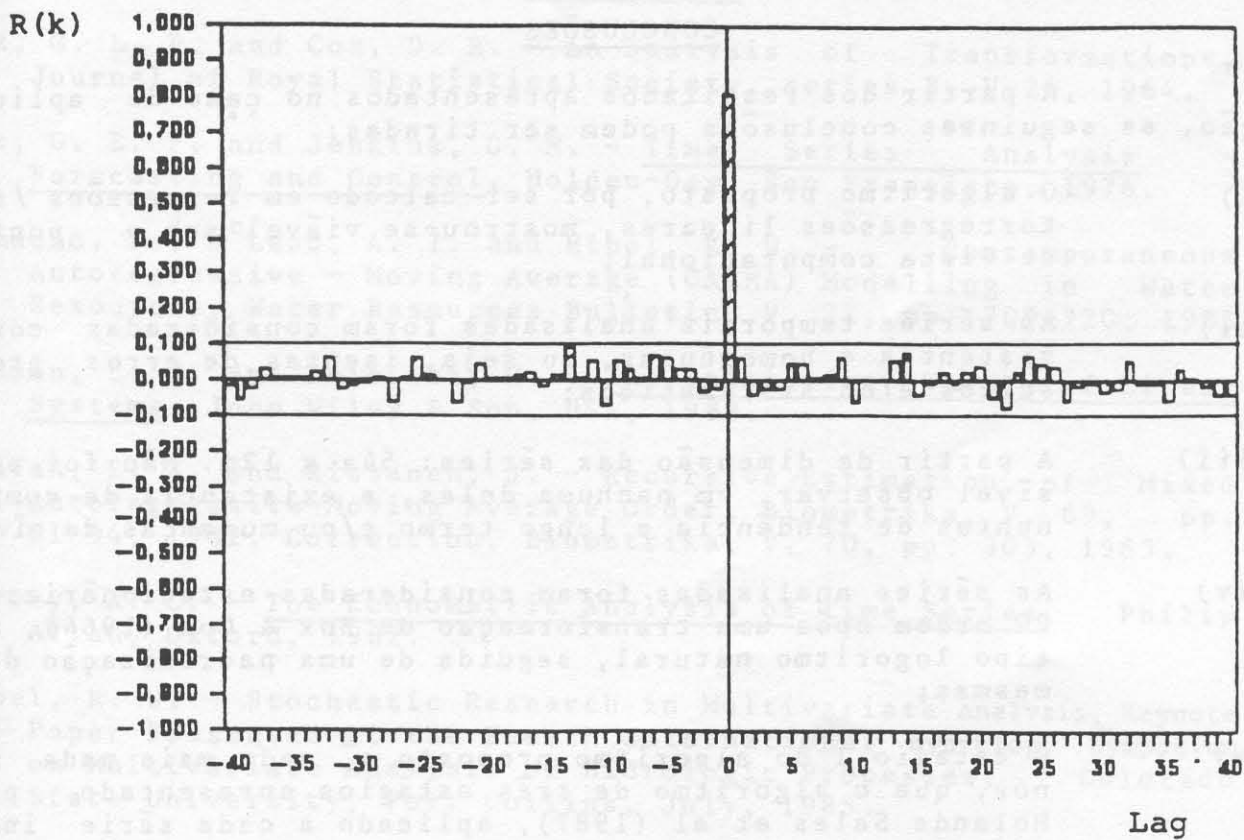


Figura 3. FUNÇÃO DE CORRELAÇÃO CRUZADA AMOSTRAL  
ITUMBIARA x ILHA SOLTEIRA

O estágio III considerou, então, de forma iterativa, as estimativas da matriz de variância-covariância,  $\Omega$ , dos resíduos obtidos no estágio I, na obtenção das estimativas finais do vetor  $\beta$  e dos respectivos erros padrão das estimativas.

A tabela 7 resume os resultados da convergência do algoritmo após 14 iterações, com precisão de  $1 \times 10^{-4}$ . Pela comparação dos dados constantes na tabela 4 nota-se, a partir da redução dos erros padrão das estimativas, entre parêntesis, a vantagem em ter-se adotado um modelo multivariado contemporâneo para as séries de Furnas, Itumbiara e Ilha Solteira, em lugar de modelos univariados.

**TABELA 7** - Estimativas Finais dos Parâmetros do Modelo Multivariado do Contemporâneo Ajustado às Séries de Furnas, Itumbiara e Ilha Solteira. Entre parêntesis, os Erros Padrão das Estimativas

SÉRIE	ESTIMATIVAS FINAIS		
FURNAS	$\hat{\phi}_1=0,7861(0,0190)$	$\hat{\theta}_1= -0,2013(0,0315)$	$\hat{\sigma}_a^2=0,4428$
ITUMBIARA	$\hat{\phi}_1=0,6661(0,0210)$		$\hat{\sigma}_a^2=0,4616$
I.SOLTEIRA	$\hat{\phi}_1=0,7572(0,0175)$	$\hat{\theta}_1= - 0,1593(0,0237)$	$\hat{\sigma}_a^2=0,4140$

#### CONCLUSÕES

A partir dos resultados apresentados no caso de aplicação, as seguintes conclusões podem ser tiradas:

- i) O algoritmo proposto, por ser calcado em regressões / autotregressões lineares, mostrou-se viável sob o ponto de vista computacional;
- ii) As séries temporais analisadas foram consideradas consistentes e homogêneas, ou seja, isentas de erros grosseiros e/ou sistemáticos;
- iii) A partir da dimensão das séries: 54a x 12m. Não foi possível observar, em nenhuma delas, a existência de componentes de tendência a longo termo e/ou mudanças de nível;
- iv) As séries analisadas foram consideradas estacionárias de 2º ordem após uma transformação de Box & Cox (1964), do tipo logaritmo natural, seguida de uma padronização das mesmas;
- v) O estágio I do algoritmo proposto é, nada mais nada menos, que o algoritmo de três estágios apresentado por Holanda Sales et al (1987), aplicado a cada série individualmente;

- vi) Em outras palavras, os modelos univariados identificados no primeiro estágio, os quais pertencem a estrutura multivariada do modelo SURARMA, não precisam ser, necessariamente, da mesma ordem;
- vii) As funções de Covariância e Correlação Cruzadas Amostrais corroboraram a hipótese de adoção de uma modelagem multivariada contemporânea, visto somente apresentarem valores significativos na ordem zero;
- viii) As estimativas finais dos parâmetros do modelo SURARMA identificado para as séries analisadas foram obtidas após 14 iterações;
- ix) A observação anterior pode, em parte, ser explicada pela adoção de uma precisão de 0,0001 para a convergência final, aliada ao fato de ter-se adotado, também, um procedimento iterativo de estimação de matriz de variância-covariância contemporânea dos resíduos;
- x) Finalmente, por estar inteiramente fundamentado em procedimentos lineares o método proposto, além de ter-se mostrado eficiente e extremamente rápido, pode substituir, com vantagens, os algoritmos não-lineares existentes, onde a intervenção do analista tenta minimizar os elevados tempos computacionais.

#### REFERÊNCIAS

- Box, G. E. P. and Cox, D. R. - An analysis of Transformations, Journal of Royal Statistical Society, series B, V.26, 1964.
- Box, G. E. P. and Jenkins, G. M. - Time Series Analysis - Forecasting and Control, Holden-Day, San Francisco, 1976.
- Camacho, F., McLeod, A. I. and Hipel, K. W. - Contemporaneous Autoregressive - Moving Average (CARMA) Modelling in Water Resources, Water Resources Bulletin, V. 21, pp. 709-720, 1985.
- Hannan, E. J. and Deistler, M. - The Statistical Theory of Linear Systems, John Wiley & Son, USA, 1988.
- Hannan, E. J. and Rissanen, J. - Recursive Estimation of Mixed Autoregressive-Moving Average Order, Biometrika, V. 69, pp. 81-94, 1982. Correction, Biometrika, V. 70, pp. 303, 1983.
- Harvey, A. C. - The Econometric Analysis of Time Series, Philip Allan, Oxford, 1982.
- Hipel, K. W. - Stochastic Research in Multivariate Analysis, Keynote Paper Presented at the Fourth International Hydrology Symposium on Multivariate Analysis of Hidrologic Processes, Colorado State University, Fort Collins, July, 1985.
- Holanda Sales, P. R., Pereira, B. de B. e Vieira, A. M. - Procedimentos Lineares de Identificação e Estimação de Modelos ARMA



## CAPÍTULO IV

### PROCEDIMENTOS LINEARES PARA IDENTIFICAÇÃO E ESTIMAÇÃO DE PARÂMETROS DE MODELOS PARA SÉRIES HIDROLÓGICAS TEMPORAIS MULTIVARIADAS

POR

PAULO R. H. SALES<sup>(1)</sup>, BASÍLIO DE B. PEREIRA<sup>(2)</sup>, AILTON M. VIEIRA<sup>(1)</sup>

RESUMO - - O presente artigo propõe uma extensão natural para o caso multivariado dos procedimentos lineares de identificação e estimação propostos por Holanda Sales (1989), Holanda Sales, Pereira e Vieira (1987) e (1989, a, b e c) para séries temporais. Como principal objetivo, pode-se destacar a enorme flexibilidade do método, a partir da minimização da intervenção do analista. Assim é que, de forma quase automática, pode-se passar do modelo multivariado completo para o restrito, onde restrições são impostas aos elementos das matrizes dos parâmetros  $\phi$  e  $\theta$ , bem como para algumas subclasses do modelo multivariado, tais como o SURARMA, o Função de Transferência e o MARMAX.

O método proposto, por ser calcado em regressões / autor-regressões lineares é, também, bastante rápido e eficiente. Basicamente, em seu primeiro estágio são considerados, apenas, modelos multivariados autorregressivos de ordem  $k$ ,  $MAR(k)$ , onde  $k$  é selecionado de tal sorte que minimize  $AIC(k)$ . O fim precípuo deste primeiro estágio é, tão somente, estimar as inovações,  $a_t$ . Com os vetores  $z_t$  e  $\hat{a}_t$  são, então, procedidos no segundo estágio, a identificação das ordens  $p$  e  $q$  e a estimação preliminar dos parâmetros do modelo  $MARMA(p,q)$ . Agora, o critério de seleção utilizado é o  $BIC(p,q)$ . Finalmente, com a matriz de variância-covariância estimada dos resíduos,  $\hat{\Omega}_{p,q}$ , são obtidas, a partir de regressões lineares, as estimativas finais das matrizes  $\phi$  e  $\theta$  e do vetor  $\sigma_a^2$ .

Como caso de aplicação foram utilizadas as séries hidrológicas temporais de vazões naturais médias mensais aos reservatórios de Furnas, no rio Grande, Três Marias, no rio São Francisco e Itumbiara, no rio Paranaíba. Os dados vão de janeiro de 1931 a dezembro de 1984 e foram obtidos no Departamento de Operação Energética das Centrais Elétricas Brasileiras S.A. (ELETROBRÁS).

---

<sup>1</sup>Engenheiros, Departamento de Operação Energética, ELETROBRÁS, Rio de Janeiro, RJ

<sup>2</sup>Professor Titular, COPPE/Universidade Federal do Rio de Janeiro, RJ.

## INTRODUÇÃO

Considere-se, inicialmente, algumas relações de dependência que podem ser encontradas entre duas ou mais séries temporais. Os três tipos mais fundamentais de relacionamento poderiam ser classificados como contemporâneo, unidirecional ou triangular e de retroação. Duas ou mais séries apresentam um relacionamento contemporâneo somente se os valores atuais são relacionados, isto é, as séries mostram uma casualidade instantânea. As vazões a diferentes estações de medição de uma região representam um exemplo deste caso se a dependência é apenas contemporânea não estando, portanto, relacionados a modificações naturais ou do homem. A maior dependência neste caso reside na precipitações, isto é, nas séries de entrada das bacias hidrográficas.

A casualidade unidirecional, por exemplo para duas séries X e Y, pode ser definida como se X causa Y então os valores de Y são melhor explicados a partir do passado de seus próprios valores e os de X. Um exemplo disso seria a de que as vazões de um posto de um rio são importantes (causam) na explicação das vazões afluentes a um outro posto, a jusante do primeiro, no mesmo rio, mas não vice-versa. Um outro exemplo de relação unidirecional seria o da "chuva x vazão".

Por outro lado, uma relação de retroação é dita existir se e só se a relação de casualidade existe nas duas direções. Neste caso, o relacionamento é simétrico, isto é, X causa Y e Y causa X. Como exemplo, poder-se-ia citar as séries de precipitação e evapotranspiração de uma bacia hidrográfica ou, ainda, a precipitação, a evaporação e o armazenamento em um reservatório.

Finalmente, não muito fácil de se encontrar mas possível, também, seria um relacionamento misto entre três ou mais séries. Em outras palavras, duas poderiam possuir um relacionamento apenas contemporâneo entre si mas um relacionamento unidirecional com uma terceira e, assim por diante.

Um número muito grande de modelos multivariados tem sido propostos recentemente na literatura. Entretanto, a grande maioria é bastante similar e pode ser classificada de acordo com os mecanismos de dependência e relacionamento anteriormente citados. Isto posto, prender-se-á a uma descrição mais minuciosa das seguintes subclasses de modelos multivariados: SURARMA ou contemporâneo ARMA; Função de Transferência; ARMA vetorial e ARMA exógeno ou ARMAX vetorial. Modelos outros como o de regressão múltipla, modelos agregativos, modelos desagregativos e o autorregressivo médias móveis espacial ou STARMA não serão, portanto, enfocados.

A forma estrutural de cada um dos modelos citados pode ser descrita como se segue. Seja um conjunto de m séries temporais representadas pelo vetor,

$$\underline{z}_t = (z_{1t}, z_{2t}, \dots, z_{mt})'$$

sejam, também,

$$\underline{\mu}_t = (\mu_1, \mu_2, \dots, \mu_m)'$$

o vetor das médias de  $\underline{z}_t$  e,

$$\tilde{\underline{z}}_t = \underline{z}_t - \underline{\mu}_t$$

O modelo multivariado ARMA (p,q) para representar as m séries pode ser escrito como,

$$\tilde{\underline{z}}_t = \phi_1 \tilde{\underline{z}}_{t-1} + \dots + \phi_p \tilde{\underline{z}}_{t-p} + \underline{a}_t - \theta_1 \underline{a}_{t-1} - \dots - \theta_q \underline{a}_{t-q} \quad (1)$$

onde,  $\phi_i$  e  $\theta_j$  são, respectivamente, matrizes m x m dos parâmetros autorregressivos e médias móveis. Os elementos  $a_{it}$ , do vetor  $\underline{a}_t$ ,

possuem média zero, são mutuamente não correlacionados no tempo mas são contemporaneamente correlacionados entre eles, ou seja,  $\langle a_{ij} a_{jt} \rangle \neq 0$  e  $\langle a_{it} a_{js} \rangle = 0$ ,  $t \neq s$ .

A equação (1) pode, alternativamente, ser escrita como,

$$\tilde{\underline{z}}_t = \sum_{i=1}^p \phi_i \tilde{\underline{z}}_{t-i} + \underline{a}_t - \sum_{i=1}^q \theta_i \underline{a}_{t-i} \quad (2)$$

onde,  $\phi_i$  é uma matriz de parâmetros m x m de elementos autorregressivos,  $\phi_{ijk}$ ;  $\theta_i$  é uma matriz de parâmetros m x m de elementos médias móveis,  $\theta_{ijk}$  e,  $\tilde{\underline{z}}_t$  e  $\underline{a}_t$  são os mesmos definidos anteriormente.

Entretanto, como observado por JENKINS (1979) a maneira de se representar o modelo multivariado, como indicado pela equação (2), apresenta alguns inconvenientes. Primeiro, há restrições nos modelos univariados para as séries individuais  $z_{it}$ , de tal forma que

os operadores autorregressivos sejam da mesma ordem e possuam os mesmos valores para os parâmetros. Segundo, as m séries são supostas estar em equilíbrio estatístico sobre médias fixas, isto é, assume-se a estacionariedade.

Estas restrições são altamente indesejáveis e podem ser contornadas, escrevendo-se o modelo como,

$$\begin{bmatrix} \phi_{11}(B) & \phi_{12}(B) & \dots & \phi_{1m}(B) \\ \phi_{21}(B) & \phi_{22}(B) & \dots & \phi_{2m}(B) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{m1}(B) & \phi_{m2}(B) & \dots & \phi_{mm}(B) \end{bmatrix} \begin{bmatrix} \omega_{1t} \\ \omega_{2t} \\ \vdots \\ \omega_{mt} \end{bmatrix} = \begin{bmatrix} \theta_{11}(B) & \theta_{12}(B) & \dots & \theta_{1m}(B) \\ \theta_{21}(B) & \theta_{22}(B) & \dots & \theta_{2m}(B) \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{m1}(B) & \theta_{m2}(B) & \dots & \theta_{mm}(B) \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \\ \vdots \\ a_{mt} \end{bmatrix} \quad (3)$$

ou,

$$\phi_p(B) \underline{\omega}_t = \theta_q(B) \underline{a}_t \quad (4)$$



onde,

- i)  $\phi_{ij}(B)$  e  $\theta_{ij}(B)$  são matrizes autorregressivas e médias-móveis, respectivamente, cujos elementos  $\phi_{ij}(B)$  e  $\theta_{ij}(B)$  são polinômios em B dos graus  $p_{ij}$  e  $q_{ij}$ ;
- ii) B é o operador atraso, isto é,  $B^h \omega_t = \omega_{t-h}$ ;
- iii)  $\tilde{\omega}_t = (\nabla_{s_1}^{d_1} \nabla_{s_1}^{D_1} \tilde{z}_{1t}^{(\lambda_1)}, \dots, \nabla_{s_m}^{d_m} \nabla_{s_m}^{D_m} \tilde{z}_{mt}^{(\lambda_m)})'$ ;
- iv)  $\nabla_i^{d_i} = (1 - B)^{d_i}$  é o operador diferença de ordem  $d_i$  da série  $i$  ( $i = \overline{1, m}$ );
- v)  $\nabla_{s_i}^{D_i} = (1 - B^{s_i})^{D_i}$  é o operador diferença sazonal de ordem  $D_i$  e período  $s_i$ , da série  $i$  ( $i = \overline{1, m}$ );
- vi)  $\lambda_i$  é uma transformação do tipo BOX & COX, para série  $i$  ( $i = \overline{1, m}$ ), e,
- vii)  $\tilde{a}_t$  é o vetor de variáveis aleatórias independentes e identicamente distribuídas com média zero e matriz de variância-covariância  $\Omega$ .

Com respeito ao modelo multivariado completo expresso por (4), as seguintes observações podem ser destacadas:

- i) os graus dos operadores  $\phi_{ij}(B)$  e  $\theta_{ij}(B)$  podem, durante a fase de construção do modelo, ser ajustados de tal forma que os modelos univariados para cada série descrevam, com mais precisão, seus comportamentos individuais;
- ii) o modelo multivariado SURARMA ou ARMA contemporâneo resulta quando os coeficientes  $\phi_{ij}(B)$  e  $\theta_{ij}(B)$  são nulos para  $i \neq j$ , isto é, as matrizes dos parâmetros são diagonais;
- iii) o modelo de Função de Transferência é obtido quando os coeficientes  $\phi_{ij}(B)$  e  $\theta_{ij}(B)$  são nulos para  $i < j$  ou, em outras palavras, quando as matrizes dos parâmetros são triangulares;
- iv) o modelo ARMA vetorial com variáveis exógenas ou MARMAX é obtido ao se delir uma ou mais linhas da matriz de parâmetros autorregressivos,  $\phi(B)$ , e uma ou mais colunas da matriz de parâmetros médias-móveis,  $\theta(B)$ .

Esta última classe de modelos, cujas propriedades matemáticas e estatísticas têm sido estudadas por Hannan (1985), (1987, a e b), Hannan & Deistler (1988), Hannan & Kavalieris (1984) e Hannan, Kavalieris & Mackisack (1986), são definidos como MARMAX(p, q, r) e podem ser escritos como,

$$\tilde{z}_t - \sum_{i=1}^p \phi_i \tilde{z}_{t-i} = \sum_{i=0}^r \psi_i u_{t-i} + \tilde{a}_t - \sum_{i=1}^q \theta_i \tilde{a}_{t-i} \quad (5)$$

onde,  $\tilde{z}_t$ ,  $\phi$ ,  $\theta$  e  $a$  são os mesmos definidos anteriormente;  $u$  é um vetor de dimensões  $k$  de variáveis exógenas e,  $\delta_i$  é a correspondente  $m \times k$  matriz de parâmetros.

### IDENTIFICAÇÃO E ESTIMAÇÃO LINEAR

Considere, inicialmente, modelos puramente autorregressivos. A equação (1) pode, então, ser escrita como,

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \dots + \phi_k \tilde{z}_{t-k} + a_{k_t} \quad (6)$$

onde,

$$\langle a_{k_t} \ a_{k_s} \rangle = \begin{cases} \Omega_k, & t = s \\ 0, & \text{caso contrário} \end{cases}$$

A equação (6) pode ser vista como uma regressão de  $\tilde{z}_t$  em  $\tilde{z}_{t-i}$  ( $i=1, \dots, k$ ).

Entretanto, a menos que haja restrições nos elementos das matrizes  $\phi_i$ , um conjunto eficiente de estimadores pode, então, ser obtido simplesmente pela regressão de cada elemento de  $\tilde{z}_t$  no vetor  $\tilde{z}_{t-i}$  ( $i=1, \dots, k$ ). Isto se prende ao fato de que as covariâncias entre resíduos, nas diferentes equações, não conduzem a melhores estimadores visto, cada equação, possuir exatamente o mesmo conjunto de variáveis explicativas.

A equação (6) pode, alternativamente, ser escrita sob notação matricial, como

$$\tilde{z}_t = B X_t + a_t, \quad t = 1, \dots, T \quad (7)$$

onde,  $\tilde{z}_t$  é um vetor ( $m \times 1$ ),  $B$  é uma matriz ( $m \times mk$ ) de parâmetros,  $X_t$  é um vetor ( $mk \times 1$ ) de regressores e,  $a_t$  é um vetor ( $m \times 1$ ), seriamente não correlacionados com média zero, e matriz de variância-covariância,  $\langle a_t \ a_t' \rangle = \Omega$ .

O estimador multivariado de mínimos quadrados para a matriz  $B$  pode ser escrito como,

$$\hat{B} = \left[ \begin{array}{c} T \\ \Sigma \\ \hline \end{array} \begin{array}{cc} X_t & X_t' \end{array} \right]^{-1} \left[ \begin{array}{c} T \\ \Sigma \\ \hline \end{array} \begin{array}{c} X_t \\ Z_t' \end{array} \right] \quad (8)$$

A  $i$ -ésima coluna de (8) fornece o vetor dos estimadores de mínimos quadrados ordinários da regressão de  $\tilde{z}_i$  ( $i=1, \dots, m$ ) em  $X_t$ .

Pela definição da matriz ( $m \times T$ ),  $Z' = (\tilde{z}_1, \dots, \tilde{z}_T)$  e da matriz ( $mk \times T$ ),  $X = (X_1, \dots, X_T)$ , uma forma mais compacta de apresentação matricial é dada por,

$$\hat{B} = (X'X)^{-1} (X'Z) \quad (9)$$

ou,

$$\hat{B} = Z'X (X'X)^{-1} \quad (10)$$

Por outro lado, a matriz  $(m^2 k \times m^2 k)$  de variância-covariância dos estimadores de mínimos quadrados é dada por,

$$\text{VAR}(\hat{B}) = \Omega \otimes (X' X)^{-1} \quad (11)$$

onde, o estimador de  $\Omega$ , vale

$$\hat{\Omega} = T^{-1} (Z' Z - \hat{B} X' Z) \quad (12)$$

e,  $\otimes$  representa os produtos de Kronecker.

Para modelos MARMA (p,q) um algoritmo de identificação e estimação linear pode finalmente ser estabelecido como:

### Estágio I - Preliminar

Estime (6) para todos os possíveis inteiros k. A ordem da autotregressão é tal que minimiza,

$$\text{AIC}(k) = \ln_n |\hat{\Omega}_k| + \frac{2k m^2}{T} \quad (13)$$

onde, T é o menor número de observações entre as m séries.

O critério AIC é, em geral, conservador, no sentido de que tende a sobreestimar o modelo. Entretanto, o propósito maior desse estágio é estimar as m séries de resíduos e não o verdadeiro modelo.

As inovações podem, então, ser estimadas por,

$$\hat{a}_t = \tilde{z}_t - \hat{\phi}_1 \tilde{z}_{t-1} - \dots - \hat{\phi}_k \tilde{z}_{t-k} \quad (14)$$

### Estágio II - Identificação e Estimação Inicial

Obtenha, para vários valores de p e q, as regressões de  $\tilde{z}$  em  $\tilde{z}_{t-i}$  ( $i=1, p$ ) e  $\hat{a}_{t-i}$  ( $i=1-q$ ), onde os  $\hat{a}_t$  são obtidos de (14).

Os modelos estimados neste estágio são, portanto, da forma,

$$\tilde{z}_t = \phi_1 \tilde{z}_{t-1} + \dots + \phi_p \tilde{z}_{t-p} - \theta_1 \hat{a}_{t-1} - \dots - \theta_q \hat{a}_{t-q} + \hat{a}_{p,q,t} \quad (15)$$

onde,  $\hat{a}_{p,q,t}$  representa o novo vetor de resíduos.

A grande vantagem do estágio anterior é que a equação (15) pode facilmente ser estimada por mínimos quadrados ordinários.

Finalmente, selecione p e q de tal forma que,

$$\text{BIC}(p,q) = \ln_n |\hat{\Omega}_{p,q}| + (p+q) m^2 \frac{\ln T}{T} \quad (16)$$

seja o menor possível.

Então, após a seleção das ordens p e q, tem-se

$$\hat{a}_t = \tilde{z}_t - \phi_1 \tilde{z}_{t-1} - \dots - \phi_p \tilde{z}_{t-p} + \theta_1 \hat{a}_{t-1} + \dots + \theta_q \hat{a}_{t-q} \quad (17)$$

Estágio III - Estimação Final

Seja, agora, a partir de (17),

$$\langle \hat{a}_t, \hat{a}_s \rangle = \begin{cases} \Omega_{p,q} & , t = s \\ 0 & , t \neq s \end{cases} \quad (18)$$

onde,  $\langle \cdot \rangle$ , representa valor esperado.

Estimativas aproximadas de máxima verossimilhança para o vetor dos parâmetros do modelo MARMA (p,q),  $\beta_i$  ( $i = 1, 2, \dots, k$ ), dado  $\Omega_{p,q}$ , são as soluções de,

$$T^{-1} \sum_{t=1}^T \hat{a}_t^{i'} \Omega_{p,q}^{-1} \hat{a}_t = 0, \quad (i = \overline{1,k}) \quad (19)$$

onde,  $\hat{a}_t^i$  representa a derivada parcial de  $\hat{a}_t$  em relação a  $\beta_i$  ( $i = \overline{1,k}$ ),

O procedimento consiste, então, em linearizar  $\hat{a}_t$  no entorno de  $\beta_{o_i}$  ( $i = \overline{1,k}$ ), onde  $\beta_{o_i}$  é uma estimativa inicial de  $\beta_i$  obtida por exemplo, no estágio anterior. Logo tem-se, aproximadamente,

$$\hat{a}_{t,o} = \sum_{i=1}^k \hat{a}_t^i (\beta_i - \beta_{o_i}) + \hat{a}_t \quad (20)$$

Substituindo (20) em (19), as equações lineares para  $(\beta_i - \beta_{o_i})$  são dadas por,

$$C (\beta_i - \beta_{o_i}) = g \quad (21)$$

onde, C é uma matriz  $k \times k$  cujos elementos são dados por,

$$c_{ij} = T^{-1} \sum_{t=1}^T \hat{a}_t^{i'} \Omega_{p,q}^{-1} \hat{a}_t^j \quad (22)$$

e, g é dado por,

$$g_i = T^{-1} \sum_{t=1}^T \hat{a}_t^{i'} \Omega_{p,q}^{-1} \hat{a}_{t,o} \quad (23)$$

Deve-se ressaltar que, ambos, C e g são avaliados em  $\beta_{o_i}$ .

Das equações (21), (22) e (23), vem que a expressão do estimador de  $\beta_i$  pode, alternativamente, ser escrita como,

$$\hat{\beta}_i = \left( \sum_{t=1}^T \hat{a}_t^{i'} \Omega_{p,q}^{-1} \hat{a}_t^i \right)^{-1} \sum_{t=1}^T \hat{a}_t^{i'} \Omega_{p,q}^{-1} \hat{a}_{t,o} \quad (24)$$

O estimador da variância assintótica de  $\beta_i$  é, então, dado por,

$$\text{VAR} (\hat{\beta}_i) = \left( \sum_{t=1}^T \hat{a}_t^{i'} \Omega_{p,q}^{-1} \hat{a}_t^i \right)^{-1} \quad (25)$$

e, o estimador da covariância entre  $\beta_i$  e  $\beta_j$ , por,

$$\text{COV} (\hat{\beta}_i, \hat{\beta}_j) = \left( \sum_{t=1}^T \hat{a}_t^{i'} \Omega_{p,q}^{-1} \hat{a}_t^j \right)^{-1} \quad (26)$$

No caso do modelo não possuir termos médias-móveis então,  $a_t$  é linear em  $\beta$  e a solução para (19) é dada, simplesmente, por,

$$\hat{\beta} = \hat{\beta}_0 + (\hat{\beta} - \hat{\beta}_0) \quad (27)$$

Entretanto, se o modelo não é autorregressivo puro, um novo vetor de parâmetros, dado por (27) é utilizado no lugar de  $\hat{\beta}_0$  e, o processo de linearização é repetido até que a convergência final seja obtida.

### APLICAÇÃO

Com o objetivo precípuo de exemplificar os procedimentos lineares propostos na modelagem multivariada ARMA, foram utilizadas as séries de vazões naturais médias mensais aos reservatórios de Furnas, no rio Grande, Itumbiara, no rio Paranaíba e Três Marias, no rio São Francisco. As séries hidrológicas temporais analisadas possuem 648 observações cada, compreendendo o período de janeiro de 1931 a dezembro de 1984 e foram obtidas no Departamento de Operação Energética (DEOP) das Centrais Elétricas Brasileiras S.A. - ELETROBRÁS.

A tabela 1, a seguir, resume as principais características dos aproveitamentos considerados.

TABELA 1 - Principais Características do Reservatórios de Furnas, Itumbiara e Três Marias

CARACTERÍSTICAS	FURNAS	ITUMBIARA	TRÊS MARIAS
ÁREA DE DRENAGEM (km <sup>2</sup> )	50.464	95.000	50.560
NÍVEL MÁXIMO NORMAL (m)	768,00	520,00	568,30
NÍVEL MÁXIMO MAXIMORUM (m)	769,30	521,20	568,50
VOLUME MORTO (km <sup>3</sup> )	5,733	4,573	4,250
VOLUME ÚTIL (km <sup>3</sup> )	17,217	12,454	15,278
CAPACIDADE DE VERTEADOR DE SUPERFÍCIE (m <sup>3</sup> /s)	13.000	16.270	8.700
CAPACIDADE DO VERTEADOR DE FUNDO (m <sup>3</sup> /s)	-	-	-
DESCARGA DE RESTRIÇÃO PARA CÁLCULO DE VOLUME DE ESPERA (m <sup>3</sup> /s)	4.000	7.000	2.500

As principais estatísticas mensais das séries analisadas estão listadas nas tabelas 2, 3 e 4.

TABELA 2 - Estatísticas Mensais da Série de Vazões Naturais  
Médias Mensais do Aproveitamento de Furnas

LOCAL	MÊS	MÉDIA (m <sup>3</sup> /s)	MÁXIMO (m <sup>3</sup> /s)	MÍNIMO (m <sup>3</sup> /s)	DESVIO PADRÃO (m <sup>3</sup> /s)	COEFICIENTE DE ASSIMETRIA
FURNAS	JANEIRO	1714	3529	571	704	0,641
	FEVEREIRO	1679	3264	355	683	0,361
	MARÇO	1520	3650	463	659	1,059
	ABRIL	1028	2367	379	366	1,013
	MAIO	747	1679	288	247	1,078
	JUNHO	630	2392	310	289	4,151
	JULHO	520	1368	275	171	2,237
	AGOSTO	431	1082	200	139	1,870
	SETEMBRO	449	1989	196	262	3,935
	OUTUBRO	543	1934	276	249	3,312
	NOVEMBRO	768	2018	324	318	1,351
	DEZEMBRO	1285	3117	252	504	0,930

TABELA 3 - Estatísticas Mensais da Série de Vazões Naturais  
Médias Mensais do Aproveitamento de Itumbiara

LOCAL	MÊS	MÉDIA (m <sup>3</sup> /s)	MÁXIMO (m <sup>3</sup> /s)	MÍNIMO (m <sup>3</sup> /s)	DESVIO PADRÃO (m <sup>3</sup> /s)	COEFICIENTE DE ASSIMETRIA
ITUMBIARA	JANEIRO	2654	4938	680	959	0,294
	FEVEREIRO	2876	7004	725	1275	0,724
	MARÇO	2723	5320	916	1132	0,555
	ABRIL	2038	4210	798	744	0,939
	MAIO	1316	2320	494	398	0,579
	JUNHO	1012	1750	423	300	0,435
	JULHO	797	1331	310	223	0,205
	AGOSTO	622	1047	271	184	0,262
	SETEMBRO	541	1375	263	193	1,714
	OUTUBRO	697	1576	254	239	0,911
	NOVEMBRO	1141	2530	483	448	1,019
	DEZEMBRO	1899	3818	554	720	0,558

A aplicação da transformação de BOX & COX (1964) às séries originais conduziu aos resultados mostrados na tabela 5. Nesta tabela pode-se notar que, em todas as séries, a transformação selecionada foi do tipo logaritmo natural.

As médias e variâncias mensais das séries transformadas eram periódicas. Procedeu-se, então, uma padronização mensal de todas as séries no sentido de se obter estacionaridade de segunda ordem das mesmas. O algoritmo proposto foi, por conseguinte, aplicado as séries resultantes.

TABELA 4 - Estatísticas Mensais da Série de Vazões Naturais Médias Mensais do Aproveitamento de Três Marias

LOCAL	MÊS	MÉDIA (m <sup>3</sup> /s)	MÁXIMO (m <sup>3</sup> /s)	MÍNIMO (m <sup>3</sup> /s)	DESVIO PADRAO (m <sup>3</sup> /s)	COEFICIENTE DE ASSIMETRIA
TRÊS MARIAS	JANEIRO	1489	3249	259	662	0,526
	FEVEREIRO	1405	3859	199	789	1,219
	MARÇO	1164	2949	330	594	1,172
	ABRIL	789	1910	201	382	0,886
	MAIO	474	1150	129	185	0,933
	JUNHO	357	1100	157	148	2,291
	JULHO	287	681	115	100	1,086
	AGOSTO	232	507	92	77	0,781
	SETEMBRO	218	539	92	80	1,273
	OUTUBRO	296	874	103	121	2,227
	NOVEMBRO	606	1520	234	281	1,285
	DEZEMBRO	1150	2290	137	488	0,423

TABELA 5 - Transformações de Box & Cox selecionadas para as séries de Vazões Naturais Médias Mensais dos Aproveitamentos de Furnas, Itumbiara e Três Marias.

SÉRIES	$\lambda_1$	$\lambda_2$
FURNAS	0	- 179
ITUMBIARA	0	- 179
TRÊS MARIAS	0	- 90

O estágio I do algoritmo linear proposto consistiu em se estimar, para as séries analisadas, modelos multivariados autorregressivos de diversas ordens. Em outras palavras, nesta fase do algoritmo, os modelos estimados foram do tipo MAR(k). Com auxílio do critério AIC(k), cuja expressão é dada por (13), o modelo selecionado foi o MAR(3). A tabela 6 mostra os valores de AIC(k) considerando, entre parêntesis as variâncias dos resíduos das séries de Furnas, Itumbiara e Três Marias, respectivamente.

TABELA 6 - Valores de AIC(k) para Modelos MAR(k) Estimados para as Séries de Furnas, Itumbiara e Três Marias. Entre Parêntesis, as Variâncias dos Resíduos Respectivos.

k	AIC(k)	VARIÂNCIA DOS RESÍDUOS		
		FURNAS	ITUMBIARA	TRÊS MARIAS
1	- 3,7804	0,4472	0,4533	0,4590
2	- 3,8371	0,4338	0,4485	0,4521
3	- 3,8490	0,4266	0,4459	0,4458
4	- 3,8349	0,4252	0,4419	0,4406
5	- 3,8169	0,4254	0,4392	0,4405
6	- 3,7990	0,4250	0,4376	0,4411
7	- 3,7747	0,4255	0,4377	0,4410
8	- 3,7615	0,4235	0,4374	0,4392
9	- 3,7400	0,4220	0,4364	0,4373
10	- 3,7214	0,4206	0,4361	0,4348

A seguir, com os resíduos das três séries, obtidos no estágio anterior, foram estimados diversos modelos multivariados do tipo MARMA(p,q). O critério BIC (p,q), dado por (16) indicou, neste segundo estágio do algoritmo, o modelo MARMA (1,1), como mostrado na tabela 7, a seguir.

TABELA 7 - Valores de BIC (p,q) para Modelos MARMA (p,q) para as séries de Furnas, Itumbiara e Três Marias.

p \ q	0	1	2	3
0	-	- 2,3650	- 2,6677	- 2,9218
1	- 3,7804	- 3,8578	- 3,8445	- 3,8269
2	- 3,8371	- 3,8391	- 3,8330	- 3,8142
3	- 3,8490	- 3,8255	- 3,8123	- 3,7990

As estimativas iniciais dos parâmetros do modelo selecionado no estágio II constam da tabela 8. Entre parêntesis, os erros padrão das estimativas.

TABELA 8 - Estimativas Iniciais para os Parâmetros do Modelo MARMA (1,1) para as Séries de Furnas, Itumbiara e Três Marias. Entre Parêntesis, os Erros Padrão das Estimativas.

SÉRIE	MATRIZ $\phi$			MATRIZ $\theta$			VARIÂNCIA DOS RESÍDUOS
	FURNAS	ITUMB.	T.MARIAS	FURNAS	ITUMB.	T.MARIAS	
FURNAS	0,8693 (0,0355)	-0,0516 (0,0382)	0,0151 (0,0429)	-0,3296 (0,0496)	0,0814 (0,0509)	0,0513 (0,0577)	0,4301
ITUMB.	0,0188 (0,0362)	0,7517 (0,0390)	0,0199 (0,0438)	0,0466 (0,0506)	-0,2004 (0,0520)	0,0939 (0,0589)	0,4479
T.MARIAS	0,0452 (0,0362)	-0,0617 (0,0391)	0,8201 (0,0439)	-0,0423 (0,0507)	0,1055 (0,0521)	-0,2094 (0,0590)	0,4498

Uma análise minuciosa dos resultados apresentados no estágio II permitiu, para o caso particular das séries de Furnas, Itumbiara e Três Marias, que o algoritmo proposto fosse explorado ao máximo. Assim é que, no estágio III, tentou-se duas classes de modelo MARMA (1,1).

Em uma primeira etapa, o processo iterativo do terceiro estágio levou em consideração o modelo multivariado completo, isto é, sem que restrições fossem impostas aos seus parâmetros. As estimativas finais para os mesmos foram obtidas, para uma precisão de  $1 \times 10^{-4}$ , após 5 iterações. A tabela 9 sumariza os resultados encontrados.

A seguir, em uma segunda etapa, foram impostas restrições aos parâmetros do modelo multivariado completo MARMA (1,1). Em outras palavras, foi considerada consistente a hipótese de que nem todos os parâmetros do modelo eram significativamente diferentes de ze-



ro. Com efeito, se considerarmos os erros padrão das estimativas, indicados entre parêntesis na tabela 8, a hipótese contrária não-la poderia ser aceita. Destarte, a adaptabilidade do algoritmo foi testada a partir da imposição  $\beta_i = 0$ , para  $|\hat{\varepsilon}_i| \leq 1DP(\hat{\sigma}_i)$ . Em parte, o que se pretendem, apenas, foi seguir o recomendado por JENKINS (1979), o qual afirma que em aplicações práticas os graus dos operadores  $\phi_{ij}(B)$  e  $\theta_{ij}(B)$ , em (3), podem ser ajustados de tal forma que os modelos univariados para as séries individuais descrevam, com mais precisão, seus comportamentos.

TABELA 9 - Estimativas Finais para os Parâmetros do Modelo MARMA (1,1), sem restrições, para as séries de Furnas, Itumbiara e Três Marias. Entre Parêntesis, os Erros Padrão das Estimativas.

SÉRIE	MATRIZ $\phi$			MATRIZ $\theta$			VARIÂNCIA DOS RESÍDUOS
	FURNAS	ITUMB.	T.MARIAS	FURNAS	ITUMB.	T.MARIAS	
FURNAS	0,8587 (0,0209)	-0,0477 (0,0243)	0,0204 (0,0260)	-0,3769 (0,0359)	0,0994 (0,0366)	0,0643 (0,0415)	0,4297
ITUMB.	0,0094 (0,0214)	0,7444 (0,0248)	0,0288 (0,0266)	0,0106 (0,0366)	-0,2072 (0,0373)	0,1141 (0,0424)	0,4477
T.MARIAS	0,0359 (0,0214)	-0,0548 (0,0248)	0,8232 (0,0267)	0,0868 (0,0367)	0,1324 (0,0374)	-0,2083 (0,0425)	0,4493

No caso específico das séries analisadas, ao se impor ao modelo MARMA (1,1) as restrições  $\phi_{13} = \phi_{21} = \phi_{23} = 0$  e  $\theta_{13} = \theta_{21} = \theta_{31} = 0$ , novas estimativas iniciais para os demais parâmetros, assim como para as séries de resíduos, foram obtidas. A tabela 10 apresenta estas novas estimativas para os parâmetros do modelo MARMA (1,1) restrito.

TABELA 10 - Estimativas Iniciais para os Parâmetros do Modelo MARMA (1,1) Restrito para as Séries de Furnas, Itumbiara e Três Marias. Entre Parêntesis, os Erros Padrão das Estimativas.

SÉRIE	MATRIZ $\phi$			MATRIZ $\theta$			VARIÂNCIA DOS RESÍDUOS
	FURNAS	ITUMB.	T.MARIAS	FURNAS	ITUMB.	T.MARIAS	
FURNAS	0,08688 (0,0205)	-0,0555 (0,0348)	-	-0,3382 (0,0366)	0,0815 (0,0523)	-	0,4299
ITUMB.	-	0,7389 (0,0327)	-	-	-0,2214 (0,0569)	0,0967 (0,0449)	0,4478
T.MARIAS	0,0458 (0,0213)	-0,0677 (0,0363)	0,8213 (0,0242)	-	0,1052 (0,0566)	-0,2177 (0,0436)	0,4497

As estimativas finais dos parâmetros do modelo MARMA (1,1) restrito foram, então, obtidas após 4 iterações do estágio III do algoritmo proposto para uma precisão de  $1 \times 10^{-4}$ . A tabela 11 resume os principais resultados do processo final de convergência;

entre parêntesis são apresentados os erros padrão das estimativas.

TABELA 11 - Estimativas Finais para os Parâmetros do Modelo MARMA (1,1) Restrito para as Séries de Furnas, Itumbiara e Três Marias. Entre Parêntesis, os Erros Padrão das Estimativas.

SÉRIE	MATRIZ $\phi$			MATRIZ $\theta$			VARIÂNCIA DOS RESÍDUOS
	FURNAS	ITUMB.	T.MARIAS	FURNAS	ITUMB.	T.MARIAS	
FURNAS	0,8685 (0,0206)	-0,0551 (0,0354)	-	-0,3397 (0,0363)	0,0827 (0,0521)	-	0,4299
ITUMB.	-	0,7380 (0,0333)	-	-	-0,2239 (0,0565)	0,0971 (0,0445)	0,4478
T.MARIAS	0,0459 (0,0215)	-0,0665 (0,0369)	0,8202 (0,0243)	-	0,1093 (0,0562)	-0,2219 (0,0431)	0,4497

### CONCLUSÕES

- i) o algoritmo linear de três estágios proposto mostrou-se amplamente viável computacionalmente;
- ii) o tempo total de execução nos computadores IBM 4381 R 14 da ELETROBRÁS, para as séries de Furnas, Itumbiara e Três Marias, cada uma com 648 observações, foi de 34,15 s, para o modelo completo e de 24,38 s para o restrito;
- iii) para todas as séries analisadas a família de transformações BOX & COX conduziu a uma transformação do tipo logaritmo natural;
- iv) a estacionaridade de segunda ordem foi obtida após se proceder a uma padronização mensal das séries transformadas;
- v) o modelo identificado no estágio I foi o Multivariado Autorregressivo de ordem 3, ou seja, MAR (3);
- vi) no segundo estágio foi identificado o modelo Multivariado Autotregressivo - Médias Móveis de ordem (1,1), ou seja, MARMA (1,1). As estimativas iniciais dos parâmetros do modelo necessárias na última fase do algoritmo foram, também, obtidas neste segundo estágio;
- vii) no estágio III foi considerado, inicialmente, o modelo Multivariado Completo. As estimativas finais dos parâmetros do mesmo foram obtidas após 5 iterações, para uma precisão de  $1 \times 10^{-4}$ ;
- viii) a seguir, com o fim precípuo de se testar a enorme flexibilidade do método proposto, foi considerado, no estágio III, o Modelo Multivariado Restrito, onde alguns  $\phi$ 's e  $\theta$ 's eram iguais a zero. As estimativas finais para os parâmetros do modelo restrito foram, então, obtidas após 4 iterações, para

uma precisão de  $1 \times 10^{-4}$ ;

ix) deve ser destacado que tanto em (vii) quanto em (viii) houve reestimação, a cada iteração, da matriz de variância-covariância dos resíduos.

#### REFERÊNCIAS BIBLIOGRÁFICAS

BOX, G.E.P. and COX, D.R. - An Analysis of Transformations, Journal of Royal Statistical Society, series B, V.26, (1964).

HANNAN, E.J. - Multivariable ARMA Systems and Practicable Calculations Manuscript, (1985).

HANNAN, E.J. - Approximation of Linear Systems, In Time Series and Econometric Modelling, Ed. I.B.MacNeill, pp. 1-12, (1987).

HANNAN, E.J. and KAVALIERS, L. - Multivariate Linear Times Series Models, Advances in Applied Probability, V.16, pp. 492-561, (1984).

HANNAN, E.J., KAVALIERS, L. and MACKISACK, M. - Recursive Estimation of Linear Systems, Biometrika, V.73, pp.119 - 133, (1986).

HOLANDA SALES, P.R. - Procedimentos Lineares para Identificação e Estimação de Parâmetros de Modelos para Séries Temporais Uni e Multivariadas, Tese de D.Sc., Univeridade Federal do Rio de Janeiro, COPPE, Programa de Engenharia da Produção, Rio de Janeiro, maio de 1989.

HOLANDA SALES, P.R., PEREIRA, B. DE B. e VIEIRA, A.M. - Procedimentos Lineares de Identificação e Estimação de Modelos ARMA para Séries Hidrológicas Temporais, VII Simpósio Brasileiro de Hidrologia e Recursos Hídricos, Salvador, novembro(1987).

HOLANDA SALES, P.R., PEREIRA, B. DE B. e VIEIRA, A.M. - Um Procedimento Linear de Identificação para Modelos de Função de Transferência para Séries Hidrológicas Temporais, IV Simpósio Luso-Brasileiro de Hidráulica e Recursos Hídricos, Lisboa, Portugal, (1989).

HOLANDA SALES, P.R., PEREIRA, B. DE B. e VIEIRA, A.M. - Identificação e Estimação Lineares em Séries Hidrológicas Temporais Multivariadas Contemporâneas, VIII Simpósio Brasileiro de Recursos Hídricos, Foz do Iguaçu, Brasil, (1989).

JENKINS, G.M. - Practical Experiences with Modelling and Forecasting Time Series, GLP Publication, Londres, (1979).

**APOIO: CNPq – FINEP – FAPESP – FAPERJ**